DOCUMENT RESUME

ED 155 602 CS 004 110

AUTHOR McLeod, John; Anderson, Jonathan
TITLE Development of a Standard Reading Test Designed to Discriminate Effectively at the Adolescent Level.
PUB DATE [77]
NOTE 21p.; Study prepared at University of Saskatchewan; Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Adolescents; *Cloze Procedure; Elementary Secondary Education; Reading Ability; *Reading Comprehension; *Reading Tests; *Standardized Tests; Substitution Drills; *Test Construction

ABSTRACT

To investigate the drastic reduction in the discriminating power of reading tests with children above the ages of 10-11 years, a test was devised based on the GAP test, itself based on the cloze technique shown to be effective in assessing reader comprehension. The GAP tests requires the reader to supply words deleted from a passage of continuous prose. Deletions were selected from words shown to be totally redundant from experiments with fluent readers (628 education students at Kedron Teacher's College, Australia, who were given copies of 192 different Test forms). Further experimentation was conducted with 5,505 students, grades 2-12, which showed that even when prose passages of quite high readability (difficult reading) were used in cloze-type tests, discriminatory power deteriorated in children aged 11-12 years when words whose redundancy was virtually 100% (totally unambiguous to fluent readers) were selected for deletion. Further research indicated that if words were deleted whose estimated redundancies were less than 100%, discrimination between readers was maintained up to ages 16-17 years. (Tables of experimental data, references, and test instructions are included.) (DF)

DEVELOPMENT OF A STANDARD READING TEST DESIGNED TO DISCRIMINATE EFFECTIVELY AT THE ADOLESCENT LEVEL

John McLeod, Ph.D. University of Saskatchewan, Saskatoon, Canada.
Jonathan Anderson, Ph.D., University of New England, Armidale, Australia.

A remarkably general and ubiquitous phenomenon which is found in the measurement of reading ability is a dramatic reduction in the tests' discriminating power with children above the age of ten or eleven. This stems from the leveling off of mean raw scores at these ages. (University of Reading, 1971).

The GAP Test (McLeod, 1966) which is no exception to this phenomenon, was constructed on the basis of Taylor's (1954) Cloze technique, a method which numerous research studies (e.g. Bormuth, 1967; Anderson, 1969) have demonstrated to be a valid method of assessing reading comprehension. The method requires the reader to restore words that have been deleted from a passage of continuous prose. Deletions in the GAP Test were selected from words that had been shown empirically, from experiments with fluent readers, to be totally redundant. (McLeod and Anderson, 1966).

However, further experimentation revealed that even when prose passages of quite high readability levels — i.e. passages that are difficult to read — were employed as cloze-type tests, discriminatory power still deteriorated with children aged eleven or twelve years when words whose redundancy was virtually one hundred per cent (i.e. totally unambiguous to fluent readers) were selected for deletion.

Further pilot studies showed however that if text were mutilated by deleting words whose estimated redundancies were significantly less than one hundred per cent, discrimination between readers was maintained up to the sixteen or seventeen year level.

Selection of Prose Passages:

Twenty-four prose passages were selected, each of approximately two hundred words, as a pool from which the final test passages would be chosen. The prose passages themselves were extracted from sources ranging from children's books and periodicals to adult novels, historical and biographical works (McLeod and Anderson, 1972). The general principles on which the passages were selected were that the English should be beyond reproach and each passage should constitute a complete unit in itself.

....2

Eight cloze-type versions of each passage were prepared: for the first version, the tenth and every subsequent eighth word was deleted; for the second version the eleventh and every subsequent eigth word was deleted, et cetera, so that in the eigth version of a given passage, every word,apart from the introductory ten words, was deleted in one version or another. In all, therefore, multiple copies were prepared of 192 (8 versions of 24 passages) different test forms, each test form being printed on a separate indexed sheet. Test booklets were made up containing a version of sixteen of the twenty-four passages. The serial order in which prose passages appeared in different test booklets was balanced in order to offset any possible constant error effects due to fatigue and/or practice.

The try-out test booklets, therefore, were heterogeneous in nature, differing from each other in the passages sampled, the words deleted within the passages, and the order in which the passages were presented. The front page of each try-out test, however, was identical for all test booklets, consisting of the same demonstration and practice exercises.

Initial Try-Out:

The deleted words in the final version of the test were to be selected according to their redundancy as estimated from the responses of fluent readers. With the kind permission of the Director General of Education for Queensland and the cooperation of the Principal and staff of Kedron Teacher's College, Brisbane, the authors — ably assisted by colleagues of the Fred and Eleanor Schonell Educational Research Center, University of Queensland — administered the try-out tests to all 628 teacher trainees in the College.

In order to ensure that all students attempted all sixteen passages in the hour available for each session, they were informed at three minute intervals, that they ought to have completed another page. In the event, it turned out that the time available was more than adequate and most of the students had completed the test well before the end of the allotted period. (For test instructions see Appendix A).

Data Processing of Try-Out tests:

Because of the scrambled nature of the try-out test booklets, it was first necessary to take the test booklets apart and reassemble the ten thousand-odd pages into 192 sets according to passage (24), and particular deletion version (8). The number of completed test forms in each set ranged between forty and fifty.

....3

Responses were recorded on punched cards and the data processed by computer. The resulting analysis included a print out of the proportion of correct responses (i.e. restorations of the original deleted word) for each deletion, together with the estimated redundancy of each work. The relative uncertainty, and redundancy, of each passage was also computed. The distribution of fluent readers' responses to one version of passage A9 (see Appendix), together with calculated redundancies, are set out in Table 1.

Table 1

Pattern of Responses by Fluent Readers to Try-Out Test Passage No.A91

Deletion Number:

1		2		3		4		5		6		7		8		9	
Soldiers	1	always	5	and	34	nation	6	a	11	and	27	as	24	the	40	court	
Londoner	19	caught	1	would	1	spirit	3	the	17	with	13	when	16	been	1	palace	
nobility	1	working	2	both	1	hard	6	their	5	like	1	feeling	1			cross	2
nil	2	then	3	nil	1	race	10	his	1							castle	
clear	1	right	2	but	2	king	3	winter	1							nil	
airman	1	first	1	yet	1	favourt	1	this	6							road	
royalty	1	seen	4	as	1	the	1									place	
French	2	found	3			people	4									square	
gentry	1	hungry	1			air	1									folies	
kings	1	really	4			history	1										
lords	1	consider	1			pride	1										
royality	1	both	1			feeling	1										
king	1	never	1			foe	1										
lads	1	now	2			men	1										
ambulanc	1	finally	1			city	1									*	
rich	1	not	1													▷	
Scots	2	nil	2													*	
nobles	1	definite	1														
upper	1	living	1														
aristocr	1	known	1														
		undoubte	1														
		particul	1														
		acting	1														
Redundancy of passage:																	
38.56		19.63		79.66		37.22		61.30		80.35		79.23		96.91		56.18	
Percentage Correct:																	
46.34		9.76		82.93		9 76		14.63		31.71		39.02		97.56		51.22	

Note: Words originally deleted are indicated by underlining.
For all responses only the first eight letters are printed.

Table 1

(Continued)

Deletion Number:

10		11		12		13		14		15		16		17	
empty	15	weary	2	would	27	the	40	death	5	with	4	understa	1	being	23
clear	10	hungry	9	might	1	heros	1	fear	1	like	6	amount	1	courageo	1
move	2	aged	2	nil	2			winter	7	and	2	passion	1	the	1
resound	1	cold	3	always	1			darkness	2	that	1	feeling	15	inner	1
nobles	1	poor	1	should	3			nil	8	fear	1	sense	10	to	2
crowd	2	old	6	could	1			night	5	suddenly	4	as	1	thus	1
nil	4	always	1	had	3			this	1	in	4	pity	1	borne	1
shake	1	sad	3	never	1			war	1	inside	1	twang	1	now	3
stir	1	working	3	will	1			France	1	at	1	thought	1	in	1
quieten	1	busy	1	was	1			cold	1	then	3	anguish	1	bravely	1
fill	3	homeless	1					destruct	1	even	1	nil	3	then	3
		apatheti	1					however	1	sharply	1	comparis	1	soon	2
		patient	2					they	1	as	4	sensatio	1	never	1
		work	1					the	1	but	1	emotion	1		
		the	1					danger	1	nil	3	concern	1		
		dogged	1					all	1	quietly	1	pang	1		
		revere	1					friend	1	certainl	1				
		liston	1					mist	1	slightly	1				
		civilian	1					rain	1	such	1				

Redundancy of passage:								
45.79	29.27	62.78	96.91	20.43	24.70	41.80	53.52	
Percentage correct:								
36.59	0	65.85	97.56	12.20	9.76	24.39	56.10	

Number of subjects: 41

Item Selection:

From the data yielded by the fluent readers, it was necessary to decide (a) which passages to use in the final test; and (b) which words should be deleted within each passage. Four primary principles were established:

(1) The passages comprising the test must be of appropriate interest and difficulty level.

(2) As far as possible, early items in the test should be easy and the later items should be more difficult.

(3) Distractors — or more correctly, error responses in what was to be an open ended test — should be as evenly distributed as possible.

(4) There should be at least four words of context between deletions.

1. Appropriateness of Material. The suitability of all passages used in the try-out test had been assessed and discussed from the first stages of designing the test, before the results of the try-out testing were known, but there could be no certainty that all passages would stand up to quantitative analysis. As it happened, an adequate number of word deletions, covering a range of item difficulty, could have been selected from any of the twenty-four passages, but it was estimated that six passages would yield a sufficient number of items to produce a test of acceptable reliability. Two equivalent tests were therefore prepared, using twelve of the original passages. The passages to be used in the test were selected so as to make for a variety of prose types (e.g. children's literature, novels, descriptive and scientific) and, for the most part, passages of easier overall readability, according to empirically determined redundancy, were chosen.

2. Gradation of Item Difficulty. Unlike a test where individual item placement can be changed freely, the test under construction was to be made up of items which are themselves embedded in continuous prose. This constraint is an impediment to smooth gradation of difficulty. Passages that had been selected for inclusion in the final test were arranged in increasing order of readability level as estimated from the responses of the fluent readers. Within the passages, deletions were chosen in such a way that the proportion of difficult items was high in the later passages and low in the early passages.

....7

Each deleted word was classified according to difficulty. If ninety to one hundred per cent of fluent readers reproduced the correct word, then difficulty level was designated as A. If sixty to ninety per cent of fluent readers reproduced the correct word the item was rated as of B difficulty, while if fifty to sixty per cent of fluent readers reproduced the correct word, the item was rated C. The distribution of difficulty levels of deleted words in each form of the final test are set out in Table 2.

Table 2

Distribution of Item difficulties in GAPADOL

	Form G Difficulty level			Form Y Difficulty level		
Passage No.	A	B	C	A	B	C
1	7	2	0	5	7	0
2	3	12	0	3	9	2
3	5	6	3	2	6	3
4	4	7	2	3	8	3
5	2	7	4	5	8	2
6	2	7	7	1	10	6
1 - 3	15	20	3	10	22	5
4 - 6	8	21	13	9	26	11
Total	23	41	16	19	48	16

3. Equi-distribution of Error Responses. In a multiple forced-choice test, it is generally desirable that no distractor should attract a significantly high number of error responses. The Gapadol tests are open ended so "distractors" are generated by the respondees themselves. For example, the first deletion in try-out passage A92 (see Table 1) elicited a wider variety of low incidence error responses and is, therefore, a potentially good but difficult item. On the other hand, deletion number 10 attracts the response "clear" almost as frequently as the response "empty", and is therefore not satisfactory.

The criterion according to which a deletion was deemed to be acceptable was that the estimated redundancy should not be higher than the proportion of correct responses. The validity of this rule of thumb is demonstrated in Appendix C.

....8

4. Context between deletions. Traditional test theory requires that the item in the test should be statistically independant. It has been empirically determined (Anderson 1969) that this condition is satisfied if an every-fifth deletion system is used (i.e. if there are at least four words between successive items). There is in the final test only one exception to this general rule: the first word in the third paragraph of the second passage of Form Y.

In summary, words to be deleted in the final tests were selected in such a way that:

(1) they elicited an acceptable number of correct responses;

(2) a larger proportion of words that are easy to restore were included in the early passages, and the proportion of more difficult items was increased in the later passages;

(3) the proportion of correct responses by fluent readers was higher than the redundancy according to the fluent readers; and

(4) there are never fewer than four words between successive deletions.[1]

Pilot administration of GAPADOL Tests:

Before standardization testing itself, the proposed GAPADOL tests were adminstered to one class from each grade level in an elementary school and in a high school of the Saskatoon public school system. (The public school system and the separate school system are independent of each other, so that no child involved in standardization testing would be involved in any testing carried out in a public school). The purpose of the pilot run was to check the adequacy of test instructions, to establish guide lines with regard to timing and to confirm that a test of six passages would yield an appropriate spread of scores. Some estimate was also required as to the likely discriminating power of the test at higher grade levels.

These objectives were achieved, with significant discrimination being obtained through Grade Eleven. Half an hour was deemed to be an appropriate time limit for test administration as most of the children had completed the test within this period; in fact, many children in higher grades completed the test in a much shorter time, so that, for children in higher grades particularly, the test is essentially a power test.

---

1. There is one exception to this condition; the first word of the third paragraph in "Polluted Beaches" (Form Y).

Standardization:

The GAPADOL tests were standardized on the entire Saskatoon Separate School population of over 5,000 students, in June 1970. A preliminary meeting of school principals had been convened to discuss the test and its rationale, and was followed up by the distribution of notes concerning the background of the test and general instructions to all separate school teachers. (See Appendix D.). Appropriate numbers of tests were packaged in separate envelopes for each class, together with a specific set of instructions for each teacher relating to test administration and marking. (See Appendix D.).

Class teachers administered the tests, adjacent children answering different forms of the test to obviate the risk of copying. Each child's raw score, together with his or her date of birth, was recorded on previously prepared sheets which were returned by school principals for analysis.

Children from Grade Two through Grade Twelve were tested, 2,751 answering Form G and 2,754 answering Form Y.

The data were processed at the Institute of Child Guidance and Development (Saskatoon), on the Hewlett-Packard computer, which had been programmed to convert raw scores into percentile ranks, at five point intervals.

After smoothing, by running averages, Saskatoon local norms were produced for the tenth, fiftieth, and ninetieth percentiles. A comprehensive table of normalized standard scores, i.e. quotients, could have been derived, but such indicies — with their appearance of high precision — can be misleading. For older children particularly, the functional value of a reading test is to determine whether a child's reading level is more or less average for his age; if it is not, then teachers and reading specialists need to be able to estimate the extent and seriousness of the reading retardation. Calculating the raw score corresponding to the median percentile rank establishes the norm or average reading performance for a child of a given age and knowing the raw score which corresponds to the tenth percentile provides a reference marker which makes it possible to relate a retarded reader's performance to that of the poorest ten per cent of readers of his own age. As a reference at the higher ranges of reading ability, the raw score corresponding to the ninetieth percentile was provided, in order to facilitate identification of outstanding readers, i.e. those in the top ten per cent of the population tested.

....10

It remained to estimate the degree to which Saskatoon norms might be more generally applicable. The author has argued elsewhere (Schonell and McLeod 1962, p.8) that it is exceedingly difficult to obtain a nationally (let alone internationally) representative and valid sample and the practice of simply selecting children arbitrarily from three or four different types of area is of doubtful validity. An alternative approach is to relate the data and norms of the test to be standardized to the norms of an existing test or tests whose validity has been established. least, such a procedure ensures that norms of the test bei standardized are compatible with those of the established test.

The obvious test to use in the present instance, was the GAP test (McLeod 1965)and, for this purpose, both GAP and GAPADOL tests were administered to every child from Grade Three to Eight in an elementary school of the public school board of Saskatoon. The order in which tests were presented - i.e. whether GAP or GAPADOL was administered first — was balanced over the classes.

In view of the anticipated fact that many children in the higher grades would be approaching or achieving beyond the ceiling of the GAP test, it was decided, a priori, that GAP ar· GAPADOL raw scores would be compared only for those children scoring not more than 25 on the GAP test.

Published GAP reading ages (McLeod 1965) were correlated with GAPADOL (Saskatoon) norms, and a highly satisfactory correlation coefficient of 0.78 was obtained, with regression linear. The best fitting straight line relating the two sets of norms had a slope of almost exactly unity, indicating that the Saskatoon GAPDOL figures were almost constantly 3 months more pessimistic than the existing GAP norms.(McLeod and Anderson, 1972).

The local Saskatoon norms were modified appropriately and, at least up to a reading age of approximately eleven years, the GAPADOL tests yield reading ages which are as comparable with those from either form of the GAP test as are the two GAP tests with each other. At higher age levels, of course, no direct comparison between GAP and GAPADOL is possible.

Reliability and Validity

The validity of the Cloze-type tests has been well established over the years. As was reported in the previous section, a product moment correlation of 0.78 was obtained between GAP and GAPADOL tests. for children with reading ages up to about $10\frac{1}{2}$ years.

....11

Reliability has been estimated by computing the internal consistency coefficient (Hoyt, 1941) at alternate grade levels, from Grade two t' ough ten (Table Three). Groups of approximately fifty children were used for the calculation of reliability coefficients, which were corrected for attenuation.

Table 3

Reliability and Standard Error of Individual Scores

	Form G		Form Y	
Age Ranges	Reliability	Standard Error	Reliability	Standard Error
7.3 - 8.3 years	0.84	2.57	0.92	2.35
9.3 -10.3 years	0.91	3.09	0.87	3.44
11.3 -12.3 years	0.93	3.51	0.89	3.49
13.3 -14.3 years	0.92	3.58	0.90	3.63
15.3 -16.3 years	0.91	3.45	0.91	3.72

By administering both Form G and Form Y within a week or so, and averaging the two estimates of a child's obtained reading age, reliability of the overall assessment can be effectively increased to 0.93 for younger children and to 0.95 for older children.

Acknowledgements

Grateful acknowledgement is made to Mr. Walter Podiluk, Director of Education of the Saskatoon Separate School Board, and to all principals and teachers of that system, for their cooperation in the standardization of the GAPADOL tests. Thanks also due to Dr. Fred J. Gathercole, Director of Education of the Saskatoon Public School system, and to the principals and teachers of City Park High School and of John Lake Elementary School, Saskatoon.

For assistance and collaboration in the original calibration of reading passages used in the tests, appreciation is extended to Mr.G.K.D.Murphy, then Director-General of Education for Queensland, to the Principal and staff of Kedron Teachers' College, Brisbane and to the staff of the Fred and Eleanor Schonell Educational Research Centre, University of Queensland.

References

Anderson, J. Application of Cloze Procedure to English Learned as a Foreign Language. Unpublished Ph.D. thesis, University of New England, Australia, 1969.

Bormuth, J.R. Design of Readability Research. Proceedings of the Eleventh Annual Convention of the International Reading Association. 11.1, 485-489, 1967.

McLeod, J. GAP tests and Manual. Melbourne, Heinemann Educational, 1968.

McLeod, J. and Anderson, J. Readability Assessment and Word Redundancy of Printed English. Psychological Reports, 18, 35-38, 1966

McLeod, J. and Anderson, J. GAPADOL Reading Comprehension and Manual. Melbourne, Heinemann Educational, 1972

Schonell, F.J. and McLeod, J. Wirral Mechanical Arithmetic Tests and Manual. Edinburgh. Oliver and Boyd, 1962.

Taylor, W.L. Cloze Procedure: A New Tool for Measuring Readability. Journalism Quarterly, 30, 415-433, 1953.

University of Reading, School of Education. Reading Tests. 1971

APPENDIX A

Redundancy Estimation of Original Passages

Test Instructions

Don't open your booklets.

You are today cooperating in a piece of research, and one of the by-products of that research will be a reading test which can be used in High Schools.

Today we are interested in whether the passages we have gathered together are suitable as test material. We are not testing your reading ability, but rather using your reading ability in order to test the test material.

The booklets in front of you contain sixteen passages each of which has some words deleted in the same way as the practice example on the front of your booklets. You have to write on the lines provided at the side, the one work that is missing from each blank.

The words deleted have not been specially selected, but rather chosen at random. Some are very simple and obvious words, for example, definite and indefinite articles. You might be able to think of several words that could adequately fill some blanks. However, you may find a few blanks that are simply impossible to replace. In every case, just try to find the one word which fits the sense best.

Let's do the practice example on the front of your booklet to give you the idea.

> One of the great exports of South Vietnam has always (pause) American optimism.

What is the word that is missing. You might notice that the blanks vary in length according to the word that has been deleted. Well, what is the first missing word? (Obtain 'been'). Write the word 'been' on the line at the side of the paper. (Demonstrate). Let's do the second blank.

....13

One of the great exports of South Vietnam
has always been American optimism, but this time
I thought (pause) I returned that it would .....

What word should you write on the next line? (Obtain 'when'). Now try the rest of the blanks in that passage for yourself. When you have finished, put your pens down, and do not turn over. (Check answers. Bring out in discussion that 'illusion' - the answer to the fourth blank - is very difficult, if not impossible. Don't be afraid to guess).

Well, that is the sort of thing we want you to do. Just write the one word that you think should be in each blank, on the line at the side of the paper. Please write clearly, and please write your answer on the same line as that in the passage from which the word is missing. If you make a mistake, cross out the wrong answer and write your new answer clearly in the right place. Don't waste time with any particular blank. Make an intelligent guess.

It is important that you have a go at every passage, so try to keep up a steady pace. As a guide-line, I will tell you from time to time what passage you ought to be on.

Are there any questions before we begin?

APPENDIX B

No. A9

"LONDON CAN TAKE IT"

These were the days when the English, and particularly the Londoners, who had the place of honor, were seen at their best. Grim and gay, dogged and serviceable, with the confidence of an unconquered people in their bones, they adapted themselves to this strange new life, with all its terrors, with all its jolts and jars. One evening when I was leaving for an excursion on the East Coast, on my way to King's Cross the sirens sounded, the streets began to empty except for long queues of very tired, pale people, waiting for the last bus that would run. An autumn mist and drizzle shrouded the scene. The air was cold and raw. Night and the enemy were approaching. I felt, with a spasm of mental pain, a deep sense of the strain and suffering that was being borne throughout the world's largest capital city.

....14

APPENDIX C

Relation of Redundancy to Proportion of Correct responses R_1, and N_2, N_3, —N_k respondees give in correct responses R_2; R_3, — R_k respectively.

Consider the expression $\alpha = \frac{\text{Proportion correct}}{\text{Redundancy}}$

$$\alpha = \frac{N_1/N}{\left(\sum_{i=1}^{i=k} N_i \log_2 N_i\right)/N \log_2 N}$$

$$= \frac{N_1 \log_2 N}{\sum_{i=1}^{i=k} N_i \log_2 N_i}$$

$$= \frac{\log_2 N}{\sum \frac{N_i}{N_1} \log_2 N_i}$$

$$= \frac{\log_2 N}{\sum \log_2 (N_i)^{N_1/N_1}}$$

$$= \frac{\log_2 N}{\log_2 \left[(N_1)^{N_1/N_1} \times (N_2)^{N_2/N_1} \times (N_3)^{N_3/N_1} \times — \times (N_k)^{N_k/N_1} \right]}$$

$$\log_2 N = \log_2 \left[(N_1)^{N_1/N_1} \times (N_2)^{N_2/N_1} \times (N_3)^{N_3/N_1} \times — \times (N_k)^{N_k/N_1} \right]$$

$$\text{i.e. } N = \left[N_1 \times (N_2)^{N_2/N_1} \times (N_3)^{N_3/N_1} \times — \times (N_k)^{N_k/N_1} \right]^{\alpha} \quad \ldots.(1)$$

Ideally, for a satisfactory item

$$N_2 = N_3 = N_4 = \text{——} = N_k$$

and each error frequently is small.

$$\therefore \quad \frac{N_2}{N_1}, \frac{N_3}{N_2}, \ldots \frac{N_k}{N_1} \longrightarrow 0$$

$\therefore$ From equation (1),

$$N \rightarrow \left[N_1 \times 1 \times 1 \times 1 \text{——} \times 1\right]^{\alpha}$$

i.e. $N \rightarrow N_1^{\alpha}$ ..... (2)

(a) For an easy item,

$N \simeq N_1$

from (2) $\alpha \simeq 1$

(b) For a difficult item,

$N_1 < N$

from (2), $\log N \propto \log N_1$

$$\therefore \alpha \rightarrow \frac{\log N}{\log N_1}$$

i.e. $\alpha > 1$

i.e. proportion of correct responses is greater than item redundancy.

APPENDIX D

M E M O R A N D U M

TO: School Principals, Saskatoon Separate Schools

FROM: John McLeod, Director, Institute of Child Guidance and Development, University of Saskatchewan

DATE: April 8, 1970.

Herewith are test forms, instructions for administration, class register blanks and marking keys. There ought to be one set of materials for each room in your school.

The tests have been put in envelopes and it would be helpful if each room's completed tests could be returned in its own envelope. To facilitate counting, the tests are in lots of twenty-five or fewer. That is, if an envelope contains tests of only one colour, then there are twenty-five tests in the envelope; if the envelope contains tests of two colours, then there are twenty-five of the more numerous tests and a smaller number of the other colour.

I trust that you will recall that during the actual test administration, alternate rows of children should be doing different tests, i.e. approximately half the children in each class will be answering the yellow form of the test and the other half the green form of the test.

I hope that there are no misprints, but there are almost bound to be a few tests that have slipped through which are faulty. If so, there ought to be sufficient extra tests for you to be able to replace any faulty ones that come to your notice.

Thank you for your help.

signed John McLeod
John McLeod, Director
Institute of Child Guidance
and Development

JM:dbv
Encl.

GAPADOL READING TESTS

Saskatoon standardization, 1970

General background

The GAPADOL tests are reading tests based on the "Cloze" technique devised by Wilson Taylor, and which research has shown to provide a highly valid measure of reading comprehension. The first published test to use a modified cloze technique in a standardized instrument for measuring reading comprehension is the senior author's GAP test (Journal of Reading, April 1966, 1X, 5, p.359).

....17

The GAP tests, however, tend to level off for average children at about the age of eleven or twelve, and the GAPADOL tests represent an attempt to raise the effective ceiling for which reading ability may be validly assessed. A preliminary try-out has indicated that the present GAPADOL tests might be expected to discriminate up to the Senior High School grades.

The tests consist of passages taken from published works, with certain words deleted. Details of how and which words were deleted have been published elsewhere. (McLeod and Anderson 1970) but are too technical to describe in the present context. Briefly, the criterion for a word to be selected for deletion was that a high proportion of adult efficient readers should, under test conditions, supply the word used by the original author of the passage, without ambiguity.

The purpose of the current testing program is to provide norms for Saskatoon, based on the Separate School population. It should also be possible to compare the appropriateness of the test norms of other standardized tests which are used locally.

General Instructions

1. Students should anwer the test in pencil. If they work in ink or ballpoint, it will be difficult for them to make corrections in the answer column.

2. Test should be distributed in class in such a way that adjacent children are answering alternate forms of the test. As one form of the test is printed on yellow paper and the other on green, the teacher should be able to check at a glance that alternate rows of students are answering different forms of the test. (The administration and instructions for each form are identical, so that both yellow and green forms can be given simultaneously.)

3. Prior to the test,have the name of the school and the Grade written on the chalkboard, as required on the front page of the test.

4. This is a test standardization, which means that if the results are to have any validity at all, the test must be carried out under standardized conditions. Therefore:

Please follow exactly the "Instructions for administration" as set out on the sheets accompanying the tests. Some teachers believe (quite correctly) that they can explain to children what it is all about better than the person who designed the test instructions. The important point is, however, that all children must receive exactly the same instructions.

....18

Timing must be exact. The GAPADOL is a power test rather than a speed test, and many children will complete all they are capable of doing within the 30 minutes time limit. However, please provide yourself with either a stopwatch or a watch/clock with a second hand.

5. Many children will finish within the 30 minutes allowed, so have some reading or drawing material available for them, so that early finishers will not distract other students.

6. Do Not help students during the test. If anyone asks for help ("What is this word?") just tell them to try to figure it out, do their best and/or go on to the next item.