

A FACET DESIGN TO MEASURE
READING COMPREHENSION

THESIS FOR THE DEGREE OF Ph. D.
MICHIGAN STATE UNIVERSITY

Warren Deloy Wilde

1973

LIBRARY
Michigan State
University

This is to certify that the
thesis entitled

A Facet Design to Measure
Reading Comprehension

presented by

Warren Deloy Wilde

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Elementary
Education

Byron H. Van Rookel
Major professor

Date May 18, 1973

0-7639

High Level
BOOK BINDERY LTD.
10372 - 60 Ave., Edmonton
"THE HIGHEST LEVEL OF
CRAFTSMANSHIP"



ABSTRACT

A FACET DESIGN TO MEASURE READING COMPREHENSION

by

Warren Deloy Wilde

This study attempted to measure reading comprehension of grade ten students using a multiple choice test constructed by a facet design. The facet design concentrates on the relation of the items to the information presented in the text or elsewhere rather than on skills and abilities which are presumed to be involved in comprehension. The study also compared comprehension as measured by the Davis Reading Test and a cloze test with the results of the facet test.

The sample consisted of 186 students chosen from a grade ten population in two large high schools in Edmonton, Alberta, Canada. All students were enrolled in a basic English course. The students received the Davis Reading Test and the facet test on one day and the cloze test one week later.

The quality of the facet test was determined by means of an item analysis which yielded an item difficulty index, a bi-serial correlation, and an item reliability

index. Kuder-Richardson 20 was used to compute the reliability of the facet test. Test results on the facet test were compared with those on the Davis Reading Test and the cloze test by means of a Pearson product-moment correlation coefficient.

It was concluded that: 1. the facet test was a reliable measure of reading comprehension; 2. the facet test discriminated among achievement levels; 3. there was a significant relationship between the facet test and other measures of reading comprehension, i.e. the cloze test and the Davis Reading Test.



A FACET DESIGN TO MEASURE
READING COMPREHENSION

by

Warren Deloy Wilde

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Elementary
and Special Education

1973

625-1

Dedicated to my loving family,
Ruth, Leslie, Renee, Jana, Russell, Marlys.

ACKNOWLEDGMENTS

The writer wishes to express appreciation to his major advisor, Dr. B. H. Van Roekel, for his guidance and assistance throughout the graduate program and especially in the writing of this thesis.

The interest of my committee members, Dr. Keith Anderson, Dr. Ruth Useem, and Dr. Gerald Duffy, as well as one member who retired from the faculty, Dr. Harold Byrum, is sincerely appreciated.

Appreciation is expressed to my colleagues at the University of Alberta who encouraged me to complete the thesis and gave able counsel in time of need. Especially a thanks to Dr. Marion Jenkinson for her support.

A very special thanks to my wife and family for their support and encouragement throughout my graduate program and especially during the summer when this thesis was written.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	vii
Chapter	
I. THE PROBLEM	1
What is reading?	1
Measurement in education	2
Need for this study	4
Purpose of this study	4
Facet design	5
Hypotheses	7
Additional considerations	7
Limitations of the study	7
Definition of terms	8
Overview	8
II. REVIEW OF RELATED LITERATURE	10
Measurement in reading	10
Common complaints about tests	13
What is reading comprehension?	14
Need for new theory	19
Facet test	23
Cloze technique	29
Summary	31

	Page
III. THE DESIGN AND PROCEDURES	33
Population and sample	33
Measuring instruments	34
Pilot study	46
Administration of test instruments	48
Analysis of data	49
Summary	50
IV. RESULTS	52
Introduction	52
Results concerning hypothesis one	52
Results concerning hypothesis two	53
Results concerning hypothesis three	58
Additional considerations	60
Summary	63
V. SUMMARY AND DISCUSSION	66
Summary of study	66
Discussion	70
Further research	74
BIBLIOGRAPHY	77
APPENDIX	
A Passages for the facet test and cloze test.	83
B Facet test directions and questions	94
C Cloze test directions and questions	108

LIST OF TABLES

Table	Page
2.1 Eight distractors of a test item in which figures vary in three facets, A, B, and C	24
3.1 Title and source of test passages	37
3.2 Readability of facet test passages	38
3.3 Facet test mappings into the range	42
3.4 Example of facet test question	44
4.1 Facet test statistics	53
4.2 Frequency distribution of facet test	54
4.3 Cutting points for five groups on facet test ..	55
4.4 Item reliability on facet test for all subjects	56
4.5 Results on facet test of low, mid, high groups.	58
4.6 Correlation coefficients for facet test, cloze test and Davis Reading Test	59
4.7 Per cent response for all distractors on facet test items for all examinees	61
4.8 Range of mapping sentence matched to distractor statement	64
4.9 Tabulation of error response for students in lower 40 per cent of facet	65

LIST OF FIGURES

Figure	Page
2.1 Test item from Guttman and Schlesinger (1967)	26

CHAPTER I

THE PROBLEM

What is reading?

The teaching of reading has been emphasized more than any other curriculum area. Reading is the only part of the elementary school curriculum which is customarily preceded by a systematic readiness program. Reading is usually accorded a greater portion of instructional time than any other subject in the elementary school. Ability to read is acknowledged by many as a prerequisite to learning in most subject matter fields. Reading has been the most persistently investigated of the receptive communication processes. (Gray, 1960, estimated that nearly 4,000 scientific studies of the sociology, psychology and teaching of reading are available.) That there has been no abatement in interest among researchers in reading is evident from Harris (Ebel, 1970, p. 1140), who reports that the number of scientific investigations completed in the 1960's was nearly double that of the previous decade and many of these were conducted by authorities from a variety of fields outside of education.

In spite of the attention received and in spite of the importance which our society accords this basic

skill, the process of reading is poorly understood. The amount of research and the inconclusiveness as to a universally acceptable definition and methodology may be the reason that most recent textbooks of reading instruction have chapters explaining the components of the reading process. However, a comparison of these books reveals disagreement as to the components of reading, the sequence in which they should be taught and the method of instruction.

While disagreement exists as to definition and methodology in reading, there are also aspects of agreement. Few would argue that two basic components of the reading curriculum are word recognition skills and comprehension skills. Jenkinson (1968) says that most studies of reading comprehension have identified four factors, i.e. vocabulary, interrelationship among ideas, abstract reasoning, and factors specific to the special subject areas. There also seems to be a general acceptance that measurement of reading performance should accompany the instruction.

Measurement in reading

The measurement of achievement in reading has received attention since early in the twentieth century, even though an absolute consensus of the process and components of reading has been lacking. The primary role of measurement in reading should be to determine if learning has taken place and to determine specific strengths and weaknesses of the learner in order to facilitate the planning of future

instruction. Many authors of textbooks on reading instruction discuss evaluation and recommend the assessment of student achievement before planning a program of instruction. They argue there is very little value in planning a reading lesson if a child cannot profit from the instruction. Glaser and Nitko (Thorndike, 1971) suggest that measurement procedures need to be designed with the information requirements of a specific instructional system in mind.

Though few educators question the importance of measurement, issues involving the kinds of test items (multiple choice, true-false, essay, etc.) that should be used and the kinds of achievement that should be measured remain unresolved. Objective tests of the multiple choice variety are most common on the standardized test market, yet they are criticized by proponents of essay-type tests. Critics of the multiple choice test argue that there seems to be no absolute method of ensuring that the foils (incorrect responses) are less correct than the correct responses. Hence, producing adequate distractors may be more of an art than a science. The decision as to whether an answer to a multiple choice type question is correct or not may depend more upon who is making the choice than anything else. Bormuth (1970) points out that a score on an achievement test constructed by currently used procedures must be interpreted as the student's response to how the test writer perceives the instruction. Because so little is known about the factors that determine test writers' behavior, perhaps we should

be more concerned with the nature of what is to be measured than with the form of the measurement.

The second issue concerns the kind of achievement which ought to be measured. Lindquist (1951) notes that achievement generally has been based upon textbooks and that educators have tended to teach what is in the textbook and test for recall of textbook facts. Ebel (1969) points out that we have moved from the measurement of facts to measuring other domains such as the cognitive and affective ones but he seems unsure as to whether much has been accomplished by this change. He even forecasts a move back to measurement of the factual information.

Need for this study

Educational measurement is primarily used to provide direction and purpose to educational effort and to report the degree of success in learning. The foregoing has pointed out two problem areas: (a) how can something be measured if it is not clearly defined, and (b) if there is no assurance of content validity, how can we be sure of what is being measured? A new approach to the resolution of this dilemma seems to be appropriate and that is the purpose of this study.

Purpose of this study

This study was an attempt to employ a new method of multiple choice test construction for measuring reading



comprehension. The approach endeavored to ensure consistency within every test item while operationally building in the validity of certain aspects of reading comprehension and relating this specifically to passage content.

Facet design

This study was patterned after the work of Schlesinger and Weiser (1970) who proposed that a facet design be used to construct a test of reading comprehension. Facet design concentrates on the relation of the item to the information presented in the text or elsewhere, rather than on skills and abilities which are presumed to be involved in comprehension. What a given item measures is defined operationally in terms of how it is related to the information presented in a passage and the reader's abilities in comprehending it are measured only indirectly.

In facet testing the respondent is probably influenced as much by the distractors of the multiple choice item as by its stem. This approach presumably acts to systematize the operation of generating the distractors and ensures that the correct answer is related to the text. The facets of reading comprehension defining the relationship of the distractor-statement to the text are formulated first and then related to various parts of the distractor. While the correct distractor-statement will have to be correct in all aspects, the incorrect distractor-statements will contradict the text in some aspect or be related to information not

included in the text. Even though a distractor-statement contradicts the text it may still be attractive to a respondent since one of its constituent clauses agrees with the text or could even agree with information obtained from some other source but not stated in the text.

This factor approach provides many advantages over the intuitive approach normally used in item construction. Systematic construction of item distractors should not only provide consistency but ensure that a parallel form of the test is identical since the items are drawn from the universe of possible questions. Diagnostic information may be available and future responses should be predictable if an analysis of test errors reveals a consistent pattern among the respondent's choices.

Little previous work had been done with the facet design in the area of reading comprehension¹ and this study endeavored to determine the usefulness of this procedure. To judge the usefulness of the facet design it was decided to analyze the relative effectiveness of the distractors for each item in the test and to compare the facet test results with achievement based upon other currently used tests which have been statistically validated.

¹A research report was published: Schlesinger, I. M. and Zehavit Weiser, "A Facet Design for Tests of Reading Comprehension," Reading Research Quarterly, V, 4:566-580, 1970. Upon corresponding with the researchers it was found that little work had been done with this design in reading but most of the work had been done in intelligence testing.

Hypotheses

1. A test constructed by facet design will not be statistically reliable.
2. There will be no difference among achievement levels when a test is constructed by facet design.
3. There will be no statistically significant correlation between the facet test and other measures of reading comprehension, i.e. between (a) the facet test and cloze test, and (b) the facet test and the Davis Reading Test.

Additional considerations

1. Will an analysis of responses to items provide additional information?
 - (a) Does the form of information facet (whether information is stated in exact words or in paraphrase) have any effect upon examinee response?
 - (b) Does the source of information facet (whether information comes from the text or some other source) have any effect upon examinee response?

Limitations of the study

This research was limited to the extent that the time allotted for the administration of the cloze test was not sufficient. This resulted in this test becoming a speed test rather than a power test, which had the effect of reducing the intercorrelations between items on the four passages and introducing an undetermined error factor into the correlations between (a) the cloze test and the Davis Reading Test, and (b) the cloze test and the facet test.

Definition of terms

1. facet test: This is a multiple choice test where each item is constructed on a priori definitional grounds after Schlesinger and Weiser (1970). The items are constructed in such a way as to relate only to the content of information being tested and as such are not dependent upon previously acquired information.
2. facet: "...two sets of elements, A and B, are called facets, and their Cartesian space is the set of all pairs of elements ab where a is an element of A and b is an element of B. A Cartesian space may consist of any number of facets, or sets of elements; with n facets, any one point in the Cartesian space has n component elements." (Guttman and Schlesinger, 1967, p. 3)
Facets are arbitrary sets of elements within a space such as comprehension. These can be thought of as sets of ideas rather than constants. In this study the facets of comprehension decided upon were (a) form of information, and (b) source of information.
3. text: This refers to the passages on which the test items are based and which students read to answer the questions.
4. distractor: This refers to all distractor statements of an item, whether the correct answer or incorrect response.

Overview

A review of the literature pertaining to this study

is outlined in Chapter II. There is reference made to educational measurement, tests of reading comprehension, and a definitive explanation of facet design.

A description of the research design and test instruments, along with the results of a pilot study are provided in Chapter III.

A detailed analysis of the data and summary of the findings are delineated in Chapter IV.

The findings of the study related to the problem are discussed in Chapter V together with suggestions for further research.

CHAPTER II

REVIEW OF RELATED LITERATURE

The skill of reading is not well defined and methods to measure reading achievement are currently being questioned. Though many tests of reading comprehension are available, their construction may be based more on the intuition of the author than on a definite, predetermined set of criteria. The facet design provided a plan for the compilation of a test which could be built upon an a priori basis. Items for the test are based directly upon content purposely exposed to the examinee rather than upon information which the examinee might have been exposed to through normal experience. The validity of the test relies upon the method of item construction to the extent that the components of the test are considered facets of reading comprehension. It was felt necessary, however, to make comparisons among other tests of reading comprehension which were considered to be adequate measures of this skill.

Measurement in reading

An important aspect of our educational system is the measurement of achievement. Lindeman (1967, p.5) states that, "Measurement in education is essential if the evaluation

process is to be accurately and effectively carried out." E. L. Thorndike at the turn of the century initiated the concept of comprehension in reading which has resulted in the construction of various devices and instruments to measure reading comprehension. Thorndike (1917) said that comprehension in reading was much the same as reasoning in mathematics. Cook (Lindquist, 1951) refers to measurement in education as a means to more accurately observe the individual, and that tests merely act to quantify certain aspects of behavior. Lindquist (1951) noted that nearly all of the standardized achievement tests and informal school examinations constructed to date have been designed to measure achievement in established school subjects.

Common sense suggests that mere measurement of behavior is futile unless the information obtained can be put to good use. It is generally agreed that test data should be used to help plan educational programs for individuals or foster adjustment of instruction to individual capacity. Cook (Lindquist, 1951, p.35) believes that "the determination of pupil status in a given area of achievement and the adjustment of instruction to status should be a continuing process in every classroom."

The measuring instrument yields quantitative information about the individual but the evaluative aspect of measurement is really qualitative in nature. Even though the teacher receives a number of correct and incorrect responses from an achievement test, he must relate these to

the quality of individual progress. Sometimes this is done by comparison to what the individual has previously achieved but often it is a comparison with peers. When tests are designed to yield specific information of progress, future instruction is more easily planned. Instruction based upon exact needs of the individual should allow for more efficient progress than instruction based upon relative information. Heilman (1967) said the only justifiable purpose for using reading tests is to obtain data about an individual's reading ability so that a reading program can be built from the data secured.

In this regard, Glaser and Nitko (Thorndike, 1971) list four activities of instructional design that influence measurement requirements, i.e., 1. analysis of the subject-matter domain under consideration, 2. diagnosis of the characteristics of the learner, 3. design of the instructional environment, and 4. evaluation of learning outcomes. The learning must be analyzed in terms of subject-matter content as well as in terms of the behavioral processes that are being learned. One must also take into account the entering behavior of the student, including the extent to which he has already acquired what is to be learned along with the necessary prerequisites for further learning. Thorndike (1971) suggests that if there is lack of a theoretical framework upon which to base instruction, that measurement and evaluation will be seriously affected.

Common complaints about tests

There have been many complaints made about tests constructed to measure educational skills including reading comprehension. One general limitation, not necessarily inherent in the test, is that knowing the correct answer is often equated with education. Nonetheless we do find that what is tested for becomes important and what is not tested for becomes less important. The weakness could be eliminated if there were specified objectives known to be essential to the content or skill being measured. We usually find that tests are constructed without adequate objectives, which results in the testing of information that may or may not be relevant. Lindquist (1951) observed that all standardized tests for school examination have been constructed to measure achievement but that these were based upon textbooks as a course of study, resulting in a test for recall of what was contained in the textbook rather than being based upon clear cut objectives of what should be taught.

In examining currently available tests many limitations are evident. Hoffmann (1962) said one limitation of multiple choice tests is the introduction of ambiguities resulting from assumptions made by the test author. This stems from the author's failure to develop a plan from which to construct the distractors. Hoffmann believes that tests favor the nimble-witted, quick reading candidate who forms fast, superficial judgments. It has been pointed out by Bean (1953) and Vernon (1962) that reading is a decided

factor in any multiple choice test and that there is dependence upon the individual's facility with instructions and his ability to cope with the item format. It was stated by Hunt (1955) that most standardized high school reading tests measured few reading skills. Holland (1967) found that even in programmed instruction there was much irrelevant information in relation to the test items.

It is interesting to note the opinions of teachers regarding tests as delineated in the survey done by Austin and Morrison (1963). The study revealed that much testing was being done in the schools but that all too frequently the results were not used to determine present status or plan for future progress. Test results were often misused because of a basic misunderstanding of test limitations. There was dissatisfaction with standardized tests because:

1. tests lacked diagnostic value,
2. tests overestimated real reading ability,
3. tests did not measure the same skills taught in the classroom,
4. there was too much emphasis placed on vocabulary and not enough on comprehension and other skills, and
5. the tests did not measure higher level reading skills.

While these opinions may stem from a misunderstanding of test usage, there are serious implications for the authors whose major aim is to construct instruments that measure reading comprehension.

What is reading comprehension?

It was previously stated that a clear understanding

of what is to be measured is essential before one attempts to measure it. Even though numerous studies have been conducted with regard to various aspects of comprehension there still appears to be little agreement as to the nature and scope of the components which make up the reading comprehension act.

Thorndike (1917) examined mistakes made in paragraph reading and concluded that the potency of any word or word group in a question may be far less than its appropriate value in relation to the words which make up the rest of the question. He said that to understand a paragraph implied that the weightings of words must be kept in proper proportion in order to evoke a response which would satisfy the purpose of the reader.

Berry (1931) summarized the skills of a good reader. These individuals read for mastery of the general outline. They are able to relate subordinate details to the general outline, select key sentences or determine the topics of selection, visualize what is being read, note new or difficult terms and concepts, and group major issues and their implications. Berry noted that good readers are able to infer the meanings of unknown words from their contexts and relate the parts of a selection to its whole.

Hunt (1957) used 204 multiple choice items with 370 examinees in an effort to measure six skills related to reading comprehension. He concluded that only the vocabulary items were measuring a skill in comprehension that was

significantly different from the others.

If we assume that comprehension is what comprehension tests measure then a study by Lennon (1968) is important. He found through a survey of thirty or more studies of reading comprehension that the following components of reading ability can be measured reliably.

1. A general verbal factor related to vocabulary knowledge.
2. Comprehension of explicitly stated material, understanding literal meaning and ability to follow directions.
3. Comprehension of implicit or latent meaning, such as the ability to draw inference, predict outcomes and perceive a hierarchical arrangement of ideas within a selection.
4. Appreciation which includes the ability to see the intent or purpose of the author, judge the mood or tone of a selection and detect the literary devices which the author uses to accomplish his purposes.

Davis (1941) surveyed the literature to identify the comprehension skills deemed most important. He compiled a list of several hundred skills, many of which pertained to the mechanics of reading, many overlapped and others seemed incapable of measurement by objective test items. After careful scrutiny and removal of some skills (for reasons noted above) the remainder were grouped into nine clusters of operational skills thought to be employed in reading comprehension.

1. Remembering word meanings;
2. Selecting appropriate word meanings in the light of context;
3. Following the organization of a passage, as in identifying antecedents and references;
4. Identifying the main thought of a passage;
5. Answering questions for which explicit answers are given;

6. Weaving together the ideas in a passage;
7. Drawing inferences about the content of a passage;
8. Recognizing literary devices and identifying the author's tone and mood;
9. Drawing inferences about the author's purpose and point of view.

Davis (1964) endeavored to measure the skills in reading using multiple choice tests. He concluded that there were two general abilities, i.e. memory of word meanings and verbal reasoning; three specific abilities, i.e. following the organization of a passage, as in identifying antecedents and references, recognizing literary devices and identifying the author's tone and mood, and drawing inferences from content; one general category, i.e. answering questions for which explicit or paraphrased answers are given.

Davis (1967) found that word knowledge and reasoning seemed to account for virtually all of the variance of comprehension.

In a subsequent study, Davis (1968) endeavored to obtain an estimate of the percentage of non-chance unique variance in the reliable variance of each of the most important skills of comprehension among mature readers. He selected eight skills for measurement: 1. remembering word meanings, 2. finding answers to questions answered explicitly or merely in paraphrase in the context, 3. weaving together ideas in the content, 4. drawing inferences from content, 5. recognizing a writer's purpose, attitude, tone or mood, 6. drawing inferences about the meaning of a word

from context, 7. identifying a writer's techniques, and 8. following the structure of a passage. He concluded that comprehension among mature readers is not a unitary mental skill or operation. He decided that substantial parts of the mental abilities used in the eight skills judged to be of importance in comprehension are independent of one another.

One can plainly see that reading comprehension is not a single skill and other investigators have elaborated on the topic. Jenkinson (1968) suggests that most studies have identified four factors in reading comprehension, namely: vocabulary, interrelationship among ideas, abstract reasoning, and specific content field factors. Schubert and Torgerson (1968) say that most tests of comprehension place a high value on memory and further suggested that standardized tests of reading may not say so much about comprehension as about the speed of answering questions. While Langer (1969) says reading is really a thinking process, Zintz (1970) says comprehension skills are classified as literal comprehension and interpretative skills, and Harris (1970) believes that a minimum essential of reading comprehension is an understanding of the words used by the author.

Auerbach, (1971), after an extensive review of comprehension tests and studies of reading comprehension, concludes that "...there is no clear understanding yet of what these tests are measuring." (p.45)

Need for new theory

Disagreement concerning the nature and measurement of reading comprehension has resulted in a variety of approaches to evaluation of pupil achievement. The basic consideration in our schools should be the individual person and his errors on tests must be given equal consideration with his correct responses if adequate program planning is to result. Some specialists believe test errors are the fault of the test instrument and others suggest that errors are the result of inadequate psychological processing of information. A third group thinks that a definite sequence of steps must be prescribed for reading instruction and each step should be criterion referenced for easier evaluation.

Osburn (1968) believes that we must be just as interested in the wrong answer choice as the correct answer if we are to provide help to the student. Unfortunately, as Osburn points out, in practice we have done statistical analyses of test data in an attempt to generalize to a collection of test items. But ideally the basis of generalization should come from the operational definition of the procedures used in generating and sampling items that will make up the test. To accomplish Osburn's goal one would have to specify in advance all items that could possibly appear in the test and then sample randomly from the universe of items which are based upon the content. This is theoretically possible but would be very difficult to accomplish in practice.

Simons (1971) believes that theory based research

is necessary if basic psychological processes in reading comprehension are to be understood. The process of comprehension has often been confused with the product of comprehension since we usually measure the latter and infer the former. The fundamental problem with traditional reading tests, according to Simons, is the confusion about what is being measured. Reading tests could be measuring several skills and abilities but we are not sure because tests may lack construct validity and this is difficult to determine. He advocates more research on the processing of information since improved instruction depends upon an understanding of the basic psychological processes.

More recently, criterion-referenced tests have received more attention. Johnson and Kress (1971) indicate that in the case of reading tests, a careful analysis of the task being evaluated is a true prerequisite to the building of these tests. They suggest that this type of test requires a careful survey of all the skills and abilities which comprise reading. Once the skills and abilities are identified, each must be defined in terms of both its end result and the learning it requires and then placed in a hierarchical sequence. Test items are drawn from each level of the sequence, resulting in the determination of whether or not a particular individual can complete successfully a specified task and at what level of proficiency an individual is operating in progressing toward mastery of a particular skill or ability. Prescott (1971) feels that the common multiple choice test

item is inadequate because the correct answer alternatives "...are possible choices an individual might make in his ignorance." (p. 351) He further states that multiple choice test items as they are currently constructed function only when the individual has some knowledge on which to make his choice, but if he guesses when having no information the item measures nothing even if the guess is correct. With criterion-referenced tests there is an attempt to evaluate performance in terms of whether an individual has achieved or has failed to achieve specified objectives. This seems to have merit since the examiner would be concerned about comparing the performance of the testee with some concrete standard of reading performance rather than with other individuals. In this case one would be just as interested in an analysis of errors as in the correct answer. The errors would have diagnostic value to the extent that they indicated specific reading difficulty.

Bormuth (1970) has recently expressed his views on achievement test writing. He favors the construction of tests by operational definition. He holds the fundamental premise that:

An achievement test item cannot produce meaningful results unless it can be related to and derived from the instruction by a set of operations. The operations themselves describe a set of publicly observable manipulations by which a test item is derived from the instruction. (p.3)

Bormuth is critical of traditional methods of item writing and points out that no constraints are placed on the

writer and no distinction has been made between content and behavior. There is no limitation on how content is to be sampled or the number of test items. There is no control exercised over the relevance of test items to the content or instruction. Bormuth advocates a method of test construction which would allow the student's score to be compared with a criterion of mastery of the content and be relevant to instruction.

The items in standardized achievement tests must be relevant to the students' instruction. There is no way to interpret a test score if an unknown proportion of the items test content not included in that student's instruction. The score cannot be compared with the scores of students in the test's norm group, because many of the students in the norm group would have been subjected to many different instructional programs. Part of the variability in their scores, then, would be due to the fact that greater or lesser proportions of the items test content contained in their instruction. Thus, unless it can be demonstrated that the items in the test were relevant to the instruction of both the student being evaluated and the students in the norm group, the score must be regarded as confounded. (pp.23-24)

Traditional item writing techniques prevent inspection by others. There is no formula or specific instructions to ensure consistency which others might follow. No one working independently can actually duplicate the item except by chance, nor can they specify what must be measured by each item.

According to Bormuth's proposal, the phrasing of an item as well as other aspects of the test would be determined by definition. The content of each item would be specified beforehand and item writers working independently

could replicate each other's work. The rules for relating items to the instruction would ensure that a test is validly measuring passage content.

The domain of item writing has been opened to questioning. There are demands for improved methods. Any method of item writing must endeavor to ensure validity and items must be randomly selected from a pool which has defineable limits. The facet design is an attempt to meet these criteria.

Facet test

Schlesinger and Weiser (1970) following the work of Guttman and Schlesinger (1966, 1967) and earlier work by Guttman (1965) attempted still another approach for measuring comprehension which makes possible the systematic construction of test items. It concentrates on the relation of the item to the information presented in the text or elsewhere but not on the skills and abilities that are presumably involved in comprehending the text information. The authors point out that since we are unable to specify the abilities necessary for reading comprehension and we cannot measure the processes but only infer them, then we should concentrate on making items which are closely associated with the information exposed to the examinee. This design can be applied to multiple choice type questions and what the item measures is defined operationally in relation to the information upon which it is based.

Guttman and Schlesinger (1966) were originally

concerned with intelligence testing and reasoned that a new approach to test construction was necessary to increase the validity of intelligence tests. Prior work had been done by Guttman (1965) in defining intelligence. The result was a new concept labelled facet design. The facets appear to be sets of ideas rather than constants. Guttman and Schlesinger (1967, p.3) defined the concept of facet as follows:

...two sets of elements, A and B are called facets, and their Cartesian space is the set of all pairs of elements ab where a is an element of A and b is an element of B. A Cartesian space may consist of any number of facets, or sets of elements; with n facets; any one point in the Cartesian space has n component elements.

An example is provided in Table 2.1 below.

Table 2.1 Eight distractors of a test item in which figures vary in three facets, A, B, and C (after Guttman and Schlesinger, 1967)

Distractor	Facet A	Facet B	Facet C
1	a_1	b_1	c_1
2	a_1	b_1	c_2
3	a_1	b_2	c_1
4	a_1	b_2	c_2
5	a_2	b_1	c_1
6	a_2	b_1	c_2
7	a_2	b_2	c_1
8	a_2	b_2	c_2

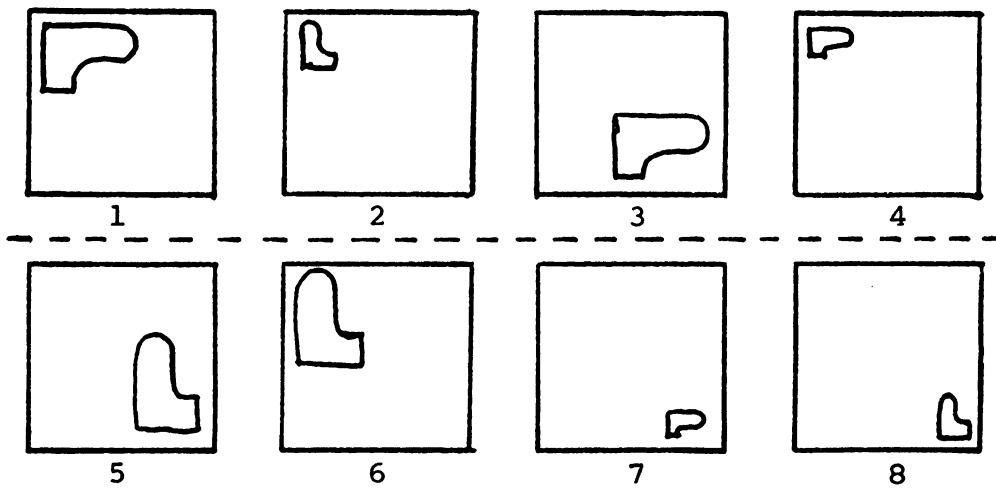
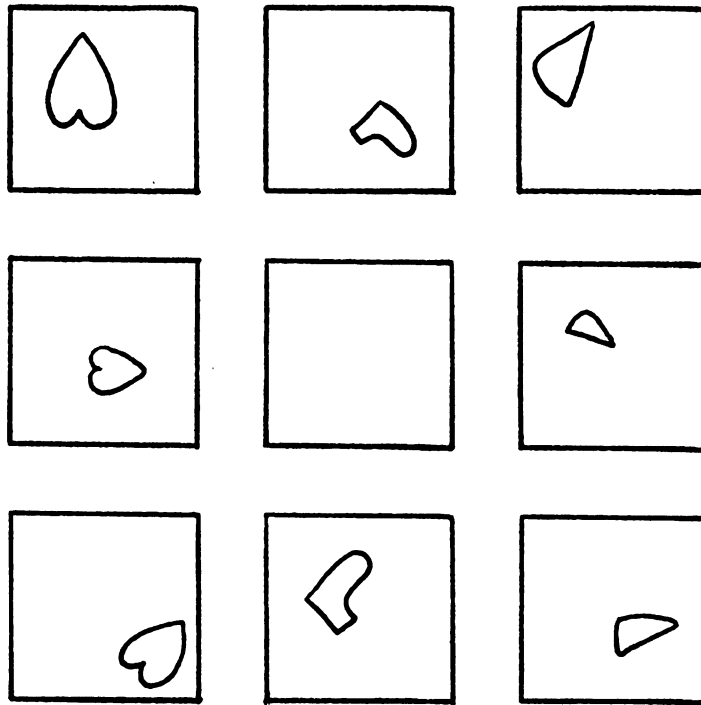
There can be only one correct answer in which all three facets represent the facts. Then there are three other distractors which differ from the correct answer on one of the three facets, three distractors which differ from the correct one on two of the three facets, and one distractor which differs on all three facets. If distractor 1 is the correct answer then distractor 2 differs from it in facet C only, distractor 3 in facet B, and distractor 5 in facet A. Distractors 4, 6, and 7 differ from it on two facets each and distractor 8 is wrong on all three facets.

An example of a test item taken from Guttman and Schlesinger (1967) is shown in Figure 2.1 on page 26. In this example the missing figure of the ninth square must be provided by the subject. Completing the item requires examinees to consider the facets, (a) shape, (b) size, (c) orientation, and (d) place.

The function of a distractor in any item is to be attractive to the respondent who does not know the correct answer, according to English and English (1958). For most test authors, the compilation of adequate distractors is not only basic but difficult, and yet it is mostly decided upon by intuition which suggests that item writing may be described as an art (Davis, 1964).

In the facet design it is argued that distractors can be constructed in a systematic manner on an a priori basis. This approach provides at least three useful features not supplied by less systematic approaches.

Figure 2.1 Test item from Guttman and Schlesinger (1967)



1. There is prediction of the relative empirical difficulties of distractors, 2. there is less variation in test results from undesired factors, and 3. there is the possibility of classifying the examinees or determining their reading difficulties according to the types of answers to which they were attracted.

The facet design is a way of devising questions systematically, thus keeping arbitrary decisions to a minimum. Then too the systematic construction of distractors for all questions is facilitated by the facet design. This assumes particular importance in the construction of parallel forms of a test. Facet design leads to a more complete definition of the universe of test items from which only a sample is drawn for a specific test. Unless some such design is employed, there could be no way of definitively predicting whether items included in two parallel forms are of comparable difficulty, even though they may belong to the same universe. In the construction of parallel forms of a test, control over the selection of distractors is of particular importance. In facet design the distractors are produced by identical criteria.

Facet design applied to a test of reading comprehension concentrates on the relationship between the test item and the text upon which it is based. The facets of comprehension are defined and then related to the text content. Schlesinger and Weiser (1970) chose four facets as follows:

1. the form of information (explicit, implicit);
2. the source of information (the text, formal instruction, informal experience);
3. the context in which the information appears (appropriate, not appropriate);
4. the frequency with which the information appears in the source. (p.571)

Guttman and Schlesinger (1967) point out, however, that one facet can be applied as well as several facets to this design as long as the test writer specifies beforehand how the distractor will be written. The facet design can best be summarized by means of a "mapping sentence" which is a formal statement of the research design. (Schlesinger and Guttman, 1969; Schlesinger and Weiser, 1970)

The mapping sentence used by Schlesinger and Weiser (1970, p.577) is as follows:

The relation of part (x) of statement (y) to
 information appearing {one
several} time(s) in {explicit
implicit}
 form in the {appropriate
not appropriate} place in {the text
formal instruction
informal experience}
 → {agreement
no information
contradiction}

The domain of the mapping sentence (the part that precedes the arrow) contains the facet by which each distractor is analyzed. The range of the mapping sentence is found to the right of the arrow and contains three basic categories of "agreement", "no information", and "contradiction". The sentence is read from left to right with one

element being selected from each facet and these are mapped into the range. Used in this manner the mapping sentence provides for each part of a distractor-statement (y), as many mappings into the range as there are combinations of elements in the domain.

While this design is still experimental, this study endeavored to assess its usefulness.¹ A prediction of the facet design made by Jordan (1968) seems appropriate at this point.

Facet theory promises to be extremely useful in determining the content of research instruments. As this technique becomes more known and available to the social scientist, it is likely that attitude research instruments will contain the kind of content that will allow systematic quantitative predictions from qualitative data.

The facet test proposes a new approach for measuring reading comprehension but there is need of comparison with other existing tests.

Cloze technique

The "cloze" procedure was first introduced by Taylor (1953) as a measure of readability. Later Taylor (1956) recognized that the cloze procedure was also suitable for use as a measure of the ability of readers to understand printed materials. Passages to be used for the cloze technique are prepared by deleting words according to an objectively specified procedure. It is customary for words to be

¹The facet design as used in this study is explained more fully in Chapter III.

deleted in one of two ways: either all of the words of a particular type, i.e., semantic or syntactic type are deleted; or every nth word is omitted. Deleted words are replaced by blanks which are of uniform length and it is the task of the subject to place in each blank the word which he thinks best fits the immediate context and the passage theme. Subjects may either read the unmutilated passage first and then fill in the blanks of the cloze test or, in most cases of readability studies, the subject is required to fill in the blanks without having first been introduced to the unmutilated passage. The completed tests are normally scored by counting as correct only those words which exactly match the deleted words. An acceptable alternative is to also count as correct the synonyms of the deleted words.

Research subsequent to Taylor (Jenkinson, 1957; Rankin, 1959; Ruddell, 1963; Bormuth, 1964, 1965, 1967, 1969) has established the usefulness of the cloze procedure as a measure of comprehension and has provided data which attests to its validity and reliability as a comprehension test.

The cloze procedure has been compared with multiple choice tests to check on the measurement of comprehension. The research cited above has been accomplished through this means to a large extent. However, though cloze has been compared with some standardized tests (Rankin, 1965), it has seldom been compared with the Davis Reading Test, as noted by Weaver (1965). This has been done in terms of a

correlation coefficient in this study and is reported in chapter IV.

Summary

While measurement has been a part of the educational scene for many years and has had great importance since E. L. Thorndike's work in 1914, there is still much disagreement as to whether many tests, especially in reading, are valid for their stated purposes. Tests have been used by classroom teachers often as an end result rather than to promote better instruction, which should be the goal. A lack of understanding of test limitations on the part of the users may be the reason for this. It also appears that many tests in reading comprehension have been constructed without a clear set of objectives since there is not consensus as to the definition of reading and all of its components. This latter fact has led to criticism of reading comprehension tests and some question about their validity and usefulness.

Recently, test conscious educators and theoreticians have proposed new models of test construction based upon a new theory base that purports to build validity into the instrument because of its operational definition. Bormuth, Guttman, Schlesinger, Weiser and others see this as an essential feature of tests.

The facet design is just one model of test construction which can be used to devise a test of reading

comprehension. With this model the test items are constructed systematically on a priori basis or from a preplanned definition, which acts to ensure validity and internal consistency.

CHAPTER III

THE DESIGN AND PROCEDURES

This research study was designed to examine the use of a facet design to measure reading comprehension. A forty item multiple choice test based on the principle of facet design was prepared by the researcher and was administered, along with two other measures of reading comprehension, to a sample of grade ten students. Scores on the facet test were correlated with scores on the two other measures of reading comprehension.

Population and sample

The sample consisted of 186 students chosen from a grade ten population who had just entered high school two weeks previous to the study. These students had completed junior high school in several feeder schools and represented a large area of the city surrounding the high school.

Three grade ten English classes were selected from each of two public high schools in Edmonton, Alberta. Since all students are required to complete an English course, these students appeared to be a representative sample of grade ten students in Edmonton. Some students were enrolled in courses leading to University entrance and others were not.

Measuring instruments

The three measures of reading comprehension used in this study include: (1) the Davis Reading Test, form 2B, designed for use in grades eight through eleven, (2) a facet test, constructed by the writer, consisting of a series of four passages with accompanying multiple choice questions and (3) cloze test based upon the identical passages contained in the facet test.

1. Davis Reading Test

The Davis Reading Test is a group test designed for grades eight through eleven and is useful in assessing the overall reading ability of individuals. The test is composed of eighty items constructed to measure the ability to use nine operational skills necessary in reading comprehension (Davis Reading Test Manual, 1962). The test items are based on the content of several passages of varying topics and lengths which high school students might be required to read. The test consists of eighty items and provides two scores: A Level of Comprehension score, computed on the first forty items, which indicates the depth of understanding displayed by the student and a Speed of Comprehension score based on all eighty of the items in the test, which indicates the rapidity and accuracy with which he understands this self-same material.

Davis computed the reliability of this test using parallel forms with a two-to-four week interval between test administrations. The reliability coefficients for grade

ten are reported as .78 for Level and .86 for Speed.

Reviewers (Buros, 1968) say that the method of construction of the Davis Reading Test seems to ensure test validity. (Buros, 1968) Davis surveyed the literature and compiled a list of several hundred skills which educators believed necessary for effective reading. Skills pertaining to the mechanics of reading and those incapable of being measured by objective test items were deleted. The remaining skills were grouped into nine clusters. After careful experimental study, Davis concluded that each of the nine operational skills is closely related to a basic ability underlying accurate comprehension in reading. He reasoned that memory for word meaning was so pervasive an ability that it was decided not to measure it separately. The eight remaining operational skills were grouped into the five categories listed below. (Davis Reading Test Manual, 1962)

1. Finding the answers to questions answered explicitly or in paraphrase in a passage;
2. Weaving together the ideas in a passage and grasping its central thought;
3. Making inferences about the content of a passage and about the purpose or point of view of its author;
4. Recognizing the tone and mood of a passage and the literary devices used by its author;
5. Following the structure of a passage.

Davis reports that this test correlated highly with other measures of verbal skills and has a predictive validity of about 0.5 when compared with high school and college English grades.

2. Choice of passages for facet and cloze tests

Passages for these tests were selected from a variety of materials that were deemed suitable for grade ten pupils. The Department of Education of the Province of Alberta has compiled, for each grade level, a list of recommended reading materials for each of the subject matter areas in the curriculum. Four books which teachers agreed would be most widely used in grade ten, were chosen from this list. These books corresponded to the subject areas of Social Studies, Science and English. One passage was selected from each of the four books as reported in Table 3.1. Each passage was self-contained in the sense that it could be read with understanding without having read preceding or subsequent paragraphs. The passages ranged in length from 586 words to 614 words. In addition, one article was chosen from a current, popular news magazine¹ that would pertain to either the Social Studies or Science curricula. (See Table 3.1) The titles of the passages and their sources are provided in Table 3.1. Some problems appearing in the pilot study (discussed later) resulted in the deletion of one passage. (See Table 3.1)

3. Readability study

The Fry formula (Fry, 1968) was used to assess the readability of the passages. The formula is reported as a reliable measure of readability by Maginnis (1969) and Pauk (1970).

¹TIME, May 17, 1971; see Appendix A.
An informal survey of teachers in high school revealed that Time magazine had often been cited for related information dealing with Social Studies.

Table 3.1 Title and source of test passages

Passage title	Source	No. of words	Curriculum area
Upgrading Neanderthal Man	TIME Canada Edition May 17, 1971, pp. 50-51.	619	Social Studies and Science
Types of Municipal Government	Greason, George K. and Roy C. King; <u>The Citizen and Local Government</u> , The Macmillan Company of Canada, Limited, Toronto, 1964.	586	Social Studies
How Snow Helps and Hinders Survival	Hardy, W. C., Alberta: <u>A Natural History</u> , M.C. Hurtig Edmonton, Alberta, 1968.	614	Science
The Balance of Nature	<u>Renewable Natural Resources in Alberta</u> , Dept. of Education, Edmonton, Alberta, September, 1962.	600	Science
Deleted from study			
Across the Pacific: I	Heyerdahl, Thor <u>The Kon-Tiki Expedition</u> , pp. 39-41.	604	English

Table 3.2 Readability of facet test passages

Passage	No. of words in each sample	No. of syllables/ 100 words in each sample	Mean no. of syllables/ 100 words in passage	No. of sentences/ 100 words in each sample	Mean no. of sentences/ 100 words in passage	Readability level
Types of Municipal Government	120 92 99	184 191 162	172	5.0 5.4 5.0	5.1	College
The Balance of Nature	112 84 110	154 144 148	149	5.4 6.0 5.5	5.6	Grade 9
How Snow Helps and Hinders Survival	107 123 115	141 132 132	135	5.6 4.9 4.3	4.9	Grade 7
Upgrading Neanderthal Man	95 115 127	167 157 169	164	4.2 5.2 3.9	4.4	College
Across the Pacific: I (deleted from study)	112 120 91	117 138 134	130	6.2 5.8 4.4	5.5	Grade 7

Three samples of approximately 100 running words each were taken from each of the passages. The mean number of words per sentence and the mean number of syllables per 100 running words was determined for each of the passages. This information, along with the readability levels computed therefrom, is reported in Table 3.2.

4. Facet test

This test was constructed in accordance with the work of Schlesinger and Weiser (1970) and Guttman and Schlesinger (1966, 1967). While Guttman's original approach was aimed at increasing the validity of intelligence tests, it has since been suggested that this approach is applicable to evaluating reading comprehension (Schlesinger and Weiser, 1970). The facet design concentrates on the relationship between the content of the test item and the information from the text upon which it is based, rather than on the skills and abilities which are presumed to be involved in choosing the correct response to the item. Although Schlesinger and Weiser (1970) suggested that four facets could be applied this study concentrates on two facets only, namely, form of information and source of information.

The form of information facet pertains to whether the test item distractor contains information explicitly stated or implicitly stated from the text or some other source. By explicit is meant that the information must be stated in the exact words used in the passage or some related passage.

Implicitly stated information is defined as a paraphrase of the information stated in the passage or some related source.

The source of information facet refers to where the information for a particular test item distractor was obtained. The test items were constructed in such a way that each distractor statement contained either information presented in the passage or information from other sources and this was expressed either in the exact words of the source or in paraphrase form.

Each distractor statement was constructed so that it conforms to one of three conditions, all of which are referred to as the range: (a) it agrees with information found in the passage, (b) it contradicts (disagrees) with information found in the passage, or (c) there is no information in the passage relating to it. To summarize the foregoing, a mapping sentence such as used by Guttman and Schlesinger (1967), London (1968), and Schlesinger and Weiser (1970), illustrates this approach.

The variable portion of each distractor
statement will be $\begin{matrix} \text{(Facet One)} \\ \left\{ \begin{array}{l} \text{a paraphrase} \\ \text{exact words} \end{array} \right\} \end{matrix}$
 $\begin{matrix} \text{(Facet Two)} \\ \left\{ \begin{array}{l} \text{in the text} \\ \text{in some other source} \end{array} \right\} \end{matrix}$
of information found
 $\longrightarrow \begin{matrix} \text{(Range)} \\ \left\{ \begin{array}{l} \text{Agreement} \\ \text{No information} \\ \text{Contradiction} \end{array} \right\} \end{matrix}$

The "domain" of the mapping sentence, that part which precedes

the arrow, contains the facet according to which each distractor will be analyzed. The "range" of the mapping sentence is found to the right of the arrow. The sentence is read from left to right with one component being selected from each facet. Thus the facet design provides that each variable part of the distractor statement will have as many possible "mappings" into the range as there are combinations of elements in the domain.

The facet test in this study was constructed according to the design of Schlesinger and Weiser (1970) except that only two facets are built into this test. If one were to let a_1 = paraphrase, a_2 = exact words, b_1 = text (passage), and b_2 = other source of information, the possible combinations of the variable component of the statement would be $a_1 b_1$ (paraphrase from the text), $a_1 b_2$ (paraphrase from some other source), $a_2 b_1$ (exact words from the text), $a_2 b_2$ (exact words from some other source). This scheme is shown in Table 3.3. Each distractor (variable component of the statement) thus includes two facets. The variable part of the domain can be mapped into the range in such a way that the total statement, including both non-variable (stem) and variable (distractor) components, is (a) in agreement with information found in the passage, (b) contradicts information found in the passage, or (c) includes information not found in the passage.

Two distractor statements for each item contain the subfacet b_2 where information is taken from a source

other than the passage. One distractor is closely related to the theme of the passage which the student reads, while the other distractor is taken from a passage somewhat unrelated to the theme of the passage. As a result, in the combinations $a_1 b_2$ (paraphrase, other source) or $a_2 b_2$ (exact words, other source), one distractor contains information related to the passage theme and the other distractor comes from a passage with little or no relationship to the theme of the original passage.

Table 3.3 Facet test mappings into the range

Facet	Parts	Symbol	Variable Combinations	Symbol	Range
1. Form of information	paraphrase (implicit)	a_1	paraphrase text	$a_1 b_1$	agreement
	exact words (explicit)	a_2	paraphrase other source	$a_1 b_2$	contradiction
2. Source of information	text (passage of test)	b_1	exact words text	$a_2 b_1$	agreement
	other source	b_2	exact words other source	$a_2 b_2$	contradiction

For example, question number one from the passage, "Upgrading Neanderthal Man" (Time, May 17, 1971), is given below as it appeared in the test.

From the first discovery of Neanderthal's bones

- a. there has been a connection with Dr. Ralph S. Solecki, who led an expedition to Lebanon last summer.
- b. his name has been associated with crudely primitive in manner or conduct.
- c. his name has been synonymous with brutishness.
- d. Darwin's theory has withstood a century of criticism.

There were ten questions constructed for each of the four passages (see Appendix A) so that the entire test contained forty items. Within each item only two kinds of distractors could possibly be accepted as correct responses. Only those distractors where the facet - source of information - came from the passage which the students read could be correct, since by definition the correct answer was either a paraphrase or exact words from the passage.

Table 3.4 explains the construction of the items used in the facet test.

Table 3.4 Example of facet test question

Passage: Upgrading Neanderthal Man Question No. One

a₁ - paraphrase, a₂ - exact words, b₁ - text, b₂ - other source

Unvariable portion (stem): From the first discovery of Neanderthal's bones

Facet combination	Distractor (variable statement)	Range	Source of information
a ₁ b ₁	there has been a connection with Dr. Ralph S. Solecki, who led an expedition to Lebanon last summer.	Contradicts	Passage given to students to read
a ₁ b ₂	his name has been associated with crudely primitive in manner or conduct.	No information	Volume 14, World Book Encyclopedia, p. 85 <u>Related*</u>
a ₂ b ₁	his name has been synonymous with brutishness.	Agrees	Passage
a ₂ b ₂	Darwin's theory has withstood a century of criticism.	No information	Hegner, R. W. and Stille, K. A., <u>College Zoology</u> , The Macmillan Company, New York, 1954, p. 797. <u>Not related**</u>

* Related means that this statement came from a passage with similar theme to the actual passage read by the students taking the test.

** Not related means that this statement came from material which was only somewhat related in content to the passage read by the students taking the test.

5. Cloze test

The cloze test, first constructed by Wilson Taylor (1953), is an instrument designed to measure comprehension and readability of materials. In recent reviews, Rankin (1965) and Bickley, et al. (1970), conclude that cloze tests are valid and reliable measures of reading comprehension.

A cloze test was constructed using the passages from the facet test previously discussed. Counting from the first word in the passage, every tenth word was deleted. Hyphenated words were counted as one word. In order to provide adequate context, Culhane (1970) recommends that the first sentence in a cloze passage remain unmutilated. In this instance, however, the unmutilated passages had previously been read during the administration of the facet first and that appeared to provide sufficient context to permit the mutilation of passages, counting from the beginning of the passage. Each deleted word was represented by a line equal in length to ten typewriter spaces. Taylor (1956) recommends that cloze tests contain at least 50 items (deletions). Each of the subtests (passages) contained from 58 to 62 items (deletions) for a total of 240 items in the test.

The scoring of these passages was modified slightly from the usual exact word system. Proper nouns with only one letter missing from the exact order, such as Soleck for Solecki, were counted as correct. Plurals were accepted if it appeared difficult to distinguish from sentence context whether a plural or singular form was required, provided number (singular and

plural) did not influence the meaning of succeeding sentences. Words at the beginning of sentences were counted correct regardless of capitalization. The administration of the cloze test followed by one week the administration of the Davis Reading Test and the facet test.

Pilot study¹

Twenty-three pupils enrolled in a grade ten mathematics course during summer school 1971 were chosen as subjects for a pilot study. The summer school was administered by the Edmonton Public School Board. There were both male and female pupils registered in the course and regular academic grade placement ranged from ten to twelve. Equal numbers of facet and cloze tests were distributed in a random order so that twelve students had the facet test and eleven students completed the cloze test.

Five passages were selected from materials that grade ten students might be expected to read. A cloze test and a facet test were constructed for each passage. The cloze test for the five passages consisted of 302 items. Each passage in the facet test was followed by ten items for a total of 50 items.

The cloze tests and facet tests were organized in a similar manner. In the cloze test each unmutilated passage was followed by a mutilated version with every tenth word deleted. The five passages with corresponding cloze tests

¹The choice of passages and test construction have previously been described.



were then put in a folder in the following order: Across the Pacific: I, Types of Municipal Government, How Snow Helps and Hinders Survival, Upgrading Neanderthal Man, The Balance of Nature. The facet test was assembled in the same order except that facet questions followed the respective passages.

Seventy minutes were allotted for the tests and students were permitted to leave when finished. All students were able to complete the facet test but three students failed to complete the cloze test within the time allotment.

The pilot study showed that some modification was in order. Because testing time to fit the school schedule, was restricted to forty-five minutes, it would be necessary to shorten the tests. Weaknesses revealed by item analysis showed that revisions in one of the passages would be too extensive to be practical. Accordingly the passage, entitled, "Across the Pacific: I. was deleted from both the cloze test and the facet test. Because some students proved to be slow workers, the passages and test items in the cloze test and the facet test were arranged in rotational order to avoid bias which might stem from the inability of some students to read the latter parts of the tests within the time limits.

The pilot study revealed some weaknesses among the distractors used in the facet test. It showed that distractors dealing with information taken from another source, although unrelated to the theme of the test passage, should appear highly plausible if they were to be attractive to the examinee. Sometimes it was necessary to revise the distractor

containing information from another source to make it more explicit, at other times it was deleted in favor of a more attractive alternate that would better fit the context identified by the invariable component of the statement.

Administration of test instruments

The tests used in the study were administered in the third week of the fall term of the academic year 1971-72. Both high schools from which subjects for the study were selected operate on the semester system. Class schedules provide for ninety minute periods with short rest periods at the mid point. Two testing sessions were scheduled for each school, one week elapsing between sessions.

In the first session students were given the Davis Reading Test during the first half of the period and the facet test in the last half of the period. Time limits for the Davis Reading Test are forty-five minutes. The mid-point rest period was used to collect and distribute testing materials.

Directions for administering the facet test specified that students were to read the passage and then respond to the multiple choice items which followed without referring again to the passage. Students were to proceed with the passages in the order in which the passages were assembled in the test folder.

One week hence the cloze test was administered to these self-same students. Students were directed to read the

first unmutilated passage and then respond to the cloze test for that passage, in that order, and then proceed to the next passage in the same manner until the test was completed or time was called, whichever came first. Students were not permitted to leave the room until the time limit expired.

The passages for both the facet and cloze tests were assembled so that each unmutilated passage was followed by the test for that passage. The sequential arrangement of the materials with each test was systematically varied so that not all students dealt with the passages in the same order. The testing materials were randomly distributed and students were instructed to proceed according to the order in which the materials were received.

Analysis of data

Data processing was done at the University of Alberta computer center.

The Kuder-Richardson Formula 20 was used to determine the reliability of the facet test. Thorndike and Hagen (1969) suggest that there is no universally accepted minimum value of this coefficient but that the higher the value the better satisfied is the test author. Davis (1964) says that many researchers set 0.50 as the minimum acceptable value of a reliability coefficient. When groups exceed 100 in number, a reliability coefficient of 0.60 was shown acceptable by Thorndike and Hagen (1969). Accordingly a reliability coefficient of 0.60 was set as the minimum

acceptable value for the facet test.

An item analysis provided measures of the difficulty and the discriminatory power of each of the items in the facet test. The difficulty index is expressed as the per cent of students choosing the correct distractor for an item. The biserial correlation coefficient and the item reliability index are used to show the degree to which an item discriminates between high and low achievers.

The Davis Reading Test was used as a criterion measure for grouping students according to achievement levels. Students whose scores exceeded a point one-half standard deviation above the mean were classified as high achievers. Students classified as low achievers earned scores which deviated more than one-half standard deviation below the mean. Students with scores falling between one-half standard deviation below the mean and one-half standard deviation above the mean were designated as the mid group. An item analysis was completed on the facet test for each group.

Finally, a Pearson product-moment correlation coefficient was computed to determine the relationship among the three measures of reading comprehension.

Summary

This study was concerned with the development of a test based on the principle of facet design for measuring achievement in reading comprehension. The test, along with two other measures of reading comprehension, was administered

to a sample of grade ten students enrolled in two of the public high schools in Edmonton, Alberta. This chapter describes these tests and how they were constructed as well as how they were administered. A pilot study which aided in the research is also described.

CHAPTER IV

RESULTS

Introduction

A multiple choice test based on facet design was prepared and administered to a sample of tenth grade students. The test was constructed on an operational plan to ensure that each item was related to information in the text. Test results were analyzed and are reported in this chapter.

Results concerning hypothesis one

A test constructed by facet design will not be statistically reliable.

A Kuder-Richardson Formula 20 reliability coefficient was computed on respondent's scores from the facet test. Table 4.1 lists the number of subjects, the test mean, standard deviation and correlation coefficient of 0.80. An acceptable value for this coefficient was previously set at 0.60¹.

It was concluded that this test was a reliable measuring instrument.

¹See Analysis of data, Chapter III.

Table 4.1 Facet test statistics

Number of subjects	test mean	standard deviation	Kuder-Richardson 20 reliability
185	25.67	6.13	0.80

Results concerning hypothesis two

There will be no difference among achievement levels when a test is constructed by facet design.

Student responses to the facet test were analyzed by means of item analysis. Table 4.2 shows the raw score distribution. A further clarification of this information is provided in Table 4.3.

Five groups were formed, each representing a twenty per cent segment of the subjects taking the test. Group one is the twenty per cent with the lowest scores on the facet test and group five is the twenty per cent with the highest scores. The other groups represent similar divisions within the range. Table 4.3 summarizes information to show that each group contains substantial numbers of scores and indicates a distribution of scores according to achievement.

Table 4.2 Frequency distribution of the scores on the facet test

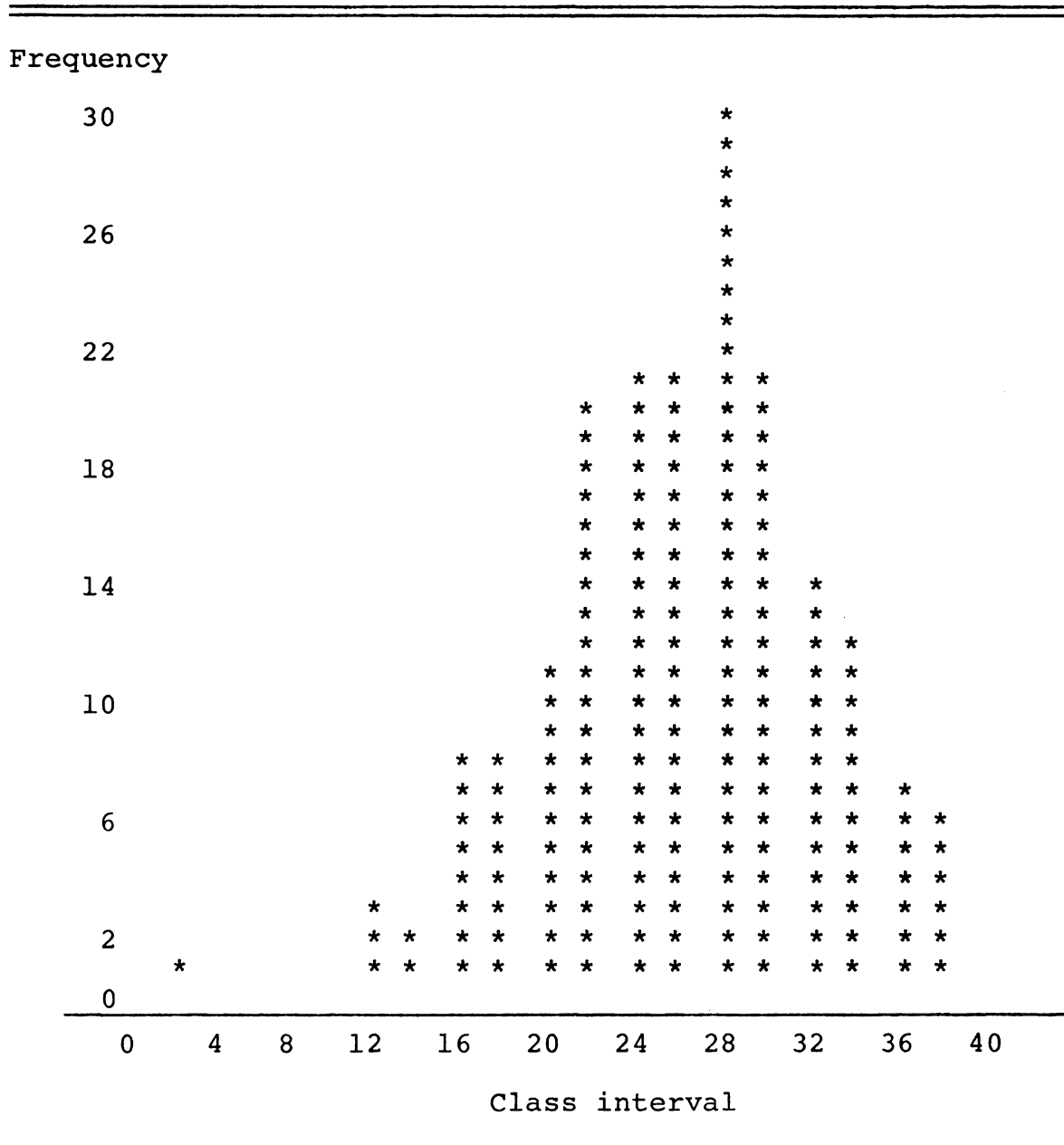


Table 4.3 Cutting scores for five groups on facet test

	1	2	Group 3	4	5
Score	21.5	21.5-24.5	24.5-27.5	27.5-30.5	30.5
number of subjects	41	33	40	32	39

Table 4.4 summarizes information computed about each item. The item difficulty index is the per cent of students choosing the correct item response. High values indicate many students selected the correct response. How well the item discriminated between higher achievers and lower achievers is indicated by the bi-serial correlation coefficient and the item reliability index. Values for the bi-serial correlation coefficient are seldom greater than 1.0, while values for the item reliability index do not exceed 0.50.

The item reliability index combines both the item difficulty index and bi-serial correlation into one number that expresses the adequacy of an item. It has been found by researchers at the University of Alberta that a reliability index less than 0.12 indicates extremely weak items.

There were thirteen items with an item reliability index less than 0.20 (Table 4.4). Six of these had an item difficulty index less than 0.50, indicating that for these

Table 4.4 Item reliability on facet test for all subjects

	item number																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
item																				
difficulty	.72	.32	.59	.62	.75	.75	.85	.77	.41	.32	.63	.85	.86	.73	.32	.75	.79	.20	.59	.88
index																				
bi-serial																				
correlation	.53	.32	.34	.54	.53	.46	.64	.55	.51	.35	.44	.41	.73	.58	.52	.74	.68	.36	.51	.65
coefficient																				
item																				
reliability	.24	.15	.17	.26	.23	.20	.23	.23	.25	.16	.21	.14	.25	.26	.24	.32	.28	.15	.25	.21
index																				
	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
item																				
difficulty	.58	.96	.95	.58	.45	.28	.50	.86	.82	.87	.69	.54	.54	.76	.52	.59	.84	.67	.69	.61
index																				
bi-serial																				
correlation	.43	.73	.50	.34	.36	.14	.25	.58	.42	.73	.42	.31	.49	.56	.54	.52	.60	.49	.44	.43
coefficient																				
item																				
reliability	.21	.14	.11	.17	.18	.07	.12	.20	.16	.25	.20	.15	.25	.24	.27	.25	.22	.23	.22	.21
index																				

six items the correct distractor was not a popular choice with the majority of students. Four items had an item difficulty index greater than 0.80 but an item reliability index less than 0.17. These latter four items, while chosen by most students, did not discriminate between the better and poorer students and must be classed as the weaker items.

Three groups were formed from the scores on the Davis Reading Test Level of Comprehension score¹. These groups were labelled low, mid, and high². An item analysis of the facet test was completed for each of these groups. The number of students, mean score, and standard deviation for each group are listed in Table 4.5. The groups have well defined limits and indicate that the facet test does discriminate among achievement levels.

A Kuder-Richardson 20 reliability coefficient was computed for each group and is reported in Table 4.5. The acceptable criterion of 0.60 for this value was surpassed for each group even though the variance was considerably reduced as compared with the total group.

The conclusion drawn from the data summarized in Tables 4.2, 4.3, 4.4, and 4.5 was that test items constructed by facet design discriminate among achievement levels and leads to a rejection of hypothesis two.

¹ The Level of Comprehension score was chosen because almost all students completed the first forty items of the Davis Reading Test upon which this score is based. This would seem to provide a more accurate measure of reading comprehension than if all eighty items were used where many of the students did not complete the test.

² See Chapter III, Davis Reading Test.

Table 4.5 Results on facet test of low, mid, high groups*

group	no.	mean	standard deviation	K-R 20 reliability
low	54	20.22	5.06	0.685
mid	66	26.12	4.34	0.617
high	65	29.74	4.99	0.762

* These groups were determined by scores on the Davis Reading Test.

Low group - $1/2$ s.d. below the mean

Mid group - between $\pm 1/2$ s.d. from the mean

High group - $1/2$ s.d. above the mean

Results concerning hypothesis three

There will be no statistically significant correlation between the facet test and other measures of reading comprehension, i.e., between (a) the facet test and cloze test, and (b) facet test and the Davis Reading Test.

A Pearson product-moment correlation coefficient was computed to determine the relationship between (a) the facet test and the cloze test, and (b) the facet test and the Davis Reading Test. The results are reported in Table 4.6.

Table 4.6 Correlation coefficients for facet test, cloze test and Davis Reading Test

	N	cloze test	Davis Reading Test a.
facet test	150	.460*	.664* (level) .712* (speed)

* correlation coefficients were significant beyond the .01 level of confidence

a. The Davis Reading Test yields two scores: (a) level, items 1 - 40, (b) speed, items 1 - 80.

The Davis Reading Test yields two scores. The score on the first forty items is designated as the level of comprehension, and the score on the total test as the speed of comprehension. The facet test correlations with both scores on the Davis Reading Test were respectively .664 and .712. These were significant beyond the .01 level of confidence.

The correlation coefficient of .460 between the facet test and the cloze test was significant beyond the .01 level of confidence.

It was concluded that a significant relationship exists between (a) the facet test and the cloze test, and between (b) the facet test and the Davis Reading Test, thus rejecting hypothesis three. These tests appear to be measuring similar aspects of reading comprehension.

Additional considerations

There seemed to be more information that could be gleaned from the study that might be of value. The questions to be answered are related to the facets of the test.

- a. Does the form of information facet (whether information is stated in exact words or in paraphrase) have any effect upon examinee response?

Information from the item analysis was useful in answering this question. The item difficulty index for the facet test summarized in Table 4.4 was compared with the per cent response for each distractor found in Table 4.7.

The twenty items having exact words from the text ($a_2 b_1$) as the correct answer had an item difficulty index greater than 0.50 and eight of these items had an item reliability index less than 0.20. This is to say that all items with $a_2 b_1$ as the correct answer had over 50 per cent response from the students taking the test, with eight items being poor discriminators between better and poorer students.

The twenty items having paraphrase from the text ($a_1 b_1$) as the correct answer had only eleven items with an item difficulty index greater than 0.50 and five of these items had an item reliability index less than 0.20.

The paraphrase items appear to be more difficult for examinees. Test items with exact words from the text seem to be easier to detect than a paraphrase from the text. This may be due to the difference in language structure used in paraphrase questions or it may be that exact words from



Table 4.7 Per cent response for all distractors on facet test items for all examinees

Distractor		item number																			
Combination		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
		*	*						*	*	*				*	*	*	*	*	*	*
paraphrase, text	a ₁ b ₁	72	31	18	21	8	16	8	77	41	31	6	9	3	9	31	75	78	20	59	5
paraphrase, other source	a ₁ b ₂	11	63	3	11	10	6	5	5	24	5	23	2	2	12	39	5	12	32	11	4
exact words, text	a ₂ b ₁	8	4	58	62	74	74	85	15	15	38	62	85	85	73	21	4	3	25	24	88
exact words, other source	a ₂ b ₂	9	1	19	5	6	2	1	1	17	23	8	3	9	5	8	15	5	21	4	1
(not the order appearing in test)																					
		21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
paraphrase, text	a ₁ b ₁	22	1	2	57	44	28	49	2	81	6	4	22	33	76	51	58	82	16	57	19
paraphrase, other source	a ₁ b ₂	17	1	3	5	7	8	39	7	1	3	17	9	9	9	10	21	3	14	9	10
exact words, text	a ₂ b ₁	* 58	* 96	* 94	36	9	26	3	86	16	86	68	54	54	5	24	1	9	65	16	59
exact words, other source	a ₂ b ₂	2	1	0	1	38	36	6	3	1	2	10	15	4	10	14	18	4	3	16	9

*Correct answer for the item

(Some columns do not total 100 since whole numbers only are given)

the text are much more easily distinguished from any other distractor.

- b. Does the source of information facet (whether information came from the text or some other source) have an effect upon examinee response?

Test responses from ten students scoring in the lowest 40 per cent of scores on the facet test were analyzed. Only the incorrect responses were chosen for analysis. The correct distractors for any item would always have information from the text as their source of information facet but the incorrect distractors would have information from another source as the source of information facet two-thirds of the time.

The answer for each distractor statement had a range of agreement, contradiction and no information¹. The range for each distractor, the facet map for each distractor and the distractor choice for each item are reported in Table 4.8. By matching student responses to the facet combinations on each item it was possible to determine if a student was constantly selecting a particular source of information. The error responses of the ten students are categorized in Table 4.9.

Every student's responses are different, with few apparent patterns developing. It appears that the better achieving students from among this group (students falling at the 40th percentile or below on the facet test) choose distractors where the source of information was related to

¹ See Chapter III

the passage theme. Better students can probably detect information that has come from another source.

An analysis of responses from grouped data and from individual responses does appear to provide additional information about the facet. There is an indication that paraphrased information is more difficult to comprehend than exact words and also that information not closely associated with the passage theme is more easily detected in the item distractors.

Summary

The results of the data analysis for this study were presented in this chapter. It was concluded that:

1. the facet test was a reliable measure of reading comprehension;
2. the facet test discriminated among achievement levels; and
3. there was a significant relationship with other measures of reading comprehension, i.e. the cloze test and the Davis Reading Test.

A further analysis of responses from group data and for selected individuals suggests that paraphrased information is more difficult to comprehend than exact words and that information not closely associated with the passage theme is more easily detected in the item distractors.

Table 4.8 Range of mapping sentence matched to distractor statement

Facet map		item number																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
a ₁	b ₁	i*ii**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
		Ab	Aa	Ca	Ca	Ca	Ca	Cb	Ad	Ab	Ad	Cc	Cc	Ca	Ca	Ca	Ac	Ad	Aa	Ac	Ac
a ₁	b ₂	Rc	Rb	Nb	Rb	Nb	Rc	Nc	Na	Nc	Nb	Ra	Rb	Rc	Nc	Ra	Na	Rb	Ra	Ra	Nc
a ₂	b ₁	Cd	Cc	Ac	Ac	Ad	Ab	Aa	Cb	Ca	Cc	Ab	Aa	Ab	Ad	Cb	Cc	Cc	Cb	Cb	Ab
a ₂	b ₂	Na	Nd	Rd	Nd	Rc	Nd	Rd	Rc	Rd	Ra	Nd	Nd	Nd	Rb	Nd	Rb	Nd	Nd	Nd	Rd
a ₁	b ₁	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
		i*ii**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
a ₁	b ₁	Ca	Cb	Cc	Ac	Ab	Ac	Aa	Ca	Ad	Ca	Ca	Cc	Ca	Aa	Ab	Aa	Ad	Ca	Aa	Cc
a ₁	b ₂	Rb	Rc	Na	Rd	Na	Ra	Rb	Rb	Ra	Rc	Nb	Na	Nb	Ra	Na	Rb	Nb	Nb	Nb	Rb
a ₂	b ₁	Ad	Aa	Ab	Ca	Cc	Cb	Cc	Ad	Cb	Ab	Ac	Ab	Ad	Cb	Cc	Cc	Ca	Ac	Cc	Ad
a ₂	b ₂	Nc	Nd	Rd	Nb	Rd	Nd	Nd	Nc	Nc	Nd	Rd	Rd	Rc	Nd	Rd	Nd	Rd	Rd	Rd	Na

*i - range, A - agrees, C - contradicts, R - information related to passage theme,

N - information not related to passage theme

**ii - distractor designation on the test

Table 4.9 Tabulation of error response for students in lower 40 per cent of facet test

Student	Error response on facet items		
	C ^a	R ^b	N ^c
1	12	8	8
2	9	10	4
3	13	5	8
4	6	9	7
5	9	6	7
6	8	7	6
7	11	7	2
8	8	8	2
9	11	5	0
10	7	6	2

- a. the answer is a contradiction of information stated in the text
- b. the information came from another source related in theme to the passage content
- c. the information came from another source with no relationship in theme to the passage content

CHAPTER V

SUMMARY AND DISCUSSION

Summary of study

The purpose of this study was to measure reading comprehension by means of a test prepared from a facet design. There were four major questions from which three hypotheses were formulated and the fourth question was considered to the extent that information was available.

The questions are stated as follows:

1. Will a test constructed by facet design be a statistically reliable measure of reading comprehension?
2. Will a test constructed by facet design discriminate among achievement levels?
3. Does a test constructed by facet design show significant relationship to other measures of reading comprehension?
4. Will analysis of responses to the items provide additional information about the examinee?
 - (a) Does the form of information facet (whether information is stated in exact words or in paraphrase) have any effect upon examinee response?
 - (b) Does the source of information facet (whether information comes from the text or some other source) have any effect upon examinee response?

A search of the literature revealed a lack of agreement as to the definition of reading. Many authorities had written about it and much research had been conducted

but there was still no consensus on the definition.

Literature pertaining to measurement in education contained controversy about testing. At question was the validity of tests as well as the type of measurement instrument to be adopted. More recently, authors were advocating a new approach to measurement that would assure a valid testing of content and instruction.

A facet design for test construction was advocated by Schlesinger and Weiser (1970) and adapted by the investigator for this study. Through this design it is possible to systematically construct items to measure reading comprehension. The facet test concentrates on the relation of the item to the information presented in the text, rather than on skills and abilities presumed to be involved in reading comprehension. The items in this test consisted of a stem, or non variable part, and four variable (distractor) portions made up of two parts. Each part of the distractor stands in the relation of agreement, contradiction, or no information to the passage.

The tests were based upon four passages taken from information that students could have been exposed to or might have been required to read. The passages were approximately 600 words in length and varied in readability from seventh grade to college level. There were ten questions prepared from each passage, yielding forty items for the test.

The facet test was administered to 185 tenth grade

students from two high schools in Edmonton, Alberta, Canada, and the results were compared to results from two other tests of reading comprehension administered to the same sample.

A cloze test was constructed on the identical passages as those of the facet test. Every tenth word was deleted from the passage. Students were given the original passages to read first and then required to fill in the blanks with the exact word in the mutilated passage. The results were compared to the facet test scores.

The Davis Reading Test was administered to all students in the sample and the results compared to the facet test scores.

The Kuder-Richardson Formula 20 reliability coefficient was calculated for the facet test.

The scores of the facet test were analyzed by means of item analysis. A frequency distribution was prepared and an item difficulty index for each item was calculated as well as the bi-serial coefficient and item reliability index. The item analysis provided evidence as to the quality of the facet test based upon respondent's scores.

A Pearson product-moment correlation was computed between the scores of the facet test and cloze test and between the facet test and the Davis Reading Test to determine the degree of relationship. It was decided that additional insights might be gained through a non-statistical analysis.

An additional consideration was whether an analysis of responses would provide further information about the test. The item difficulty index was compared to the item reliability index for each item to determine if the form of information affected examinee response. Ten students were selected from the lower forty per cent of scores on the facet test. The error responses of these students were analyzed in terms of the distractor combination chosen to determine if the source of information affected the examinee's response.

The analysis of data led to the following conclusions:

1. A Kuder-Richardson Formula 20 reliability coefficient disclosed the facet test to be a statistically reliable instrument for measuring reading comprehension.
2. The facet test did discriminate among achievement levels of examinees.
3. There was a significant relationship between other measures of reading comprehension. The Pearson product-moment correlation coefficient indicated a significant relationship between the facet test and cloze test and between the facet test and the Davis Reading Test.
4. The test data suggested that paraphrased information was more difficult to comprehend than explicitly stated information and information from another source not closely related to the passage theme was more easily detected in item distractors than information closely

related in theme to the text.

Discussion

The reliability of a test can be considered in two ways. The first way is to think of it as consistently measuring whatever it measures. This was computed by the Kuder-Richardson Formula 20 and reached an acceptable level for this research. Another concept of reliability is that related by Thorndike (1971), indicating that a test must reliably measure the content upon which it is based. This was built into the test by means of operations which systematically put passage content into the item distractors. This ensures that the test items are related to the text. The close association between item content and passage content was accomplished by means of an operational approach called facet design.

One facet of reading comprehension applied in this test was the form of information. The information in a distractor was either exact words (explicit statement) or a paraphrase (implicit statement) of information. Two distractors in every item had to contain information from the text stated in one of these ways. One of these distractors was the correct answer. The other two distractors also had either a paraphrase or exact words, though the information was taken from another source.

It was suggested from the data that paraphrased information might have caused distractors to be more diffi-

cult than when information was in exact words. Carroll (1927) provided information that was in agreement with this finding as he concluded from a study dealing with detailed directions that material explicitly stated caused less errors than material implicitly stated. However, the form of information was only one facet considered in this test.

The second facet was source of information. The information could have come from the text or some other source. Only when information came from the text was the answer correct. Information from another source was by design either closely related to passage theme or unrelated to passage theme. The data suggested that information closely related to passage theme rendered the distractor more acceptable to examinees.

The pilot study provided worthwhile information concerning the attractiveness of items. It was discovered that if information from another source were used in an item distractor it had to be closely associated with the ideas in the passage. One could paraphrase information from another source only if it seemed to bear a close relationship to the stem of the item or be closely associated to something said somewhere in the passage. The researcher found that an idea from part of the passage, not related to the item stem, might be attractive to the examinee even if the exact information about the idea came from another source. It was also discovered that if the facet combination was exact words from another source (a_2 b_2) there had to be a close relationship

of theme between the information in the other source and information in the test passage. Item compilation was an onerous task using the two facets of this study.

The item distractors, in order to be attractive to the examinee, needed to be plausible answers. The only way that an $a_2 b_1$ combination (exact words from the text) could be an incorrect answer was to select information unrelated to the stem of the item. This proved to be very difficult to accomplish and is probably reflected in the data, as this combination was chosen a high percentage of the time. The $a_1 b_2$ combination (paraphrase from another source) was always an incorrect answer since information did not come from the test passage. This distractor combination, if related to the passage theme and the stem of the distractor, could only be false if it were made so on purpose or if it was lifted from its original context so that it lost original meaning. To make such an item too attractive would defeat the purpose of the test.

The distractors had to be constructed in such a manner that there was only one correct answer. If a distractor combination of exact words from another source were true and almost identical to information found in the passage, students had no possible method by which to discriminate it from the correct distractor statement. This problem could have led to some distractors discriminating negatively between higher and lower achievers.

The facet test, as a whole, did discriminate among achievement levels but some items did not discriminate well between the higher and lower achievers. None of the items achieved a high discrimination index as described by Ebel (1965). Some of the items with a discrimination index less than 0.20 may need revision. It must be noted, however, that these items are valid in terms of test content and to dispense with the item is to reject the whole item class. Osburn (1968, p. 102) says that "any decision to exclude items based upon item analysis data must result in a redefinition of the universe of content." The item distractors, when reviewed, need to be carefully analyzed and rewritten in conjunction with the operational rules of facet design. The generalization as to the validity of the test cannot come through item analysis but only through an operational definition of procedures used in generating the items that go into the test.

The facet test was compared to the Davis Reading Test which was validated mainly on a statistical basis and the cloze test which claims operational validity. The Pearson product-moment correlation coefficient showed a significant relationship among these instruments. This indicates that the tests are measuring similar aspects of reading comprehension.

The cloze test was constructed from the same passages as the facet test and was administered to the same sample of students. The correlation coefficient, while

significant, was lower than expected since both tests are operationally defined. The cloze test proved to be more lengthy than expected, resulting in only one third of the students completing it. It was thus a speeded test and the correlation of a speeded measure with a power test where all students finished introduced an unknown error factor. This was a limitation of the study¹.

Further research

Further research is necessary to judge the worth of facet design. Some possible areas are delineated below.

1. More facets could be employed in the design. Schlesinger and Weiser (1970) suggest four facets to be used but there are more which can be applied. The generating of items, however, may become more difficult with increased facets. Use of a computer could prove helpful.
2. The scoring of responses could be changed from counting only the correct distractor to letting the student decide whether the distractors: (1) have no information related to the text, (2) contradict information stated in the text, or (3) whether the distractor agrees with information stated in the text.
3. One could determine if the facet design can be used to diagnose reading comprehension difficulties.
4. Individual respondents might be questioned on their incorrect responses to determine at what point the student

¹ See Chapter I

made his error.

5. A readability formula could be applied to the distractor statements to determine if the readability level is similar to the text information when the author is allowed to paraphrase information.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Auerbach, I. An Analysis of Reading Comprehension Tests. Harvard College, Graduate School of Education, Cambridge, Mass., Unpublished Thesis, 1971.
- Austin, M. C. and Morrison, C. The First R. New York: The Macmillan Company, 1963.
- Bean, K. L. Construction of Educational and Personal Tests. New York: McGraw-Hill Book Company, Inc., 1953.
- Berry, B. T. "Improving Freshman Reading Habits," English Journal, 20 (1931), pp. 824-28.
- Bickley, A. C., Ellington, B. J. and Bickley, R. J. "The Cloze Procedure: A Conspectus," Journal of Reading Behavior, II, 3 (Summer, 1970), pp. 232-49.
- Bormuth, J. R. On the Theory of Achievement Test Items. Chicago: The University of Chicago Press, 1970.
- Buros, O. K. (ed.) Reading Tests and Reviews. Highland Park, New Jersey: The Gryphon Press, 1968.
- Buros, O. K. (ed.) Seventh Mental Measurements Yearbook. Highland Park, New Jersey: The Gryphon Press, 1970.
- Carroll, R. P. An Experimental Study of Comprehension in Reading. New York: Bureau of Publications, Teachers College, Columbia University, 1927.
- Culhane, J. W. "Cloze Procedures and Comprehension," The Reading Teacher, 23, 5 (February, 1970), pp. 410-13.
- Davis, F. B. Fundamental Factors of Comprehension in Reading. Unpublished doctor's dissertation. Harvard University, 1941.
- Davis, F. B. and Davis, C. C. Davis Reading Test (Manual), New York: The Psychological Corporation, 1962.
- Davis, F. B. Educational Measurements and Their Interpretation. Belmont, California: Wadsworth Publishing Co., 1964.

- Davis, F. B. Identification and Measurement of Reading Skills of High-School Students. Co-operative Research Project No. 3023, U.S. Department of Health, Education, and Welfare, 1967.
- Davis, F. B. "Research in Comprehension in Reading," Reading Research Quarterly, III, 4 (Summer, 1968), pp. 499-545.
- Ebel, R. L. Measuring Educational Achievement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1965.
- Ebel, R. (ed). Encyclopedia of Educational Research. (4th ed.), American Educational Research Association, The Macmillan Company, 1969.
- English, H. B. and English, A. C. A Comprehension Dictionary of Psychological and Psychoanalytical Terms. New York: Longmans, Green and Co., 1958.
- Fry, E. B. "A Readability Formula That Saves Time," Journal of Reading, 11 (April, 1968), pp. 513-16.
- Fry, E. B. "The Readability Graph Validated at Primary Levels," The Reading Teacher, 22 (March, 1969), pp. 534-38.
- Gray, W. S. "Reading." In C. W. Harris (ed.). Encyclopedia of Educational Research. (4th ed.). New York: Macmillan, 1960, pp. 1086-1135.
- Guttman, L. "A Faceted Definition of Intelligence," Scripta Academica Hierosolymitana, 14 (1965), pp. 166-81.
- Guttman, L. and Schlesinger, I. M. Development of Diagnostic Analytical and Mechanical Ability Test Through Facet Design and Analysis. (Report of Project NO. OE - IS - I - 64). U.S. Department of Health, Education, and Welfare, 1966.
- Guttman, L. and Schlesinger, I. M. "Systematic Construction of Distractors for Ability and Achievement Test Items," Educational and Psychological Measurement, 27 (1967), pp. 569-80.
- Harris, A. J. How to Increase Reading Ability (5th Ed.). New York: David McKay Company, 1970.
- Heilman, A. W. Principles and Practices of Teaching Reading (2nd Ed.). Columbus, Ohio: Charles E. Merrill Books, Inc., 1967.

- Hoffman, B. The Tyranny of Testing. London: Collier - Macmillan, Ltd., 1962.
- Holland, J. G. "A Quantitative Measure for Programmed Instruction," American Educational Research Journal, 4 (1967), pp. 87-102.
- Hunt, J. T. "Selecting a High School Reading Test," High School Journal, 39 (October, 1955), pp. 49-52.
- Jenkinson, M. D. "Basic Elements of Reading Comprehension," Proceedings of the Second World Congress on Reading, Copenhagen, Denmark, (August), 1968, pp. 41-47.
- Johnson, M. S. and Kress, R. A. "Task Analysis for Criterion-Referenced Tests," The Reading Teacher, 24, 4 (January, 1971), pp. 355-59.
- Jordan, J. Attitudes Toward Education and Physically Disabled Persons in Eleven Nations, East Lansing, Michigan: Michigan State University, 1968.
- Langer, J. H. "Vocabulary and Concepts: Essentials in the Reading - Thinking Process," Elementary School Journal, 69 (April, 1969), pp. 381-85.
- Lennon, Roger T. "What Can Be Measured." The Role of Tests in Reading, Proceedings of Annual Education Conferences (Russell Stauffer, Ed.). Neward: University of Delaware, 9 (March, 1960), pp. 67-80.
- Lindeman, R. H. Educational Measurement. Glenview, Ill.: Scott, Foresman and Company, 1967.
- Lindquist, E. F. (ed.). Educational Measurement, American Council on Education, Washington, D. C. Mensha, Wisconsin: George Barton Publishing Company, 1951.
- Maginnis, G. H. "The Readability Graph and Informal Reading Inventories," The Reading Teacher, 22 (March, 1968), pp. 516-18.
- Osburn, H. G. "Item Sampling for Achievement Testing," Educational and Psychological Measurement, 28, (1968), pp. 95-104.
- Pauk, W. "Another Practical Note on Readability Formulas," Journal of the Reading Specialist, IX (1970), pp. 141-43.

- Prescott, G. A. "Criterion - Referenced Test Interpretation in Reading," The Reading Teacher, 24, 4 (January, 1971), pp. 347-54.
- Rankin, E. F. "The Cloze Procedure - Its Validity and Utility," The Eighth Yearbook of the National Reading Conference, Milwaukee, Wisconsin: The National Reading Conference, Inc., 1959.
- Rankin, E. F. "The Cloze Procedure - A Survey of Research," Fourteenth Yearbook of the National Reading Conference, 1965, pp. 133-50.
- Schlesinger, I. M. and Guttman, L. "Smallest Space Analysis of Intelligence and Achievement Tests," Psychological Bulletin, 71 (1969), pp. 95-100.
- Schlesinger, I. M. and Weiser, Z. "A Facet Design for Tests of Reading Comprehension," Reading Research Quarterly, V, 4 (Summer, 1970), pp. 566-80.
- Schubert, D. G. and Torgerson, T. L. (ed's.). Readings in Reading: Practice Theory Research, New York: Thomas Y. Crowell Company, 1968.
- Simons, H. D. "Reading Comprehension: the need for a new perspective," Reading Research Quarterly, VI, 3(Spring, 1971), pp. 338-63.
- Taylor, W. L. "Cloze Procedure: A New Tool for Measuring Readability," Journalism Quarterly, 30, 1953.
- Taylor, W. L. "Recent Developments in the Use of Cloze Procedure," Journalism Quarterly, 33, 1956.
- Thorndike, E. L. "The Understanding of Sentences," Elementary School Journal, 18 (1917), pp. 98-114.
- Thorndike, R. L. and Hagen, E. Measurement and Evaluation in Psychology and Education, (3rd ed.). New York: John Wiley & Sons, Inc., 1969.
- Thorndike, R. L. Educational Measurement (2nd ed.). American Council on Education, Washington, D.C., 1971.
- Traxler, A. E. "Critical Survey of Tests for Identifying Difficulties in Interpreting What is Read," Promoting Growth Toward Maturity in Interpretating What is Read, Supplementary Educational Monograph, No. 74, University of Chicago Press, Chicago, 1951, pp. 195-200.

- Vernon, P. E. "The Determinants of Reading Comprehension,"
Educational and Psychological Measurement, 22
(1962), pp. 269-86.
- Weaver, W. W. "Theoretical Aspects of the Cloze Procedure,"
Fourteenth Yearbook of the National Reading
Conference, 1965, pp. 115-32.
- Zintz, M. V. The Reading Process: The Teacher and the
Learner, Dubuque, Iowa: Wm. C. Brown Company, 1970.

APPENDIX A
PASSAGES FOR THE FACET TESTS
AND CLOZE TESTS

The Province of Alberta has six distinct types of municipal corporations - four urban and two rural. Cities, towns, villages, and urban counties are urban municipalities; municipal districts and rural counties are those found in rural areas.

As you will be aware, Alberta has a number of urban centres such as Edmonton, Calgary, Lethbridge, and Medicine Hat, whose populations have grown sufficiently to enable them to become cities. They are the largest urban municipal corporations in the province.

We are going to outline the city in detail as the basis for understanding municipal corporations.

Prior to 1951, cities were given separate charters by the provincial government. Calgary received its charter in 1893, Edmonton in 1904, Medicine Hat and Lethbridge in 1906. Each charter included such details as the name of the city and the powers its council would exercise.

The practice of issuing individual charters to cities created a number of problems. If a city's charter was lost temporarily, there was difficulty in deciding what laws or statutes governed its operation. Moreover, the separate charter system discouraged closed relations and co-operation between officials of one city and another. Realizing the weaknesses of the existing charter system, the provincial legislature passed the City Act which came into effect in January, 1952.

This statute, by ensuring that the charter of cities would be basically uniform, laid the groundwork for improved local government in large urban centres. Cities, as well as other municipal corporations, are, of course, under the control of the Department of Municipal Affairs, a department of the provincial government.

How does an urban area become a city? The answer to this question may be found in the City Act. To acquire that status a town must have a population of over six thousand people and its council must apply to the Minister of Municipal Affairs for legal recognition as a city. If the Minister so recommends, the Lieutenant Governor in Council (the cabinet) may, by proclamation, grant the town the status of city. This proclamation will, among other particulars, state the name of the new city, its area, and the date on which it is officially to become a city.

The City Act of 1952 did much to establish on a uniform basis the system of government for large urban centres. Executive and legislative powers were granted to councils composed of a mayor and aldermen. The mayor, elected for a two-year term by all the voters, is assisted by a deputy mayor who is selected by the council from among its members. If the mayor is absent from meetings, the deputy mayor assumes his duties. If neither of these municipal officers is present at a council meeting, the councillors present may appoint an acting mayor to preside over deliberations.

In addition to the elected mayor, a city council includes aldermen. By law a council must not have fewer than six nor more than twenty aldermen, and the number of aldermen must be an even one. One-half of the aldermen are elected each year for a term of two years by all the voters.

Alberta's largest cities, Edmonton and Calgary, show a few variations from the general practice in electing their councils. In Edmonton, the mayor and twelve aldermen are elected every two years on a city-wide basis. In Calgary, six wards have been established, with two aldermen elected to represent each ward. One-half of the aldermen retire each year. The mayor, however, is chosen by all the voters of the city for the usual two-year term.

Almost from the moment his bones were first discovered in Germany's Neander Valley a century ago, his name has been synonymous with brutishness: a squat, shambling creature who wooed his women with a club and sometimes ate his fellow men when he was hungry. Scientists have long doubted this harsh popular image of *Homo neanderthalensis*, or Neanderthal man. Now, as the evidence accumulates, Neanderthal man is rapidly being rehabilitated into a more attractive ancestor of modern man.

From remains found in Europe, archaeologists have already concluded that Neanderthals were skilled hunters and toolmakers, held formal burial rites that indicated a belief in an afterlife, and even practiced a primitive form of Social Security for their aged and infirm. More recently, paleontological examination of skeletons has suggested that Neanderthal man's stooped appearance may have been the result of disease rather than low evolutionary status. According to this theory, he was plagued by a dietary deficiency of vitamin D. This deficiency was aggravated by the diminished sunlight of the ice age, and eventually caused rickets. Now, the most detailed and sympathetic picture yet of Neanderthal man comes from extensive diggings by an American-led expedition in a mountain cave near the village of Shandar in Iraqi Kurdistan.

In an article in the current *Smithsonian* magazine, and in a forthcoming book, *Shanidar: The First Flower People* (Knopf: \$8.95), the expedition's chief archaeologist, Dr. Ralph S. Solecki, reports that at least one of the nine Neanderthal skeletons uncovered in the Shanidar cave was buried with flowers. Another skeleton was that of a man about 40 (equivalent to an age of 80 by modern life-spans) who had been born with a withered right arm. The limb had apparently been amputated above the

elbow by a Neanderthal "surgeon." The man's age and physical condition indicated to the scientists that he had been unable to fend for himself. They surmised that his fellows kept him alive until he met his death in an accidental rockfall inside the cave, a common peril for these communal hunters who lived from 100,000 to 40,000 years ago.

Comments Anthropologist Carleton S. Coon: "On the grounds of behavior alone, the Shanidar folk merit the title of *Homo sapiens*."

Chimp or Philosopher. Neanderthals conducted other elaborate rites besides funerals. Clues to one of these were uncovered in Lebanon last summer when an expedition led by Solecki, who is a professor of anthropology at Columbia University, found the dismembered skeleton of a small deer in a cave overlooking the Mediterranean. The 50,000 year-old bones had apparently been arranged in an orderly way and sprinkled with red ocher, a substance used for symbolic purposes by Neanderthal man. Reporting on the discovery last week, Solecki said: "These men were trying to ensure a successful hunt by the ceremonial treatment of one of the animals." In other words, Neanderthal man resorted to a form of hunter's magic.

How did so sophisticated a creature acquire such an unwarranted reputation? For one thing, the first Neanderthal bones were dug just about the time that Darwin astonished the world with his announcement that man and ape were descended from a common ancestor. Neanderthal's apish image was further enforced by the writings early in this century of the respected French paleontologist Pierre Marcellin Boule. His portrait of Neanderthal as a stunted, beetle-browed creature who walked with bent knees and arms

dangling in front of him served as the model for several generations of artists and cartoonists. While certain coarse features in Neanderthal man are undeniable, on physical considerations alone he deserves far better treatment. As the late Harvard anthropologist Earnest Hooton once commented: "You can, with equal facility, model on a Neanderthaloid skull the features of a chimpanzee or the lineaments of a philosopher.



Cold and snow are twins in the boreal forest. For many sorts of mammals, as well as for grouse, snow is a protection against a plummeting temperature. The short-legged, stout-bodied, short-tailed voles, for instance, spend almost the entire winter in the air space between the soil and the snow. So do the shrews and a few mice.

We know very little as yet about this world between the ground and the snow, except that its temperature seldom gets very much below freezing, and that the light is extremely dim. We can guess that its relative humidity is high and that the sounds of the world above are as muffled as the light. The group of voles living in this sort of quarter-world includes the meadow and red-backed voles which are also found in the rest of the province, the phenacomys vole and the bog lemming. Occasionally the voles build runways through the snow but usually, because of the air space, there is no need to do this. However, they often drive ventilation tunnels to the surface of the snow cover. They also bore through the snow itself from time to time, apparently in search of food particles fallen from the trees or trapped in the snow-cover. Their habitat has holes, like those in Swiss cheese, wherever there is gaminiq with snow too thin to give insulation.

While the voles use snow for survival, beavers and muskrats avoid the extreme cold of winter by living under the ice of ponds and rivers. This is the warmest place for them, since in water the temperature cannot fall below freezing. Both species move about as they do in summer, seldom coming onto the surface of the land unless their ponds are frozen to the bottom. Sealed off by snow and ice and protected by the frozen walls of their houses, they are relatively safe. But every Eden, even a frozen

one, has a handicap. Winter does not prevent mink and otters from invading their habitat.

Snow can hinder as well as help. In Alberta everyone knows from experience how difficult it is to struggle through knee-deep snow. The animals of the boreal forest have solved this difficulty in various ways.

One group of mammals simply wades through the snow as we do. The best adapted animal in this group is the moose. It moves easily through two feet of snow because of its long legs and a special arrangement of muscles and bones that permits the legs to work almost straight up and down. The toes of the moose seldom leave drag marks in the snow so that it is usually difficult to tell from its tracks in which direction it was moving. The wapiti and smaller deer run into trouble at somewhat lesser depths. The predator of the moose, the wolf, also wades through the snow and, as long as it is soft and fluffy, the wolf is at a considerable disadvantage. But if a crust forms, the wolf may be able to stay on the surface while the moose keeps on wading, but with difficulty. Hence, the depth and structure of the snow mean life or death to both the wolf and the moose. Animals living in herds reduce the energy needed for ploughing through snow by playing follow-the-leader. This is the technique used by the bison of Wood Buffalo National Park, by caribou, and in deep snow by deer, when they tramp down an area to form what is known as a yard.

The relationships that exist among the plants and animals that have been described are almost unbelievably complex. Each species has its food requirements, its space requirements and its requirements of temperature, light and water. It is common to refer to "competition" among organisms for these requirements. Organisms which are at any disadvantage in the competition, for any reason, decline in numbers and may become locally or universally extinct. Of course conditions may change (for instance, moist years may follow drought) and the trend to decline may be reversed. Because natural conditions are constantly changing, the numbers of living things fluctuate. As examples, for a few years there may be many hares, then, for some years, hares may be scarce. Or, around a slough willows may predominate for a period of time, to be drowned when the water level rises, and replaced by cattails. In some years sparrow hawks are abundant, in some grasshoppers thrive. In some years Canada thistle is a greater nuisance than in others. We are not always sure what causes the fluctuations.

The relationships between organisms are often called together "The Balance of Nature". This phrase means that living things occur as best they can in the face of the competition of all other living things and the changes of physical conditions (weather, soil conditions, etc.). The Balance of Nature is not something that is the same year in and year out. The balance changes all the time because new and different influences are always affecting one or another member of the natural community. This, in time, affects neighbouring members of the community so that living conditions might be bettered or worsened for different members of the community. The balance will have shifted.

Let us examine an imaginary case, to see how a balance might shift over a period of time. We shall look at a grassy area at the margin of the aspen parkland zone. The grasses there might be chiefly a rough fescue, porcupine grass and June grass. The grasses generally prevent the establishment of aspen: under the proper conditions, grasses are more successful competitors for soil than aspen. But the situation can be reversed if bare areas of soil develop. Such bare areas can be formed by burrowing Richardson's ground squirrels, badgers, pocket gophers or other animals, or by overgrazing and trampling by cattle, or in days past by bison. Bare soil areas give shrubs, such as snowberry, a chance to become established. Perhaps sharp-tailed grouse, using the bare area as a dusting ground, left seeds there in droppings. Other seed-eating birds or wind dispersal can also seed bare areas. If snowberry and wolf willow become established, their shade inhibits grass growth. Coyotes tend to den in shrubby sites further disturbing the soil and preventing the growth of grass. Some grasshoppers, preferring grass, become scarce, as do some ants. Different species of ants and grasshoppers are found which prefer shrubby habitat. Aspen seedlings appear. As aspen develops, it begins to shade and restrict snowberry and wolf willow. Shrubs come to form a fringe around the edge of an enlarging grove of aspen. Songbirds are attracted to nesting sites. Flickers and crows inhabit the woods. Snowshoe hare and Franklin's ground squirrels prefer the woods. So does the great horned owl.

As the aspen ages, spruce seedlings may begin to appear. Spruce may be held back if a lot of hares are present to feed on the tender young

plants. Fire will stop the development of spruce, while after a fire aspen can regenerate by means of suckers. But these aspen suckers can be held down by the foraging activities of rabbits or other browsers. In such cases, grass will reappear, once again to be the main vegetation, with its own distinctive population of animals.

APPENDIX B
FACET TEST DIRECTIONS
AND QUESTIONS



Research Project

NAME _____

TYPE OF SCHOOL PROGRAM (e.g., Matriculation, Diploma, etc.) _____

DIRECTIONS

This package of materials contains four passages for you to read and a set of 10 multiple-choice questions about each passage.

1. Read the first passage through thoroughly. Turn to the questions that follow and answer each one by circling the correct answer.

Do not look back at the passage you have just read when answering the questions. Answer all of the questions even though you may have to guess.

2. Choose the best possible answer to the questions based only upon information contained in the related passage.
3. When you have read the first passage and answered the questions, go on to the next passage and questions. Continue through the passages and questions in the order assembled until you are finished.
4. Do not go back to check your answers.



Questions

1. In order for an urban area to acquire city status
 - a. a resolution containing this idea is passed by the cabinet and initialled by the Prime Minister.
 - b. the Lieutenant Governor in Council may grant it upon the recommendation of the Minister of Municipal Affairs.
 - c. it must have thousands of people, much housing, stores, schools, offices, theaters and parks.
 - d. the mayor and twelve aldermen are elected every two years on a city-wide basis.
2. Choose the central theme of the passage.
 - a. The city in Alberta.
 - b. The laws that group communities into classes in relation to their population.
 - c. Alberta's largest cities, Edmonton and Calgary.
 - d. Certain rural areas of the province are more thickly populated than others.
3. The city act
 - a. allowed for separate charters to be granted by the provincial legislature.
 - b. was passed to solve the differences between municipal bodies and school boards.
 - c. came into effect in January 1952.
 - d. outlines the regulations concerning elections in cities.



4. The most recent city act

- a. granted separate charters to every city.
- b. gave the mayor the right to be an ex officio member of every board, commission, or organization where the council has jurisdiction to appoint members.
- c. did much to establish on a uniform basis the system of government for large urban centers.
- d. allowed that at no time can any individual or group in government make a "blanket" claim-to-fame for having achieved the perfect solution.

5. The proclamation granting city status would

- a. provide six types of municipal corporations.
- b. recognize, by law, the rights and privileges of citizens.
- c. provide police and fire protection, street and traffic control, health and welfare services and other services to maintain community life.
- d. state the name of the new city, its area, and the date on which it is officially to become a city.

6. According to law, the city council

- a. issues separate charters that provide details of the city's name and powers of its council.
- b. must not have fewer than six nor more than twenty aldermen.
- c. must be Canadian citizens or British subjects.
- d. are pleased when people in the community take a genuine interest in the business of the municipality.

7. In Edmonton

- a. the mayor and twelve aldermen are elected every two years on a city-wide basis.
- b. two aldermen are elected from each ward.
- c. one of the powers of government is to create and establish the municipal level of government.
- d. aldermen receive a regular yearly allowance.



8. The deputy mayor

- a. has a threefold function to perform which is legislative, executive and judicial.
- b. is chosen by all the voters of the city for the usual two-year term.
- c. may receive pay for his work in this capacity.
- d. assists the mayor and is selected by the council from elected aldermen.

9. The city act laid the groundwork for better local government in urban centers because

- a. cities were given separate charters by the provincial government.
- b. it made sure that the city charters were basically uniform.
- c. it established the eligibility of candidates and persons to vote as well as other matters related to municipal elections.
- d. it outlines clearly the mandatory and optional powers of the council.

10. One could describe the municipal corporations in Alberta as

- a. a body or unit whose inhabitants (permanent residents and ratepayers) have been united by law.
- b. being similar to the strong-mayor form of government with the mayor having broad powers.
- c. a number of urban centers such as Edmonton, Calgary, Lethbridge and Medicine Hat whose populations have grown sufficiently to enable them to become cities.
- d. being either urban or rural.

Questions

1. From the first discovery of Neanderthal's bones
 - a. his name has been associated with crudely primitive in manner or conduct.
 - b. his name has been synonymous with brutishness.
 - c. there has been a connection with Dr. Ralph S. Solecki, who led an expedition to Lebanon last summer.
 - d. Darwin's theory has withstood almost a century of criticism.
2. From remains found in Europe, archeologists have concluded that
 - a. Neanderthals were skilled hunters and toolmakers.
 - b. Neanderthals made excellent flint tools.
 - c. Earnest Hooton was correct in saying that you could, with equal ease, put the features of a chimpanzee on the Neanderthal skulls.
 - d. vitamin D is a group of about 10 fat-soluble vitamins that prevent rickets.
3. Recent studies suggest that Neanderthal Man's stooped appearance may have been
 - a. the result of sprinkling red ocher on animal bones.
 - b. the result of disease rather than low evolutionary status.
 - c. due to the fact that his brain was somewhat larger than modern man.
 - d. essential to determine the origin of inherited variations and the effects of the environment upon them.



4. Dr. Solecki reports that

- a. bones from a human skeleton were discovered over a century ago in Germany's Neander Valley.
- b. Neanderthal people probably lived between 30,000 and 60,000 years ago.
- c. rickets is a vitamin deficiency that prevents the bones from hardening properly thus resulting in deformations.
- d. at least one of the nine Neanderthal skeletons uncovered in the Shanidar cave was buried with flowers.

5. On a recent expedition to a mountain cave in Iraqi Kurdistan

- a. scientists discovered a skullcap.
- b. Darwin astonished the world with his announcement that man and ape were descended from a common ancestor.
- c. there were several scientists led by an American.
- d. it was found that making the stone axe required a number of simple skills.

6. The main idea expressed in this passage is that

- a. the ice age has been present during the majority of man's existence.
- b. Neanderthal, is the name of a race of prehistoric men who lived in caves in Europe, North Africa and western and central Asia.
- c. Neanderthal's apish image was further enforced by the writings early in this century of the respected French paleontologist Pierre Marcellin Boule.
- d. Neanderthal Man may not have been as primitive as originally speculated.

7. The Neanderthal villages probably had someone comparable to
 - a. a medical doctor.
 - b. men about five to five and one-half feet tall who did not walk erect because of curved thigh bones.
 - c. the respected French paleontologist Pierre Marcellin Boule.
 - d. shamans who speak with the voices of the dead.
8. Dr. Solecki, a professor of anthropology,
 - a. would be concerned about the origin of Shanidar folk.
 - b. astonished the world with his announcement that man and ape were descended from a common ancestor.
 - c. would be concerned about the culture of Shanidar folk.
 - d. views people as highly innovative creatures, organizing and reorganizing their field of experience at every moment.
9. Neanderthal Man got his original reputation concerning his life-style because
 - a. scientists discovered a skullcap in the Neander Gorge near Düsseldorf, Germany.
 - b. of extensive diggings by an American-led expedition in a mountain cave near the village of Shanidar in Iraqi Kurdistan.
 - c. the Neander Valley discovery was made near the time Darwin's theory was announced.
 - d. they produce, through cultivation, most of their own food, but produce also for a market.
10. Scientists have concluded that a Shanidar man of forty would be
 - a. a stooped, squat and shambling creature.
 - b. equivalent to an age of 80 by modern life-spans.
 - c. suffering from a vitamin deficiency that prevents the bones from hardening properly, thus resulting in deformations.
 - d. as erect as any modern man, he did not have a bull neck, and he was not knock-kneed.

Questions

1. The world between the ground and the snow
 - a. is protection for the grouse against cold temperatures.
 - b. is the home for many organisms who survive the harsh winters in Alberta.
 - c. has lichens and other low growing plants.
 - d. seldom gets very much below freezing.

2. Beavers avoid the extreme cold of winter by
 - a. living under the ice of ponds and rivers.
 - b. building runways through the snow.
 - c. storing an ample supply of willows, aspen poplar and birch.
 - d. becoming very fat and yielding to hibernation.

3. The moose is one of the best adapted animals for wading through snow because
 - a. it has rounded, wide-spread hoofs on the hind feet.
 - b. of special arrangement of muscles and bones that permits the legs to work almost straight up and down.
 - c. it seldom leaves drag marks.
 - d. it makes daily treks through the wintering area to keep the snow in pathways at a minimum level.

4. A yard is
 - a. the technique used by the bison of Wood Buffalo Park, by caribou and in deep snow by deer.
 - b. an area set aside for a particular business or activity.
 - c. an area where snow has been tramped down by deer or caribou.
 - d. a small enclosed area open to the sky.

5. You could describe the appearance of the vole which lives in the space between the snow and the ground as
- a. dainty and for the most part richly colored.
 - b. a plump animal with stubby legs.
 - c. twins in the boreal forest.
 - d. a relatively plump rodent with small ears.
6. "Their habitat has holes, like those in Swiss Cheese, wherever there is qaminig with snow too thin to give insulation!" The word qaminig
- a. is an Eskimo word referring to snow that reaches the ground.
 - b. means runways through the snow.
 - c. most likely means snow covering the ground.
 - d. is a hollow conduit or recess.
7. The wapiti is a member of the deer family
- a. smaller than the moose.
 - b. known as an elk.
 - c. seldom coming onto the surface of the land unless their ponds are frozen to the bottom.
 - d. with big, heavily-furred feet, which with long legs, create a rather bizarre appearance.
8. Deep, crusted snow hinders survival of the moose because
- a. even a frozen Eden has a handicap.
 - b. he is unable to forage sufficiently well.
 - c. snow crystals often cling together to form snow pellets over an inch thick.
 - d. the wolf may be able to stay on the surface while the moose keeps on wading.

9. Choose the best theme for the passage.

- a. The greater the variety of food the better is the chance of survival.
- b. Cold and snow are twins in the boreal forest.
- c. In some years, it snows only a little, but in others it snows a great deal.
- d. How snow helps and hinders survival.

10. Voles living in the space between ground and snow

- a. can reduce the energy needed for survival by playing follow-the-leader.
- b. often drive ventilation tunnels to the surface of the snow cover.
- c. have a dark, glossy-brown pelage on top and slatey grey below.
- d. of course can exist only in small numbers, since for support they require so many of the lesser creatures.

Questions

1. Which of the following would be most correct about plants and animals according to the author?
 - a. Shrubs can grow in bare areas.
 - b. The food and space requirements of each species varies with climatic conditions.
 - c. Each species has its own food requirements and its space requirements.
 - d. Predatory animals subsist on prey which in turn must capture other animals or eat plants.

2. "The Balance of Nature"
 - a. refers to the distribution of animals in Alberta.
 - b. is not something that is the same year in and year out.
 - c. means that for a few years hares may be scarce but then increase in number.
 - d. includes the change of simple chemical substances to living things and living things back into simple chemicals.

3. Under certain conditions
 - a. the relationships among all plants and animals is complex.
 - b. pocket gophers do not live in some parts of the extreme southeastern part of the province even though they inhabit surrounding areas.
 - c. willows planted on a riverbank are good protection against soil erosion.
 - d. grasses are more successful competitors for soil than aspen.

4. Which of the following could be considered the best theme of the passage?
 - a. The balance of nature is related to nutrition.
 - b. Grasses generally prevent the establishment of aspen.
 - c. The relationships among plants and animals are subject to many varied conditions.
 - d. All sorts of limiting factors can be found and often their results are unexpected.

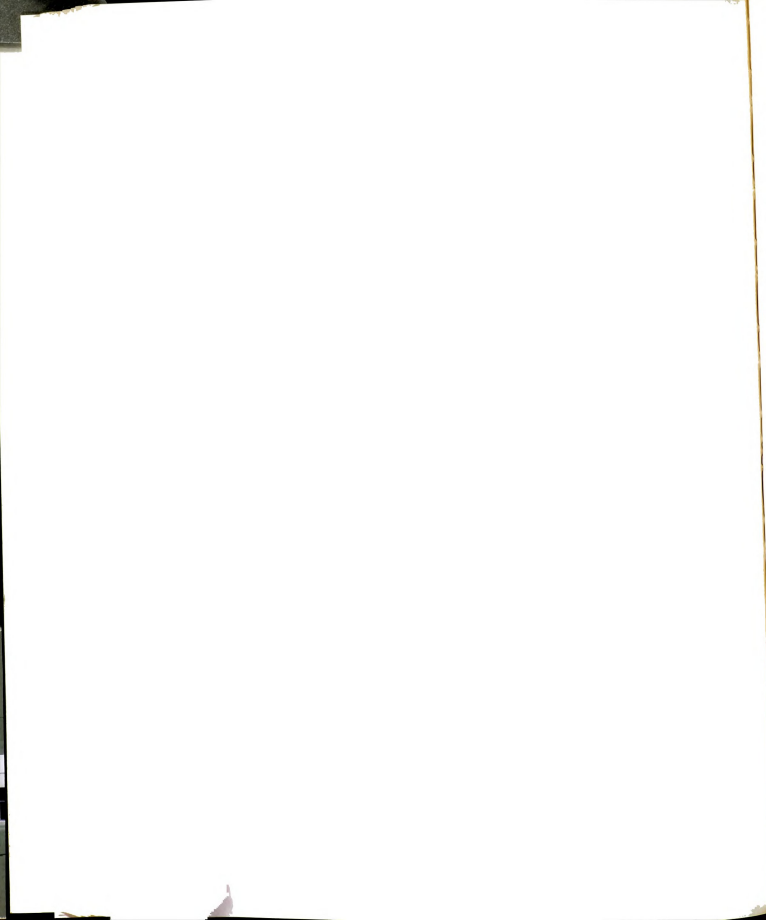
5. The author of this passage agrees that plants which do not normally grow in one area
 - a. are helpful in building up the soil with minerals and root fibres.
 - b. may flourish if competitive plants are removed.
 - c. are always affecting one or another member of the natural community.
 - d. were transported and placed there by a number of different agencies.

6. According to this author, every living situation
 - a. is undergoing change either slowly or rapidly.
 - b. could be affected by climate under certain conditions.
 - c. begins to shade and restrict snowberry and wolf willow.
 - d. is the essential first step that prepares the way for all the life that exists on earth.

7. Cattail is a plant that may grow
 - a. at the margin of the aspen parkland zone.
 - b. in the sandy areas where jack pine is common.
 - c. to a height of 5 or 6 feet, and have long, broad leaves.
 - d. where willows once predominated but were drowned out by high water.

8. According to this passage, coyotes
 - a. form bare areas which allow plants, other than grasses, to begin growing.
 - b. serve to illustrate the potential danger in introducing new species which might compete with and overcome native species.
 - c. tend to den in shrubby areas.
 - d. live in regions of western North America from Panama to Alaska.
9. Browsers, as referred to by the author, are
 - a. animals, such as the rabbit, that eat tender shoots and twigs.
 - b. predators that avoid man.
 - c. shrubs that come to form a fringe around the edge of an enlarging grove of aspen.
 - d. subject to the same basic principles of adaptation and balanced control that apply to all living communities.
10. Bare areas are necessary in the balance of nature because
 - a. small animals, especially insects, mites and other arthropods, dwell in the soil in very large numbers.
 - b. they provide a home for many living things such as bacteria, earthworms, etc., which add to the organic content.
 - c. conditions are constantly changing which results in fluctuations of living things.
 - d. these give shrubs, such as the snowberry, a chance to become established.

APPENDIX C
CLOZE TEST DIRECTIONS
AND QUESTIONS



NAME _____

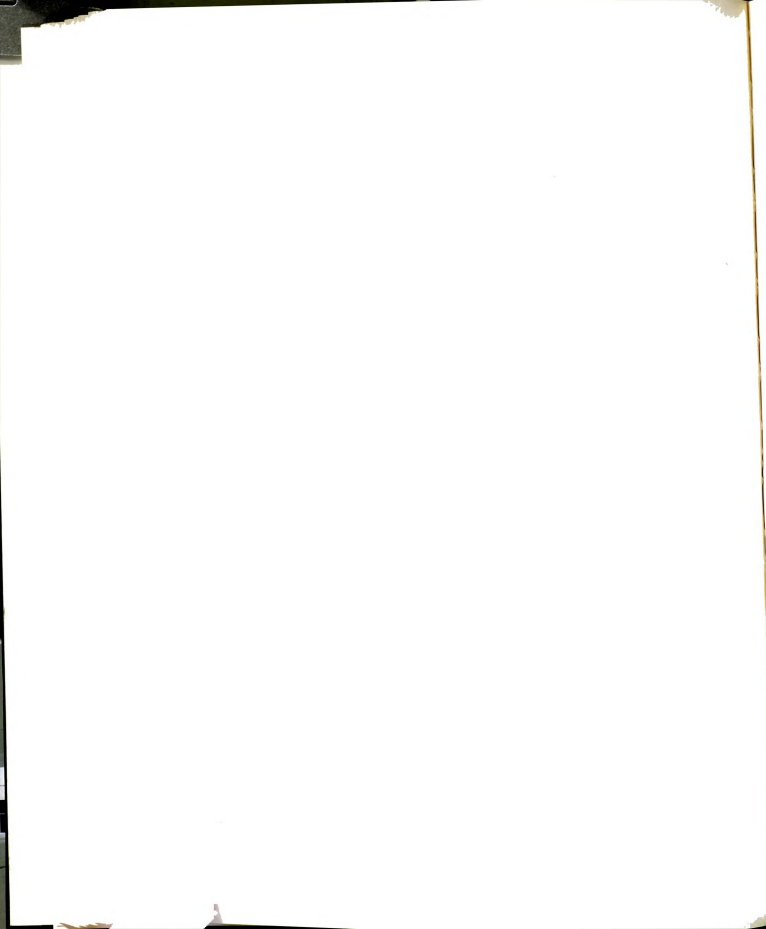
DIRECTIONS

This package of materials contains four passages for you to read and a fill-in-the-blanks test for each passage.

1. Read the first passage carefully, then proceed to the pages immediately following marked "Questions." The passage you have just read is rewritten with blanks at certain intervals. You are to fill in the blank with the exact word that was in the original passage.

DO NOT LOOK BACK AT THE ORIGINAL PASSAGE WHEN FILLING
IN THE BLANKS.

2. Try to complete every blank even if you must guess at the answer.
3. When you have read the first passage and completed filling in the blanks with the appropriate word for the accompanying test, you should continue on to the next passage and test and complete it in the same way.
4. Continue through the passages and questions as they are assembled until you are finished.
5. Work as quickly as you can. Do not spend too much time on any one blank.

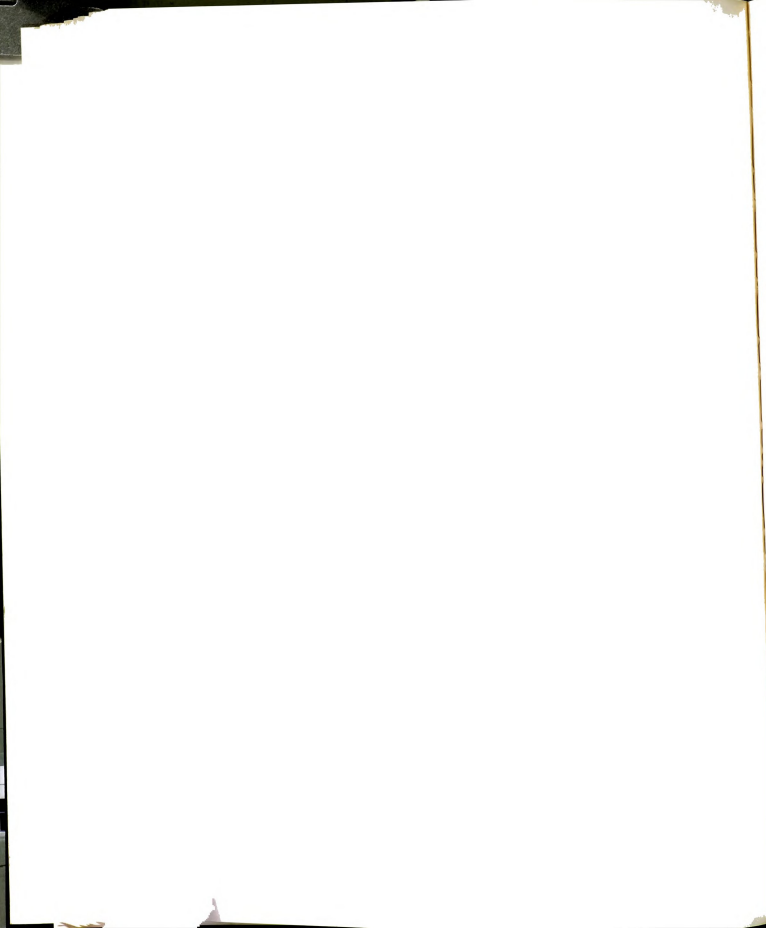


Questions

Almost from the moment his bones were first discovered _____ Germany's Neander Valley a century ago, his name has _____ synonymous with brutishness: a squat, shambling creature who wooed _____ women with a club and sometimes ate his fellow _____ when he was hungry. Scientists have long doubted this _____ popular image of Homo neanderthalensis, or Neanderthal man. Now, _____ the evidence accumulates, Neanderthal man is rapidly being rehabilitated _____ a more attractive ancestor of modern man.

From remains _____ in Europe, archaeologists have already concluded that Neanderthals were _____ hunters and toolmakers, held formal burial rites that indicated _____ belief in an afterlife, and even practiced a primitive _____ of Social Security for their aged and infirm. More _____, paleontological examination of skeletons has suggested that Neanderthal man's _____ appearance may have been the result of disease rather _____ low evolutionary status. According to this theory, he was _____ by a dietary deficiency of vitamin D. This deficiency _____ aggravated by the diminished sunlight of the ice age, _____ eventually caused rickets. Now, the most detailed and sympathetic _____ yet of Neanderthal man comes from extensive diggings by _____ American-led expedition in a mountain cave near the village _____ Shandar in Iraqi Kurdistan.

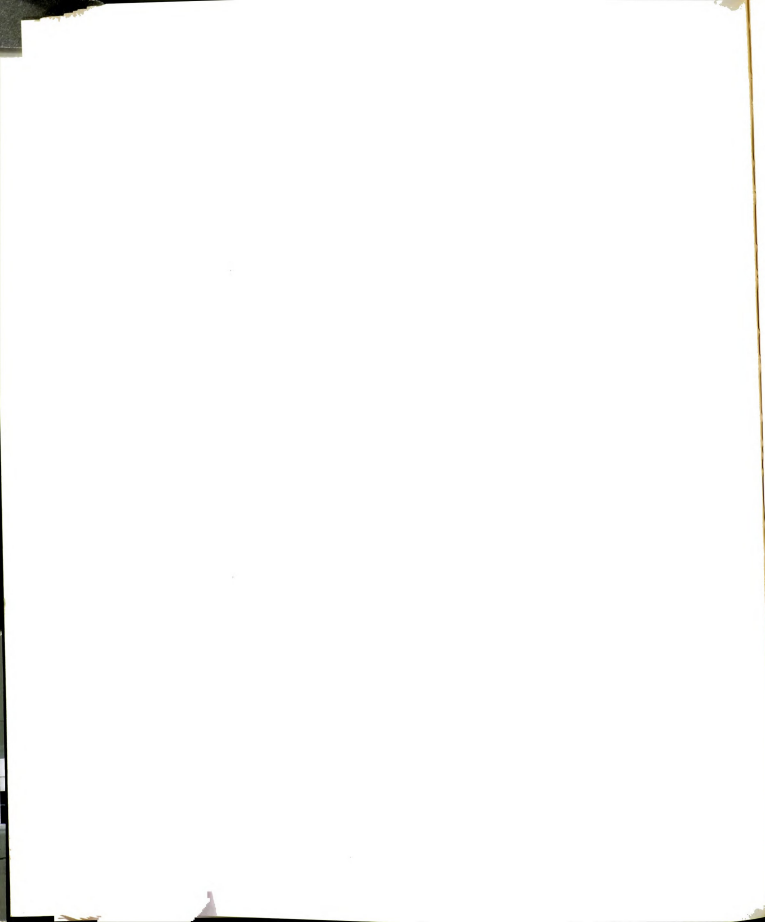
In an article in the _____ Smithsonian magazine, and in a forthcoming book, Shandiar: The _____ Flower People (Knopf: \$8.95), the expedition's chief archaeologist, Dr. Ralph S. _____, reports that at least one of the nine Neanderthal _____ uncovered in the Shandiar cave was buried with flowers. _____ skeleton was that of a man about 40 (equivalent _____ an age of 80 by modern life-spans) who had _____ born with a withered right arm. The limb had _____ been amputated above



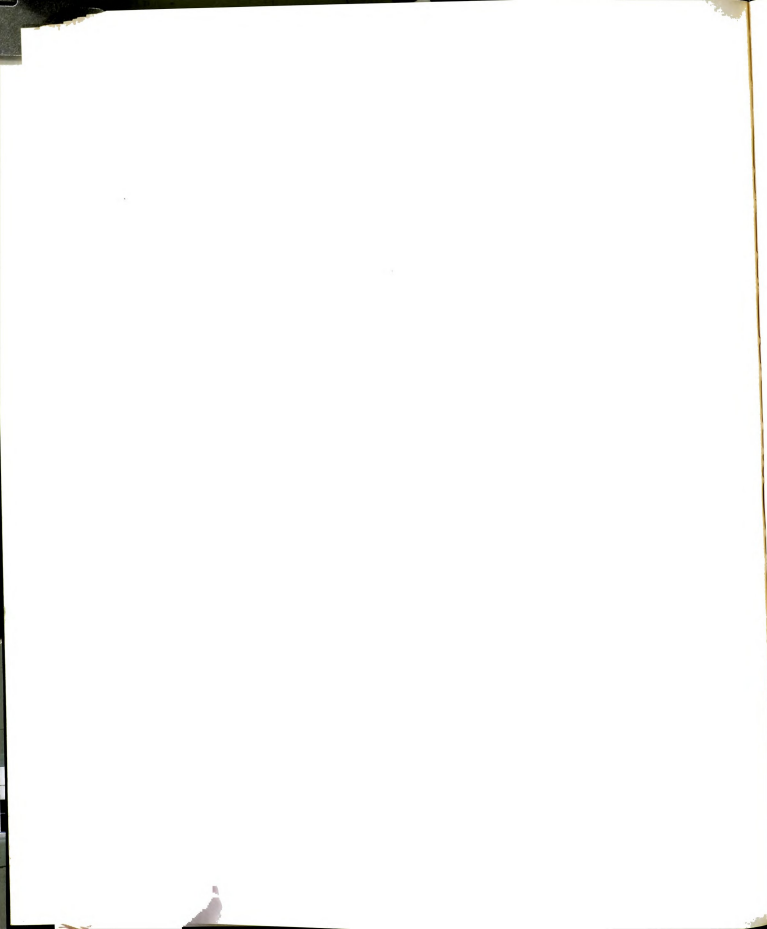
the elbow by a Neanderthal "surgeon." _____ man's age and physical condition indicated to the scientists _____ he had been unable to fend for himself. They _____ that his fellows kept him alive until he met _____ death in an accidental rockfall inside the cave, a _____ peril for these communal hunters who lived from 100,000 _____ 40,000 years ago. Comments Anthropologist Carleton S. Coon: "On _____ grounds of behavior alone, the Shanidar folk merit the _____ of Homo sapiens.

Chimp or Philosopher. Neanderthals conducted other _____ rites besides funerals. Clues to one of these were _____ in Lebanon last summer when an expedition led by _____, who is a professor of anthropology at Columbia University, _____ the dismembered skeleton of a small deer in a _____ overlook the Mediterranean. The 50,000 year-old bones had apparently _____ arranged in an orderly way and sprinkled with red _____, a substance used for symbolic purposes by Neanderthal man. _____ on the discovery last week, Solecki said: These men _____ trying to ensure a successful hunt by the ceremonial _____ of one of the animals." In other words, Neanderthal _____ resorted to a form of hunter's magic.

How did _____ sophisticated a creature acquire such an unwarranted reputation? For _____ thing, the first Neanderthal bones were dug just about _____ time that Darwin astonished the world with his announcement _____ man and ape were descended from a common ancestor. _____ apish image was further enforced by the writings early _____ this century of the respected French paleontologist Pierre Marcellin _____. His



portrait of Neanderthal as a stunted, beetle-browed creature _____
walked with bent knees and arms dangling in front _____ him served
as the model for several generations of _____ and cartoonists. While
certain coarse features in Neanderthal man _____ undeniable, on
physical considerations alone he deserves far better _____. As the
late Harvard anthropologist Earnest Hooton once commented: " _____
can, with equal facility, model on a Neanderthaloid skull _____
features of a chimpanzee or the lineaments of a _____."

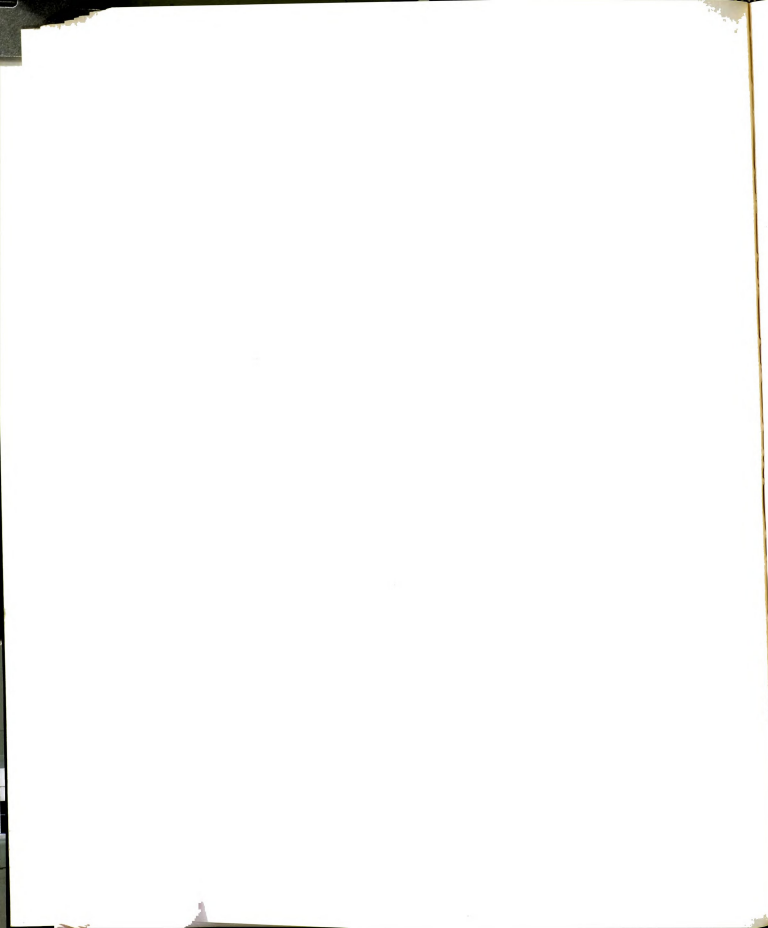


Questions

Cold and snow are twins in the boreal forest. _____ many sorts of mammals, as well as for grouse, _____ is a protection against a plummeting temperature. The short-legged, _____, short-tailed voles, for instance, spend almost the entire winter _____ the air space between the soil and the snow. _____ do the shrews and a few mice.

We know _____ little as yet about this world between the ground _____ the snow, except that its temperature seldom gets very _____ below freezing, and that the light is extremely dim. _____ can guess that its relative humidity is high and _____ the sounds of the world above are as muffled _____ the light. The group of voles living in this _____ of quater-world includes the meadow and red-backed voles which _____ also found in the rest of the province, the _____ vole and the bog lemming. Occasionally the voles build _____ through the snow but usually, because of the air _____, there is no need to do this. However, they _____ drive ventilation tunnels to the surface of the snow _____. They also bore through the snow itself from time _____ time, apparently in search of food particles fallen from _____ trees or trapped in the snow-cover. Their habitat has _____, like those in Swiss cheese, wherever there is qaminiq _____ snow too thin to give insulation.

While the voles _____ snow for survival, beavers and muskrats avoid the extreme _____ of winter by living under the ice of ponds _____ rivers. This is the warmest place for the, since _____ water the temperature cannot fall below freezing. Both species _____ about as they do in summer, seldom coming onto _____ surface of the land unless their ponds are frozen _____ the bottom. Sealed off by snow and ice



and _____ by the frozen walls of their hoses, they are _____ safe. But every Eden, even a frozen one, has _____ handicap. Winter does not prevent mink and otters from _____ their habitat.

Snow can hinder as well as help. _____ Alberta everyone knows from experience how difficult it is _____ struggle through knee-deep snow. The animals of the boreal _____ have solved this difficulty in various ways.

One group _____ mammals simply wades through the snow as we do. _____ best adapted animal in this group is the moose. _____ moves easily through two feet of snow because of _____ long legs and a special arrangement of muscles and _____ that permits the legs to work almost straight up _____ down. The toes of the moose seldom leave drag _____ in the snow so that it is usually difficult _____ tell from its tracks in which direction it was _____. The wapiti and smaller deer run into trouble at _____ lesser depths. The predator of the moose, the wolf, _____ wades through the snow and, as long as it _____ soft and fluffy, the wolf is at a considerable _____. But if a crust forms, the wolf may be _____ to stay on the surface while the moose keeps _____ wading, but with difficulty. Hence, the depth and structure _____ the snow mean life or death to both the _____ and the moose. Animals living in herds reduce the _____ needed for ploughing through snow by playing follow-the-leader. This _____ the technique used by the bison of Wood Buffalo _____ Park, by caribou, and in deep snow by deer, _____ they tramp down an area to form what is _____ as a yard.



Questions

The Province of Alberta has six distinct types of _____ corporations - four urban and two rural. Cities, towns, villages, _____ urban counties are urban municipalities; municipal districts and rural _____ are those found in rural areas.

As you will _____ aware, Alberta has a number of urban centres such _____ Edmonton, Calgary, Lethbridge, and Medicine Hat, whose populations have grown _____ to enable them to become cities. They are the _____ urban municipal corporations in the province.

We are going _____ outline the city in detail as the basis for _____ municipal corporations.

Prior to 1951, cities were given separate _____ by the provincial government. Calgary received its charter in _____, Edmonton in 1904, Medicine Hat and Lethbridge in 1906. Each _____ included such details as the name of the city _____ the powers its council would exercise.

The practice of _____ individual charters to cities created a number of problems. _____ a city's charter was lost temporarily, there was difficulty _____ deciding what laws or statutes governed its operation. Moreover, _____ separate charter system discouraged closed relations and co-operation between _____ of one city and another. Realizing the weaknesses of _____ existing charter system, the provincial legislature passed the City _____ which came into effect in January, 1952.

This statute, _____ ensuring that the charter of cities would be basically _____ laid the groundwork for improved local government in large _____ centres. Cities, as well as other municipal corporations, are, _____ course, under the control of the Department of Municipal _____, a department of the provincial government.



How does an _____ area become a city? The answer to this question _____ be found in the City Act. To acquire that _____ a town must have a population of over six _____ people and its council must apply to the Minister _____ Municipal Affairs for legal recognition as a city. If _____ Minister so recommends, the Lieutenant Governor in Council (the _____) may, by proclamation, grant the town the status of _____. This proclamation will, among other particulars, state the name _____ the new city, its area, and the date on _____ it is officially to become a city.

The City _____ of 1952 did much to establish on a uniform _____ the system of government for large urban centres. Executive _____ legislative powers were granted to councils composed of a _____ and aldermen. The mayor, elected for a two-year term _____ all the voters, is assisted by a deputy mayor _____ is selected by the council from among its members. _____ the mayor is absent from meetings, the deputy mayor _____ his duties. If neither of these municipal officers is _____ at a council meeting, the councillors present may appoint _____ acting mayor to preside over deliberations.

In addition to _____ elected mayor, a city council includes aldermen. By Law _____ council must not have fewer than six nor more _____ twenty aldermen, and the number of aldermen must be _____ even one. One-half of the aldermen are elected each _____ for a term of two years by all the _____.

Alberta's largest cities, Edmonton and Calgary, show a few _____ from the general practice in electing their councils. In _____. the mayor and twelve aldermen are elected every two _____ on a city-wide basis. In Calgary, six wards have _____ established, with two aldermen elected to represent each ward. _____ of the aldermen retire each year. The mayor, however, _____ chosen by all the voters of the city for _____ usual two-year term.



Questions

The relationships that exist among the plants and animals _____ have been described are almost unbelievably complex. Each species _____ its food requirements, its space requirements and its requirements _____ temperature, light and water. It is common to refer _____ "competition" among organisms for these requirements. Organisms which are _____ any disadvantage in the competition, for any reason, decline _____ numbers and may become locally or universally extinct. Of _____ conditions may change (for instance, moist years may follow _____) and the trend to decline may be reversed. Because _____ conditions are constantly changing, the numbers of living things _____. As examples, for a few years there may be _____ hares, then, for some years, hares may be scarce. _____ around a slough willows may predominate for a period _____ time, to be drowned when the water level rises, _____ replaced by cattails. In some years sparrow hawks are _____, in some grasshoppers thrive. In some years Canada thistle _____ a greater nuisance than in others. We are not _____ sure what causes the fluctuations.

The relationships between organisms _____ often called together "The Balance of Nature." This phrase _____ that living things occur as best they can in _____ face of the competition of all other living things _____ the changes of physical conditions (weather, soil conditions, etc.). _____ Balance of Nature is not something that is the _____ year in and year out. The balance changes all _____ time because new and different influences are always affecting _____ or another member of the natural community. This, in _____, affects neighbouring members of the community so that living _____ might be



bettered or worsened for different members of _____ community. The balance will have shifted.

Let us examine _____ imaginary case, to see how a balance might shift _____ a period of time. We shall look at a _____ area at the margin of the aspen parkland zone. _____ grasses there might be chiefly a rough fescue, porcupine _____ and June grass. The grasses generally prevent the establishment _____ aspen: under the proper conditions, grasses are more successful _____ for soil than aspen. But the situation can be _____ if bare areas of soil develop. Such bare areas _____ be formed by burrowing Richardson's ground squirrels, badgers, pocket _____ or other animals, or by overgrazing and trampling by _____, or in days past by bison. Bare soil areas _____ shrubs, such as snowberry, a chance to become established. _____ sharp-tailed grouse, using the bare area as a dusting _____, left seeds there in droppings. Other seed-eating birds or _____ dispersal can also seed bare areas. If snowberry and _____ willow become established, their shade inhibits grass growth. Coyotes _____ to den in shrubby sites further disturbing the soil _____ preventing the growth of grass. Some grasshoppers, preferring grass, _____ scarce, as do some ants. Different species of ants _____ grasshoppers are found which prefer shrubby habitat. Aspen seedlings _____. As aspen develops, it begins to shade and restrict _____ and wolf willow. Shrubs come to form a fringe _____ the edge of an enlarging grove of aspen. Songbirds _____ attracted to nesting sites. Flickers and crows inhabit the _____. Snowshoe hare and Franklin's ground squirrels prefer the woods. _____ does the great horned owl.



As the aspen ages, _____ seedlings may begin to appear. Spruce may be held _____ if a lot of hares are present to feed _____ the tender young plants. Fire will stop the development _____ spruce, while after a fire aspen can regenerate by _____ of suckers. But these aspen suckers can be held _____ by the foraging activities of rabbits or other browsers. _____ such cases, grass will reappear, once again to be _____ main vegetation, with its own distinctive population of animals.





MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03178 7546