Errors in the marking of printed tests

Australian Council for Educational Research (ACER)
AUSTRA LIAN COUNCIL FOR EDUCATIONAL RESEARCH.

INFORMATION BULLETIN

NO. 9

SEPTEMBER 1947

ERRORS IN THE MARKING OF PRINTED TESTS

147 Collins St., Melbourne C.I.
The following comments are the result of preliminary analysis of some errors made by a random sample of teachers in marking the test papers of pupils who, in 1946, took part in the testing programme in connection with the Curriculum Survey being conducted by the A.C.E.R.

All schools where children did the tests were asked to select a certain number of papers at random from each grade and forward them to the A.C.E.R. The numbers sent in from each school were roughly proportional to the size of the school. Thus large schools sent in more papers per grade than the smaller ones.

On arrival at the A.C.E.R. office, the papers for each test were sorted by States into equivalent grades, and approximately 200 papers chosen at each of the five grade levels tested in the survey. The number chosen from each state was roughly proportional to the number of pupils doing the tests in that State.

The papers taken out for re-marking from each bundle of tests were not related in any way to those from any other test. Each such set of papers was an unselected group of the total available and necessarily came from a number of different schools. Thus, the test papers in any one State in two different tests could have come from two different groups of schools, and in any one set of papers all schools in the State had an equal chance of selection.

It was hoped to obtain sufficient volunteers to re-mark all five groups; as there were not sufficient volunteers, the central grade group (corresponding to Grade 6 in N.S.W. and Victoria) was selected and the papers from that group were re-marked. The re-marking was done partly by first-year Psychology students of Melbourne University and partly in the A.C.E.R. office.

The purpose of the re-marking was primarily to ascertain -

1. Whether the average error in marking on any particular test was likely to effect the average scores calculated from the results submitted by teachers.

2. Whether there were constant errors in any of the tests due to inadequate marking instructions.

3. What types of tests produced the highest proportion of marking errors.

Errors were classified in four groups -

A. Answers right but marked wrong by the teacher.

B. Answers wrong but marked right by the teacher.

C+. Errors of addition (or of calculation where an adjustment for mistakes was required) which had the effect of lowering the child's true score.

C-. Errors similar to C+ errors where the effect was to raise the child's true score.
The net effect on the total score for all children is 

\((4+ C+)-(B+ C-)\). This total may be either positive or negative 
according to whether the net effect of the teachers' errors was 
to make the total score lower than it should have been, or higher.

2278 papers were marked in the tests listed in Table 111. 
In these papers errors were found as follows:

<table>
<thead>
<tr>
<th>Type of Error</th>
<th>Total Errors of Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>328</td>
</tr>
<tr>
<td>B</td>
<td>372</td>
</tr>
<tr>
<td>C+</td>
<td>301</td>
</tr>
<tr>
<td>C-</td>
<td>303</td>
</tr>
</tbody>
</table>

The number of separate papers containing errors was 510 and the 
percentages for the states range from 16.7% in South Australia 
(number of papers \((N) = 261\)) to 31.4% in Tasmania \((N = 261)\), the 
overall percentage being 22.4%.

Counting the papers re-marked in one State in one test 
as the unit case, the effect of the errors was as follows:

<table>
<thead>
<tr>
<th>No. of Cases</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>Score raised above true level.</td>
</tr>
<tr>
<td>29</td>
<td>Score lowered below true level.</td>
</tr>
<tr>
<td>7</td>
<td>Errors balanced out.</td>
</tr>
</tbody>
</table>

The net effect on the total score on the papers marked, considered 
by separate tests, was as follows:

<table>
<thead>
<tr>
<th>Test</th>
<th>No. of papers.</th>
<th>Net effect</th>
<th>Effect on average score</th>
<th>No. of papers with errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed of Reading.</td>
<td>187</td>
<td>+6</td>
<td>.03 mark</td>
<td>7</td>
</tr>
<tr>
<td>Reading for Meaning.</td>
<td>181</td>
<td>-20\frac{1}{2}</td>
<td>.11 mark</td>
<td>23</td>
</tr>
<tr>
<td>Word Knowledge.</td>
<td>194</td>
<td>+2</td>
<td>.01 mark</td>
<td>36</td>
</tr>
<tr>
<td>Sentence Structure.</td>
<td>189</td>
<td>+69</td>
<td>.37 mark</td>
<td>58</td>
</tr>
<tr>
<td>Word Usage.</td>
<td>193</td>
<td>-31</td>
<td>.15 mark</td>
<td>118</td>
</tr>
<tr>
<td>Addition</td>
<td>200</td>
<td>-2</td>
<td>.01 mark</td>
<td>16</td>
</tr>
</tbody>
</table>
**TABLE III, contd.**

<table>
<thead>
<tr>
<th>Test</th>
<th>No. of papers</th>
<th>Net effect</th>
<th>Effect on average score</th>
<th>No. of papers with errors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(+ = Score raised above true mark)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(- = Score lowered below true mark)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtraction</td>
<td>202</td>
<td>-18</td>
<td>.09 mark</td>
<td>25</td>
</tr>
<tr>
<td>Multiplication</td>
<td>199</td>
<td>+13</td>
<td>.06 mark</td>
<td>17</td>
</tr>
<tr>
<td>Division</td>
<td>198</td>
<td>+7</td>
<td>.04 mark</td>
<td>36</td>
</tr>
<tr>
<td>Arithmetic Terms and Relations</td>
<td>197</td>
<td>-17</td>
<td>.08 mark</td>
<td>63</td>
</tr>
<tr>
<td>Spelling</td>
<td>141</td>
<td>-19</td>
<td>.13 mark</td>
<td>70</td>
</tr>
<tr>
<td>Arithmetical Processes</td>
<td>190</td>
<td>+35</td>
<td>.18 mark</td>
<td>38</td>
</tr>
</tbody>
</table>

It will be seen that -

(a) In six (6) tests the effect was to raise the score.

(b) In six (6) tests the effect was to lower the score.

(c) In no test was the effect equal to a difference of 1 in the average mark awarded to the papers marked, and in only five of the 12 tests did the difference equal or exceed, .1.

The tests in which the error was greatest were Reading for Meaning, Sentence Structure, Word Usage, Arithmetical Processes and Spelling.

In the Reading for Meaning test, where two answers were required for each of 30 paragraphs, an evident lack of clarity in the instructions led some teachers to award marks on the basis of paragraphs and not separate answers. The effect was to halve the true score. The net effect of such errors in the 181 papers marked was to lower the total score on these 181 papers by 20½ marks.

In the Sentence Structure test, teachers were required to subtract one third of the wrong answers from the total right answers. Inaccuracies in this process, or failure to make the adjustment, gave this test the highest error percentage of any.

In the Word Usage test, where pupils were required to underline the right words, out of five possible, which fitted a gap in a sentence, and teachers had therefore to mark every response given, finally subtracting wrongs from rights, the errors were of all types. This type of test is likely to be a fruitful source of error in marking.

In the Spelling test, the most common type of error made was to take a wrong maximum for the number of right answers. Teachers who marked errors only and subtracted this from the maximum possible therefore affected the marks of all their pupils if the maximum used was incorrect. There was evidently a lack of clarity in the instructions for this test.
In the Arithmetical Processes test, the errors were made chiefly in a section where children were asked to relate a series of measures (ton, foot etc) to a series of commodities (firewood, ribbon etc). In some cases the method used to indicate the correct answer was not that asked for, but there was no doubt that the correct answers were given. As this section carried 8 marks, failure to credit a child with his answers meant that the recorded score was often appreciably lower than the real score. The fault probably lay here in the instructions given to teachers to adhere to the answers given in the key. Some teachers interpreted this to the very letter - even to the extent of disallowing a plural although there could be no doubt of the child knowing the right answer.

The general conclusions arrived at from this brief analysis are:

1. The net effect of errors in marking has not affected the average score by a whole mark in any test at the 6th grade level.

2. The large number of papers with marking errors shows the need for great care in marking a pupil's papers where such marks are part of his permanent school record and where tests are used for clinical and diagnostic purposes.

3. In tests where some system of marking is adopted other than the simple totalling of correct responses, great care must be taken to make the instructions crystal clear. Teachers marking such tests must make sure that they understand the instructions and apply them correctly.

4. In those cases where some latitude is possible too literal adherence by the teacher when marking to the exact form of answer given in printed answer booklets may penalize the individual child.

Several points arise from the first of these conclusions -

1. There will be, in a test covering five successive grades, fewer questions answered by the 4th Grade than the 6th. It is, however, considered improbable that the number of errors in marking in proportion to the scores would vary appreciably from one grade to another.

2. It is not possible to determine whether the net effect of the errors would differ because of the difference in number of questions answered by say, the fourth and eighth grades. The probable effect would be to lower the incidence of each type of error proportionately, so that the net effect should remain approximately the same.

3. The papers re-marked in the tests discussed were from a random group of schools and therefore from a random group of teachers. Conclusions drawn from their marking may reasonably be considered valid for other similar random groups even if those groups taught in different grades.
4. The effect of adding random errors of measurements to recorded scores is generally to leave the mean comparatively unaltered, but to somewhat increase the standard deviation of the scores. The errors of measurement due to marking errors, however, are not all strictly random errors. Further analysis is needed to ascertain the effect of the errors on the standard deviation of the scores and therefore on the standard error of measurement.

Owing to the limitation of the re-marking to the papers from one Grade, there were insufficient papers marked in each test in each State to justify presenting detailed interstate comparisons test by test.

For each state however, an examination was made of the net effect of marking errors on each test. In those cases where this appeared to give an average error of 1 mark or more to the papers marked, a further check was made. In all but one of these, it was found that the large error was due to misinterpretation of instructions for scoring in one or two schools. This misinterpretation had been noticed in the checking of the result sheets in the A.C.E.R. office and the scores had either been revised after referring to the school, omitted from the tabulation of marks for the State, or revised by the A.C.E.R. staff where it was obvious what the real score was (e.g. when a score in the Reading for Meaning test was shown as $8\frac{7}{12}$ and it was apparent that the teacher had halved the true score). In the remaining test - Sentence Structure, Tasmania, it was found that the large error in the sample of papers marked was due to the action of one teacher in adding to the right answers one third of the number of wrongs, instead of subtracting. A check of other papers submitted from Tasmania showed that in 16 other schools from which papers were examined, no other school had made this error. As this school had only seven pupils in the tested groups with no more than 3 in any grade, and as the net effect of all other errors in the re-marked papers was no more than .1 of a mark per paper, it was thought that the effect of the error of this school on the totals would be small.

In three other tests where in Tasmania the net effect was nearly 1, more papers were also marked. In the original sample, the number of papers re-marked for this State was small and the effect of the errors made by one teacher in a large school from which two or three papers were included, was therefore proportionately greater. In each of these cases, the average error dropped appreciably with the addition of more schools and more papers to the sample re-marked.

Supplementary Note.

An analysis of 200 papers (100 from Grade 4 and 100 from Grade 8) of the General Ability test, remarked for other purposes, gave the following information.

Grade 4 - 25 papers with errors. Net effect, to raise score on the average by .11 mark above true figure.

Grade 8 - 27 papers with errors. Net effect, to lower score on the average by .30 marks.
of an amazing uniqueness of a child, indicative of some distinctive aptitude, interest or ability, suddenly manifests itself so that it can be visible to the naked eye or assessed by the tape-measure of those who crave so much for measurement today, and made to form a basis for the classification of children for secondary education. I would like to see us review this position anew. I am not convinced that we are yet sure of the right age at which to divide children for secondary education.

There is something rather menacing in the story that I have told you - this story of the philosophic rationalisation of a situation - because, although there are doubts in the minds of many as to whether the separation at 11 is sound, as to whether we ought not to think of some other basis, the general Hadow thesis is even now in most areas, being carried over into the new Act, and as I see it, in many an area, the Butler Act is being launched in a sea of philosophic insincerity.

Although the picturesque phrases may, to some extent, be discounted, it seems evident that the doubts of the psychologists are now shared by some educational administrators; whether or not this speech caused others to reason in a similar way, or whether it was merely a more vivid expression of a widely hold view, it has been the fore runner of similar expressions of opposition to transfer at 11 plus.

This change in view is reflected in several recent Development Plans of the Education Authorities. In its issues of 22/29 August "Education" published a report of the Northumberland Education Committee on the approach of the Authority to the problem of assessment of children at 11 plus for the purpose of transfer to the most suitable type of secondary school. The Report states:

"It is clearly impossible to forecast at the pre-adolescent stage of 11-12 years the changes which will take place during the next few years. It is clear also that the testing programmes carried out at the age of 11 plus can only be limited as there is insufficient evidence to show that special abilities and interests which may be of the greatest importance inadolescent and adult life, have developed sufficiently at this early age to be capable of being measured with any validity."

In spite of these views it is apparently impossible to change the existing practices for the present time, for the Report continues,

"The situation, however, remains that the stage when this decision has to be made is between the ages of 11 and 12 years, and however easy is made the process of transfer between schools at a later stage, it is obvious that it is in the best interests of the child to be transferred to the most appropriate school at the recognised stage of transfer." In other reports which recognise this problem a compromise is found through the Norwood suggestion that the first two years of all secondary courses should contain a common core of basic subjects so that transfers could be made from one type of school to another if special abilities were revealed. It seems that the administrative difficulties have prevented the more radical proposal of raising the age of transfer. However, though such a change would involve difficulties at present the position will be more complicated once the new schools have been built; if a change is not made soon, it may be deferred for a generation.
B. THE ORGANISATION OF SECONDARY EDUCATION

In its report of secondary education the Norwood Committee envisaged three types of school, the grammar, the technical and the modern, a tripartite division which has been favoured by the Ministry of Education in its official reports. The functions of these schools are best described in Pamphlet 9 of the Ministry of Education, *The New Secondary Education*.

The Modern School will cater for the majority of children who learn most easily when dealing with concrete things and following a course rooted in their own day-to-day experience. At the age of 11 few of these children will have disclosed particular interests and aptitudes well enough marked for them to require any other course. The majority of such children will do best in a school which provides a good all round education in an atmosphere which enables them to develop freely along their own lines.

The Technical School will cater for those who at an early age have decided to make their careers in branches of industry or agriculture requiring a special kind of aptitude in science or mathematics. Others may need a course longer and more specialized than that provided in the modern school, with a particular emphasis on commercial subjects or art.

The Grammar School will cater for those whose ability and aptitude require the kind of course with the emphasis on books and ideas. Such students who are attracted by the abstract approach to learning will normally stay at school long enough to benefit by the sixth form work and a high proportion may be expected to go on to the university.

The Norwood Committee considered combinations of these types.

The idea of a multilateral school combining three types of education was rejected, as it was held that the success of the technical school depended on its close association with local industry, a liaison which would be difficult to maintain unless the school were an independent unity. Limited approval was given to a combination of grammar and modern schools but it was held that larger schools of this nature prevented the personal contact between headmaster and each boy.

Although the Norwood tripartite division was criticized by those who feared that a class system of education was being preserved, the accession to office of the Labour Government in 1945 did not alter the implicit approval of this division by the Ministry of Education; only after a number of L.E.A.s had declined to accept the Norwood recommendations did the Ministry in Circular 144 (June, 1947) recognize the other types of organisation.

This pamphlet succinctly defined terms which had been loosely used in controversy. In view of the general approval accorded to these definitions they are reproduced here.

Bilateral School means one which is intended to provide for any two of the main elements of secondary education, i.e. modern, technical or grammar, organised in clearly defined sides.

Multilateral School means one which is intended to cater for all secondary education of all the children in a given area and includes all three elements in clearly defined sides.

Comprehensive School means one which is intended to cater for all the secondary education of children in a given area without an organisation into three sides.
The suggested advantages of the multilateral schools have been well summarized by W.T. Stevenson in the *Journal of Education*, November 1946. He suggests that they are:-

(a) The child would not be "labelled" at the age of 11 years and requirements for examinations at the end of the primary school stage would disappear.

(b) Special abilities at 13 and 14 years would be known and provided for.

(c) Transfers of pupils would be easy at any stage.

(d) Each block would have facilities for the highest achievements in its sphere.

(e) Each child, however poorly endowed, would mix at times with pupils of higher capacity. "Any social activities would be common to all.

(f) The staff would be large enough to cover all child interests and abilities.

(g) The work of the teachers would be more varied and interesting.

(h) Special education for all children would find its rightful place.

A survey of other statements indicates that the demand for multilateral and comprehensive schools seems to rest on social grounds and administrative expediency as much as on purely educational grounds. Though a system of scholarships opened the schools to all classes to a limited extent the existence of fees closed them to some section of the community but left them open to less able children of wealthier families. Many feared that the tripartite division would continue that system; the best alternative seemed to be a non-selective secondary school, which, while preserving a common core, curriculum for all, would, at appropriate stages, provide a variety of choices for those of widely differing abilities. This view has been expressed in a resolution (6) carried at a Labour Party Conference as early as 1942 and similar resolutions were agreed to in subsequent years.

Multilateral schools have been favoured on administrative grounds by those L.E.A. which have been faced with the need for extensive rebuilding programmes. Swansea (7) which has decided to build six large multilateral schools, gives as its main reasons for doing so, the need to start from scratch, the desire to experiment and its ability to find the sites (50 acres for each school of 1,500 pupils). London, the prime advocate of large comprehensive high schools (8) advances the reason that from them,

"will flow in time a healthy mutual regard and understanding between persons of different kinds of ability." (T.E.S. 10/5/47)

but it is equally possible that the demand for large new schools is, in part, due to the extensive bomb damage to London schools.

(6) "All schools for children over 11 to be brought under a common code of regulations for secondary schools, with common standards of accommodation, staffing etc., and for the Board to encourage as a general policy, the development of a new type of multilateral school which would provide a variety of courses suited to children of all normal types." (T.E.S. 1/2/47.)

(7) "The existing buildings of the four secondary grammar schools are such as to necessitate their replacement whatever form of organisation is adopted... Secondly, there are no new secondary modern schools while the junior technical school is housed in temporary buildings." (T.E.S.15/1/47)

(8) The plan includes 67 comprehensive high schools, which when the school age is raised to 16 years, will house up to 2,000 pupils each.
In the published reports varying reasons (10) are adduced to support the tripartite system but usually these L.E.A. feel that their existing grammar schools are adequate; where there are secondary school buildings too good to be scrapped, yet unsuitable for multilateral schools or where there is not the room for the large sites required, the existing organisation has been retained.

During the controversy in the early months of this year, the Times Educational Supplement declared itself very plainly. The editor wrote (15/2/47):

"what we have supported and continue to support is the long overdue rebellion against the emphasis, almost exclusive emphasis ... on intellectualism in the education of the young."

"We see no reason for thinking that the common secondary school inevitably leads to grave social, educational and cultural evils. We are convinced that it is merely a matter of organisation ... to secure that the scope and pace of education are adapted to the individual, that there is denial of opportunity to no one, nor any retardation of the intellectually ablest. Positively, we see real and substantial advantages ensuing from the common secondary school."

The charges and counter blasts which filled the columns of the journals in the early months of this year have now faded away; it seems to be realized that the opponents and advocates of the multilateral and comprehensive schools have few facts on which to base their arguments and that the next step will be to build schools of those varying types and to examine the results. The present position is that the shortage of buildings seems to have forced the L.E.A. to base their policies on non-educational grounds to a greater degree than they would probably admit. Most have realistically faced the present situation, and, in accordance with the English tradition of working on an empirical basis rather than to a rigid plan, in different parts of the same county, different systems of organisation will be used.

(10) The Nottinghamshire education committee's belief in the importance of the head teacher's influence led them to regard a total roll of 500-600 as a desirable minimum for any one school (T.E.S. 5/7/47). The decisive argument in Warwickshire has been apparently that it would be unwise to tamper with the grammar schools. This means that the multilateral experiment must be confined to areas where there is as yet no grammar school. (T.E.S. 12/4/47)

In Devon, the largest single administrative unit in the country, the education committee believes that there is scope for schools of more than one type and has planned for one multilateral and several bilateral schools. (T.E.S. 24/5/47).
Bibliography

Those references marked x are cited in the text.

Times Educational Supplement (T.E.S.)

Most of the quoted references are to short reports; the following are longer statements.

4. 1. 47. The Grammar School : (L. Hollingworth)

18. 1. 47. Development Proposals : Wiltshire and Swansea

1. 2. 47. The Challenge to Grammar Schools : (E. James)

10. 5. 47. Developmental Plans: Secondary Education
(A summary of proposals submitted to the Ministry)

13. 9. 47. Tripartitism's Weakness : (J. Hill)

The Journal of Education

November, 1946. The Multilateral School : (W.T. Stevenson, Chief Inspector of Schools, Manchester.)

Education

10. 1. 47. Methods of Selection for Different Types of Secondary Education : (W.P. Alexander)

7. 2. 47. Secondary and Technical education under the 1944 Act : (A.L. Binns, Chief Education Officer, Lancashire.)

14. 2. 47. The Northampton Development Plan.

7. 3. 47. The London Development Plan.

8. 4. 47. Address by Sir F. Mander to the L.E.A.


12. 9. 47. Allocation of Children for Secondary Education : (L. Smith)

The British Journal of Educational Psychology

November, 1943, The Education of the Adolescent (C. Burt)

Occupational Psychology


October, 1947. Significance of Ability Differences at 11 plus.

(F.M. Earle)

Dr. Earle, principal of Kirkcaldy High School and formerly head of the N.I.E.P. Vocational Guidance Department, discusses the practical implications of his own investigations into the classification of secondary school pupils.

Ministry of Education

Circular 144: Organisation of Secondary Education.

Pamphlet 1: The Nation's Schools.

Pamphlet 9: The New Secondary Education.