TESTING PRACTICES

IN

CHRISTCHURCH PRIMARY SCHOOLS

---

A thesis

presented to the Faculty of the Department of Education

University of Canterbury

---

in partial fulfilment

of the requirements for the Degree

Master of Arts in Education

---

by

Douglas John Callander

March 1961
ACKNOWLEDGEMENTS

The writer is greatly indebted to many headmasters and teachers, who gave freely of their time and knowledge. Without their willing co-operation, this investigation would not have been possible.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th></th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>THE PROBLEM AND THESIS ORGANIZATION</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>The Problem</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Statement of the Problem</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Importance of the Investigation</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Organization of the Thesis</td>
<td>4</td>
</tr>
<tr>
<td>II.</td>
<td>REVIEW OF THE TESTING MOVEMENT</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Historical Divisions</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Early beginnings to 1800 A.D.</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>1800 to 1900</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>1900 to present time</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Present Status in New Zealand</td>
<td>12</td>
</tr>
<tr>
<td>III.</td>
<td>METHOD OF INVESTIGATION</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Types of Schools Selected</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Analysis of Teachers and Classes</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Factors Affecting the Sample</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Characteristics of Teachers</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Obtaining the Information</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td>28</td>
</tr>
<tr>
<td>IV.</td>
<td>INTELLIGENCE TESTING</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Types of Tests Used</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Teachers' Qualifications for Testing</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Uses of Tests</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Testing Programmes</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td>44</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>ATTAINMENT AND DIAGNOSTIC TESTING</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>Criteria of Tests and Test Plans</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Tests in Current Use (Summary)</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Arithmetic Tests</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Spelling Tests</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Reading and Comprehension</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>Comparisons and Criticisms of Tests</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>Uses of Tests</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Characteristics of Testers</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Test Programmes</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td>74</td>
</tr>
<tr>
<td>VI.</td>
<td>TEACHER CONSTRUCTED TESTS</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Criteria of Tests</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Arithmetic Tests</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Spelling Tests</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>Comprehension Tests</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>Composition Tests</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Social Studies Tests</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>Writing Tests</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>Marking Techniques</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>Headmasters' Influences</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>Summaries</td>
<td>114</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>VII. SUMMARY AND CONCLUSIONS</td>
<td>123</td>
<td></td>
</tr>
<tr>
<td>Replies to the Statement of the Problem</td>
<td>123</td>
<td></td>
</tr>
<tr>
<td>Conclusions</td>
<td>138</td>
<td></td>
</tr>
<tr>
<td>Implications</td>
<td>143</td>
<td></td>
</tr>
<tr>
<td>Further Research</td>
<td>147</td>
<td></td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>149</td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>I</td>
<td>Location and Type of School in Sample</td>
<td>16</td>
</tr>
<tr>
<td>II</td>
<td>Teachers in Relation to Classes</td>
<td>17</td>
</tr>
<tr>
<td>III</td>
<td>Composite Classes</td>
<td>18</td>
</tr>
<tr>
<td>IV</td>
<td>Analysis of Teachers According to Schools</td>
<td>20</td>
</tr>
<tr>
<td>V</td>
<td>Bases of Streaming by Schools</td>
<td>22</td>
</tr>
<tr>
<td>VI</td>
<td>Roll Numbers of Classes</td>
<td>23</td>
</tr>
<tr>
<td>VII</td>
<td>Relationship between University Qualifications of Teachers and Types of Schools</td>
<td>24</td>
</tr>
<tr>
<td>VIII</td>
<td>Testing Qualifications of Teachers</td>
<td>25</td>
</tr>
<tr>
<td>IX</td>
<td>Teaching Experience of Teachers</td>
<td>26</td>
</tr>
<tr>
<td>X</td>
<td>Analysis of Certified Service</td>
<td>27</td>
</tr>
<tr>
<td>XI</td>
<td>Reasons for Use of Otis Test</td>
<td>31</td>
</tr>
<tr>
<td>XII</td>
<td>Reasons for Non-Use of Intelligence Tests</td>
<td>33</td>
</tr>
<tr>
<td>XIII</td>
<td>Standard Tests in Current Use</td>
<td>48</td>
</tr>
<tr>
<td>XIV</td>
<td>Teachers Using Standard Tests</td>
<td>69</td>
</tr>
<tr>
<td>XV</td>
<td>Experience and Test Usage</td>
<td>69</td>
</tr>
<tr>
<td>XVI</td>
<td>Reasons for Use of Text Book</td>
<td>82</td>
</tr>
<tr>
<td>XVII</td>
<td>Reasons for Part or Non-Use of Text Book</td>
<td>83</td>
</tr>
<tr>
<td>XVIII</td>
<td>Comprehension Test Construction</td>
<td>95</td>
</tr>
<tr>
<td>XIX</td>
<td>Frequency of Revision Testing in Social Studies</td>
<td>101</td>
</tr>
<tr>
<td>XX</td>
<td>Types of Assessments and Marking Methods</td>
<td>106</td>
</tr>
<tr>
<td>XXI</td>
<td>Tests and Users</td>
<td>124</td>
</tr>
<tr>
<td>TABLE</td>
<td>PAGE</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>XXII. Uses of Attainment Tests</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>XXIII. Uses of Intelligence Tests</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>XXIV. Examiners in Arithmetic</td>
<td>137</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER I

THE PROBLEM AND THESIS ORGANIZATION

The increasing use of standardized tests and the refinement of teacher constructed tests in recent years has provoked considerable discussion among teachers and has been an issue debated at refresher courses and in magazine articles.

Little is known, however, about the extent to which tests are applied or the value of these tests as they are currently used in schools in this country. There does not appear to be any specific data available on current practice in the use of either standardized tests or teacher planned tests, while neither the construction, nor marking of teachers' own tests seem to have been systematically investigated in New Zealand.

This investigation is planned to meet these needs.

I. THE PROBLEM

Statement of the Problem

The problem then, was to investigate a sample of teachers and headmasters to find out (1) the uses made of standardized testing and the purposes to which they are put, together with (2) the use, content, and marking of teacher-constructed tests.

Specifically answers were sought to the following questions:-

**Standardized Tests**

(a) What proportion of teachers use standardized tests of attainment, diagnosis, and intelligence?
(b) How are the tests used?

(c) What relationship exists between the experience and qualifications of the teachers, and the testing done?

(d) To what extent do schools follow a definite, adequate testing programme involving principles suggested by authors of text books on testing?

**Non-standardized or Teacher-constructed Tests**

(a) From what sources do teachers draw materials for tests that they construct?

(b) With what frequency do teachers assess attainment with tests they have made?

(c) How soundly based are the construction and content of these tests?

(d) How do the various marking systems compare with modern theory?

(e) What influence do headmasters exert on teachers, tests, and examinations in their schools?

(f) How adequate is the recording of test results?

This information was sought from ninety-three headmasters and teachers interviewed in seventeen Christchurch schools, and is confined to standardized tests in use in those schools, as well as tests in arithmetic, spelling, comprehension, composition, social studies, and writing.

**Importance of the Investigation**

At present there do not appear to be any figures on the
amount of standardized testing done in New Zealand primary schools.

Personal communications were made with likely sources\(^1\) of information, but no classified data was available. The New Zealand Council for Educational Research sold approximately sixty-five thousand Otis Intermediate tests during the year June 1959–June 1960, but the proportion sold to primary schools is not readily available. This, and other relevant information could be sifted from the files held by the N.Z.C.E.R., and the major suppliers of tests in this country, but such a task is outside the scope of this investigation.

As well as the advantages of using such tests, there are for teachers untrained in their use, certain dangers. It is important then to find out if in fact these are known, and how in current practice, standardized tests are used in schools.

Allowance is made in the present primary teachers' salary scale for university qualifications and in many cases, teachers are studying education or psychology as a part requirement for this. This study will analyse, for the sample investigated at least, any uses being made of these qualifications, as could reasonably be expected by the monetary inducements offered to obtain them.

\(^1\)Personal letters written to: Dr R. Wintemberg, Professor of Education, University of Auckland. Mr J. Watson, Acting Director, New Zealand Council for Educational Research, Wellington.
Periodically, criticism is voiced by teachers of the source material available to them, and so some importance could be attached to actual figures showing sources and in some cases criticism of these.

From parents', children's and teachers' viewpoints, marks, which are products of tests, are a measure of children's progress throughout their education. It is important therefore, to know how these marks are obtained, by analysing test construction and evaluation in current practice, against the criteria of modern educational theory.

II. ORGANIZATION OF THE THESIS

The present introductory chapter leads on to Chapter II which contains a brief summary of the history of the testing movement together with some information on its present status, as related to the problem being investigated.

Sources of data, the procedures followed for obtaining it, and an analysis of the sample constitute Chapter III.

Then follows Chapter IV, which shows the amounts and types of intelligence testing done within the sample, the reasons for use or not of intelligence tests, and an assessment of current practices in intelligence testing. Chapter V sets out the information obtained on standardized tests of attainment and diagnosis, showing the types and frequency of usage, uses and misuses as related to current thought, the applicability of tests used, and relevant criticisms.
In Chapter VI teacher constructed tests in the subjects already mentioned (arithmetic, spelling, comprehension, composition, social studies, and writing) are examined to show the amount, frequency, and type of test used, sources of test data, techniques of marking and recording of results, and headmasters' control over such tests.

The final Chapter VII contains the summary and conclusions.
CHAPTER II

REVIEW OF THE TESTING MOVEMENT

Greene, Jorgensen and Gerberich\(^1\) divide the history of measurement of educational attainments and capacities of school children into roughly three periods - the first being to about 1800 A.D., the second covering the 19th century, and the third from 1900 to the present.

During the first period, from beginning of historical records down to about the nineteenth century A.D., educational measurements were rather crude. Individual differences in capacities and abilities have been recognised for several thousand years but relatively little progress in educational testing was made until the opening of the present century.

According to Ballard\(^2\) the first examiner was Chinese and the first examination was a Civil Service Examination. As early as 220BC, China had an elaborate national system of examinations for the purpose of selecting her public officials, and these examinations have been known down through the ages for their unusual severity. The candidates were confined in isolated cells for hours a time, and were compelled to write lengthy papers for treatises on special topics assigned to them.

---


Other evidences of examinations are to be found in ancient literature, such as the Holy Bible\(^3\), when forty-two thousand Ephraimites were killed because they failed to pass an objective test of pronouncing the word "Shibboleth". Socrates subjected his pupils to exhaustive and searching questioning. Oral quizzing has undoubtedly been part of classroom procedure from the beginnings of teaching activity, and Ballard cites the example of the Earl of Eldon, who went to Oxford in 1766, testifying that an examination for the degree at Oxford was a farce. The Earl was examined in Hebrew and in History, and was asked what is the Hebrew for "the place of a skull", and "who founded University College?" On answering both these questions correctly, the degree was granted.

The first tests in school seem to have been in the medieval times. By 1219 A.D. the University of Bologna and the University of Paris before the close of the thirteenth century, required degree candidates to defend their theses orally. It would seem, however, that written examinations probably, first made their appearance for educational use at Cambridge, England, in 1702.

The second period embraces approximately the nineteenth century. Educational measurement began to assimilate ideas from various sources including scientific and statistical

\(^3\)Judges 12 : 5-7
techniques which were later to result in the modern objective testing movement.

In U.S.A., it seems that the first examinations of note were those of Boston in 1845. These examinations covered the subjects of arithmetic, astronomy, geography, grammar, history and natural philosophy. The details of preparation, administration, and interpretation of results, were intended to make the examinations as fair as possible. Questions were graded carefully in terms of difficulty and rules for scoring were prepared. On these bases, the schools were ranked in order of merit by the Boston School Committees. It was also emphasised that the test measured intellectual activity and requirements only, but that the inclusion of other "Immeasurables" such as work habits, respect for order etc., could modify the rankings thus obtained.

Early objective tests were apparently first used by the Rev. George Fisher, an English schoolmaster who designed a series of "Scale books" as early as 1864, which provided model answers of varying degrees of correctness numbered one to five against which the answers obtained in the tests could be measured.

The work of Galton, brought the scientific study of individual differences into focus, and also the derivation of

\[4\] Greene, Jorgensen and Gerberich *op cit* p.49
\[5\] Greene, Jorgensen and Gerberich *op cit* p.40
statistical methods. He devised a system of "Standard scores" and furnished tools essential to the development of educational and mental testing, and scientific method in education.

About 1887 Dr. E. S. Chaille an American physician developed simple tests for judging the mental levels of children to the age of three. Cattell in 1890, together with Wissler and Jastrow were prominent among the American experimenters devoting attention to intelligence. In France, Binet and Henri were also working on this problem. Dr J. M. Rice in 1894, administered a list of spelling words to pupils in many schools' systems, analysed the results, and found that pupils who had studied spelling thirty minutes a day for eight years were no better spellers than children who had studied spelling fifteen minutes a day for eight years. This was one of the first of investigations that resulted in significant attention being given to the objective method in educational testing.

The third brief period dating from about 1900 to the present is characterised by remarkable advances in statistical techniques in the measurement of achievement, intelligence and personality, and by increased analysis of the nature of ability, the tools used in its measurements, and the proper use and further improvement of these tools.
The third period saw the biggest step forward when Binet and Simon brought out the first intelligence scale in 1909 with a view to selecting mentally retarded pupils who required special instruction. They introduced the mental age concept, which was developed by subsequent workers in the intelligence quotients. Revisions of the Binet-Simon scale are basic to most individual intelligence tests of today, and this work was followed up by Terman and his collaborators who made the Stanford revisions of the Binet Scale in 1916, 1937 and again in 1959.

It was only about 1917 that the first group intelligence test was published and used. This was developed by Otis in the form of the Army Alpha test, used for Personnel Selection by the American army during the First World War. This verbal test was accompanied by the Army Beta, a non-verbal test which could supposedly be used with illiterates and persons of foreign extraction who could not speak English.

In these first two decades of this century, persons such as McDougall, Thorndike and Spearman were developing various aspects of theory and practice in the testing movement. Fleming maintainsthat the educational clinics established in America in the twenties led the way to a

Fleming, Research and the Basic Curriculum (London - University of London Press, 1946) P. 3 ff.
fuller realisation of the necessity for better adaptation of schools to the needs of pupils, and schemes such as the Delton Plan and Winnetka Technique were evolved. Studies on heredity and environment were consolidated, and group tests of attitudes and values began to absorb as much attention as tests of intelligence and achievement.

In the fourth decade, efforts were made to evaluate what had been accomplished before, and plans such as those of Terman's genetic studies of genius were formulated for studies of large groups of children over long periods of years, as well as highly elaborated case studies of individuals. Much of the results of this work, is now only reaching publication, but there is evidence which seems to show that many tests are highly specific, that there is much variability in the growth of individuals, and that a profound influence is exerted by a person's interest, ambitions, and wishes, upon his present performance.

Emphasis is also being put on the value of educational guidance, of the exploration of the whole personality in its setting, and the acknowledgement of the limitations of testing as such.

Rodgers⁷ states that the case in favour of scientific measurement in the social sciences has been established.

---

securely by research and investigation over the last fifty years, and for this reason such organisations as the National Institute of Industrial Psychology in Britain and the American Council on Education in America are continually concerned with test construction, validation and improvement.

The present status of standardised testing of intelligence and attainment in New Zealand has been well put by Rodgers\(^8\) when he states that teachers here have few New Zealand standardised tests at their disposal. This has led to an almost exclusive reliance on the Otis Intermediate Intelligence Test for group intelligence testing, and the use of a variety of attainment tests, only a few of which have New Zealand norms.

According to Schonell\(^9\) there are five different forms of standardised tests. These are tests of (1) general intelligence, (2) attainment in school subjects, (3) diagnosis of backwardness in a school subject, (4) special abilities, (5) prognosis in respect of future progress of school subjects. Of these five types of tests, the first is dealt with in chapter IV, while of the remaining four, only tests of attainment and diagnosis were investigated in this survey and are elaborated in Chapter V. No tests of special

\(^8\)Ibid.

abilities or prognosis were encountered among the teachers seen. It is in fact, most unlikely that these tests would be used in primary schools, as they are not readily available, and the subjects taught are too general to warrant their use at this level.

From personal observations made before and during this investigation it seems that there is a desire among teachers in primary schools to use standardised tests, but except for those who have attended an appropriate course at University, the ranges and uses of tests available are not well known.

Although no specific data was available before this survey was commenced, it seemed that little variation existed in the construction of non-standardised tests. The variety of questions available for new type tests as expounded by Vernon\textsuperscript{10} were not, in the writer's experience very widely used, mainly due to a lack of knowledge in test construction.

This was the climate then, in which this investigation was carried out.

CHAPTER III

METHOD OF INVESTIGATION

This survey was undertaken in the last four months of 1959 so that most teachers would have available lists covering at least two terms' work, apart from the standardized tests normally given at the beginning of the school year. In September, invitations to co-operate in this investigation were sent to nineteen schools, giving an approximate total of one hundred and twenty four teachers including headmasters. These schools were in three major groups - the North Western, South Eastern, and Eastern postal zones of Christchurch.

The main reason governing the selection of the sample was to include schools in which a considerable amount of testing would probably be found. So far as standardized testing was concerned, it seemed logical that because of the age limit of nine years on the Otis Intelligence Test, little could be done below standard four.

Types of Schools Selected

Thus having fixed on the classes of standard four upwards, the most likely places to find a concentrated amount of testing of all types seemed to be in the intermediate schools. One intermediate school from the Eastern and one from the North Western part of Christchurch were selected.
Because testing practices are often co-ordinated within each intermediate school's vicinity, the standard four classes of each of the four primary schools contributing to these intermediate schools were also selected.

So that the sample would not be restricted to intermediate and their contributing schools, an approximately equal number of non-contributing schools was included. For reasons of convenience these were selected from the Eastern postal zones of the city. It was thought that these would be representative of other Christchurch schools.

The different zones were chosen to allow for different socio-economic backgrounds of pupils. This aspect may not be particularly relevant as the sample is of teachers, not children. There are some teachers who teach in their own localities, but it does not follow that their socio-economic status is thus governed. Other factors such as availability of housing and transfers to other positions, influence where they live, but a teacher moving from one position to another in this city does not necessarily shift residence also. The status of the children could possibly affect the type of teaching required and thus the type of testing undertaken by the teachers.
Table I shows the location etc. of schools visited. Among the contributing schools invited, two declined, and are marked "x" in the table.

### Table I

**LOCATION AND TYPE OF SCHOOL IN SAMPLE**

<table>
<thead>
<tr>
<th>Location</th>
<th>Contributing</th>
<th>Non-contributing</th>
<th>Intermediate</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N.W.</td>
<td>$2^x$</td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>N.E.</td>
<td>1</td>
<td>1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>E.</td>
<td>$2^+$</td>
<td>6</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>S.W.</td>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>City</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>6</td>
<td>9</td>
<td>2</td>
</tr>
</tbody>
</table>

$x$ 2 refusals

+ Includes one school retaining forms I and II
The refusal of two contributing schools allied to the same intermediate school upset the ratio of intermediate to contributing schools but still left a sample of over one fourth of the total number of primary schools in the Christchurch Metropolitan Area. Thus the final sample consisted of six contributing schools, nine non-contributing schools, and two intermediate schools. One of the schools exercised the option of sending pupils to the intermediate school, or retaining them in Forms I or II, but for this survey was classified as a contributing school. A further analysis of the final sample in relation to classes taught is shown in Table II.

<table>
<thead>
<tr>
<th>Headmasters</th>
<th>Form II</th>
<th>Form I</th>
<th>Std. 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male 26</td>
<td>Male 15</td>
<td>Male 20</td>
<td></td>
<td>93</td>
</tr>
<tr>
<td>Female 6</td>
<td>Female 4</td>
<td>Female 6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

+ Includes 1 teaching headmaster
x Includes 1 teaching headmaster
Excluding headmasters, and the non-teaching first assistant, there were sixty male teachers and sixteen female teachers. Apart from intermediate schools, the senior mistresses were the only women seen, as common practice among the seventeen schools visited was to place men, particularly senior men, with the upper classes. The predominance of men in the sample is thus related to schools' practices of placement of male teachers. Some composite classes were seen, as shown in Table III following.

<table>
<thead>
<tr>
<th>Classes</th>
<th>No. of Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stds. 2-4</td>
<td>1</td>
</tr>
<tr>
<td>Stds. 3-4</td>
<td>7</td>
</tr>
<tr>
<td>Std.4-Form I</td>
<td>2</td>
</tr>
<tr>
<td>Std.4-Form II</td>
<td>1</td>
</tr>
<tr>
<td>Form I-Form II</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>17</strong></td>
</tr>
</tbody>
</table>
In the case of the standards two to four and three to four, work affecting the standard four part of the class only, was selected. For classification purposes in the remaining cases, the higher standard was usually taken. Whichever standard of the composite class had been selected, the results would have little effect on the survey, as these classes were normally treated as one group for method, planning, and other details affecting this investigation. If differing methods were used, the class was placed in the standard which had the greatest number of pupils.

Teachers and Classes

In September, October, November, and December of 1959 seventy-six teachers, one non-teaching first assistant, and sixteen headmasters, from seventeen different primary schools, were interviewed. Of the seventy-six teachers, one was a teaching headmaster. Table IV shows the totals of teachers, in relation to the schools seen.
**TABLE IV**

**ANALYSIS OF TEACHERS ACCORDING TO SCHOOLS**

<table>
<thead>
<tr>
<th>Type of School</th>
<th>No. of teachers (including Headmasters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contributing</td>
<td>Male 19</td>
</tr>
<tr>
<td></td>
<td>Female 3</td>
</tr>
<tr>
<td>Non-contributing</td>
<td>Male 33(^{+})</td>
</tr>
<tr>
<td></td>
<td>Female 8</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Male 24</td>
</tr>
<tr>
<td></td>
<td>Female 5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>92</strong></td>
</tr>
</tbody>
</table>

\(^{+}\) Includes 1 teaching headmaster who is classed as standard 6 teacher in later tables, and as a headmaster in others.
Table IV shows that the overall balance of numbers of fifty-one teachers and headmasters of intermediate and contributing schools against the non-contributing schools total of forty-two is not even, but some of this is accounted for as most headmasters requested that probationary assistants and relieving teachers be not interviewed. The sample then is confined to teachers established in the schools seen. The numbers were also upset by the two contributing schools who did not take part, the illness of some teachers, and non-availability of others. Thus from a possible one hundred and fourteen teachers in the seventeen schools, eighty-one per cent was interviewed.

Factors Possibly Affecting the Sample

In the sample there were only two single-sex classes, and as may be seen from Table V below, only three schools used intelligence tests as a part basis for streaming.
**TABLE V**

**BASES OF STREAMING BY SCHOOLS**

<table>
<thead>
<tr>
<th></th>
<th>I.Q. plus ability in class</th>
<th>Classroom ability only</th>
<th>Nil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate</td>
<td>2 schools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contributing</td>
<td>1 school</td>
<td>5 schools</td>
<td></td>
</tr>
<tr>
<td>Non-contributing</td>
<td></td>
<td>6 schools</td>
<td>3 schools</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>3 schools</td>
<td>6 schools</td>
<td>8 schools</td>
</tr>
</tbody>
</table>

**No. of classes involved**

|                | 29 | 24 | 23 |

It can be seen that the number of classes streamed on I.Q. plus ability, on scholastic performance in the classroom only, and those not streamed at all, are approximately even. This does help to preserve the balance of the sample as the types of tests set could be affected by the spread of ability in a class.

The work required in marking tests is liable to affect the frequency of this usage, particularly so when a class is large.

Table VI gives the distribution of roll numbers among the sample.
### TABLE VI
ROLL NUMBERS OF CLASSES

<table>
<thead>
<tr>
<th>Under 30</th>
<th>31-35</th>
<th>36-40</th>
<th>41-45</th>
<th>46 plus</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of classes</td>
<td>11</td>
<td>5</td>
<td>27</td>
<td>16</td>
</tr>
</tbody>
</table>

Thus the most frequent class roll was from thirty-six to forty children, but thirty-three teachers (or forty-three per cent) had classes in excess of the staffing ratio of forty children to one teacher. This could be a factor influencing the amount and type of testing done. An analysis of the full figures gave a mean of 39.2 and standard deviation of 7.3.

Further factors in the sample which could influence the results are the experience and qualifications of the teachers concerned.

It is possible, because of the more specialist organization of the intermediate school, that the more qualified teachers tend to be appointed there. If this was so, such a concentration could affect the normality of the sample as thirty-six per cent of the teachers seen were drawn from the two intermediate schools. Following from this, is the assumption that the courses covered at university could
influence a teacher's work, and hence his testing. It is a possibility not proven, but one which should not be overlooked. On the other hand, a not so academically qualified teacher may have other interests, experiences, and general knowledge, which could be a balance to academic qualifications.

Table VII sets out the relationship between the university qualifications of teachers and types of schools in which they are located.

**TABLE VII**

**RELATIONSHIP BETWEEN UNIVERSITY QUALIFICATIONS OF TEACHERS AND TYPES OF SCHOOLS**

<table>
<thead>
<tr>
<th></th>
<th>Under 6 Units</th>
<th>Over 6 Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate School</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>Non Intermediate School</td>
<td>35</td>
<td>16</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 4.258 \]

\[ \rho = < 0.05 \]

This shows that more qualified teachers, as measured by the above classification, are to be found in the intermediate schools of the sample. Reasons for this may include the more specialized teaching done in the intermediate schools, and the selection of teachers (as their positions are advertised as "Special") thus overruling to some extent
appointment by grading alone). There seems to be considerable competition for these positions.

The sample could also be affected by the number of teachers qualified to administer standardized tests. Superficially, it could be expected that there should be more standardized tests where the persons qualified to do it are located.

If, therefore, the sample has a concentration of qualified testers in any one type of school, the results could be affected. Table VIII sets out the location of testers.

**TABLE VIII**

**TESTING QUALIFICATIONS OF TEACHERS**

<table>
<thead>
<tr>
<th></th>
<th>Qualified</th>
<th>Unqualified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate Schools</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>Non Intermediate Schools</td>
<td>12</td>
<td>39</td>
</tr>
</tbody>
</table>

\[
\chi^2 = 2.208
\]

\[ p \geq .05 \]

So far as the qualification for testing with standardized tests is concerned, there is no concentration other than normal in intermediate schools.
The investigation results could also be affected from normal, if the more experienced teachers were grouped in one type of school. Younger teachers may rely more on tests, where experienced teachers have a background which serves instead. Table IX shows the teaching experience of the sample, in relation to school groupings.

### TABLE IX

**Teaching Experience of Teachers**

<table>
<thead>
<tr>
<th></th>
<th>Under 20 years</th>
<th>Over 20 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>Non Intermediate</td>
<td>28</td>
<td>23</td>
</tr>
</tbody>
</table>

= 4.544

= .05

It is evident that younger teachers are more frequently present in the intermediate schools. It must be borne in mind however that the younger teachers in non intermediate schools tend to be given lower classes than standard four.

From Tables VII and IX it is evident, that for this sample, the intermediate school teachers were younger, and possessed more academic qualifications, than those of non-contributing schools. As they could not rely so much on experience, they might test more, and having an academic
background might be more willing to try tests for themselves. This is offset, however, by their having test results supplied, as will be seen in Chapters IV and V.

The general experience of teachers over the whole sample would be relevant, when the results of the survey are considered. New ideas and experiment could be more likely from younger teachers. Table X sets out an analysis of certificated service of the sample.

**TABLE X**

**ANALYSIS OF CERTIFICATED SERVICE**

<table>
<thead>
<tr>
<th>Years of Teaching</th>
<th>Under 9</th>
<th>10-19</th>
<th>20-29</th>
<th>30 years plus</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Totals</strong></td>
<td>23</td>
<td>25</td>
<td>17</td>
<td>11</td>
</tr>
</tbody>
</table>

\[ M = 14.1 \]
\[ sd = 8.7 \]

Thus sixty-six per cent of teachers are in the group having from five and one half years to twenty-three years experience. The majority then have not completed half their anticipated service of forty years, and could be expected to still be trying new methods, and receptive to ideas and experiments. This survey then consists of mainly younger teachers.
All the information for this investigation was obtained by personal interview with the teachers and headmasters in their own schools. Each interview took from thirty to sixty minutes. Interviews were arranged in advance with the teachers concerned.

Several attempts were made to design a satisfactory interview form to be filled in by the investigator, so that the one finally used would cope with the wide variation of answers expected. In order to keep the interviews somewhat standardized, a question framework was developed. A trial was made on four teachers who were not included in the final investigation.

Head teachers of the schools concerned were interviewed to find out their school policy in regard to examinations and testing in general, and these results were checked against the information supplied by the class teacher. Naturally both headmasters and class teachers were assured of the confidential nature of any information that was given to the investigator.

SUMMARY

The sample of teachers from which the information was gained, was drawn from intermediate and related contributing schools on one hand, and an approximately equal number of teachers from non-contributing schools on the other. Although socio-economic factors were not likely to have affected the sample, as it concerned teachers and not children,
allowance was made for this possibility. In its final form, the sample covered over one quarter of the schools in the Christchurch Metropolitan Area. The teachers were predominantly male, probably due to many schools' practices of placing men teachers in upper classes. It was considered that more standardized testing would be carried out in the upper levels of the primary school, than in the lower ones, as most test norms started from about age nine years plus.

Factors which could affect testing, and thus the results of this survey, included the spread of ability in the classes, but an approximately equal distribution of classes graded on intelligence, classroom ability, or not graded at all, was found.

It was shown that the teachers at intermediate schools were younger and had better academic qualifications than did those at non-contributing schools; however, there was no greater concentration in either type of teacher with testing qualifications. Furthermore it might have been thought that the intermediate teachers with their higher educational qualifications would be willing to experiment more with standardized tests, but in fact they had much of their testing done for them.

Finally all the information was obtained by personal interview in the teachers' own schools.
CHAPTER IV

INTELLIGENCE TESTING

The intelligence test most widely used in the sample, was, as expected, one of the forms of the Otis Intelligence Tests, but there were also a variety of others. In two different schools, the headmaster and one teacher in each, made use of the Raven Progressive Matrices Test as well as the Otis for survey purposes. One of the teachers, although working in a school seen, was outside the limits of the sample. Three teachers (two in one of the above schools) used the A.C.E.R. Junior Non-Verbal Test in addition. These tests were used with the Otis, to attempt a balance between verbal and non-verbal performance. One headmaster (who was qualified) used the Revised Stanford Binet Intelligence Scale for special cases. Other tests were administered as part of practical work for university courses. These were the Wechsler Intelligence Scale for Children, and the Primary Mental Abilities Test, each used by one teacher.

Excluding the two teachers who used tests for university course training, ten headmasters and nine class teachers were involved in intelligence testing. Of these nineteen, only two had less than fourteen years teaching experience.

Of the 93 headmasters and teachers seen, 18 used one of the forms of the Otis Intelligence Test during the year 1959; this is nineteen per cent of the sample. Six of the eighteen users were headmasters, who, in their schools, frequently used the Otis Test and encouraged other teachers
on their staff to do likewise. In addition to using the tests themselves in some cases, they trained a team of testers where this was required.

Apart from these regular users there were a number of the sample who used Otis Tests infrequently, usually for special cases such as a suspected dull child. In this category are three headmasters and five class teachers. Thus eleven per cent of the sample were regular users, nine per cent occasionally apply them, and eighty-one per cent do no intelligence testing.

The eighteen headmasters and teachers using the Otis Test gave reasons as shown in Table XI for their use of it.

TABLE XI

REASONS FOR USE OF OTIS TEST

<table>
<thead>
<tr>
<th>No. of Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Streaming and grouping of classes (4 schools)</td>
</tr>
<tr>
<td>2. As a check on school performance</td>
</tr>
<tr>
<td>3. To obtain A.Q. and for own information</td>
</tr>
<tr>
<td>4. Grouping within classes</td>
</tr>
<tr>
<td>5. Complementing other tests</td>
</tr>
<tr>
<td>6. Miscellaneous</td>
</tr>
</tbody>
</table>
Apart from streaming and complementary testing, some of the ten regular users obtained achievement quotients and used the individual results of the tests in other specialised ways. All the normal uses for which intelligence tests are used, seem to be present in the above reasons.

In a class peculiar to themselves involving four schools, are seven teachers and four headmasters, who tested quite widely for the benefit of other schools or persons. In one such school the headmaster and five teachers applied a full battery of California Tests on behalf of a Fulbright Scholar, and at the time of this investigation had not received the results of these, but could reasonably expect to do so later. These results would not be of any use of the teacher of Form II as the end of the year was near. Two other schools tested children for intermediate school and one tested for a high school, a few weeks prior to the children leaving their home schools.

The small numbers of teachers using intelligence tests, did however, provide results for forty-three other teachers, who were then relieved of the necessity of testing for themselves. An analysis of reasons advanced for not using intelligence tests is set out in Table XII.
## TABLE XII

**REASONS FOR NON-USE OF INTELLIGENCE TESTS**

<table>
<thead>
<tr>
<th>Reason</th>
<th>No. of Teachers (including H.M.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results available from other sources (including 6 testing for other agencies)</td>
<td>49</td>
</tr>
<tr>
<td>Discouraging attitude of headmasters in three schools.</td>
<td>11</td>
</tr>
<tr>
<td>Equivalent results available from progress cards plus teacher's experience.</td>
<td>5</td>
</tr>
<tr>
<td>Time needed</td>
<td>2</td>
</tr>
<tr>
<td>Norms too old for many Std. 4 children.</td>
<td>2</td>
</tr>
<tr>
<td>Lack of training in testing.</td>
<td>8</td>
</tr>
<tr>
<td>Not long enough in present class (normally tested)</td>
<td>2</td>
</tr>
</tbody>
</table>
Some teachers advanced more than one of the above reasons.

How valid are these objections? The cost of the tests is a very real reason, as some school committees may object to the spending of their money on this material, so the tests (whatever type) must have some practical advantages when used. A headmaster who does not favour testing can also put many obstacles in the way of a teacher who wishes to use them. The three headmasters and two class teachers who combined their experience and the progress card results, to assess a child's ability saw no need for using intelligence tests. This idea has some merit. Vernon\(^1\) in discussing the prediction of primary school children's success at high school allows that "A good teacher should have a fuller knowledge of children's persistence, intellectual promise and other qualities relevant to successful secondary school work, than can be given by a limited battery of tests or purely written examinations." If teachers' judgements are reliable in this matter, it must be conceded that they could group classes on an ability basis from this data.

There is though, the serious difficulty which Vernon later points out that however accurate the teachers'

---

assessments of the relative merits of children within his class, his judgements of standards may be highly inaccurate. After all some value of intelligence testing is in its predictive nature thus giving the teacher an idea of the work of which the child is capable. Peel and Butler\(^2\) findings to quote only one of the many experiments in this field, found that the best single predictor of school certificate passes in England, was the intelligence test. Teachers' predictions of success against objective criteria are not good as McClelland (1942) shows. The intelligence test then is a better predictor of academic success measured against objective criteria, such as success in high school, than are teachers ratings. It follows then that the teacher grouped class may be classified according to the childrens' apparent working ability but not necessarily grouped according to their capacity.

Further, many children of high ability go unnoticed under the teacher rating system. These children offer work at an acceptable standard and are often not noticed. In a class it is often the neat conforming child with accurate and ready answers, plus perhaps a bright personality, who is likely to be assessed as bright, while a persevering

\(^2\)Ibid, Vernon Secondary School Selection, P.73.
child of low ability may appear to have more ability than is the actual case.

A local survey made in Christchurch in 1960 by the Psychological Service of the Department of Education\(^3\), in which the writer took part, invited eighteen schools to submit names of children considered to have an I.Q. of 130 plus. Some of these children had already been assessed on the Otis Test by the schools. The eighteen schools included six already seen in the writer's investigation. Of the 210 children tested on the W.I.S.C. Verbal scale, only seventy-seven or thirty-seven per cent gained I.Q.'s of 130 or better. Teachers ratings in these cases were far from accurate.

The time taken to administer tests is not however a valid objection. One hour of school time is sufficient to administer most group tests and most marking keys are quick and easy to use. Furthermore, although it is conceded that the Otis Test has its disadvantages of assessing intelligence on a limited number of questions at the age of nine years, it can be a guide for even a part of a standard four class who have attained this age.

\(^3\)Dept. of Education - Psychological Service, Christchurch; Results of Gifted Child Survey - August 1960, mimeographed.
Eight persons would have tested had they not felt their lack of training, but there are many more among those who had results supplied to them who are in a similar predicament. It does show, however, that some teachers are aware of the difficulties and dangers of tests being used by untrained personnel.

Of the eighteen administering intelligence tests, only six of the nine headmasters had studied mental testing in University courses, such as Education II or Psychology II. The remaining three had either studied the manual or talked to others about it. Of the nine class teachers using the Otis Tests, five had University qualifications and three had personal tuition from other sources. Only four used the tests regularly.

There were in the sample, twenty two teachers excluding headmasters, who had passed University courses involving intelligence testing. Eighteen of these twenty two teachers were male; the four female teachers so qualified did not test at all. Thus seventeen teachers were leaving idle, skills which they could usefully have applied to their school and class. School organisation, headmasters' policies, lack of test equipment, and little time available, contributed to this. In at least one school, academically unqualified teachers were testing the complete school for streaming purposes, when teachers with degrees in education in that
school were not used at all for this purpose.

Some teachers did not realize the dangers of using the results of one group verbal test as a means of estimating intelligence, and apart from other factors, the poor reader is automatically hindered. Again many of the teachers reduced I.Q.s to a single figure and some did not understand the dangers of this. On the other hand the Otis Test has certain advantages: it is easy to administer and mark; the wording has been altered to suit New Zealand speech; New Zealand norms are available.

Intelligence tests are the basis for obtaining educational and achievement quotients. The few teachers who used this were qualified to do so, although the concept of achievement quotients has disadvantages. If for example, the child exceeds his supposed potential of learning at school as measured by an intelligence test, and gains an achievement score giving an A.Q. above one hundred, this means that he has on paper achieved more than he is capable of achieving. Anastasi maintains the A.Q. technique is practically extinct, as the fallacy underlying it is that

---


the pupil's "innate capacity" is represented by his mental age, and that a single unitary capacity accounts for performance in all fields. She suggests that the most satisfactory norms for the evaluation of achievement test performances are the "percentile-within-grade" norms, which indicate the individual's own position in his own standard level.

In their latest text Remmers, Gage and Rummel\(^7\) give samples of minimum and optimum testing programmes for schools. They state that in an elementary school, or junior high school, which approximates the New Zealand intermediate school, the minimum testing programme should include at least one measure of intelligence at some time, while an optimum plan would include intelligence tests about once a year. This minimum criterion for testing was met by both the intermediate schools and two others, where children were assessed for intelligence once during their time at these schools. In the intermediate schools these tests were given during the final weeks of the year before entry, so that the results would be available at the beginning of the school year in which the new pupils entered the intermediate school. As previously mentioned this testing was done either by the contributing school concerned or selected teachers from the intermediate schools. In two

---

cases the testing teams contained members trained by their headmasters only. The final selections of the class groupings however were undertaken by the headmasters, who both had training at University in this.

The two other schools normally surveyed the children at the beginning of the standard three year. Those children who were not old enough for the norms, were often tested later, or allowances made for their youth. These allowances were clearly marked by the use of "I.Q. plus" and of course, were treated with caution.

These four schools, had the maximum intelligence testing programme seen in the sample, but this is in fact the minimum mentioned by Remmers.

There are reasons for this, the most important being the lack of teachers trained in testing. Little instruction in testing appears to be given during teachers' training, although for simpler group tests such as Otis, only a limited background is required to give confidence, and a knowledge of the limitations of the tests and testers. Considerably more training, normally of a University Stage II type, is needed for giving individual tests.

Furthermore there is not the tradition of testing in New Zealand schools that is found under some other overseas education systems.

It would seem then that from the very limited number of teachers using intelligence tests, and the somewhat vague ideas
on their uses and limitations on the part of others, more might be done while teachers are being trained. Theoretical knowledge on the principles of testing and interpretation, practice in administration, use and interpretation of the simpler types of intelligence tests could be given.

Greater use could be made of trained teachers already in the schools, for testing, to spread the load already borne by a few, some of whom are not trained. As only four schools measure up to the minimum criteria of a testing programme, the use of these qualified teachers at present making little use of their training, could extend such a minimum programme to more schools, and perhaps build up the existing ones.

The results obtained from the widespread use of the Otis test could be weighted by practice effects and coaching. Vernon's experiments\(^8\) demonstrated gains in these respects, from four to five I.Q. points under various circumstances; thus the need for alternative tests suited to New Zealand conditions, as Rogers\(^9\) show, is important. Roger's standardisation of the S.R.A. Primary Mental Abilities Test fills such a need, although the initial cost is rather high for many schools.

---


The S.R.A. Primary Mental Abilities Test is criticized by Kurtz\textsuperscript{10}, on grounds of its reliability which he says is low at 11 - 17 years, and probably very low at the younger ages. It correlates well with achievement test scores and high school grades, but poorly with vocational training ratings and college grades.

Anastasi\textsuperscript{11} also criticizes it on similar grounds, plus what she considers is an undue emphasis on speed, and the high intercorrelations of the tests at lower age levels, which raise a real question on the feasibility of interpreting P.M.A. profiles for younger subjects.

Kurtz sums up by stating that at ages 7-11 almost any reliable intelligence test with a large verbal component should do as well as the P.M.A. in predicting school achievement.

The New Zealand P.M.A. standardisation does however, provide another useful and much needed tool for teachers.

Where teachers were merely testing agents for other schools, it seems a better idea to test those children likely to attend the intermediate or high school concerned, at the beginning of their final year. The results should not differ from those obtained at the end of the year if allowance

\textsuperscript{10}The Fifth Mental Measurements Year Book - Highland Park, New Jersey, - The Gryphon Press 1959 - P. 717, Criticism by Albert J. Kurtz.

is made for the standard error of measurement, and would be available for use by the teachers for at least one year.

Achievement Quotients are still in use in New Zealand, even though in some cases, of doubtful value. It would be useful, as suggested by Anastasi, if more emphasis was placed on the "percentile - within-grades" norms. Few of these are available however, for New Zealand conditions. Providing the weaknesses of the A.Q. are understood, they do seem to have a place in testing.

So far as prediction and intelligence tests are concerned, Vernon\textsuperscript{12} thinks that our present group tests particularly, are too generally constructed. He postulates the idea of selecting or constructing specific tests which would give the best predictions of whatever kinds of ability we are interested in, such as intelligence for grammar work. Until these are available however, the best single prognostication of academic success, seems to be the intelligence test as we have it. An even better one is this combined with experienced teachers' estimates plus results entered on the cumulative record cards.

As Burt\textsuperscript{13} aptly sums up however, "Tests can only be the beginning, never the end, of the examination of the child".

\textsuperscript{12} Philip E. Vernon - \textit{op cit}, P. 100.

SUMMARY

As could be expected, the Otis Intermediate Intelligence Test was by far the most widely used test, but was administered by only nineteen per cent of the sample. Eleven of the eighteen users had university qualifications for testing. The few teachers testing for intelligence however, made this information available to fifty-seven per cent of the teachers, but there is the danger of unwise use of this data by persons unfamiliar with its significance.

There were however, seventeen other teachers qualified who did not use intelligence tests, and in some cases they were in schools where testing programmes were administered by teachers with lesser qualifications.

This showed a wastage which could be overcome. Intelligence testing was not applied widely, and much more use could be made of it as useful teaching aid, providing its merits and limits are understood.

The main uses to which it was applied, were for grouping and checking scholastic performance. But there were some who were actively opposed to intelligence testing and maintained that an equally good estimate could be gained from cumulative records. Although this idea has merit it fails when an objective standard, as opposed to a class one, is required.

Attention was drawn to the number of schools who tested for other agencies, such as intermediate and high schools.
Had the intelligence tests administered been advanced by nine months, the results would have been available to teachers with the consequent benefit to the children.

The Otis test has been for many years, the only group verbal test with New Zealand norms, available to teachers. The need for an alternate test standardised for this country is now met by the S.R.A. P.M.A. test. Neither it, however, nor the New Zealand standardised A.C.E.R. Junior Non Verbal Test, which has been available for many years, were much used among the teachers seen.

As there is not the tradition of testing in this country that there is in others, this lack of tests was not particularly noticed by most of the teachers. Thus it was not surprising that minimum testing programmes set out by overseas writers, were met by four schools only.
CHAPTER V
ATTAINMENT AND DIAGNOSTIC TESTING

During the year 1959, the teachers and headmasters seen, used thirteen different standardized tests for varied reasons. Remmers, Gage, and Rummel1 state that standardized tests usually provide comprehensive measures of basic knowledge, skills, and abilities widely recognized as important. Conversely the tests may not fit closely the specific instructional objectives of a particular classroom, school, or school system. In basic subjects, such as those examined in this study, the differences among teachers, classrooms, and school systems may not be great, but the users of tests should make certain that the test content is in agreement with instructional objectives. The proper use of standardized tests depends upon selecting those that fit the instructional objectives of a particular school.

Apart from a few exceptions, the general planning of testing in the schools appeared haphazard. Traxler2 suggests that an adequate test plan should meet the following five criteria:

---


2 Traxler and Others - Introduction to Testing and the Use of Test Results in Public Schools (New York - Harper and Brothers 1953) P. 14.
(1) The tests should be selected and administered for specific purposes which are stated in advance.

(2) The programme should be undertaken co-operatively by all the teachers.

(3) A comprehensive list of the procedures involved in carrying out the programme, must be included in the overall plan.

(4) The programme should be practical and definite.

(5) The programme should be continuous and long range in scope.

These standards make a convenient measure against which to assess the testing that was encountered.

An analysis of the survey shows that many tests are in current use. These are shown in Table XIII.
### TABLE XIII
STANDARD TESTS IN CURRENT USE

#### ARITHMETIC
1. Schonell  
   Diagnostic Arithmetic Tests (1936).
2. A.C.E.R.  
   Arithmetic Test (in the four fundamentals only). Forms C or D (1946).
3. Seville  
   Easy Steps in Arithmetic (1952).
4. Brueckner  
   Diagnostic Tests in Fractions (1926).
5. Burt  
   Arithmetic Tests (Four Fundamental Rules) (1921).

#### SPELLING
1. Schonell  
   Graded Spelling Tests - Regular and Irregular Words (1938).
2. Burt  
   Graded (Spelling) Vocabulary Test (1921).
3. A.C.E.R.  
   Spelling Tests (1936).

#### COMPREHENSION
1. A.C.E.R.  
   Reading Tests Part 3 Reading for Meaning.
2. Watts F.A.  
   Metropolitan Reading Scale (1942).

#### READING
1. Burt  
   Graded (Reading) Vocabulary Test (1921).
2. Schonell  
   Graded (Reading) Vocabulary Test (1945).
3. N.Z.C.E.R. Fieldhouse  
   Oral Word Reading Test (1952).
Just under half of the above tests were published before 1940, but their use does not appear to be related to teaching experience, as of the ten class teachers using the Schonell Arithmetic Test, exactly half were above the mean teaching experience of the sample. A similar result was gained for all the Burt tests, where, of the thirteen class teachers using these, seven were above the teaching experience mean. Four of the teachers were in two schools, and the remainder scattered. It was not, therefore, a school policy to use the older attainment tests. Schonell's Spelling Tests were not thus analysed, as their widespread use was closely related to the spelling list then in vogue in most of the schools seen. These older attainment tests are dated by their norms, which in Burt's tests are almost forty years old. The "age" concept of the norms are, however, appealing to many teachers, although the danger of their being based on the English school system and not our own, may not be appreciated by some users.

Diagnostic tests do not suffer to the same extent from age. Two of the diagnostic tests (both in arithmetic) in use are old, but the processes do not alter; thus the diagnostic parts of the Schonell Test and Bruckner's Test are still useful. Hamley, in discussing the contents of a good

diagnostic arithmetic test holds that once such items as counting by enumeration, counting without objects, recognition of numbers in groups and so on, have been learned, the pupil's mastery of various number combinations in the four processes, for all computation and problem work depends largely on efficiency in this direction. Then must be found how far he has progressed in applying these in mechanical examples in each of the four processes.

**ARITHMETIC TESTS**

**Schonell's Diagnostic Arithmetic Tests**

These were published about 1936 and fill the above requirements. It was the only test in use that gave a systematic survey of basic number combinations in the four processes. In fact it was only this part of the test that was used by any of the teachers. Schonell aimed at providing a means of both systematic survey and accurate diagnosis in arithmetic. It is an attainment and diagnostic test combined, using arithmetic age norms. After the pages on basic combinations, there follow tests for age groups seven to fourteen years, of various types of examples in the four processes.

Of the sixteen users in the sample, thirteen teachers in nine schools used it as an attainment test and three of these

---

*Fred J. Schonell - Diagnosis of Individual Difficulties in Arithmetic Second Edition (Edinburgh - Oliver and Boyd Edinburgh 1942) P. VI*
later retested, with the same test, to judge the children's progress. Three teachers used it purely as a diagnostic test and five used it as a combination attainment and diagnostic test. Just over half the teachers using this test, that is a total of nine, used the results for classifying and grouping children in arithmetic, while of the eight using it for diagnostic purposes, only one took any remedial action based on the results. Five teachers used it for their own information, that is, they seemed to make no use of the results. The figures are noticeable for the omissions in that no headmasters, or female teachers in forms I and II, used this test, although it is easily obtained and widely known. This arithmetic test was one most frequently used, probably because of its availability and reputation. Most schools seem appear to have copies of Schonell's books on attainment and diagnostic testing available. Many times the test is produced on a duplicating machine and the result given as a direct arithmetic age and not as a range.

*Easy Steps in Arithmetic*

Seville's diagnostic tests published in 1952 are contained in "Easy Steps in Arithmetic". They provide pages of examples in each process of addition, subtraction, multiplication, and division for numbers and a separate book.
for money, for diagnostic purposes.

Although Seville's book was supplied to all the schools seen by the Department of Education, only six teachers used it. Here again this was confined to four schools and was misused by two teachers as an attainment test, when it was designed purely as a diagnostic one. Seville makes it quite clear in the instructions that this test is intended for diagnosis only. It is not meant as a teaching test, but as a guide to helping poor workers. One of the teachers using it as an attainment test also used it as a diagnostic test, while a further two teachers applied its results for grouping. The misuse of this test is perhaps an indication of lack of training in the uses of standardized tests, and also of a rather vague idea of their application. One teacher used it as a retest at a later time in the year to judge progress, one used it for grouping, while only two out of the four who used it for diagnostic purposes, did any remedial work with it at all. The other two simply tested and recorded the results. In one case the headmaster supplied Seville to a class teacher, but no use appeared to have been made of it

---

6Ibid P. 5
at the time of the survey. Only male teachers used this test, predominantly in standard four. In some schools it had not been heard of although, as previously mentioned, it had been supplied. In others it was regarded as just another arithmetic text book.

The A.C.E.R. Arithmetic Tests

These were published in 1946 and are provided in two parallel forms, designed to measure attainment in six subdivisions of arithmetic, and to a limited extent to diagnose particular difficulties within these subdivisions. The total number of possible combinations available have been reduced to a representative "pool" of questions in which combinations are evenly distributed and adequately covered.7

In the five schools which used it, only the first four subdivisions covering the four rules, were administered. Each subdivision is set out on a separate page.

There were eight users of this test, two of whom were headmasters, one of an intermediate and the other of a contributing school. The former used it in the initial grouping of classes. The latter grouped within classes with it, and also used it as a check on teachers' work by treating it as

---

an objective test so that class progress could be measured from teacher to teacher. This idea had its merits in that the test is an objective one and standardized as such, providing the testing took place at a time when the classes had covered all the work included in the test. Due regard would also have to be given to the pupils' intelligence, which was not always assessed in this school. Anastasi's ideas previously mentioned in Chapter IV could be applicable here. Coaching factors could also influence the result as the same test is used yearly.

In seven cases the test was used for its primary purpose, attainment, and five of these seven teachers grouped within classes on its results. One teacher applied it for diagnostic purposes and supplied remedial work on the basis of errors found. The remaining two users apparently did little with their findings.

Its use was evenly spread among all the male teachers only. It was included among the battery of screening tests given by one intermediate school, and the other had made use of it in times past, for similar purposes.

It is perhaps unfortunate that it is not more widely known and used, as New Zealand norms are available for it, and it is convenient to use.
The Brueckner Diagnostic Test in Fractions

This is an analysis of the different types of examples that are found in working with the fundamental processes in fractions. It contains types of the four processes and allows for a detailed analysis of difficulties in each process. The only teacher who employed it did follow his findings with corrective actions based on the results.

Burt - Arithmetic Tests

The remaining test of arithmetic used was Burt's Fundamental Rules (speed) Test, published about 1921. A uniform level of difficulty was aimed at throughout the four processes and a coverage in all different processes in each section attempted. Again one teacher only used it, in its correct sense, as an attainment test to check that of Schonell's.

In this capacity it is useful, but the remarks made previously on age and English norms are applicable here also.

SPELLING TESTS

Burt states that for general purposes the most

---

8 J. Brueckner, H. W. Distad, Abbie Chestek - Brueckner Diagnostic Test in Fractions Manual of Directions - Revised (Minneapolis Educational Test Bureau, Educational Publishers Inc. 1943) P. 3.

9 Cyril Burt - Mental and Scholastic Tests (Westminster (England) - P.S. King and Staples Ltd, Westminster 1939) P. 301.

10 Ibid P. 287.
serviceable test of spelling is supplied by a graded vocabulary. The point then arises whether this vocabulary should be based on the basic spelling list or on a general list of words likely to be encountered in everyday work. The selection of these everyday words could be affected by the pupil's environment, and school system.

**Schonell Graded Spelling Test**

Of the three spelling tests used, Schonell's tests of regular and irregular words were by far the most frequently used. These are intended primarily for use in obtaining a comparative estimate of the child's ability to spell phonetic and non-phonetic words. Each list contains sixty words of graded difficulty and English "age" norms are available for each. The regular word list may also be used diagnostically.

The twenty-four users in thirteen schools included all classes except those of standard six female teachers. All employed it as an attainment test, and no diagnostic work was attempted, possibly because Schonell's instructions on its use for this are rather too general in nature. Of the total users, three retested later as a check on progress. Fifteen used it as a basis for grouping their classes for spelling and of these fifteen, two used it as a check on the A.C.E.R.

---

Spelling Test. This, of course, is a wise precaution and showed that they realized the pitfalls attached to screening on one test only. No particular use seemed to be made of the results by six teachers.

These high figures for users, approximately 31.5 per cent of the sample of class teachers, are to be expected as, at the time this survey was made, the spelling list in use in schools was overwhelmingly "Schonell's Essential Spelling List."

The handbook supplied as a teacher's guide for this list contains these tests, and could be conveniently administered at the beginning or any part of the year to obtain some idea of a child's spelling ability, particularly if the teacher wished to form his class into spelling groups, as did thirty-nine teachers in the sample. There was a tendency to attach too much weight to spelling ages obtained from this test, particularly where a reasonably accurate intelligence quotient of the child was not known. This is a danger with any form of standardized test, but in the case of the Schonell Spelling, which is widely used, the danger is greater, as a child may be expected to work at a level above his capacity, or, conversely, below it. The tests are a useful guide if kept in perspective.

Burt Graded (Spelling) Vocabulary Test

Burt's spelling list was encountered on ten occasions in six schools. It is a graded vocabulary of increasing
difficulty, classified on an age basis. All the teachers concerned used it correctly as an attainment test. Of the ten users, eight grouped with it (including one who retested for progress) and two made little use of the results. The Burt Test is a considerably older (1921) test than the Schonell equivalent, and was not so closely related at the time of this survey to the current spelling list.

**A.C.E.R. Spelling Tests**

The spelling tests produced by the Australian Council for Educational Research in 1936 are designed to test the ability of pupils to spell when tested under strictly standardized conditions.

The word to be spelled is given, repeated in a standardized sentence, and said again.

Schonell suggests the same system, but does not put the words into sentences. Thus in the A.C.E.R. test, every child has the word in the same sentence, making for a more standardized delivery of the test. Norms are provided for various Australian States. These tests were used by nine teachers in six schools, for its designed purpose, attainment. Although no one stated it was used for diagnostic purposes, three of the teachers took remedial action on the incorrectly

---

12 Cyril Burt - *Mental and Scholastic Tests* op cit P. 287.

spelled words. This misuse, for teaching purposes, detracts from its value as an objective standardized test and is commented on in the manual. Four of the nine teachers used it for grouping and classification purposes, while two used it as a check on Schonell Spelling Tests. Again, two headmasters used this test in conjunction with their A.C.T.R. battery of tests.

READING

According to Hamley\(^{14}\) again, diagnostic tests in reading depend on a clear conception of the factors involved in this process, namely:

(a) Perceptual ability - recognition visual and auditory.
(b) Analytic ability - visual and auditory.
(c) Synthetic ability - visual and auditory.
(d) Articulatory ability 'in so far as it is linked with auditory units).
(e) Associative ability - linking symbols and signs with meaning.
(f) General language background (influencing comprehension).

Hamley goes on to say that the most difficult diagnostic tests are those for detecting weaknesses in word recognition, as speed of reading and power of comprehension do not exist without

\(^{14}\)H.R. Hamley - Ibid P. 86.
some degree of facility in word recognition.

Thus the two comprehension tests used are, according to
the above author, useless unless there is some facility in
word recognition. On the other hand, recognition does not
imply comprehension.

A.C.E.R. Reading Tests

The tests were published in 1946. Part III, Reading
for Meaning, was used in five schools by six teachers and
two headmasters. It is the last test in a series of three
related reading attainment tests. It consists of paragraphs
of varied content and increasing difficulty, on each of which,
two questions testing comprehension are asked. It has a
time limit of twenty minutes. The two headmasters who have
been using the A.C.E.R. series of tests used this one, and
in all cases the people concerned used it correctly as an
attainment test. In four cases, the results were used for
grouping combined with other tests, as part of a reading
programme. In one case little use was made of it at all,
and one headmaster used it again later in the year as a retest
for progress. The New Zealand norms were not used.

Metropolitan Reading Scale

The other comprehension test is the second part of

---

15A.C.E.R. Reading Tests, Form C. Manual - Description of Tests
"The Metropolitan Reading Scale for Infants and Juniors", published about 1942. It is an attainment test to gain a comprehension age from questions set opposite the sentences of the model scale. There are multiple choice answers. Its use was confined to six teachers in two schools, all of whom used it for attainment purposes, and together with other tests, for grouping in reading. In one case, diagnosis was attempted but met with little success, as this test is not designed for it.

Schonell Graded Vocabulary Test

The Graded Reading Vocabulary Test devised by Schonell in 1945, consists of a scientifically selected sample of words of increasing difficulty, that will (according to Schonell) give an accurate estimate of a pupil's power of word recognition, which the author says is the very basis of his ability to read and understand any printed material. The test words increase by a known amount of difficulty from one word to the next. It is an individual, oral, untimed test of attainment, which with care can be used for some diagnostic purposes, providing, as Harley points out, the basic reading skills are understood. The reading ages are gained direct from the number of words correct, but again

16 F. A. Watts - The Metropolitan Reading Scale for Infants and Juniors (From Duncan - The Education of the Ordinary Child - 1942 - Nelson) mimeo - Psychological Service, Department of Education, Christchurch.

17 Fred J. Schonell - The Psychology and Teaching of Reading (Edinburgh - Oliver and Boyd Ltd, 1949) p. 32.
are based on an English sample.

This reading test was the most widely used of any standardized test in the sample. Twenty-nine teachers (or 38 per cent) used it in fifteen different schools, and of these, twenty-eight applied it for attainment purposes while two used it for diagnosis. Diagnostic use presupposes that the teachers concerned were most careful in their recording of results, and understood the fundamentals of reading as mentioned previously, so that they could recognise the errors made by the pupils, and group these errors for remedial action. Four of the twenty-nine teachers retested later in the year to measure improvement. Among the twenty-three teachers using it as a basis for grouping, seven combined one of the comprehension tests already enumerated, to offset the purely mechanical nature of this test. Four teachers made no use of the test other than to record the marks, while in two cases teachers used the Graded Vocabulary as part of their reading material. It is the only test of which women teachers of all classes were among the users.

The popularity of the Schonell Graded Vocabulary is probably due to its being one of the best known standardized tests among the primary school teachers. It is readily available in most schools and although rather time consuming if properly administered, does give a reasonable basis for grouping children in reading.
Burt Graded Vocabulary Test

Burt's Graded Reading Vocabulary, published about 1921, is a parallel test to Schonell's. The words are arranged in average order of increasing difficulty. Regular and irregular words alike were included, plus words outside the normal scope of children at the different age levels. It is an oral, individual attainment test, which by a trained tester can be used for diagnosis. Except for word content, it is virtually the same as Schonell's and one suspects that Schonell bases his own test closely upon this earlier one. It also is useful for grouping, but was used with English norms.

Altogether only three teachers, all female, used this test in three different schools, as an attainment test, two grouping their readers upon it. One made no further use of it after obtaining the results.

N.Z.C.E.R, Oral Word Reading Test

Fieldhouse's Oral Word Reading Test, published in 1952 by the New Zealand Council for Educational Research, represents the only New Zealand word recognition test available. It has been designed to determine the level of a child's attainment in word recognition, and also provides

\[18\] Cyril Burt - op cit P. 271.
a means of comparing his attainment with that of other children throughout the age range of the test. In administration it is the same as Surt's and Schonell's tests. Of the whole sample, it was only used by one teacher, who grouped his readers on its results.

**COMPARISONS OF TESTS**

Reynolds' criteria of importance of testing is met by some of the sample in the basic subjects, but the applicability of some tests is doubtful, so the tests used will be compared in their respective groups.

In comparison with Schonell's tests in arithmetic, Seville assumes a knowledge of fundamental combinations, and parallels the later parts of Schonell's test, except that no time limits are imposed. For diagnostic purposes, unlimited time is essential; to some extent, the lack of this feature is one of the disadvantages of the Schonell test. A time limit is however an essential part of an attainment test. Schonell was faced therefore, with its necessity if the test was to include attainment. He does, however, attempt to overcome this by setting a limit on part only of each section of the test. Nevertheless it is a drawback so far as the diagnostic part of his test is concerned.

---


20 Edward W. Seville - *op cit* p. 4.
Both diagnostic tests set out examples in the different processes of addition, subtraction, multiplication and division, and their format follow the step by step system of graded difficulty. Schonell provides four examples in each type, while Seville gives two. On the other hand, Seville's work has more steps, and is set out in greater detail. Both suggested remedial work for faults, but Seville provides four extra specific examples paralleling each test problem, thus saving the teacher's time considerably. The Seville test was designed for the New Zealand syllabus in arithmetic and of the three is thus the most applicable, and the best of these tests, providing it is used according to instructions. It is detailed, covers types fully, fits the syllabus, provides examples, and is easily interpreted.

Of the three spelling tests encountered, the most useful was Schonell's. Burt's test is outdated so far as its English norms are concerned. None of the sample used Winterbourn's Christchurch norms. As with the A.C.T.R. the Burt list does not offer any diagnostic provisions. Neither of these two tests were related to the spelling lists in use at the time, but on the other hand could be usefully employed as a check on everyday words known. No single spelling list is complete in its own right however. There is at present no spelling test available with New Zealand norms, based on New Zealand lists, but it is hoped that once the new N.Z.C.E.R. list compiled by Arvidson has
been established, that some work along this line may be completed, as spelling seems to be one of the major gaps, so far as full New Zealand norms are concerned, in standardized tests available to teachers in this country.

Of the five reading tests administered by teachers in the investigation, both the comprehension tests were used in conjunction with one of the word recognition tests. This shows that in ten cases, teachers were creating reading groups on a sounder testing basis than were the remaining twenty-five who also used tests for grouping. Both comprehension tests are alike in construction, but of the two, the Metropolitan is simpler to work. New Zealand norms for the A.C.R.R. have been compiled.

As shown all three word recognition tests are similarly constructed. All tests in the main, measure mechanical reading ability only, and do not take into account that a high score does not necessarily mean the child understands what he has read. This is a danger that is easily overlooked. They are however easy to use and convenient for grouping. Fieldhouse 21 reminds his users, that as skill in reading involves attainments of kinds other than word recognition, several tests measuring these skills are needed. The

21 A.E. Fieldhouse - op cit P. 1.
"Reading Age" given on word recognition tests is in fact a
"word recognition age" — a facet only of reading.

Again, R. Winterbourne's Christchurch norms for the Burt
test, were not known by any users, and thus the English ones
were employed.

The Fieldhouse test has been designed for New Zealand
conditions, and is worthy of wider use than at present
seems to be made of it. It does not, however, appear to be
well known, but is more applicable for its age range in this
country than the better known Burt and Schonell tests.

USES OF TESTS

In arithmetic, twenty-seven per cent of the sample used
tests, and of these seven used two or more tests of arithmetic.
This total of twenty-five testers is spread among eleven
schools. Actual misuse of these tests was apparent in only
two cases, although the use of a purely diagnostic test is
preferable to using a basically attainment one for both
purposes. The diagnostic test is systematically designed
to find weaknesses, whereas an attainment test is not. Grouping
within classes for arithmetic, was undertaken by twenty-one
teachers, sixteen of whom based these groups on their test
results. Thus sixty-four per cent of the testers used their
results for this purpose, although two of these based groups
on the results of a purely diagnostic test. Of the ten who
preferred to use tests for diagnosis, however, later
information showed that only two took any specific remedial
action. Nine teachers tested and took no action on the results, which is a sheer waste of time.

The surveys made by the intermediate schools, however, provided ten form one teachers with results, and so to a degree freed them from the necessity of testing.

Overall, fifteen schools used spelling tests involving thirty-seven of the sample or about forty per cent. Of this number two were headmasters, so in effect forty-six per cent of class teachers used some form of spelling test. All used it for attainment purposes. It is noted that although some tests can be used for diagnostic purposes, not one of the sample used them then, although three did some remedial action on the incorrect words, which is the only misuse noted in these tests.

There were eight (or twenty-two per cent of arithmetic test users) who tested and made no use of the test results.

Six of the thirty-seven applying spelling tests used two or more tests of spelling. Of the seventy-six teachers seen thirty-nine used groups in spelling, twenty-four of these based them on test results.

Reading tests of one of the types mentioned were given by thirty-five teachers, while ten gave two or more. This represents forty-nine per cent of the sample, six of whom can be discounted as they made no use of it after testing. Two teachers diagnosed from these tests, but it was found in the following information gleaned from them, that little
was actually done as a direct result. Of the total users, twenty-three used the results for grouping purposes. An overall view of the diagnostic and attainment testing undertaken reveals that five headmasters and fifty-six teachers used these tests, a total percentage sixty-six. A more detailed analysis follows in Table XIV.

**TABLE XIV**

**TEACHERS USING STANDARD TESTS**

<table>
<thead>
<tr>
<th>No. of tests used</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of users</td>
<td>27</td>
<td>20</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

It is possible that younger teachers, having less experience to rely on, may test more than the more experienced. Table XV sets out the relationship between teaching experience and numbers of teachers using standardized tests.

**TABLE XV**

**EXPERIENCE AND TEST USAGE**

<table>
<thead>
<tr>
<th></th>
<th>Under 14 yrs</th>
<th>Over 14 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of teachers using tests</td>
<td>31</td>
<td>25</td>
</tr>
</tbody>
</table>
This Table shows that experience does not affect the numbers of teachers testing, as the difference is small. The volume of testing was not affected by experience either as the thirty-one teachers with less than fourteen years experience administered fifty-six tests, while the twenty-five more experienced ones gave fifty-five standardized tests.

The far greater portion taking advantage of these tests than those using intelligence tests can be accounted for by several reasons. Firstly standard tests are more related to class work, and secondly do not require extensive courses to use and apply. It was noted, however, that all the teachers academically qualified for testing purposes, were included in the total who used attainment and diagnostic tests.

There are many more types of standardized tests readily available than there are intelligence tests, and their uses are more readily accepted. All schools encouraged some type of testing to a more or less degree and teachers felt obliged to do some even if no great use was made of the results.

This was owing accounted for thirty-eight per cent of the total testers, which reflects lack of understanding of what the tests have to offer. As with intelligence testing, it shows a gap in the teacher's training.

Misuse of tests other than this was low being confined to those few previously mentioned.

It does seem, however, that more use might be made of the
results other than grouping. Retesting later in the year for progress, and the wider use of diagnostic tests could be undertaken, provided the necessary small amount of training needed was given to the teachers. More use could be made of the tests with New Zealand norms, as these, particularly when retests are done, give the teacher a national standard, as opposed to his own, against which to reassure both his pupils and his own progress with them. Due care should be taken that testing results are kept in the proper perspective however.

Watts warns that age is the only variable for which standardized tests make due allowance. They afford no compensation for instance to those whose schooling has been interrupted or have been victim of indifferent teaching, or an uncongenial home environment.

While agreeing on their value he believes there is a danger in that teachers may fail to recognize the occasions on which they should not be used.

Test Programmes

The nearest approach to Traxler's type of programme was seen in the intermediate schools, but here again, much of it was done by the standard four teachers whose children were to

\(^{22}\text{Watts W.F. - Can We Measure Ability (London - University of London Press 1953) P. 4ff.}\)
go to intermediate schools, and so no definite testing procedure was invoked. All the marking, however, was done by teachers in the intermediate schools concerned, under the control of the headmasters. Their programmes were practical and definite, but many of the programmes carried out by individual teachers in the remainder of the sample were haphazard, and often only one test was given for some indefinite purpose.

A testing plan should provide information to assist in the aims of the class or school, and also as a check as to how far these goals have been met. Here again the standards required have to be clearly defined in advance. A considerable amount of time has been used by these teachers in administering, marking, and organizing the results of the tests. Apart from the intermediate schools and one of the other schools, however, there was no organized or co-ordinated testing throughout a school. Many schools in the sample have conferences regular or otherwise, of all teachers, and set out for the year, the work to be covered in various subjects, so that over-lapping will not occur, and some standardization of content will be maintained, but very few plan school-wide checks or tests to see if this is in effect being coped with.

Traxler's second point is an important one in that any testing programme should be a co-operative one. Here,
those who are suited by training to administer tests can help those who do not know very much about it, which was done in four cases. This is closely related to the third point of delineating the procedures for testing. Some form of standardized control of administration, marking, and grouping could then be established through the school, bearing in mind always of course, the rights of teachers to organize their classes as they wish, within limits of the headmaster's policy. Again only the intermediate schools did so.

Such a policy, however, would also stop the useless testing encountered in over one third of the testers.

The fifth point of the programme, being continuous and long range, is one which could bear closer inspection in the schools seen. Many of the test results were not entered in the class register of marks, and apart from a few schools, who recorded intelligence tests scores, only one school made a permanent record of standardized tests administered. It seems that if these tests were given correctly, it is rather a grave loss on the child's cumulative record card if they are omitted. Many teachers do not pass on results of their tests when children are removed from their class, either for promotion or transfer. Where a child is transferred to another school a considerable amount of time could be saved if these results were available to the new teacher. Again if results followed the child on promotion, the new teacher would have some basis on which to judge the child's progress from the year before if the same tests were reapplied.
Many teachers did not know of the large number of standardized tests available. A wider knowledge of such tests would give teachers more choice in selecting appropriate tests or better still, batteries of tests, which would give them a quick idea of children's ability and so serve for better grouping, if that is what the tests were wanted for.

The factor of cost again arises, but it is common practice in most schools for tests to be duplicated. In some cases teachers were using tests whose titles they did not know, but had merely been given to them by another teacher. Occasionally norms were not available, but this was overcome by using the test on a comparative basis.

It is quite apparent therefore that all five points made by Traxler are not being met by any of the schools in the sample.

On the other hand, the majority of tests used, were applicable for the purpose, and very few were misused.

SUMMARY

Of the three attainment tests seen in arithmetic, the Schonell test was the most frequently used, mainly because of its sections testing tables. The New Zealand standardized test, the A.C.E.R., was not as widely used.
For diagnostic purposes, the Seville tests, produced in this country for the New Zealand syllabus, was used by seven per cent of teachers only, while eleven per cent used Schonell's test for diagnosis, although it suffers by being a dual purpose (attainment and diagnostic) test. The two tests could be used as a complement to each other.

Arithmetic attainment tests used by thirty-three per cent of the teachers were applied for the assessment of scholastic levels, and by twenty per cent for grouping purposes. Little use of the results was made by twelve per cent of teachers.

The Spelling test most frequently used was Schonell's as the Schonell Essential Spelling List was in current use, when this survey was undertaken. With the advent of the N.Z.C.E.R. (Arvidson) List, this spelling test in current use will not be as widely used although words outside any list need to be tested. It is hoped some test, based on the new list, may be forthcoming. Non-use of results was again noted in eleven per cent of cases and a singular lack of diagnostic use was evident. The main use was again for attainment and grouping.

In reading, the favoured test was the Schonell Graded Vocabulary, which was in fact the most frequently used of all the standardized tests seen. It was largely used for grouping. As with the Burt Graded Vocabulary the mechanical side of reading only is measured; few teachers tested for
comprehension in addition to gain a balanced view of their pupils' capabilities. Christchurch norms were not used, and the only test designed for this country, the Fieldhouse Oral Word Reading Test, was used in one case only.

It was noted that sixty-six per cent of the teachers used one or more standardized tests; a considerably higher proportion than those who used intelligence tests. This could be accounted for in that standardized tests are more directly related to teachers' current work and give an immediately applicable result. There is not the "mystery" surrounding them (that some persons at least) attach to intelligence tests and they are more concrete in results from a teaching point of view. Further, a more limited training in their administration and interpretation will suffice.

Measured against overseas criteria, standardized testing programmes were not met in full by any of the schools seen, but the programme provided by the intermediate schools was the nearest.
CHAPTER VI

TEACHER CONSTRUCTED TESTS

The most common method of evaluating school work and school progress, is with tests constructed and given by the teachers themselves. Some teachers and schools dislike giving tests or examinations, but evaluation of children's progress and attainment is necessary for the teacher's own information as well as the child's records throughout the school.

Remmers and Gage¹ defend the right to evaluate both by standardized and teacher constructed tests. They state that schools are the places where the process of fitting pupils to curriculum, and citizens to social role is carried out with the best results for individual adjustment and social welfare. In schools this fitting process is guidance, and guidance requires the evaluation of pupils so that their specific capacities, both strengths and weaknesses, may be determined. They go on to say that as pupils proceed up the educational ladder in elementary school, which provides in its general curriculum the common core of skills, that everyone in our civilization must acquire, the process of evaluation must operate upon the children to reveal all their relevant attributes, so that at the end of primary education,

the knowledge gained about the pupil as it affects his choice of secondary education should be the result of continuous evaluation and comprehensive evaluation. These ideas are also put forward by Wiles who states "The evaluative process is a carefully developed method for making more valid judgements concerning ways of improving the learning situation". He agrees with Remmers and Sage that evaluation must be continuous and comprehensive.

This survey is concerned only with evaluative procedures contained in normal classroom tests in the basic subjects of arithmetic, spelling, comprehension, composition, social studies, and writing, although evaluation of the pupil naturally covers a wider range than these.

Greene, Jorgensen and Gerberich advise that a test should not only be so constructed as to measure the degree of attainment of the pupils in the desired outcome of the test, but it should do so by means of test situations which involve the ability to apply and use facts, as well as the knowledge of facts.

Care should be taken in revision tests to sample

---


course content widely, and they say, recall items, which demand highly of the pupil's memory of specific facts, should be used only for important facts. Greene suggests that recognition items will fulfil the need in many cases.

The tests examined were broadly categorized into source, frequency, construction and content, as applicable to each subject, while the marking systems were also analysed. The percentages given are based on the seventy-six class teachers' tests only, as headmasters did not normally give regular tests of this type. The percentages are rounded to the nearest whole number.

ARITHMETIC TESTS

Of the seventy-six teachers in the sample, all gave some form of unprepared test in arithmetic. As the name implies, this type of test is used as a normal routine test, requiring little or no preparation by the teacher. The terms "routine" and "unprepared" are both used for this type of test.

Seventy-four of the seventy-six teachers constructed their own tests. Two only, obtained their tests from other sources. Forty-nine of the teachers (sixty-five per cent) gave unprepared tests daily. The next highest total is nine teachers (or fourteen per cent) who administered such tests weekly. The remainder varied with eight teachers giving them at irregular times. Thus for this sample of seventy-six class teachers at least, the majority of teachers (that
is sixty-four in number or eighty-four per cent) administered regular weekly or more frequent, routine tests of this nature.

In content the tests were mainly of a cumulative type, whereby a teacher could revise work that had been covered throughout the year at any stage. Sixty-one teachers (eighty per cent) followed this pattern. (Twenty-four (thirty-one per cent) of teachers placed emphasis on current work combined with cumulative work, while no teachers sought to teach new work by means of this sort of testing. New work here is used in the sense of teaching in advance, such things as tables of measurement, before these are actually met in the normal arithmetic syllabus. Five however did combine this idea in current work and cumulative work. It is noticeable that only one teacher gave routine tests on revision alone.

Mental arithmetic does lend itself to this type of routine testing. The frequency of testing showed that a reasonable coverage was given. No great amount of time is needed to give such routine tests, and providing prepared tests are at times given as a check on work completed, these types of tests serve the useful purpose of a quick, readily available testing tool.

Peculiar to themselves, are a number of arithmetic tests used for attainment, which three teachers have constructed themselves. They had administered this test for
several consecutive years and classified their class on
a resultant comparison basis. There were no norms, but
the tests were used for the particular needs of the teachers.
One which was examined, contained all the number combinations
up to the nine times table, and hence could just as easily
have been gleaned for Schonell's Diagnostic Arithmetic Test.
It is difficult to justify the work involved in the
composition of such tests, when others, such as the
scientifically constructed and standardized A.C.E.R. series,
are readily available, more reliable, and more informative.

As distinct from the term "routine" or "unprepared"
tests, the type of test requiring thought and preparation is
referred to as a "prepared" test.

Ten sources of prepared tests in arithmetic were
encountered. The two main sources are those tests
constructed by teachers themselves and those selected from
the text book appropriate to the class. Fifteen teachers
however made no use of the text book at all, twenty-six used
it occasionally, and thirty-five (including two who used the
advanced standard six text book) referred to it often.
Reasons given for using the Department of Education text
book in arithmetic are set out in Table XVI.

Those who used it seldom or not at all, gave as their
reasons, those listed in Table XVII.
# TABLE XVI

REASONS FOR USE OF TEXT BOOK

<table>
<thead>
<tr>
<th>Reason</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Inexperienced</td>
<td>1</td>
</tr>
<tr>
<td>2. As an assured means of covering the syllabus.</td>
<td>2</td>
</tr>
<tr>
<td>3. Considered quite adequate, no extra work needed.</td>
<td>10</td>
</tr>
<tr>
<td>4. No expressed opinion, but quite satisfied.</td>
<td>13</td>
</tr>
<tr>
<td>Reason</td>
<td>Total</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>1. Used for revision only.</td>
<td>8</td>
</tr>
<tr>
<td>2. Used for alternative work only.</td>
<td>2</td>
</tr>
<tr>
<td>3. Used as a basis only.</td>
<td>3</td>
</tr>
<tr>
<td>4. Standard too high.</td>
<td>4</td>
</tr>
<tr>
<td>5. Exercises inadequately graded.</td>
<td>24</td>
</tr>
<tr>
<td>6. Too few examples.</td>
<td>13</td>
</tr>
<tr>
<td>7. Standard too low.</td>
<td>7</td>
</tr>
<tr>
<td>8. Miscellaneous.</td>
<td>2</td>
</tr>
<tr>
<td>9. School Policy.</td>
<td>13</td>
</tr>
</tbody>
</table>
The most common reason therefore for not using the supplied text is the complaint that the examples are inadequately graded. Many specific examples of this were cited by the teachers and in some circumstances they were justified. Closely related to this is the felt lack of examples in some types of sums. This however, can easily be overcome by the teacher preparing the types himself, relative to the needs of the class. The question of high or low standards, depends on the type of class which the teacher has. Rather obviously the normal text book does not cater adequately for a very low stream in a school or, for that matter, the opposite extreme. It appears to be designed for the average class, and here again the teacher is expected to adapt it to his own needs and those of his class.

The majority of the teachers, as seen in Chapter III, were experienced and although the text has definite shortcomings, it is a basis from which work can be expanded. As three teachers said, it is a guide. Some teachers were not limited to one complaint concerning the text and often refused to use it on a basis of several of those listed above.

The total of thirty-two teachers who preferred to make up their own tests, rather than accept those in the text book, or from other sources listed above, could show that they recognize the need to adapt any test to their particular class. It is obvious that to test that which has not been taught would be futile. Tests therefore should be designed to sample work that
has already been taught and learned. Although a test already available from some source has to be checked before it is given to a class, the extra effort on the part of the teacher in constructing a suitable test for the class's requirements, has the beneficial effect of letting him examine what he has taught and the goals to which he is aspiring in the test. An overall impression of the figures shows that teachers have more than one source of tests, even though they may not use the textbook.

Twenty-nine per cent of teachers prepared tests at fortnightly or more frequent intervals. Some teachers used a combination of test intervals, the most popular (used by seven teachers) being a weekly plus monthly cycle. Monthly tests only were administered by twenty teachers. The most frequent form of test used was that which covered one process or topic, when the work on it had been completed.

Irregular testing was done by eight teachers, and here again this was done mainly to have some record, because records as such are demanded. Only one teacher tested once a term.

The principle of continuous and comprehensive evaluation is then being followed by the vast bulk of teachers in this sample, as far as prepared tests in arithmetic are concerned.

The content of prepared tests was classified into four types as follows:

(1) The tests containing one process whether current or not.
(2) Revisionary block of work covering, for example, first half term's work.
(3) Two or more topics or processes - usually more limited in scope than (2) above.

(4) Cumulative work to date, including current work done.

The most popular method of testing here was testing one process at a time. This of course is closely allied to the method of teaching a process and testing it, but some teachers who were included in this figure preferred to test one process which perhaps was not current work at the time of testing. It is in effect a form of revision. Forty-six per cent of the sample used this frequently.

The revisionary block of work which fifteen teachers preferred to have as the content of their tests, is the least used in this sample, but is still twenty per cent. This type of revision as defined here has its disadvantages. It is not related to the children's immediate work, but is an attempt to reach back and revise previous material which may well have been forgotten, and is normally used in work prior to examinations.

The third heading is also allied to the revisionary block type of test, as it covers two or more processes. When a test is made up of several processes it does constitute a type of "revisionary block", but was kept separate in this survey, as the "revisionary block" is meant to be more a test on a time basis of work done, whereas this is a more selective type which is often used by teachers to find weaknesses in processes upon which to base remedial action.
The cumulative test was used by twenty-seven teachers (thirty-five per cent), and is a common method of testing work that has just been taught together with a constant amount of revision. It is a convenient way to revise work which would otherwise become forgotten. It is also a means for the teacher checking not only his current work, but the amount of past work, that the pupils can still cope with.

An analysis of the construction of prepared tests in arithmetic, revealed that fifty of the teachers (sixty-six per cent) preferred to keep problem work separate from mechanical. This seems to be rather an artificial type of relationship except perhaps for lower streams of classes. Problems are based on mechanical work and therefore to divide the two seems arbitrary, as one is really the application of the other. Many teachers justified this separation on the grounds that in the lower classes seen, particularly standard four, it was the mechanical aspect, which was the most important, as this was the foundation of later work. It should then be this mechanical part that was tested. They appeared to think that problems were easier coped with at an older age.

This argument has some surface merit at least, but is probably not applicable in higher streams where children do have more intellectual ability and can relate mechanical work
to problems. A number of teachers did put in a few problems, but amongst the fifty mentioned this was comparatively rare. As will be seen later, allowance was made by many teachers for a weighted mark for problems as compared to mechanical work. Thus in another aspect the importance of problems was realized. It also is much easier to set and mark purely mechanical tests rather than a mixture including problems. In the lower ability groups, there is more difficulty in doing a sum and then working out an application, than there is in doing a complete set of mechanical sums followed by a separate set of problems, as two distinct tests.

It was noted that only three teachers gave tests of graded difficulty, starting with the more simple types and ending with the harder examples. One reason could be that these tests take considerably more time to prepare but on the other hand they do extend the better pupils in a class. If the test was graded, it had to be properly so, and not merely have one or two hard examples put in at the end to absorb the time of the better children, a practice quite common among the sample.

Emphasis was placed on mechanical work in tests by thirty teachers while eleven tended towards problems. Thirty-five therefore had no particular trend when constructing tests. The eleven who commonly set problems were teachers of upper stream children, and the arguments presented before in this section on the suitability of mechanical versus problem work
to various streams in schools, applies here also.

It is significant that the three teachers who graded their tests, where those who practised diagnostic and remedial work. By grading a test it is possible to do some diagnostic work with the results. The opposite extreme of setting tests of equal difficulty was used by two teachers in the lower streams, so that all children could complete the test, rightly or wrongly, and feel some satisfaction from this. Five teachers made fully integrated tests by giving some mechanical work and basing problems upon it, as opposed to the twenty-six who merely gave a mixture of mechanical and problems, one not particularly related to the other.

**SPELLING TESTS**

The commonest source for spelling words was Schonell's "The Essential Spelling List", plus an error list of class or pupil mistakes. Thirty-two teachers relied on this source. A total of twelve teachers used this Schonell list, together with another, other than the class or child's error list. The Schonell list solely was used by twenty-eight teachers. This total of seventy-two teachers represents ninety-five per cent of the class teachers seen.

This conformity to the Schonell list by ninety-five per cent of teachers is understandable as it has been regarded as the basis of school spelling over the last few years. Furthermore, it should be noted that this survey was completed in 1959, before the new N.Z.C.E.R. Spelling List (Arvidson) was introduced into schools.
Of the remaining four class teachers of the sample, two each used the de Berry and Thorndike lists. It was noted that the de Berry list was used exclusively in one school, and likewise the Thorndike List, mainly on the Headteachers' instructions.

Routine testing was normally given daily or on a weekly cycle. A typical weekly cycle consisted of teaching the week's words on Monday, using them in English and sentences on Tuesday, meanings and kindred words on Wednesday, extensions on Thursday, and a retest on Friday. There were of course variations to this, but the pattern was basically similar. Occasionally a pre-test is used, so that the most difficult words for the whole of the class can be ascertained and concentrated on. Only four teachers tested at intervals of longer than one week for their routine spelling.

Of all the tests this survey investigated, spelling was the most regularly and frequently tested subject, showing that from a comprehensive view its place in the curriculum was well recognised. Perhaps some of this stems from the continual criticism levelled at spelling in the primary school. It is also an easily tested subject.

Revision testing was mainly administered weekly, while only one teacher was irregular in giving revision spelling tests. Six teachers gave revisionary tests at fortnightly as well as four to six weekly intervals.

The content of spelling tests as confined to the revisionary
tests, revealed that the largest group (thirty-nine teachers) preferred to do their work in revisionary blocks (as defined in the section in arithmetic). This usually calls for a series of tests so that the children might learn and be tested on the first three lists, then the next three lists, and so on.

Twenty-six preferred to set only words selected from the list, these words being ones in which the class had found most difficulty during the routine testing. Only eleven tested on a cumulative basis. This is probably because cumulative revision does not give such good results as learning a shorter number of words.

Thirty-five teachers compiled their dictation from selected list words, while twenty-one included all the set words in theirs. Thus almost seventy-four per cent of the teachers usually applied the children's knowledge to dictation. Of these, twenty-nine (thirty-eight per cent) tested at regular intervals.

**COMPREHENSION TESTS**

The majority of teachers in the sample made comprehension tests from their own sources. Thirty teachers (thirty-nine per cent of the sample) did this, while the next highest total (twenty teachers), as far as sources are concerned, obtained their tests from miscellaneous material. The wide variety of source material used by teachers for comprehension tests, is reflected in the eight categories plus miscellaneous, under which the results were classified, but if the total of thirty teachers
who make up their own tests, is taken from the overall sample there is no one particular source which is greatly favoured apart perhaps from the Teachers' Monthly Guide, which was used by fifteen teachers.

One of the difficulties in finding a comprehension test suitable for a class is the wide variety of material taught. If the comprehension is taken as a normal part of reading then it is usually correlated with the reading material being covered by the class. Unless comprehension tests are included in the text used, it is often much more convenient for the teacher to make up his own tests. In this way the teacher can test either generally or specifically, and if he wishes put any type of question in the test.

The disadvantages of using a test already constructed from some other source often means that the test is not particularly related to the work in hand. At times of course this is not a necessity, but the question then arises of what exactly is the teacher trying to test. Comprehension is the understanding of what has been read, but there is a danger of a comprehension test being turned into a sort of quiz or intelligence test, when it is not related to some work that has been taught or is current.

A good comprehension test takes time and effort to construct correctly. The simplest construction used by fifty-three per cent, consisted of direct questions. These are usually questions which require short answers of direct
information, which can be gained from the text preceding the question. This does not involve very much thought on the children's part as they merely have to find facts. Thus it is not a true comprehension test.

Burt states that a comprehension test is best framed on the principles of a "directions" test. The criterion of intelligent reading is then not the power to reproduce from memory the substance of some story or abstract argument, but the ability to carry out from a printed order, some concrete practical instruction.

Thirty-one teachers however preferred tests containing a miscellaneous mixture of questions.

Such tests would include selection from questions requiring the following types of answers: direct, yes-no, multiple choice, short essays and completions. These tests will usually allow the less intellectually gifted children to gain satisfaction from being able to do some of the questions, while sufficient questions of a reasoning type will extend the better pupils in the class. It is a more balanced type of test suitable for a mixed ability group.

Only one teacher favoured putting many "yes-no" type of

\(^{1}\)Cyril Burt, *Mental and Scholastic Tests* (Westminster: P.S. King & Staples 1921) P.275.
questions in the test, or those of a type requiring sentence completions. The impression gained was that the teachers had a dislike of compensating in a yes-no type of question by subtracting marks, if the answers were wrong. This attitude is quite reasonable as most children cannot see the reason for this, and although it is supposed to discourage guessing seldom does.

Fourteen teachers favoured essay type questions including the following of instructions as the basis of their tests. These types are probably satisfactory for high ability classes, but for the lower streams, if the instructions are too complicated or too high a standard of essay writing is required, the test may fall down under the weight of these side effects and so miss the main aim of testing in comprehension. At least twenty-four teachers used combinations of these types of construction mentioned above.

Closely allied to the construction of the test is its content and one is interdependant on the other to a large degree. The general trend of content in comprehension tests is shown in Table XVIII.

These figures show a fairly even distribution of trends, but those teachers requiring mainly direct information, were generally teaching lower ability groups. They were not, however, testing comprehension as defined in the statement from Burt, previously quoted.
TABLE XVIII
COMPREHENSION TEST CONSTRUCTION

<table>
<thead>
<tr>
<th>Content Tendencies</th>
<th>No. of Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reasoning</td>
<td>21</td>
</tr>
<tr>
<td>Direct Information</td>
<td>15</td>
</tr>
<tr>
<td>Both Reasoning and Information</td>
<td>25</td>
</tr>
<tr>
<td>Uncommitted</td>
<td>14</td>
</tr>
<tr>
<td>Oral Tests Only</td>
<td>1</td>
</tr>
</tbody>
</table>
Testing of routine comprehension was most often done on a weekly basis; forty-one teachers tested in this way. Those who tested more frequently confined their tests to very short periods of time.

Lengthier and more comprehensive tests than routine ones, requiring the same type of preparation on the teacher’s part as those classified as "revisionary", for other subjects were not done so regularly. These tests are difficult to construct for there is a danger of them becoming merely memory tests of what has been read, and not tests of understanding, and skills taught earlier.

These more "comprehensive" tests were not so regularly applied as the more routine ones. Three teachers in fact gave survey tests only for this type of test. There was a pattern amongst sixteen teachers to give a regular weekly routine test, and a longer comprehensive test at monthly or six weekly intervals. Two teachers tried to combine comprehension tests in relation to social studies and English. The danger here again, of course, is that direct information rather than comprehension would be the result.

COMPOSITION TESTS

Composition topics in general use were classified into eight broad groups: concrete, abstract, imaginative, everyday occurrences, children’s interests, descriptive, original stories, and an overall selection from all of these. It is
realized that most of the above topics overlap to some degree. There is also the composition which is done as an integral part of other work, which is not included in the figures for topics. Twenty-seven teachers used other written work as a form of composition, thereby integrating it with other subjects conserving and developing flexibility of effort to give an overall use of many skills, as will be required in higher classes.

There is however, the effect of children having to concentrate on several things at once, not only on style and grammar as required for composition, but also on the subject that is being tested at the time.

Apart from this figure, the next highest is eighteen teachers who prefer to give their children concrete subjects with which to work, and if the total of two more for everyday things and three for children's own interests are added, the general heading of "concrete" could cover topics given to twenty-three teachers. Again these figures are indications only, as many teachers used more than one type of topic, but are indicative of the trend of the particular teacher.

Topics are closely tied to the type of class the teacher has as it would be little use asking children of low ability to write many original stories and hope for a very satisfactory response. In the main the more concrete type of subjects were predominantly found in the lower streams while the imaginative type were found more often in the higher ability class of worker.
C.T. Ford\textsuperscript{5}, from the results of an investigation of pupils in Canterbury primary schools, found that a composition programme should be flexible and as far as possible, individualized. He also states that, especially for younger children, topics should be related to pupils' experience, and that there is a marked correlation between success in the expressionional aspects of composition, and high verbal intelligence. In general then, teachers are relating topics to the abilities of their classes.

As in the section on comprehension, the term "comprehensive" test is used here. It signifies a composition of more length than a routine one, and is used to bring out skills in writing which have been previously practised. Routine tests are those frequently given to exercise and develop specific composition skills.

The weekly routine testing seems, again, to be quite consistent as twenty-four teachers (thirty-two per cent) did composition tests of some form regularly once a week. Those who made it an integral part of other work, did so at their convenience, and could well be combined with the teachers who are found to give composition irregularly. Thus only fifty teachers out of the seventy-six (sixty-six per cent) gave

regular routine composition, and forty-two (fifty-five per cent) gave irregular comprehensive tests at intervals of more than one month.

Composition then does not seem to be a particularly favoured subject as far as regular testing is concerned. This is not surprising, in that the amount of work involved in marking compositions is quite considerable, although it is the best means of providing opportunities to test effective writing.

SOCIAL STUDIES TESTS

The general impression gained during the investigation on sources of social studies tests, was that teachers were rather vague about where their material for both teaching and testing came from. Fifty-five teachers out of the sample, (seventy-three per cent) appeared to have no definite source. This is probably because there is no one particular book from which all the social studies material to be taught in the present syllabus, can be gained. The nearest text which would approach this is the duplicated material published by the Christchurch Teachers' Training College in 1948 entitled "Social Studies". There are some booklets on social studies topics, issued by the Education Department, from which much valuable material can be gained. Although they are issued for children to read, many teachers state that a considerable number of their books are too difficult for the level at which they should be introduced.
A set of books to cover the syllabus is virtually unobtainable, and so teachers have had to rely very heavily on their own personal knowledge. Source material is literally "where you can find it". Tests, therefore, apart from some in the teachers' monthly guide, are largely non-existent. Some of the current affairs sessions from the Broadcasts to Schools, sometimes provide a short quiz, and at times a headmaster or other teachers will make up one test to cover a certain part of the syllabus, which has been taught to several classes. Of the sample seen, six teachers used the current affairs broadcasts and four teachers, all in one school, combined to devise tests for classes other than their own. This naturally presupposes a very similar curriculum. Thus it is not surprising that the bulk of teachers made their own tests, from their own sources.

Only one teacher gave routine tests and the frequency of revision testing is summarized in Table XIX.

Although social studies is often tested at irregular intervals, the forty per cent who teach a topic and then test it, are making a good test coverage. As this subject is normally taught in topics, this method of testing is a most convenient and natural one.

The twenty-eight per cent who tested twice or less in a year favour other methods of assessment, which have merit, particularly when a project system is used.
<table>
<thead>
<tr>
<th>Intervals</th>
<th>No. of teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly - 2 monthly</td>
<td>14</td>
</tr>
<tr>
<td>Term</td>
<td>1</td>
</tr>
<tr>
<td>Term plus projects</td>
<td>1</td>
</tr>
<tr>
<td>Surveys only</td>
<td>12</td>
</tr>
<tr>
<td>As part of other subjects</td>
<td>9</td>
</tr>
<tr>
<td>Mark books and projects only</td>
<td>9</td>
</tr>
<tr>
<td>Test at completion of topic</td>
<td>30</td>
</tr>
</tbody>
</table>
The construction of social studies tests showed an emphasis on the direct type of question requiring a short and often one word answer. Social studies tests can at times be too long and this form of answer does make for easy marking on the part of the teacher. Ten teachers liked essay type questions and a few put in yes-no, and multiple choice questions.

Apart from the one who tested regularly and also used part marks, eighteen others allotted part of their assessment on the work done in projects and social studies books. This seems a very fair thing to do, as a good deal of work is often required in the children's social studies books, and these plus test marks should give a good overall assessment of the children's ability in this subject.

WRITING

The Department of Education Syllabus for schools\textsuperscript{6} sets out the type of letters to be used in both printing and writing and gives models for these. Certain methods of teaching writing are also explained but it is left largely to the individual teacher, to decide upon the style of writing to be taught, and the method of teaching it. In some schools, the

differences between letter formations are apparent from
class to class, but others however do have conferences on
letter formation so that uniformity will be achieved.
throughout the school. Thus pupils passing from one class
to another will not be expected to relearn different letter
shapes.

The requirements of writing are a combination of speed
and neatness, but the amount of emphasis on either of these
is usually a matter for the particular teacher or the policy
of the Headmaster within a school. A generally acceptable
standard of writing among the teachers seen (as far as
generalizations can be made), was writing that was neat and
tidy, uniform and appealing. There was not so much emphasis
on speed in most of the lower classes seen. More speed was
expected in forms I and II. In many form II classes, writing
was not taught at all as a separate subject.

The standards for examinations then were an individual
matter for the teacher concerned. It was quite apparent that
routine writing consisted largely of copying cursive writing
from the blackboard. Slightly more teachers tested writing
by setting copy from printed material, such as a book, thus
making the cases of testing by transcription a total of
fifty-seven (seventy-five per cent). Some teachers in fact,
used a combination of blackboard copy and copying from printed
material.
Although only one teacher used routine writing tests from normal written work, nine regularly tested revision from this source. Seven teachers used corrections of other work as writing exercises. As expressed before, the predominant place for testing writing is in the standard four classes, as during forms I and II, writing lessons tend to be dispensed with, particularly in the upper streams, although it is suggested in the syllabus that rhythmic exercises be continued in forms I and II. Routine testing in writing was done by nineteen teachers daily; nine gave no tests at all; forty-eight tested irregularly. This is perhaps indicative of the fact that the cliche "Every lesson is a writing lesson" is taken fairly seriously by many teachers. The tedium of marking writing very frequently was a factor involved as well.

Twenty-two teachers in the sample entered no marks in the register, but a few took an assessment from any other written work. Again all these teachers involved were in forms I and II. Those who tested only irregularly made the next highest total of nineteen, but they did in fact give a separate writing test. Monthly testing was done by twelve teachers, several of whom were in schools in which regular monthly tests were done as part of school policy.
MARKING OF TESTS

Remmers, Gage and Rummell state that among the purposes of marking systems, the more common are to provide a basis for information for parents on pupil status and progress; promotion, and graduation, motivation of school work, guidance of learning, guidance of educational and vocational planning, guidance of personal development, prizes, participation in many school activities, reports and recommendation to future employers, data for curriculum studies, and reports to a school the pupil may attend later.

In the schools seen, the majority of these purposes were applied in one form or another, but the marking methods differed widely.

In all the subjects investigated, marks were used more than five point scales (which for convenience will be referred to as grades). Grades, however, were used by forty-eight per cent of teachers in composition and thirty-five per cent in writing. These subjects lend themselves more to an impression type of marking than an analytical one and grades are much more applicable in the case of marking by impression.

Table XX sets out in more detail the proportion of teachers using different types of assessment and the allied methods of marking in each case.

---

# TABLE XX

## TYPES OF ASSESSMENTS AND MARKING METHODS

<table>
<thead>
<tr>
<th>Assessments</th>
<th>Letter grades (%) of class teachers</th>
<th>Marks (% of class teachers)</th>
<th>Methods of Marking</th>
<th>Analytical (%) of class teachers</th>
<th>Analytical (%) of teachers</th>
<th>Impression (%) of class teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic</td>
<td>8</td>
<td>92</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Spelling</td>
<td>13</td>
<td>87</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Comprehension</td>
<td>14</td>
<td>87</td>
<td>95</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Composition</td>
<td>48</td>
<td>52</td>
<td>14</td>
<td>71</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Social Studies</td>
<td>21</td>
<td>79</td>
<td>92</td>
<td>8</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td>35</td>
<td>65</td>
<td>21</td>
<td>63</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>
From Table XX it can be seen that in composition and writing analytical schemes were little used. Six teachers who had previously used them stated they were too cumbersome and slow to use, and that the resultant total of marks was often not consistent with their impression of the work as a whole. They therefore let it fall into disuse. The majority of teachers, when marking writing and composition, use a combination of some analytical plans, plus an overall impression. This allows for the best of both systems. The eight per cent who did this in social studies were among those who give essay type questions, in which some marks are allowed for such things as style, clarity and grammar.

Remmers\(^8\) states that a raw score on tests has practically no significance without additional data for interpreting it. Even if raw scores are transformed into percentile ranks or similar values, they merely compare the performance of one pupil with another and in themselves have no particular meaning. Remmers further states that marks or grades imply that value judgements have been made with the assistance of other criteria, and in terms of some type of value, objective, goal, or standard.

Of those teachers in the sample using marks, thirty-eight per cent entered raw scores only in the marks register, except

\(^8\)Ibid P. 266
in composition and writing. These entries then have little significance according to Rammers. The best that could be done from these is to rank pupils in class order of merit for that particular test, unless what was considered a pass mark was indicated.

Early systems of marking, according to Rammers usually marked out of one hundred, or converted marks to a percentage. He criticizes this on the sound basis that, for example, available measurement instruments and techniques cannot adequately differentiate to the fineness of one single mark, as this system implies.

Thirty-three per cent of teachers, however, did use the percentage system (except in composition and writing), although of these thirteen per cent of the sample of teachers attempted to offset its dangers by using intervals of five points between groups. In this case however, marks should be given in a range, but were not. This latter system was really equivalent to marking out of twenty, not one hundred, and the converting to a percentage.

Marking consistently out of one hundred, or for that matter, any arbitrary number, does allow for easy and quick comparison of a pupil's relationship to class rankings, but does not allow for the ease or difficulty of the test involved.

The percentages of teachers using grades in the various subjects offset the disadvantages of the numerical system, but Rammers again points out that this system often falls down due
to the meanings various teachers attach to the different
categories. These, however, are defined in a Department
of Education instruction E-19/23 of 1959, available to all
teachers, and this was the standard followed in most cases.
There is the difficulty, however, of different teachers
having different standards of what is average work.

To overcome this in two schools, the headmasters (whose
schools were streamed), allocated each class the numbers of
each grade to be awarded, prior to the tests, on the basis
of a normal distribution. This allocation was, however,
flexible enough to allow for unforeseen results. In schools
(as these ones were) where there are large numbers in each class
this system has merit. On small numbers, however, where a
normal distribution would probably not be found, it could be
statistically unsound. Similarly teachers attempting to
apply a normal distribution of grades on a class basis alone can
find themselves in the position of giving a top rank to work
which may gain only an average grade in another class, which has
a higher ability group of children.

There were a few, varying from eight per cent in writing
to nil in spelling, who gave neither marks nor grades (except
for half yearly surveys), relying rather on their own knowledge
of pupils' work, oral questioning, and remarks on answer sheets,
to guide them. If competition is desired this is not a
satisfactory method, but it is encouraging for low ability
children.

Remmers again points out that for some pupils marks often
become the end and aim of education, for their primary goal is a "good" mark regardless of how much they have learned, or how they have gone about getting the mark. Frequently low marks unduly discourage pupils at the lower end of the achievement scale, although it is these pupils who need encouragement. At times bright pupils are not encouraged to do well, as they are able to obtain satisfactory marks with little effort. These are not inherent defects of marks, but rather an outcome of the way they are often used.

All teachers were asked what they considered an acceptable standard of work, both in tests and routine matters. With minor variations in different subjects sixty-two per cent based their standards on what they considered each individual child capable of producing. This capability was based in some cases on intelligence ratings (with all the consequent dangers already pointed out in Chapter IV), but normally on a general overall impression built up from the child's past performance. It does though allow for individual differences. Of the remainder, twenty-nine per cent worked on a minimum standard they thought the class as a whole should attain, and set their pass mark accordingly, while the balance of nine per cent set an objective standard. The total of thirty-eight per cent then were setting rather rigid standards a little reminiscent of "proficiency days" of times past, and were inclined to disregard individual differences.

No teacher seen used the marking system involving
standard scores\(^9\), whereby the resultant mark does not alter the order of merit of the class, but does enable one child's marks to be compared with his own or another's in the same or different subjects, irrespective of the test standard. Such marks may also be summed, and the results used to give an overall idea of the pupil's standard, or ranking over a range of subjects.

This system requires time and a small mathematical knowledge to apply, but the results in terms of meaningful marks, more than justify this.

Some form of statistical treatment of marks was carried out by approximately twenty-nine per cent of teachers mainly to scale results to fit predetermined marking schemes, or regrade the results of tests considered difficult or too easy. More use of this was made in arithmetic, where thirty-five per cent made statistical adjustments, normally a multiplication of the existing mark by a given figure to bring the range of marks up or down, so that the bulk of the class gained what were considered satisfactory marks.

Arithmetic was one of the most frequently tested subjects and the results lend themselves to this type of manipulation.

---

\(^9\)Douglas W. McIntosh, David A. Walker, Donald Mackay, *The Scaling of Teachers' Marks and Estimates* (Edinburgh: Oliver and Boyd 1949), mimeo extract – University of Canterbury, Education Department.
Allied to this, to some degree, was the practice of weighting some questions in all tests except writing. Comprehension was weighted in marks by forty-nine per cent of teachers mainly on questions of a reasoning type. Arithmetic had the least number of teachers (twenty-eight per cent) using this technique, and here it was mainly used on those questions considered harder, or more important.

In spelling, the weightings took a slightly different turn, in that forty-two per cent deducted more than one mark for each error.

This is a reminder of the times of the proficiency examination, when out of twenty-five spelling test words, worth one mark each, three were subtracted for each error. This legacy of having to get half or more of the words correct before scoring any marks is in the writer's opinion to be deplored, as poorer spellers are greatly discouraged by this system.

To take some of the work out of testing, many class teachers let pupils mark other children's papers. Seventy-six per cent did so in arithmetic for routine work, but only fifty-five per cent checked the results at all. Composition naturally enough was always marked by the teacher only, but spelling was marked by pupils in fifty-one per cent of cases and checked only by thirty-one per cent of the teachers.

The lack of any check at all seems a rather doubtful practice. Some, however, did have pupils check on the marking,
but a small amount of time spent by the teacher on even a random check would be of much value.

A further feature of the survey was an enquiry into the extent to which headmasters exercised checks on the standards maintained within their schools. Apart from standardized testing, already discussed in Chapter V, sixty-two per cent of teachers claimed that apart from infrequent visits and signing reports, no formal control was exerted, and this was confirmed by the headmasters concerned. In ten per cent of the sample, the headmaster tested the class regularly, but this was mainly in those classes controlled by inexperienced teachers. The only other regular control, in five per cent only, was where the headmaster set and marked examinations.

In the remainder of cases (twenty-three per cent), some control was exercised through collecting exercise books occasionally, giving irregular tests, and visiting classes for special cases of backwardness. Thus the majority of teachers were free to organize, teach, and test their classes in their own manner.

As was pointed out earlier, however, in this sample of teachers, seventy per cent had ten or more years experience, and should therefore need little supervision and guidance from headmasters. Some, indeed, invited the headmaster to help when difficulties were encountered, or when they wanted an objective assessment of some work.

Every teacher is required to keep a register of marks
and forty-six per cent entered all tests undertaken, while twenty-six per cent recorded what they considered a fair sample of the children's work. Thirteen per cent put in marks at regular intervals. Thus eighty-five per cent recorded comprehensively and systematically. Of the remainder, only six per cent entered survey marks only. Thus fifteen per cent were not fulfilling part of their role as teachers, for these registers kept in the school for seven years after the pupil has left, do form a useful measure of a pupil's attainment, to supplement that on the progress card.

SUMMARY

ARITHMETIC

In arithmetic, eighty per cent of teachers tested mental arithmetic or similar routine matters, weekly or more, mainly with a cumulative type of test, showing that good coverage was given to this type of work.

Prepared tests were administered regularly by fifty-five per cent at monthly, or more frequent intervals. Thus frequency of testing was not as high for revision tests, but still provided a good sampling of work.

The most common type of test was a single topic one, which has the advantages of testing what has been taught immediately before. It does not include constant revision, which is often needed by most classes, if processes are to be remembered. Only twenty per cent used this type of current plus revision test.
In testing arithmetic half of the teachers tended to divorce problems from mechanical work, justified mainly on the difficulties associated with classes of lower intelligence. As a result of analysing this argument it was concluded that it is rather difficult to justify.

Considerable disagreement was found over the use of the arithmetic text book supplied by the Department of Education and much criticism was levelled at its construction, especially with regard to gradation of examples. In at least one school the headmaster encouraged the staff to disregard the text book completely.

In defence of the text book it could be said that it does offer a guide for complete coverage of the syllabus, and is not meant to provide all the data for teaching arithmetic. It is being constantly improved by periodic revision.

**SPELLING**

In spelling, ninety-five per cent of teachers used Schonell's *Essential Spelling List* as a basis for words to be learned. Twenty-eight per cent of teachers set words only from this list. Any one single spelling list does not aim to cover all the words encountered, and so some extension is required.

Generally, routine spelling tests were administered with a weekly or more frequency, by ninety-five per cent of teachers. This high test rate could be due to its ease of testing and marking, apart from the fact that it is a basic
subject. Revision spelling tests were based by most on blocks of words; only fifteen per cent favoured cumulative revision, probably because of the large number of words to be learned by pupils.

The application of spelling is in written sentences and seventy-four per cent of teachers used dictation tests to assess their pupils' ability to apply spelling knowledge in this manner. Of these only thirty-eight did so regularly, showing again some gaps between learning and application of knowledge.

Although no exact figure was obtained, a number of teachers were using spelling tests, as writing tests also. Whether this should be done or not is arguable, but an excessive pressure on writing, during a primarily spelling test, sometimes detracts from the concentration of pupils of lower ability. This may end in confusion yielding neither good spelling nor good writing, and so be an unfair assessment of either.

COMPREHENSION

The nineteen per cent of teachers encouraging essay type questions allowed their pupils more scope in selecting, organizing and integrating comprehension facts. This, however, is of more use to children of higher ability.

The best all round test, in the writer’s opinion, are those of the mixed question type, used by forty-one per cent of teachers, as they do provide material for all levels of ability.
From the figures given it seems that there is an adequate coverage in comprehension, but that it is less systematically done at lengthy intervals than at short ones. If, of course, it is tested at short intervals with some regularity, the necessity for lengthy tests, apart from survey or record purposes is unnecessary, unless techniques of answering have been taught throughout the year and it is these techniques that are being tested. The danger has been pointed out of a comprehension test being just another test of memory for facts. There is also the possibility of tests becoming rather stereotyped and somewhat unimaginative.

The results of the survey suggest that there appears to be scope for considerable rethinking about the construction and content of comprehension tests, on the part of many teachers.

COMPOSITION

In composition the type of topics most favoured were generally related to the kind of class the teacher had - "concrete" topics for lower streams, and more imaginative for upper ones. This subject was regularly tested for routine purposes by sixty-six per cent of teachers, but as it is an arduous test to mark, the irregular testing by the remainder is understandable. As composition is the best means to test effective writing however, this one third of the sample may not have been doing justice to their class in this subject.

Seventy-one per cent marked this work purely on impression.
Composition standards are however difficult to set, and maintain between one teacher and another.

As with social studies, spelling and writing, care must be taken if composition is used as an integral part of other subjects, as did thirty-five per cent of teachers. Remmers\textsuperscript{10} warns against multiple testing, and shows the pitfalls already mentioned, associated with it. For higher ability pupils however, it is a most useful practice, as it requires more effort, and makes them use many skills in an integrated effort. For the slower learner it could be expecting too much at a primary school level.

C.T. Ford's idea of keeping a flexible and if possible individualized programme in composition sums up this argument.

**SOCIAL STUDIES**

Social studies was most irregularly tested. Fifty-five per cent of teachers gave no regular tests in this subject. The amount done, however, could be sufficient, for overtesting could dull this subject which can be made most interesting to children.

Source material is not easily found from any single book, or even a series of tests. This does allow for variety of content, but poses a problem of conveniently covering the syllabus. Brighter pupils can be encouraged

\textsuperscript{10} H.H. Remmers, N.L. Gage and J. Francis Rummell \textit{op cit} P. 272.
to find material, and it is a useful exercise in locating information, not readily at hand.

Test construction consisted mainly of questions requiring short, factual answers, but thirty-six per cent of sample of class teachers allocated marks for work done on projects, researches, and books. Providing instruction has been given to the children on research methods, and possible location of material, this project work should be of great value to pupils of higher ability, but could be sheer drudgery for others. The projects and results acceptable would need to be graded according to the child's level of ability.

WRITING

It was found that systematic teaching of writing falls off rapidly in forms I and II, although the Primary School Syllabuses state that rhythmic exercises should be continued and remedial teaching undertaken. The latter, although often only "rewrites" was frequently undertaken, but the former is largely disregarded. Informal teaching was generally the case with these upper classes.

Regular testing and teaching of writing was done mainly in standard four; in the upper part of the school it was confined chiefly to the lower ability classes. Tests were mainly transcriptions from the blackboard or from printed material, and the results were marked on the teacher's impression. The standards of criteria of course vary from teacher to teacher.
There have been many "fashions" of handwriting in the New Zealand primary schools, such as the Palmer system and at the moment Italic script.

The findings of Ussher\textsuperscript{11} on the comparison between "italic" and "cursive" handwriting in a Dunedin school in 1959 showed that "italic" writing was faster, and at the greater speed just as legible as the cursive method. In addition it seemed a much easier system to teach and gave more pleasure in writing to both pupils and teachers. Fleming\textsuperscript{12} makes the point that handwriting skill is complex and develops slowly compared to the development of other forms of manual dexterity and suggests the need for personal adjustment and mental health as important precursors of good handwriting. There seems to be some need for co-ordination within a school on the type of writing taught, as major variations from teacher to teacher cause emotional upsets, and considerably retard the pupils progress in writing. For some it is a difficult motor skill to acquire and poor writers need all the assistance that they can get. They need to be secure in their already established patterns, and not subjected to continual change.

\textsuperscript{11} John S. Ussher, \textit{A Study of Some of the Relative Merits of Cursive and Italic Handwriting in the Primary School} (Thesis for M.A. in Education; University of Otago, New Zealand, November 1959) P. 72 ff.

\textsuperscript{12} C.M. Fleming, \textit{Research and the Basic Curriculum} (London: University of London Press 1946) P. 94
MARKING SYSTEMS

An analysis of teachers' marking systems in various subjects showed that composition and writing were assessed mainly on a combination of impression plus some analytical scheme, while the remainder of the subjects were marked predominately on a purely analytical basis.

The results of tests were expressed in raw scores by over one third of the teachers. As no additional data was usually supplied, these marks did not reveal very much about the pupils' work. Thirty-three per cent of teachers used a percentage system. This was criticized on the basis of the fineness involved in distinguishing one result as better than another, by one single mark, and also from the lack of indication of a test's difficulty.

A few teachers (maximum eight per cent in writing) relied on remarks only, which do encourage lower groups. The balance of teachers used grades on a five point scale, usually on a class or school basis of normal distribution. This has the advantage of abolishing fine distinctions in work, but may be criticized on the fact that the different grades may mean different standards to various teachers. No indication of test difficulty is indicated either. One of the latest suggestions for marking, the use of standard scores, was not used at all.

Acceptable standards of work were judged by most teachers, on what they considered, in the light of their own
experience and the pupils' past performances, the children were capable of accomplishing.

Twenty-nine per cent set a minimum standard which they expected the whole class to attain, taking into account the classes' past work, while the remaining nine per cent set an objective standard. These last two ideas may have merit for University Entrance classes, but seem rather rigid for primary schools, and do not allow for individual differences, or social promotion.

The practice of "weighting" marks was also examined. This was done in all subjects except writing, and was used to give emphasis to questions considered more difficult than others. Spelling marks suffer from this practice more so than any others, possibly as a reminder of the times in which the Proficiency Examination was set. It does tend to discourage poorer spellers.

There were also fifteen per cent of teachers making little systematic record of test results, which could perhaps cause difficulty for their pupils, should some assessments be required in later years.

The degrees of direct control exercised on class teachers by headmasters, affected only twenty-eight per cent of the sample of teachers seen, but as seventy per cent of teachers had ten or more years experience, this could have been anticipated.
CHAPTER VII

SUMMARY AND CONCLUSION

SUMMARY

In Chapter I, questions were posed for which this investigation was designed to find answers. These questions are re-presented and answered in this section.

Standardised Tests

Practices in the use of standardised tests of attainment, diagnosis, and intelligence, were considered in four questions.

1. What proportion of teachers use standardised tests of attainment, diagnosis and intelligence?

Standardised tests of attainment and diagnosis were used by sixty-six per cent of the sample of ninety-two headmasters and teachers. Of these five were headmasters and fifty-six class teachers. Twenty-nine per cent of teachers used one test only, twenty-five per cent used two tests, while the remaining fifteen per cent used three or more such tests. The most frequently used tests, and the numbers using them are set out in Table XXI.

Intelligence tests were used by ten headmasters, and eleven class teachers, a total of twenty-three per cent of the sample. Of these, two teachers administered tests purely as a requirement of university courses. Apart from one headmaster who occasionally used the Revised Stanford Binet
<table>
<thead>
<tr>
<th>Test</th>
<th>Used by (per cent of sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schonell Graded (Reading) Vocabulary Test</td>
<td>38%</td>
</tr>
<tr>
<td>Schonell Graded Spelling Tests</td>
<td>32%</td>
</tr>
<tr>
<td>Schonell Diagnostic Arithmetic Test</td>
<td>21%</td>
</tr>
<tr>
<td>Burt Graded Spelling Test</td>
<td>13%</td>
</tr>
<tr>
<td>ACER Spelling Tests</td>
<td>12%</td>
</tr>
<tr>
<td>ACER Arithmetic Tests</td>
<td>10%</td>
</tr>
<tr>
<td>ACER Reading Tests (pt. III) Reading for Meaning</td>
<td>8%</td>
</tr>
<tr>
<td>Seville Easy Steps in Arithmetic</td>
<td>8%</td>
</tr>
</tbody>
</table>
Intelligence Scale, the balance of nine teachers and nine headmasters, all used the Otis test. One headmaster and one teacher used the Raven Progressive Matrices Test in addition to the Otis.

The most frequent users were those applying Otis tests, numbering nineteen per cent of the sample.

2. How are these tests used?

Attainment tests were used by fifty-four class teachers and headmasters (fifty-nine per cent) and Table XXII shows the uses made of the results.

**TABLE XXII**

**USES OF ATTAINMENT TESTS**

<table>
<thead>
<tr>
<th>Use</th>
<th>Percentage of 54 Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results not used</td>
<td>22%</td>
</tr>
<tr>
<td>Results used for grouping</td>
<td>69%</td>
</tr>
<tr>
<td>Results used for retesting progress</td>
<td>11%</td>
</tr>
<tr>
<td>Miscellaneous uses</td>
<td>3%</td>
</tr>
</tbody>
</table>

Although several of the tests used by teachers could be applied for diagnostic purposes, as well as attainment, little use was made of this aspect of these tests. Seven teachers used purely diagnostic tests, one of whom misused his as an attainment test. Only three of the seven however, took any
remedial action on the results. Of the thirty-five teachers who used the combination attainment and diagnostic tests, only fifteen per cent used it diagnostically. Even less, (three per cent of the thirty-five users to be precise), attempted remedial work on the basis of faults found.

It seemed clear that much more use could be made of the results, particularly in the diagnostic field.

New Zealand, or local norms were available for some tests but were not used, probably because of their existence not being known in many cases.

Intelligence tests, used by eighteen of the sample, were put to uses shown in Table XXIII.

No specific misuse was noted, but a wider use and understanding of the results would be beneficial.

3. What relationship exists between the experience and qualifications of the teachers, and the testing done?

In diagnostic and attainment testing, there was little relationship between either the qualification or the experience of teachers, and the amount of testing undertaken. Some testing was done by an approximately equal number of teachers above (total 25) and below (total 31) the mean teaching experience of fourteen years. The volume of testing by the same groups averaged two tests each, as the thirty-one in the younger group administered fifty-six tests, while the twenty-five more experienced teachers gave fifty-five tests.
<table>
<thead>
<tr>
<th>Uses of Intelligence Tests</th>
<th>% of Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grouping and Streaming within Schools</td>
<td>38%</td>
</tr>
<tr>
<td>Grouping Within Classes</td>
<td>10%</td>
</tr>
<tr>
<td>Check on Scholastic Performance</td>
<td>33%</td>
</tr>
<tr>
<td>Obtain Achievement Quotients</td>
<td>19%</td>
</tr>
<tr>
<td>Complementing Other Tests</td>
<td>19%</td>
</tr>
<tr>
<td>For University Courses</td>
<td>10%</td>
</tr>
</tbody>
</table>
The younger and more academically qualified teachers
were found more in intermediate schools, but there was less
need for them to test as they had results supplied to them.
There was also no significant relationship between the types of
tests used and experience of the teachers, as it was shown
that of the users of the older pre-war tests exactly half
were above the mean teaching experience of the sample.

A similar pattern was formed for intelligence testing.
Disregarding the two teachers who administered tests purely
as part of their university courses, of the remaining nineteen,
seven of the ten headmasters and five of the nine class
teachers were qualified by university courses, to administer
and interpret intelligence tests. Thus academic qualifications
for testing were held by only sixty-three per cent of those
using intelligence tests. Intelligence testing is not
therefore particularly related to academic testing qualifications.
A definite relationship was shown in that experienced teachers
tested more than those below the mean experience of the
sample. Only two testers had less than fourteen years
experience. This was due to the predominance of headmasters
testing, which is advantageous in that the administration
of the test would be the same for all groups, if done by
one person.
4. To what extent do schools follow a definite, adequate testing programme involving principles suggested by authors of textbooks on testing?

In intelligence testing only four of the seventeen schools (two intermediate and two non-intermediate) understood a testing programme which reached the very minimum programme suggested by Remmers, Gage and Rummel.

Troxler's five point programme suggested as a criterion for attainment and diagnostic testing, was not met in full by any single school visited. The nearest approach was found in one intermediate school.

It is quite apparent that schools lagged behind current overseas opinions on testing programmes.

Teacher Constructed Tests

Tests constructed by class teachers were considered in six questions.

1. From what sources do teachers draw materials for tests they construct?

Material for the tests came from a variety of sources, according to the subject. In arithmetic, the two main sources of supply of examples was balanced evenly between the text book, and the teacher's own examples, which in spelling words were drawn mainly from the Schonell or a teacher augmented Schonell list.

The largest source of work for comprehension tests came from the teacher's own ideas. A somewhat small number
selected tests from either the Teachers' Monthly Guide" or the "Reading for Meaning" series. Material for social studies tests was gleaned from many indefinite sources, difficult to analyse, while writing tests were composed mainly of transcription material supplied by the teacher. Composition tests were based on topics, supplied again in most cases, by the teacher.

2. **With what frequency do teachers assess attainment with tests they have made?**

Generalizations are difficult, but it may be said that tests are used as frequently as is consistent with comprehensiveness, continuity and pertinence to the subjects tests. Testing however, is more frequently, more efficiently and more reliably done in some subjects than in others. Some subjects such as arithmetic and spelling, lend themselves to simple quick testing, while others such as composition are tedious to mark.

Daily testing in Arithmetic was carried out by sixty-five per cent of class teachers. Eighty-five per cent tested at weekly or more frequent intervals. Arithmetic revision tests were given by fifty-five per cent of teachers at monthly or less intervals.

In spelling, daily testing was given by half the teachers but ninety-five per cent were included among those who tested at weekly or less intervals.
The most common comprehension test was a short one given weekly by fifty-five per cent of the class teachers, while longer and more searching tests were most irregularly applied; some twenty-two per cent did use them at monthly intervals. There was however, overlapping between these categories.

Short, routine composition exercises were set weekly by thirty-two per cent of class teachers, but this type of work was done only irregularly by thirty-four per cent. Fuller and longer compositions were administered as an integral part of other subjects by thirty-five per cent, although the majority of teachers, (fifty-five per cent) tested such work irregularly. There was definite evidence of irregular rather than regular marking of composition.

Social Studies was routinely tested only by one person. It was the least frequently tested subject investigated. Forty per cent of the teachers tested irregularly, and only one teacher tested at more frequent intervals than one month. Sixteen per cent tested for surveys only.

The practice of discontinuing writing lessons and tests in the upper school affected its test frequency in this subject. Twenty-five per cent tested routine writing daily, mainly in Standard Four classes. Sixty-six per cent tested no routine writing at all. The most common revision test
was the monthly one, where sixteen per cent set tests. Twenty-nine per cent of class teachers entered no marks in the register for writing, but some on occasion assessed writing from other written work.

Standards of assessment seem to be largely individual to the teachers' concerned.

3. How soundly based are the construction and content of these tests?

Ideally, the thoughtful planning of a test, according to Thorndike and Hagen\textsuperscript{1}, involves several steps; to define the objectives that are to be appraised, analyse these objectives into more specific components, and then translate them into terms of pupil behaviour so that it is possible to indicate the skill, knowledge and understandings that a pupil should show if he really has achieved those objectives of instruction.

More detailed work then follows:— plan what the student is expected to do as a result of these objectives, outline the content of the test, and then prepare a test blue-print. Next the teacher must decide on the relative emphasis to be given for item difficulty. Then the actual test items are prepared bearing in mind the advantages of essay or new type

\textsuperscript{1}Robert L. Thorndike and Elizabeth Hagen - Measurement and Evaluation in Psychology and Education - (New York, John Wiley and Sons Inc., 1957) p.29 ff
tests, and the principles involved in each of these.

They go on to say that essay questions should not be too vague, general or comprehensive. It should be decided in advance what factors are to be measured, model answers should be prepared showing what points are derived and credits to be allowed for each. All answers to one question must be marked before going on to the next, and factors such as handwriting should not affect the value of the answers.

"New type" tests, should be prepared with more questions than are required, and reviewed later, the layout should be well organised, similar items should be grouped, and the directions clear. Finally test results should be analysed.

If all the tests made by the teachers seen were measured against these criteria, not one would meet them. To prepare even one test under such conditions would take an immense amount of time and effort, not generally warranted in most cases at a primary school level.

Although Thorndike and Hagen's full demands are rather unrealistic for most teachers in a primary school here, they do point out a need for some thought to be given to test construction.

In general the tests produced by the teachers of the sample, although routinely constructed, and often unimaginative in test items, did assess attainment sufficiently well for class
placements and promotions. The relative difficulty of items in the tests were usually emphasised by the distribution of marks. Some teachers attempted question variety, and an increasing number are issuing separate test papers to each pupil.

Tests although important, should not dominate or detract from teaching. More thought could be put into test construction; increased motivation of the pupils would probably result. More time could be spent in their preparation if one teacher could set the same examination for several classes, and thus have fewer tests to prepare, and so construct a better test. This implies however, some form of co-ordination of curriculum content, but could lead to better testing, than is at present the case.

4. How do the various marking systems compare with modern theory?

Apart from composition and writing the most commonly used marking system in the subjects investigated was the recording of raw scores, which in themselves have no significance without additional data for interpreting them. Thirty-eight per cent of class teachers did no more than this.

Percentages were used by thirty-three per cent of all class teachers in the same four subjects. A few teachers gave no marks at all, relying on written comments as encouragement to low ability pupils.
Composition and writing were marked by forty-eight per cent, and thirty-five per cent of class teachers respectively on a five point scale.

In two schools (for survey purposes) a five point scale was applied to all subjects over each class level throughout the school, on a normal curve distribution, and the number of "A's" to "E's" allocated for each class in advance of testing, with due allowances made for exceptional cases. Other individual teachers attempted to allocate a five point scale, on a normal distribution within their own class.

Allocating marks as a percentage, according to Remmers, is a very early technique, and without an analysis of the results, is most unsatisfactory. Unless the distribution is analysed, the difficulty of the test cannot be known, and the fallacy of differentiating one result from another by one point, is obvious.

The five point scale, which Remmers states is the most common system in use, has the merits of stopping this impossibly fine distinction, but often suffers from a lack of uniform definition of each grade among various teachers.

The number of teachers using a five point scale, varied in different subjects. These users were however, applying a marking system, nearer to modern theory, than were the remainder.
The practice of conforming marks to a normal curve, with predetermined percentages of students allocated to each grade, may be satisfactory with a large number of unselected children, but a very real difficulty could be met when this system is applied to one class only.

It is quite possible under this system, for a pupil to receive an "A" assessment for work which may gain only a "C" rating in another class of the same grade, in the same school, due to the differing levels of ability in the classes. Thus the allocation of grades on a normalized scale could be false.

Not one teacher seen, converted raw scores to standard scores for any record purposes. This is unfortunate, as it does provide the most fair means of comparing marks.

5. What influence do headmasters exert on teachers, tests and examinations in their schools?

Direct, formal influence by headmasters was exerted on only ten per cent of the sample of teachers. This took the form of regular tests being given by the headmaster, and was applied mainly to younger teachers, thus setting standards for the teachers' on tests.

Among twenty-three per cent of the teachers seen, indirect control was exercised through the headmaster collecting books and visiting classes.
The balance of teachers, sixty-two per cent, reported that apart from infrequent visits and occasional sampling of pupils' work, headmasters did not exert formal control.

From this it appears that considerable autonomy is given to teachers, but as only thirty per cent of teachers had less than ten years' experience, this freedom could be expected.

In examinations for the half yearly surveys, headmasters often exercised control in arithmetic as Table XXIV shows.

**TABLE XXIV**

| Exams set and marked by Headmasters | 14 |
| Set and marked by Teacher, checked by Headmaster | 18 |
| Set by Headmaster, marked by class teacher | 10 |

In other subjects, little direct influence was exerted by the headmasters. All schools reports, however, must be signed by them, and this in itself is a check on teachers' ratings.

6. **How adequate is the recording of results?**

Forty-six per cent of teachers entered the results of all tests undertaken, twenty-six per cent recorded their idea of a fair selection, and thirteen per cent marked at regular time intervals. The remaining fifteen per cent recorded
irregularly and inadequately.

In so far as it is important to have a number of marks on which to assess a pupil, the vast majority of teachers recorded adequately as a good coverage of test results were entered in the register of marks. This sampling of pupils' work gave sufficient data for reports to parents, promotions, reteaching where necessary, and assessments of the pupils' standards of work.

The fifteen per cent who did not keep an adequate record of work, failed in these respects, for they probably had to rely excessively on their own memories, and their general impressions of the pupils' mastery of various subjects. Furthermore, no marks records, which could be of later use to the pupils concerned, would be available, should they be required.

CONCLUSIONS

In standardised testing of all types, much more could be accomplished if teachers were adequately trained in the principles and techniques of testing. The small numbers using intelligence tests were not all academically trained to do so. However, this need not mean that the tests were misused. Simpler group intelligence tests, as well as other standardised tests can be administered by persons with limited training in their use. (This is done in the New Zealand Armed Services where such people are supervised by fully trained personnel). Such tests are useful tools to teachers who understand them, but their potential is not being exploited.
Some of the tests used do have New Zealand or local norms, but the existence of these norms is little known. More New Zealand standardisations would be useful. Attainment norms, based on the pre-war English school system are not so applicable here, and should not be used when well standardised New Zealand ones, such as some of the A.C.E.R. series, are available. Local norms may have to be treated a little more cautiously, as the sample numbers may not be large.

Little use is being made of diagnostic tests. Many are simple to administer and of considerable value in accurately tracing errors, thus saving time on the teacher's part. Even the set of Sevilles' supplied by the Education Department to schools, is in many cases not used. Although some skill may be needed in diagnosing some reading faults, very little is required in arithmetic. A careful perusal of most manuals, is often sufficient to enable the teacher to diagnose the general types of errors in arithmetic and spelling. Much more could be done with even the work of Seville, which is particularly suited to New Zealand conditions.

There has been noted, that among the teachers trained in testing, a large proportion make no use of this qualification. This could be due to the organisation of the schools in which they teach, or an indication of the attitude of these teachers to testing. There are many cases of specific difficulties in basic subjects which are at present referred
to specialist agencies, that a teacher qualified in testing, could deal with in the school.

The most common one is reading difficulty. The writer estimates that of all cases referred from schools to the Education Department Psychological Service in Christchurch during 1960, at least one-quarter had reading difficulties. This estimate was confirmed by other staff members.

During the mid-term of this, and past years', under the auspices of the District Senior Inspector of Schools, a course in remedial reading has been given to about fifteen teachers in Christchurch. There is also a course on reading incorporated in the Diploma of Education, at the University of Canterbury. The effects of these courses does not seem apparent at present, perhaps due to lack of opportunity to use their skills, and the comparative recency of the latter course.

Such trained teachers, should be able to locate reading disabilities early, and be available to other less qualified staff members, for advice on treatment.

With a small amount of study on the skills involved in reading, teachers with testing qualifications, given adequate time, could diagnose and advise on cases within their own schools, thus giving a better service than is at present available, and lifting the burden placed on special services.
Related to this is the lack of co-ordination of test programmes seen among schools. Systematic programmes were the exception but from the experience gained by four schools who have current programmes, schemes could be built up to meet needs of other schools. It would be most desirable of course, for teachers to be in sympathy with such plans if they were to be involved in testing. Lack of training appears to be the key to this.

Those teachers who do use standardised tests, appear to make little permanent record of the results. Assuming that the tests are given correctly the results are part of the pupils' school performance, and so are valuable to other teachers. Intelligence test results should be recorded to see the scores, dates of testing and types administered. These results could save time on other teachers' parts, and provide useful data should further testing be given at a later date.

In teacher constructed tests it was seen that there was a definite tendency to separate knowledge from its application, particularly in arithmetic and spelling. Knowing how to do a mechanical sum is not in itself enough, for in daily life one seldom meets problems in such a form. Similarly spelling is normally met in sentences. Words have to be spelled correctly as they are used in writing. The testing of lists of words seems artificial.
Over one-third of the teachers marked composition as an integral part of other tests or subjects. Writing and composition were used likewise by a smaller number.

It is argued that this practice teaches the pupils to meet the needs of completing work accurately in all respects, as will be demanded at high school, and in the community. A child may have many accurate answers, but unless these can be expressed in a form easily read and understood, he is at a very big disadvantage. To ask for completely integrated work, means exercising a combination of skills at one time, and it is only by practice that this will be accomplished.

On the other hand except for pupils of high ability, it may be asked whether the primary schools is the level at which this should be attempted, as it demands the child's concentration on several skills at one time. This does not seem reasonable for low ability children, as the constantly poor results would discourage them from further efforts. For the average pupil, Standard Four may be a little soon for much of this, but if suitable combinations are selected, it could be worth while. The feature which is being primarily tested, should however, be the major issue, with the others as more incidental. Thorndike and Hagen warned against allowing factors, such as writing to affect the value of the answers to tests.
Care should be given to do suitably related topics, if this type of test was favored. In the writer's opinion, social studies topics, and composition could go together as one is the medium of expression of the other. Writing and spelling however, do detract from each other and require a divided attention, and are not as applicable.

In general, this type of programme hinges on the ability of the individual child, and the selection of topics.

Test construction by teachers is rather stereotyped and lacks variety. Questions are frequently too factual, often requiring little thought on the pupil's part. Some teachers attempted to grade questions and vary the form, but mostly no great thought appeared to have gone into test construction.

In spite of these facts, however, teachers do in general provide adequate assessments for primary school purposes. To construct tests such as those set out by Thorndike and Hagen would give magnificent tests, but leave no time for teaching. With more planning involving a lesser number of tests on the part of each teacher, better tests could be produced, and given to several classes, as was done in one school seen.

IMPLICATIONS

During teachers' training more could be done to enable teachers to use and interpret simpler group intelligence tests,
together with the most useful attainment and diagnostic tests. Some time would have to be given to the principles as well as the techniques of tests and testing, so that tests not met with in such a course, and those yet unpublished, could be evaluated, and correctly used. If these matters were dealt with at the beginning of a teacher's career, they would become a part of his teaching techniques, and not an appendage to be studied later.

For those teachers already certificated, but possessing no testing qualifications, encouragement could be given in some form, by the Education Department, for them to study suitable courses in education or psychology offered by the various Universities. The number eligible for these courses is however, often limited by academic prerequisites, held only by a small number of teachers.

A further alternative would be "in-service" courses, paralleling in content, the suggestions made concerning teacher trainees. Those teachers already fully qualified should be available to help and supervise those who took such courses.

It seems that a wider dissemination of available New Zealand standardised norms, (A.C.E.R., N.Z.C.E.R. series) for some attainment tests in common use, could be made, as many teachers are unaware of their existence. If these were published in booklet or pamphlet form, and widely publicised, a much needed service would be rendered.
The value of diagnostic tests as a means of finding pupil’s errors, and teacher’s weaknesses, needs stressing. This would be covered in courses on testing, but apart from this, the easier ones, such as Seville’s arithmetic, are self-explanatory, if the manual is carefully read. All teachers could use it without special training. It would seem that a good deal of money will be wasted if more use is not made of the free issue made to schools of this test. Some education on the advantages of such tests could well be done, in teachers’ groups.

Teachers qualified in testing could be used to considerably more advantage than is currently done. Those who do not test, although qualified to do so, could be used to help and supervise other teachers who wish to test, but lack knowledge of testing principles and techniques. The tests used would naturally be of a simple type. Encouragement from headmasters would aid this, for on them depends the amount of time which could be devoted to testing. If a school could be so organised that a qualified teacher could be released from classroom duties, to undertake limited case work, in reading and intelligence testing for example, and so be responsible for some diagnostic work and remedial suggestions, schools would meet pupils’ needs better than they are now.

Testing programmes could be extended if qualified and interested teachers were allowed to conduct them. The results
of such yearly tests could be valuable to teachers and provide a yardstick on school standards, that would be of interest to the community, provided it was done with discretion, and was not used as a measure of a school's teaching efficiency. Time, together with trained personnel are again factors, which although not insuperable, create difficulties.

Test material, such as the Wechsler Intelligence Scale for Children, or the Revised Stanford Binet Intelligence Scale are expensive. There could well be a central pool from which qualified teachers could obtain such test materials, as required. At present the University, and the Psychological Service appear to be the only lending sources. More testing means more tests will be required. Teachers could not fairly be expected to purchase these themselves.

Some more specific direction could be made on the recording of standardised test results, as few are at present preserved in permanent form on a cumulative record card. Should standardised testing become more prevalent, details of this nature would be necessary.

There is a tendency to segregate parts of subjects when they should be integrated, such as mechanical and problem type arithmetic. This depends somewhat on the ability of the class taught, but where these subjects are met in life as integral wholes, it seems that they would be better taught
in this manner. More work could be conducted in this manner. Preparation and thought would be required in lessons and test construction, but the extra trouble would be reflected in a more practical and meaningful syllabus.

Teachers could well be given more exercises in test construction and marking during their training. Very few even approached the standards set by Vernon for "new type tests", and marking systems generally do not give a true idea of the relative value of the test result. For older teachers, in-service training could help here also.

It would seem then, that much could be done to correct these shortcomings by adequate training in the Teachers' Colleges, while an in-service training scheme, refresher course or similar schemes, could well be provided for practicing teachers not familiar with tests and testing. Tests, while by no means the most important part of teaching, are measures and helpful tools, and as such deserve more recognition, use and thought than is at present given.

FURTHER RESEARCH

As a result of this investigation, several further projects seem to be necessary.

Some research needs to be done to provide attainment and diagnostic tests of spelling, based on Arvedson's N.Z.C.E.R. list. This would then provide a test for the spelling list now peculiar to this country.
Further New Zealand standardisations could be made of tests in common use. As no New Zealand wide figures are available on the types and amounts of standardised testing done in this country, some form of national survey would help to show trends, and give backgrounds against which the more segmental investigations could be collated. Such raw information is available from R.Z.C.E.R. and other test suppliers.

More detailed evidence could be collated about the usefulness or otherwise of school arithmetic text books. Something similar should be done to collect social studies material, at present rather scattered, into a single volume, to supplement or replace the now unobtainable text published by Christchurch Teachers' College. A series of books suited to the curriculum would be of assistance to many teachers.
BIBLIOGRAPHY


Primary School Syllabuses


Rodgers, Cyril A. Measuring Intelligence in New Zealand. Auckland University College, monograph series, No. 2. 1956.


