0-315-08930-X

National Library of Canada — Bibliothèque nationale du Canada
Canadian Theses Division — Division des thèses canadiennes
Ottawa, Canada
K1A 0N4

56847

PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

- Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur

DOUGHTY ANTHONY EDWARD

Date of Birth — Date de naissance	Country of Birth — Lieu de naissance
1940 10 21	ENGLAND

Permanent Address — Résidence fixe

11715 - 145 Ave
EDMONTON, ALBERTA
T5X 1M2

Title of Thesis — Titre de la thèse

The Use of Reading Difficulty Indices for Predicting Test Item Difficulty

University — Université

University of Alberta

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée

Master of Education

Year this degree conferred — Année d'obtention de ce grade	Name of Supervisor — Nom du directeur de thèse
1982	Dr M.W. Petruk

Date	Signature
1982 04 26	[signature]

THE UNIVERSITY OF ALBERTA

THE USE OF READABILITY INDICES FOR PREDICTING TEST ITEM DIFFICULTY

by

© ANTHONY EDWARD DOUGHTY

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

IN

VOCATIONAL EDUCATION

DEPARTMENT Industrial and Vocational Education

EDMONTON, ALBERTA

SPRING, 1982

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Anthony Edward Doughty

TITLE OF THESIS The Use Of Reading Difficulty Indices For Predicting Test Item Difficulty

DEGREE FOR WHICH THESIS WAS PREPARED Master of Education

YEAR THIS DEGREE WAS GRANTED 1982

PERMANENT ADDRESS:

11715 145 Avenue

Edmonton, Alberta

T5X 1M2

DATED April 21, 1982

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled The Use of Readability Indices for Predicting Test Item Difficulty submitted by Anthony Edward Doughty in partial fulfilment of the requirements for the degree of Master of Education in Vocational Education.

....

Supervisor

....

....

Date April 22, 1982

ABSTRACT

A 40 item multiple choice examination was produced by selecting items from an existing gasfitting test item bank maintained [illegible] the Pipetrades section of the Northern Alberta Institute [illegible] Items were selected to stratify the predetermined [illegible] indices with a platykurtic distribution within th[illegible] 0.95. The items were evaluated to determine th[illegible] ty indices using a BASIC language programme writt[illegible] I microcomputer. The examination was administer[illegible] graduating examination, to 52 apprentice plumb[illegible] who were unaware of the experimental nature of the exam[illegible] who were within one week of graduating from their final [illegible]aining.

An item difficulty index was determined f[illegible]st item from the experimental groups, and these indices were correlated with the predetermined item difficulty indices used to select the items, to determine if the item difficulty indices were replicable. The data indicates that although there are some group variation in the item difficulty indices, statistically the indices are replicable.

Three readability indices were determined for each item, using the Gunning Fog Index, the Automated Readability Index and the Grade Reading Level. These readability indices were correlated to determine if the three indices measured the items consistently. The data indicates that the three readability formulae produced consistent readability indices.

The three readability indices were then correlated to the

predetermined item difficulty indices to determine if the readability index could be used to predict the test item difficulty index. The data indicates the item difficulty index could not be predicted from the readability index.

This research demonstrated that a microcomputer can be used effectively to determine an index of readability for multiple choice test items, and that where the average readability index of the test items approximates the average education grade level expected of the examinees, readability contributes insignificantly to the test item difficulty index.

Different item difficulty indices between different groups of examinees is attributed to other variable factors such as students' lack of knowledge, students' misconceptions about the subject being examined, or other components of the test item construction such as homogeniety of the alternative distractors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. Introduction

Few tasks undertaken by an instructor are as demanding as that of writing valid, reliable, objective, content referenced multiple choice test items. The instructor in post secondary technical and vocational education has a responsibility to the students to ensure that the examination instrument adequately evaluates the students' competence, but at the same time the instructor has an obligation to society at large to ensure only safe, knowledgeable practitioners are graduated.

In most trades areas in the Province of Alberta, the Alberta Apprenticeship Board has the authority to certify tradesmen graduating from the Technical Institutes, but individual instructors have the responsibility of preparing the apprentices for the Apprenticeship Board examinations, and must compile their own evaluation tests to certify the apprentices ready to attempt the Tradesmans Qualifying examinations.

These responsibilities place the instructor in a difficult position, taxing his knowledge of the subject matter being taught, and his ability to express himself clearly and succinctly in examination test items. Unfortunately there are reasons to believe these dual responsibilities are not being fulfilled. Robert Ebel identified the significance of this problem when he stated:

> Far more harm is currently being done to student learning by the shortcomings of classroom tests by which student's educational efforts are largely stimulated, directed and evaluated, than is being done by all the faults of external testing programs. ... I am fully persuaded that the current problem in testing that most urgently requires the attention of all professional educators is that of improving the tests made and used by the classroom teacher. (Ebel, 1980).

Many factors have been identified as contributing to the shortcomings of classroom tests, and many authorities describe methods for analysing and determining defective test items (Ebel, 1980; Mehrens and Lehman, 1973; Popham, 1978; Sax, 1974). Their recommendations are based on evaluation and interpretation of classical item analysis data obtained, after the fact, from students performances on the test, and from norm referenced criteria that distribute people and their performances under a normal distribution curve.

Other authorities (Bloom, 1971; Tyler, 1971; Stodala and Stordahl, 1967; Finley and Berdie, 1970; Lindvall and Nilko, 1975) have described methods of improving test items by careful analysis of both the validity and reliability of each test item.

Validity of a test is an indicator of the degree to which a test actually measures that which it is supposed to measure. Bloom, (1971) defines validity as a function of four concepts of validity:

1. Content validity - the degree to which an instrument is consonant with the content, skills, objectives it is supposed to measure.
2. Construct validity - refers to the extent to which the results of a measure permit inferences about underlying traits.
3. Concurrent validity - the degree to which results of a measuring instrument match the results of an alternative measuring instrument, and:
4. Predictive validity - refers to the extent to which the results of a measuring instrument permit prediction of results of future performances.

Reliability measures the extent to which the instrument provides consistent and stable indications of the characteristics being measured. Bloom also views reliability as a function of six different components:

1. Reader reliability - represents the agreement whith which different users of the instrument will interpret the measurement of a particular product.
2. Internal consistency - the extent to which a measuring instrument contains items that measure a common trait or characteristic.
3. Instrument stability - represents the extent to which errors occur with the instrument as a result of fluctuations in the performance of the respondents.
4. Examinee reliability - describes the degree to which an instrument measures a particular respondent accurately.
5. Sampling reliability - another form of content validity. The reliability with which an instrument samples the different types of content and behaviour stressed.
6. Congruence reliability - indicates the extent to which items on a measuring instrument indicate the same traits.

Although many authors have demonstrated the importance of both the validity and reliability components of written test items (Bloom, 1971 Ebel, 1980; Sax, 1974; Denova, 1979; Tyler, 1971; Miller, 1972; Storey, 1970) final evaluation of several aspects of both validity and some components of reliability, such as sampling reliability, rely on subjective evaluation of the instrument by either the test writer, by colleagues who are presumed to have the same training and expertise as the test writer, or by peers with different training and expertise.

One factor that most researchers agree upon as contributing to the shortcomings of classroom tests, and therefore contributing to poor validity and reliability of any written measuring instrument, is the readability of the written test. As Benson and Crocker (1979) state:

> Failure to master essential reading skills at lower grade levels has particularly compounded the problems of instruction and evaluation in special subject areas at higher grade levels. Although it might seem ideal to improve basic skills before proceeding with additional instruction it is unrealistic to suppose that instruction in critical areas of ... vocational subjects can be postponed until students' cumulative reading deficits are remedied.

Several authors of texts on test construction techniques include a generalised statement that tests should be written at the language level of the student being evaluated (Sax, 1974; Denova, 1979; Ebel, 1980), but no attempt is made to suggest how this can be objectively determined or achieved.

In a research project (Clark, 1977) to determine the readability of textbooks used in industrial education in the province of Alberta, it was found that the mean grade reading level of 25 highschool texts was 14.0 years. Clark stated:

> Unfortunately there is reason to believe the average reading level of vocational students is appreciably lower than that of the academic students at the same grade level.

The implication is that if the readability of written material is too high for the students' reading ability level, then the material may discriminate against some students because of readability. This implication was forcefully expounded by E. F. Gardiner in his presidential address to the National Council on Measurement in

Education in 1979. Gardiner stated:

> A test designed to measure achievement in elementary science, and in which the items are presented at a language and syntax far above the level of the pupils tested is biased, especially against the poor reader ... The score on the test should measure knowledge of science with appropriate scientific vocabulary, but the score should not be dependent upon general reading ability. (Benson and Crocker, 1979).

Statement of the problem

Although writers on test construction, testing and performance evaluation agree that the readability of test items, and the reading ability of examinees can, and does contribute to bias in test results, none of these authorities describe methods by which readability can be objectively evaluated before the test is administered to the students. Furthermore very few instructors are known to make an objective evaluation of any of the instructional materials they produce.

Purpose of this study

This study is being conducted to determine whether a microcomputer can be effectively used to evaluate an index of readability for content referenced multiple choice test items. And, based on the assumption that a measurable component of the item difficulty index is dependent upon the readability of the test item, this study is being conducted to determine whether a readability index obtained from the microcomputer could be used as a valid method of predicting the item difficulty of individual multiple choice test items, and thus provide an objective method of contributing to an improvement in classroom tests, constructed and administered by educators.

Need for the study

With the current emphasis in vocational and technical education on performance based instruction, and the resulting content referenced evaluation procedures, instructors want their evaluation test results to indicate how much the students know about the content domain being evaluated.

Furthermore, with the rapid introduction of microcomputer technology into the education system and training institutions, many groups of instructors with common pedagogic responsibilities may be cooperating to develop test item banks using these microcomputers.

A simple method of determining the readability index of test items could be used by educators to improve the quality of examination and other test items, and permit the selection of test items that approximate the reading ability level of the students being evaluated. This would lead to an improvement of both the validity and reliability of classroom tests, by minimizing bias introduced by complexities in the reading difficulty of the test items.

Also, a simple programme to evaluate the readability of written material, and readily available for a classroom microcomputer, may encourage educators to subject their written material to readability analysis before storing the information onto the disk storage. In the case of test items this would enable the readability data to be stored with the item for later use in selecting test items.

Questions to be answered

This research study is being undertaken to determine if an index of readability for individual items contained in a content referenced multiple choice examination can be used to predict the individual test item difficulty index.

Specifically the research will attempt to answer the following questions:

1. For content referenced multiple choice test items, is the item difficulty index of individual test items replicable when the test items are administered to different groups of examinees?
2. Can an index of readability be established for multiple choice test items, using a microcomputer? And
3. Can an index of readability be used to predict the item difficulty index for individual test items?

Delimitations

The test items used in this research were selected from over 2000 test items in the gasfitting item bank maintained in the Pipetrades section at the Northern Alberta Institute of Technology (N.A.I.T.). Items contained in this bank are currently used for N.A.I.T. testing of students prior to writing the Alberta Tradesmans Qualifying examination administered by the Provincial Apprenticeship Board, under the Alberta Tradesmans Qualifications Act. For the purpose of this research the test items were selected on the basis of measuring a specific training objective, and stratified to represent a full range of item difficulty from 0.30 to 0.95

Two classes of N.A.I.T. graduating pipetrades students were used as the experimental group, and the experimental test items were administered to the classes within 3 days prior to writing the

Apprenticeship Board examinations.

The experimental test items were administered to the apprentices as a final N.A.I.T. graduation examination, and the students were not informed of the experimental nature of the examination.

This research study is investigating only the readability of the experimental test items contained in this N.A.I.T. graduating examination, as defined by mean sentence length, mean word length, and the item difficulty index of the items as defined by the proportion of the students who answer the item correctly.

Limitations

This research was limited to evaluating multiple choice test items that measure the general training objective of the students being able to interpret the Canadian Standards Assocation Gas Codes - B 149.1 and B 149.2 (CSA, 1976 and CSA, 1978). This study was further limited by the selection of the student sample, representing students trained in the Pipetrades section at the Northern Alberta Institute of Technology, using the teaching techniques and unique training aids used there.

The results of this study are only generalizable to the extent that similar training institutions adopt similar training practices, and receive similar student populations.

Assumptions

To undertake this research it has been necessary to make the following assumptions:

1. It is assumed the test questions selected from the

gasfitting test item bank have content, construct, concurrent, predictive and criterion related validity.

2. It is assumed the questions selected from the test item bank also have reader, sampling, congruent, and examinee reliability, and that the test possesses internal consistency and instrument stability.
3. It is further assumed that the graduating classes of gasfitting students used as the experimental group, are representative of all students graduating from the training programme.
4. The reading ability of all students in the sample tested in this study is normal and representative of the reading ability level of all graduating students from the programme.
5. It is further assumed that the classical item analysis performed on previous examinations, and which was used to determine the item difficulty index of the questions contained in the test item bank, is valid and reliable and has been obtained from a control group of students with similar training and reading ability characteristics as the experimental group.

Definition of terms

For the purpose of this study, the readability terms and phrases used are defined as follows:

1. Automated Readability Index:

A measure of the readability of a passage of text, derived from the average word length and the average sentence length, without considering sentence load, sentence structure or vocabulary load. The formula is:

ARI = AWS + (9 X AWL)

Where: ARI = Automated readability index

AWS = Average Words per Sentence

AWL = Average Word Length

2. Cloze Procedure:

A technique used in testing, practice work and evaluation for determining readability of a text passage. It involves deletion of words from the text and substitution blank spaces of equal length. The measurement is made by rating the number of blanks which can be correctly filled

3. Gunning Fog Index:

A measure of the readability of written material based on the average sentence length and the number of three syllable words the passage contains. the formula is:

GFI = 0.4 (ASL + %TSW)

Where: GFI = Gunning Fog Index

ASL = Average Sentence Length

%TSW = Percentage of Three Syllable Words

4. Index of Readability:

Another term for readability index. A numerical value that describes the mathematical relationship between predetermined characteristics of written text, and which ranks that text relative to all other text passages in order of the difficulty of reading as defined by the characteristics being measured

5. Readability:

The characteristics of a text passage represented by the average sentence length and the average word length, that facilitate or obstruct reading.

6. Readability Index:

Another term for Index of Readability.

7. Readability Level:

An indication of the difficulty of reading material in terms of grade level at which it might be expected to be read successfully (Gunderson, 1969).

8. Reading Ability:

The human characteristic of being able to read and understand printed material. It is generally assumed reading ability increases in proportion to the amount (or level) of education the reader has achieved.

Reading ability involves factors such as a persons cognitive style, response set, word and vocabulary knowledge, comprehension and interpretational ability

The ability to percieve and understand how written symbols refer to whatever context the language carries. (Gunderson, 1969).

9. Rudolph Flesch Index:

A readability index of a text passage that considers the average number of words per sentence and the average number of syllables per 100 words. The formula for calculating the index is:

RFI = 206.835 − (0.846 X AWL) − (1.015 X ANS)

Where RFI = Rudolph Flesch Index

AWL = Average Word Length

ANS = Average Number of Syllables per 100 Words

10. Sentence Load:

Complexity of sentences caused by the choice of words, syntax, position of words and punctuation, and which contribute to reading difficulty.

11. Vocabulary Load:

The total number of lexical terms in a particular persons repertory, or in a particular (section of) Text. (Gunderson, 1969).

Other terms and phrases are defined as follows:

12. Classical Item Analysis:

The process of examining students' responses to each test item and specifically to determine the item difficulty index of the test item.

13. Content Referenced Test Item(s):

Multiple choice test question(s) specifically selected to measure the content of a predefined general training objective.

14. Item Difficulty Index:

The proportion of the total group of students who answer a particular test item correctly. The item difficulty index ranges from 0.0 (nobody answered the item correctly) to 1.0 (everybody answered the item correctly) and is obtained by counting the number of examinees who answer the item correctly, and dividing by the total number of examinees attempting the examination. For this research the item difficulty index was obtained from classical item analysis procedures.

15. Item Reliability Index:

A measure of the effectiveness of a test item to discriminate between high and low scoring students when simultaneously considering the point biserial correlation, and the test item difficulty index. The item reliability index is calculated from:

IRI = PBC X SQR (DIFF X (1 - DIFF))

where: IRI = Item Reliability Index

PBC = Point Biserial Correlation

DIFF = Item Difficulty Index

SQR = Mathematical operation of taking the square root.

Summary

Considerable agreement exists between researchers that the quality of classroom tests needs to be improved. One possible method of improving quality may be in measuring and recording the readability of individual test items. Because of the rapid introduction of the microcomputer technology into classrooms of Alberta schools it was decided to use a microcomputer to evaluate readabilty of written material.

Predetermination of an index of readability for test items, and storage of this information on disk files together with the test items, could be used for future selection of questions of appropriate reading difficulty for the students being evaluated. This would contribute to an improvement in the quality of classroom tests by diminishing the bias introduced into test scores by students' lack of reading ability. The next chapter will outline some of the readability formulae that are used to determine readability of text material.

II. Review of the Literature

This study is concerned with the readability of multiple choice examination items, and its' relationship to the item difficulty index. This review of the literature includes a review of both the literature on readability formulae, and on multiple choice test item difficulty.

<u>Readability formulae</u>

The term readability has different interpretations and is most frequently used in three different contexts:

1. Ease of understanding or comprehension of written text due to the syntax or style of writing.
2. As a measure of legibility or typographic layout of the text. And,
3. Ease or difficulty of reading due to the word length and sentence length.

The definition of readability used by Dale and Chall (1948) embodies these three components. They stated:

> In the broadest sense, readability is the sum total (including interactions) of all those elements within a piece of printed material that effect the success which a group of readers have with it. The success is the extent to which they understand it, read it at optimum speed and find it interesting.

Thus readability has three major components: that influenced by human factors; that attributed to mechanical-typographical features; and that caused by the lexical content.

The human component of readability

The human component of readability involves the psychological and physiological attributes individuals contribute to their reading assignments. Many variables have been identified and investigated, and are generally classified under the rubric of reading ability.

Physiological factors involved in reading ability include sensory ability, strength, and reaction time. Research has indicated that decreased visual acuity in adults may cause altered perception, and adults cannot read as fast as they could when younger (Norris, 1977; Botwinick, 1978). These physiological components cannot, however, be isolated from the psychological factors involving such variables as recognition speed, reading efficiency, memory span, comprehension, motivation, etc. (Lennon, 1970; Chall, 1970; Miller, 1962; Lorge, 1949; Gilliland, 1972; Botwinick, 1978). These factors involve information processing.

Differences in the way people process information is referred to as cognitive style. Two major cognitive styles have been identified: Global, and Analytic. Global cognitive style is characterized by descriptive and functional modes of abstraction, short attention span, and distractedness. Whereas the analytic cognitive style is characterized by longer attention spans, greater reflectivity and deeper concentration (Rowley, 1974).

Cognitive style represents the way in which an individual selects, organizes and processes information. If there are differences in these cognitive styles between individuals, then these differences challenge the validity and reliability of testing procedures.

Another factor involved in information processing is response set. Mehrens and Lehmann (1969) explain response set as the 'Tendency of an individual to reply in a particular direction, almost independent of content. An individual exhibiting response set will answer identical questions (but presented in a different format) differently.' Three major response sets have been delineated, they are: acquiescence, social acceptability, and evasiveness. (Sax, 1974; Huck, 1978).

Shuyler Huck (1978) demonstrated that by providing students with a prompt of the predetermined item difficulty index of test questions response set could be modified. Using five most difficult multiple choice, five most difficult true/false, and five most difficult fill-in-the-blank items, Huck found that students performed better on these items when they were presented with the item difficulty information.

Although the research has not yet demonstrated an interdependence, cognitive styles and response styles have been extensively investigated. Rowley (1974) states:

> The use of multiple choice tests can produce scores which favour certain types of examinees and penalize others for reasons not explained in terms of knowledge of material being tested. ... High risk taking, testwise examinees score more highly than others.

<u>The mechanical component of readability</u>

The mechanical component of readability involves the practical, mechanical and environmental conditions which contribute to the ease or difficulty encounterd in reading printed material.

Harold Burtt (1949) outlines nine factors which contribute to this mechanical variant in readability, including type face; upper vs. lower case characters; line length; distance between lines; page (spatial) layout; use of colour; type of paper; and quantity and type of illumination. Most researchers recognize these variants, and they are often referred to in the literature under the headings of visibility, legibility or typography (Gilliland, 1972; Dale and Chall, 1949; Clark, 1977).

Although the importance of the human and mechanical components of readability cannot be discounted, the present study is concerned only with the lexical component of readability. That is the ease or difficulty of reading due to the word length and sentence length of the text.

The lexical component of readability

The lexical component of readability has been investigated since the time of the Talmudists who, 'in compiling and studying the body of laws called the Talmud, counted the occurences of words and ideas in seeking to distinguish differences in meaning.' (Gilliland, 1972. P.16).

Five methods by which the lexical component of readability could be analyzed have been identified. These are:

1. Subjective assessment
2. Objective question and answer techniques
3. Sentence completion
4. Formulae
5. Tables and Charts

1. Subjective assessment is considered by Gilliland (1972) to be the most unreliable method of determining readability, but one of the most frequently used. Librarians, teachers, instructors or anyone else choosing reading material on behalf of other people haphazzardly cast a glance through the pages using clues such as content, style, vocabulary load, format, page layout, print style and size, quantity of pictures, and the use of colour, make value judgements about the reading difficulty of that text.

2. Objective question and answer techniques essentially measure comprehension of content by quizzing readers about what they have read. This procedure, while more impartial and controlled than subjective assessment still has severe limitations restricting its utility.

3. Sentence completion is also known as the Cloze procedure (Bormuth, 1966; Milagros, 1969; Rankin, 1970).

The Cloze readability test consists of preparing a copy of the passage to be tested, with every fifth word deleted and replaced by under-lined blanks of a standard length. The test is administered without time limits, with the students being told to fill in each blank with the word that they think has been omitted.

Responses are scored by counting exact matches of deleted words, disregarding spelling errors. The student's score is interpreted as a measure of how well the student understood the material from which the test was compiled. (Bormuth, 1968).

The Cloze procedure has been used to determine the readability of different texts, and it has been shown that the Cloze procedure has a high correlation with results of intelligence tests. Rankin (1970) points out that:

> Although the Cloze procedure appears to be a valid and useful measure of readability, intelligence, knowledge, and reading comprehension, only a few studies have been carried out to assess the validity and usefulness of the technique under varying conditions.

It appears that the Cloze procedure has many potential uses, but as a research tool it still needs further validation to differentiate between its ability to measure readability and reading ability.

4. Formulae have been used extensively in determination of the readability of text material since they were first introduced in 1923. Many formulae have been developed since that time.

George Klare (1963) provides a most complete synthesis of the many readability formulae used since 1923, describes in detail over 31 different formulae suitable for use at different grade levels primer through college, provides detailed methods for using these formulae, and outlines many validation studies performed on these formulae over many years. He also provides an extensive annotated bibliography.

Although many different formulae are now extensively used, to determine readability of written text material, many different researchers reiterate the principle that all formulae are empirical measures that cannot be derived mathematically, nor proven scientifically. At best readability formulae provide rough estimates of the reading difficulty level of the text based on the predefined mathematical relationships between predetermined characteristics of the text being analyzed. (Powers, et al. 1958; Gilliland, 1972; Clark, 1977)

Readability formulae attempt to express the mathematical relationship between variables such as: word length; sentence length; vocabulary load; sentence load; polysyllabic words; number and types of nouns, personal pronouns, affixes, prepositional phrases; or number punctuation marks. (Klare, 1968).

The major advantages of using formulae to determine an index of readability are: each formula will treat every passage, analyzed using that formula, in exactly the same way. And, using the criteria identified, will rank the text passage relative to all other text passages analyzed using the same formula, in order of the relative difficulty of that passage.

The three most commonly used, and most frequently validated formulae are: Flesch reading ease formula; Dale and Chall readability formula; and the Gunning Fog index. Although both the Dale and Chall and the Flesch reading ease formulae do use charts or tables to obtain a final readability index.

The Gunning Fog index appears to be one of the easiest formulae to use, and has been used in computerized readability programmes (Baker, 1981; Noonan, 1981; Carlson, 1980). The Gunning Fog index represents the reading grade level required for understanding the material, and is computed from the average sentence length and the number of words containing three or more syllables. The formula is:

$$GFI = 0.4 \times (ASL + \%TSW)$$

Where: GFI = Gunning Fog Index

ASL = Average Sentence Length

%TSW = Percentage of Three Syllable Words

A syllable is defined as a: 'vocal sound or set of sounds uttered with a single effort of articulation and forming a word or part of a word; each of the elements of spoken language comprising a sound of greater sonority with or without one or more sounds of less sonority. (Oxford English Dictionary, 1971).

The unique characteristic that identifies a syllable is that it contains a vowel, pseudo vowel, or vowel combination in conjunction with other consonants. Thus by counting the number of vowel or vowel combinations a reasonable accurate determination of the number of syllables the passage contains, can be made.

Smith and Senter (1976) demonstrated that a relationship exists between the number of letters in a word and the number of syllables it contained. They demonstrated that as the number of syllables in words increased the length of the word increased, they also found that the relationship was consistent and statistically significant.

Because of the ease of using a computer programme for counting the number of vowels, pseudo vowels, or vowel combinations in a text, the Gunning Fog index was selected to be programmed into the computer for use in the current research.

Smith and Senter (1976) developed two new automated readability formulae intended for use with modified typewriters. The modification would cause counters to be incremented whenever keys were depressed. These counters tabulated the number of characters, the number of words, and the number of sentences contained in any text passage written on that typewriter. Using multiple regression techniques, Smith and Senter computed an assigned grade level formula to be:

$$GL = (0.5 \times ASL) + (4.71 \times AWL) - 21.43$$

Where: GL = Assigned Grade Level

ASL = Average Words per Sentence

AWL = Average Characters per Word

They further simplified this formula to be:

$$ARI = ASL + (9 \times AWL)$$

Where: ARI = Automated Readability Index

ASL = Average Words per Sentence

AWL = Average characters per word

Because of the simplicity and ease of use of the Smith and Senter formulae, both the Grade Level and the Automated Readability Index formulae were selected for use in the current study.

5. Tables and charts are frequently used in conjunction with readability formulae. Identifiable aspects of the text are quantified and converted to a readability index or grade level using tables of data or conversion charts. Typical examples of readability indices using tables and/or charts include:

1. Dale and Chall formula (3000 word list)
2. Fry readability graph (Utilizes a conversion chart)
3. Rudolph Flesch index (Uses conversion table)

The Dale and Chall formula, intended primarily for adult reading material, uses a word list of 3000 words. Words which do not appear on this Dale list are classified as 'hard' words. The readability index is computed using the average sentence length and the percentage of hard words. The formula is:

$$RGL = (0.1579 \times \%HW) + (0.0496 \times ASL) + 3.6365$$

Where: RGL = Reading Grade Level

%HW = % Words not on Dale's 3000 word list

ASL = Average sentence length

The Dale word list is repoduced in Dale and Chall (1948).

Although Fry (1967) stated 'I find their (Dale and Challs') readability formula loaded with fussy rules, a tedious vocabulary, and decimal figures carried to the forth place, a bit overly precise ... ', their formula is now one of the most frequently used (Klare 1963).

Fry first introduced his readability graph in 1967 with the belief that his method avoided some of the drawbacks he criticised in the Dale and Chall formula. However he stated:

> The readability graph was first developed in Uganda ... The original version appeared in print read by British readers ... The readability graph presented in this article is aimed at the United States Education scene. (Fry 1967).

In validating his graph Fry found:

> The Dale Chall formula correlates quite highly with the readability graph (.94). The correlation with the Flesch formula (.96) was expected ...

The Fry readability graph has been validated in various research studies. In comparing the Dale and Chall, Fry, Flesch and Gunning Fog index reliability indices Clark (1977) stated of the Fry readability graph:

> ... (It) appears to be the best formula for general use. It is fast and easy to use, it correlates highly with other formulas; and its slightly higher (Grade reading level) value ... gives some margin of safety without the inflated range of the (Gunning) Fog index.

Although the Fry readability graph has a high correlation with other readability formulae and has the distinct advantage of not requiring complex calculations, the process still requires counting the number of sentences, and the total number of syllables in the passage.

When refering to the syllable count, Fry simply stated:

> There is a syllable for each vowel sound; for example cat (one), blackbird (two) ... Don't be fooled by word size, for example polio (three), through (one). Endings such as y,ed,el,or le usually make a syllable, for example, ready (two), staffed (two), bottle (two).

Fry made no mention of how to handle proper names, numerals, contractions or abbreviations. For this information we must turn to Laubach and Koschnick (1977). The Fry readability graph is reproduced in Clark (1977).

The Rudolph Flesch reading ease formula has also been used extensively since it was introduced in 1949. Based on the average sentence length and the number of syllables per 100 words, the formula provides an index of reading ease and not a grade reading difficulty level. The grade reading level must be determined from the reading index by using a conversion scale. The Rudolph Flesch reading ease formula is:

$$RFRI = 206.835 - (SYL \times 0.846) - (ASL \times 1.015)$$

Where: RFRI = Rudolph Flesch Readability Index

SYL = Number of Syllables per 100 Words

ASL = Average Words per Sentence

Although these readability indices are popular among researchers, and have been validated in many studies, they were rejected for use in the current research because of their limitations in using tables or charts that would be impractical to programm into a microcomputer.

Test Item Difficulty

Historically, test item difficulty has been considered from two perspectives. The first defines test item difficulty as the proportion of examinees who answer the test item correctly. Using this definition a large item difficulty value indicates an easy test item.

The second perspective defines the test item difficulty as the proportion of students who answer the test item incorrectly. Using this definition a small item difficulty value inducates an easy test item. Most researches (Mehrens and Lehmann, 1973; Sax, 1974; Ebel, 1972; TenBrink, 1974; Huck, 1978) adopt the first prespective, and it is this definition that is being used in the current study.

The item difficulty index is calculated by dividing the number of examinees answering the test item correctly by the total number of examinees attempting the examination. The maximum value of 1.0 occurs when every examinee attempting the item answers it correctly, and a minimum value of 0.0 occurs when every examinee answers the item incorrectly or missed the item (Sax, 1974).

Although the item difficulty index indicates the proportion of the examinees who answered the item correctly, it does not indicate if the item is valid and/or reliable because the item difficulty index is not solely a property of the test item. It reflects the ability of the group responding to it.

A test item may have validity and reliability but produce a high item difficulty index. This is especially true when tests are used to measure mastery of subject matter. As Sax iterates:

> ... Mastery test items usually have high values of (item difficulty indices), which means that scores tend to pile up at the high end, creating a negatively skewed curve.

Alternatively, test items may display validity and reliability but produce low item difficulty indices. This has been demonstrated to be true when examinees have not been instructed in the subject matter being evaluated, or because students cannot read adequately, or because of a particular response set or cognitive style. TenBrink (1974) states:

> If an item is part of a mastery test, you would expect the items to be quite easy. In a test designed to measure knowledge prior to instruction, the items will all be quite difficult.

TenBrink also states:

> Other factors besides the difficulty of the subject matter can influence the level of difficulty of a given item. For example, the reading level of an item can make an item more difficult. When this is true the information obtained will reflect not only the subject matter being tested, but also the individuals ability to read.

Thus item difficulty is a function of several different factors, among these different authorities have identified: ambiguity; use of more obscure or less important factual information; unintentional clues to the correct response; complexity in phrasing; reasonableness of wrong alternatives; student misconceptions of material being covered; instructional changes; practice effect; memory for answers given (Myers, 1962; Stodala and Stordahl, 1967; Storey, 1970; Thorndike and Hagen, 1977).

Marshall and Hales (1972) summarize factors influencing test item difficulty as follows:

> The difficulty of an item is the function of the learning experience of the examinees, the questions asked, and the responses offered. The complexity of the item stem, as well as the extent of the ambiguity in the stem, influences item difficuly. Perhaps even more influential, however, are the alternatives offered. If the alternatives are quite homogeneous, the item will be more difficult than if the choices are relatively heterogeneous.

Shuyler Huck (1978) conducted research to determine whether the presentation of item difficulty information would assist students in earning higher scores on objective tests. Huck found that when students were presented with this information they did score higher results. However, Huck was investigating aspects of response style and concluded that the presentation of the item difficulty information exchanged the influence of one response style with that of another. Huck did not consider cognitive styles of the examinees, and it may be possible that presentation of the item difficulty information exchanged the influence of one cognitive style for that of another.

Item difficulty indices are considered by many researchers to be valid measures in assembling and analysing teacher made classroom tests, and are important because of their relationship with the test item's index of discrimination. Any test item that is extremely difficult or very easy cannot discrimimate effectively among students.

Ebel (1972) describes an experiment to determine the relationship beween the spread of test item difficulty indices and the distribution of the test scores. In analysing the results Ebel states:

> ...(there is an) inverse relationship between the spread of item difficulties and the spread of test scores. The wider the dispersion of difficulty values, the more concentrated the distribution of test scores. Note, too, the very low reliability of scores on the test composed only of very easy or very difficult items and the somewhat higher reliability of the scores when the items are concentrated near the midpoint in difficulty than when they are distrubuted in difficulty.

The question of what is an ideal item difficulty index has been addressed by several researchers. Ebel's experimental data suggests that a 0.5 difficulty index produces the highest reliability of scores, but other researchers indicate that the values should be somewhat higher. According to Sax:

> The optimal difficulty level of a test depends on the chance score and the number of items making up the test. One way of estimating optimal difficulty level is to compute the chance score and to add to that value one-half the difference between a perfect score and the chance score.

Using Sax' criterion, the optimal item difficulty index of a four distractor multiple choice test item should be approximately 0.63.

Sax (1974), suggests that where guessing can occur, far more reliable results can be obtained by making the test slightly easier. By reducing the possibility of chance guessing the test reliability is increased. Sax suggestes that for a four distractor multiple choice examination the ideal aveage difficulty index for the test should be approximately 0.74 to produce a maximally discriminating test.

Although there is considerable agreement among the authorities that item difficulty indices are important, the literature also indicates that the degree of importance has to be tempered by the application of the test. Where test scores produce a skewed distribution (either positive or negative) the items will not discriminate in the area under the curve where the scores pile up. The most accurate discrimination occurs at the tail of the distribution where few scores duplicate. Sax stated 'whether items should be difficult or easy and should discriminate or not depends upon the purpose of the testing.'

Summary

Readability is composed of three major factors, the human component, the mechanical component, and the lexical component. The component that is of concern in this research is the latter.

Many different formulae have been introduced to quantify the mechanical component of readability, and Klare (1968) outlines over 31 formulae and some of their derivations. After an extensive review of the readability literature three readability formulae were selected for use in the present study. These formulae were selected partly on the basis of their ease of use, and ease of programming into a BASIC microcomputer programme, in addition to their popularity among other researchers.

There is considerable agreement among researchers that the item difficulty indicies are a valid measure for assembling, analysing, and validating test items. Although there is also considerable agreement that the item difficulty is effected by many variables, of which the item readability is only one.

The research has shown that where students are aware of the test item difficulty they obtain higher scores on the questions of greater difficulty.

The following chapter will detail the method used to determine if an index of readability of test items could be used to predict the item difficulty index under test performance conditions.

III. Methodology and Instrumentation

For approximately 10 years, the pipetrades section of the Northern Alberta Institute of Technology has maintained multiple choice test item banks in the trades of plumbing, steamfitting and gasfitting, and faculty members have been involved in evaluating the test items through classical item analysis procedures. This has resulted in an item bank of approximately 2000 multiple choice test items being established for the gasfitting programme, and many of these test items have been subjected to classical item analysis several times.

Selection of items

For the purpose of this research 40 test items were selected from the gasfitting test item bank. All test items were selected on the basis of measuring the general training objective of the students being able to "interpret the Canadian Standards Association B 149.1 and B 149.2 Gas Codes" (Canadian Standards Association, 1976).

Because items in the test item bank have previously been subjected to classical item analysis, an item difficulty index was known for each item. The research test items were selected from the test item bank using the following criteria:

1. Each test item selected evaluated an aspect of the objective that the student would be able to interpret the gas codes.
2. The correct interpretation was contained in the gas codes.
3. Mathematical items, or test items involving any computation were rejected.

4. The predetermined item difficulty index was in the range 0.30 to 0.95 with a stratified, platykurtic distribution throughout the range.

5. Test items containing a large number of words were selected in preference to shorter items.

These criteria resulted in a selection of 40 multiple choice test items with an average predetermined item difficulty index of 0.67, and a mean word count of 71.5 words per item. The distribution of items within the predetermined item difficulty index range, and within the range of word count, is indicated in figures I and II.

The test items were compiled into a typical multiple choice examination, and administered to two different graduating classes of gasfitting apprentice students, taught by two different instructors. The examination was administered as a final graduating examination in code interpretation, and was administered within three days of the students writing the Alberta Apprenticeship Board Tradesmans qualifying examination.

Readability Analysis of the Items

A programme was written in the BASIC computer programming language for the Compucolor II microcomputer to analyse each test item for the following data:

1. Number of words contained in the item.
2. Number of sentences.
3. Average number of words per sentence.
4. Mean word length per item.
5. Number of three or more vowel words.

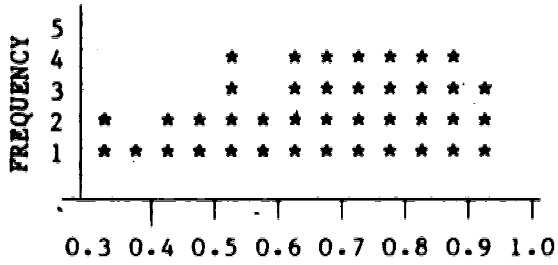

Figure I - Histogram of Predetermined Item Difficulty Indices

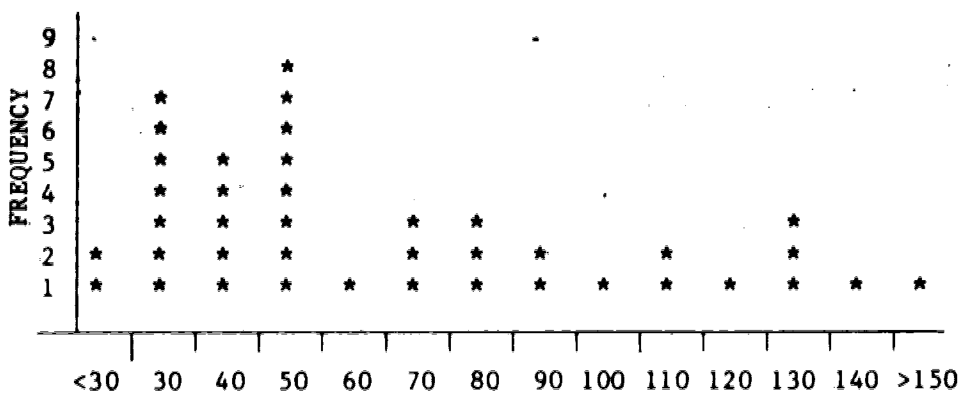

Figure II - Histogram of Number of Words in Items

6. Number of words containing nine or more characters.

7. Gunning Fog Index.

8. Automated Readability Index.

9. Reading Grade Level Index.

This BASIC programme was validated using passages of text selected from passages analysed using similar techniques, by different authors. Passages were selected from Larry Noonan (1981), Ronald Carlson (1980), and from Laubach and Koschnick (1977).

Because each of these different authors used a slightly different analysis technique, and measured a different aspect of readability, the results from each passage could not be compared exactly with the current BASIC programme. However where the programmes analysed similar traits the results from the current programme where the same.

The criteria established by Dale and Chall (1948) for counting the number of words and sentences, were used. But for the purpose of this research the following interpretations were made:

1. For counting words in each test item, abbreviations contractions and other hyphentated words were each treated as a single word. The item numbers and distractor identifiers were not counted as words, nor counted in the letter count. Numeric values were counted as a single word, and not as separate words for each individual number. Compound names and other logical combinations of words or words and numbers were treated as single words except where a blank space occured between two such logical combinations, then the space delineated a separate word.

Table I

Typical Word Count Examples

Typical Example	Word Count	Reason
51 C	1	compound name
0.25 inches	2	numeric value
Btu's	1	abbreviation
Btu hr	2	not hyphenated
a) 1000	1	distractor identifier not counted
Clause 6.15.3	2	logical combination

2. When counting the number of characters a space between words was not counted as a character. Every alphabetic or numeric character was counted. A decimal marker (.) was counted as a character, as was a slash (/), a period in a logical word or number combination the inch (") and feet (′) marker, and any other logical abbreviation, symbol or non-punctuation mark. Punctuation marks were not counted as characters in the letter count.
3. For counting sentences a period (.), semicolon (;), colon (:), question mark (?) or an exclamation mark (!) delimited a sentence. Each item stem was treated as a separate sentence, even where it requires a distractor to be appended to complete the grammatical syntax. Each distractor was counted as a complete sentence, even when it needed to be appended to an item stem to complete the grammatical syntax.
4. To determine the number of three or more vowel words, each occurence of one of the vowels (a,e,i,o and u), or of the pseudo vowel (y) was counted, except where a blend of two or more consecutive vowels occured, then only one vowel was counted.
5. Each word of nine or more letters was counted every time it occured in an item.

Mean number of words per sentence, and the mean word length was computed for each item using the data obtained, and this new data was employed to compute the Gunning Fog Index, the Automated Readability

Index, and the Reading Grade Level. The following formulae were used:

The Gunning Fog Index was calculated from:

$$GFI = (W/S + ((0.5 X (TVW + NLW)) / TNW) X 100) X 0.4$$

Where: GFI = Gunning Fog Index

W/S = Mean Words per Sentence

TNW = Total Number of Words in test item

TVW = Three Vowel Words

NLW = Nine (or more) Letter Words

This formula represents a slight modification of the original formula proposed by Gunning.

The Reading Grade level was computed from the formula:

$$GL = 0.5 X (W/S) + 4.71 X (C/W) - 21.43$$

Where: GL = Reading Grade Level index

W/S = Mean Words per Sentence

C/W = Mean Word Length

This formula is the multiple regression equation for predicting the grade reading level from the two ratios, first developed by Smith and Senter (1967).

The Automated Readability index was calcuated from the formula:

$$ARI = W/S + 9 X (C/W)$$

Where: ARI = Automated Readability Index

W/S = Mean Words per Sentence

C/W = Mean Word Length

This formula represents a simplification of the Smith and Senter Grade Reading Level formula.

Instrument Readability Data

Using the criteria outlined above the data summarized in table II was obtained for each item.

Administration of the Instrument

The research examination composed of the preselected test items, was administered to two different classes of graduating plumber/gasfitter apprentice students. The examination was administered as a final graduating examination in code interpretation, and the test subjects were not informed that the examination was being used for the purpose of this study.

Item Difficulty Analysis

The examination results obtained from both experimental classes of students were analysed using classical item analysis procedures to obtain the item difficulty index for each of the test items. The data shown in table III, indicates both the predetermined item difficulty indices, and the item difficulty indices obtained from the experimental groups.

Statistical analysis was performed on the data using the Pearson Correlation Coefficient programme available on the Statistical Package for the Social Sciences Programme (Nie, et al. 1975).

Summary

40 test items were selected from the N.A.I.T. gasfitting test item bank. Items were selected on the basis of their known item

difficulty index, and that they measured a predefined objective.

The test items were compiled into a multiple choice examination which was administered to 56 apprentice plumber/gasfitter students who were attending the final week of their training. The examination was administered as a final graduating examination, without the students being informed that the examination was being used for this study.

The test items were analysed for reading difficulty using three different readability formulae. And three indices of reading difficulty were established for each test item.

The examination papers were subjected to item analysis procedures to determine the item difficulty index for each test item. The next chapter will outline the analysis of the data obtained.

Table II

Item Readability Data

Test Item Number	Number of Words	Mean Word Length	Number of Sentences	Mean Sentence Length	Number of Long Words	Number of Three Vowel Words	Gunning Fog Index	Reading Grade Level	Automated Readability Index
1	50	4.240	6	8.333	4	3	6.133	2.707	46.493
2	75	4.253	8	9.375	2	6	5.883	3.291	47.655
3	44	5.091	6	7.333	7	10	10.660	6.215	53.151
4	86	5.023	6	14.333	8	28	14.105	9.396	59.543
5	36	3.944	6	6.000	3	6	7.400	0.148	41.500
6	137	4.737	8	17.125	14	27	12.835	9.445	59.760
7	104	4.356	7	14.857	9	17	10.943	6.514	54.059
8	138	5.348	6	23.000	29	31	17.896	15.258	71.130
9	88	5.205	6	14.667	16	22	14.667	10.417	61.507
10	47	4.660	6	7.833	3	8	7.814	4.433	49.770
11	55	4.527	6	9.167	5	7	8.030	4.477	49.912
12	54	5.463	6	9.000	11	17	13.970	8.801	58.167
13	54	5.315	7	7.714	12	14	12.715	7.460	55.584
14	59	4.441	6	9.833	5	10	9.018	4.402	49.800
15	71	4.986	6	11.833	7	17	11.494	7.970	56.707

Table II - continued

Item Readability Data

Test Item Number	Number of Words	Mean Word Length	Number of Sentences	Mean Sentence Length	Number of Long Words	Number of Three Vowel Words	Gunning Fog Index	Reading Grade Level	Automated Readability Index
16	57	4.667	7	8.143	3	4	5.713	4.621	50.143
17	113	5.239	6	18.833	25	33	17.799	12.662	65.984
18	136	4.949	7	19.429	22	33	15.859	11.591	63.965
19	123	4.870	8	15.375	12	27	12.491	9.195	59.204
20	71	4.803	7	10.143	8	8	8.564	6.236	53.368
21	149	4.698	8	18.625	19	30	14.027	10.010	60.907
22	93	5.430	7	13.286	19	20	13.701	10.789	62.157
23	27	5.445	6	4.500	1	9	9.207	6.463	53.500
24	97	4.464	7	13.857	10	22	12.141	6.524	54.032
25	32	4.594	6	5.333	3	4	6.508	2.873	46.677
26	36	4.750	6	6.000	0	6	5.733	3.943	48.750
27	116	4.991	7	16.571	19	27	14.559	10.365	61.494
28	34	5.176	6	5.667	5	10	11.090	5.785	52.255
29	17	4.588	6	2.833	2	2	5.839	1.597	44.125
30	52	4.000	6	8.667	4	9	8.467	1.743	44.667

Table II - continued

Item Readability Data

Test Item Number	Number of[a] Words	Mean Word Length	Number of Sentences	Mean Sentence Length	Number of Long Words	Number of Three Vowel Words	Gunning Fog Index	Reading Grade Level	Automated Readability Index
31	86	4.570	6	14.333	6	16	10.849	7.260	55.461
32	53	3.981	6	8.833	2	7	6.929	1.738	44.664
33	49	4.918	6	8.167	8	15	12.654	5.519	52.432
34	41	4.732	6	6.833	2	10	8.587	4.273	49.419
35	68	4.941	6	11.333	8	18	12.180	7.510	55.804
36	38	4.526	6	6.333	2	7	7.270	3.056	47.070
37	34	5.058	6	5.667	1	8	7.561	5.230	51.196
38	48	4.208	6	8.000	4	7	7.783	2.391	45.875
39	158	5.019	8	19.750	22	33	14.862	12.084	64.921
40	36	4.400	6	5.056	3	7	7.518	5.381	51.500

Table III

Test Item Difficulties for all Items

Item Number	Item Difficulty Indices: Predetermined	Experimental (Groups A & B)	Experimental (Group A only)	Experimental (Group B only)
1	0.871	0.577	0.750	0.429
2	0.524	0.115	0.000	0.214
3	0.548	0.692	0.750	0.643
4	0.790	0.442	0.292	0.571
5	0.855	0.904	0.917	0.893
6	0.774	0.769	0.875	0.679
7	0.540	0.019	0.042	0.000
8	0.847	0.654	0.708	0.607
9	0.685	0.538	0.625	0.464
10	0.742	0.308	0.417	0.214
11	0.903	0.654	0.667	0.643
12	0.581	0.557	0.542	0.607
13	0.895	0.418	0.792	0.214
14	0.573	0.423	0.417	0.429
15	0.339	0.058	0.125	0.000
16	0.476	0.346	0.458	0.250
17	0.444	0.519	0.458	0.571
18	0.605	0.36[illegible]	0.667	0.107
19	0.839	0.577	0.750	0.429
20	0.911	0.692	0.833	0.571
21	0.685	0.365	0.417	0.321
22	0.605	0.288	0.333	0.250
23	0.395	0.212	0.125	0.286
24	0.605	0.288	0.417	0.179
25	0.887	0.673	0.792	0.571
26	0.468	0.288	0.292	0.286
27	0.315	0.250	0.292	0.214
28	0.847	0.692	0.708	0.679
29	0.798	0.558	0.708	0.429
30	0.734	0.615	0.583	0.643
31	0.613	0.269	0.417	0.143
32	0.439	0.442	0.667	0.250
33	0.715	0.538	0.542	0.536
34	0.795	0.538	0.667	0.429
35	0.683	0.481	0.792	0.214
36	0.810	0.731	0.875	0.607
37	0.500	0.173	0.208	0.143
38	0.731	0.577	0.708	0.464
39	0.669	0.288	0.375	0.214
40	0.943	0.635	0.875	0.492

IV. Analysis of Results

This research sought to determine if the item readability index of multiple choice test items can be used to predict the item difficulty index of the test items.

The data therefore must be analysed to determine three different measures. First are the item difficulty indices replicable when the items are administered to different experimental groups with similar training backgrounds.

Second, to determine if the item readability indices obtained from the microcomputer BASIC programme are consistent, reliable and replicable between different measurement techniques. And finally to determine if the readability indices can be used to predict the individual test item difficulty indices.

Item Difficulty Analysis

The item difficulty index data collected from the examination results obtained from the two experimental classes of graduating gasfitting students was correlated using the S.P.S.S. Pearson Correlation Coefficient, to determine if the two classes represented similar samples.

The null hypothesis - that the correlation between the item difficulty indices of the two experimental groups is not significantly different from 0.0 would be rejected at the 0.005 level of probability. The correlation results are shown in table IV. The data indicates a high correlation between the two groups, and the null hypothesis was rejected.

Evaluation of the item difficulty indices for the joint

Table IV

Pearson Correlation Coefficient for Item Difficulty Indices

	Group A	Group B
Groups A and B Combined	0.7486 (P=0.000)	0.6806 (P=0.000)
Group A		0.6386 (P=0.000)

experimental groups indicated that the indices were within the range 0.058 to 0.731, slightly lower than the predetermined item difficulty indices range of 0.315 to 0.943. The data is summarized in table V.

Table V

Summary of Item Difficulty Indices

	Item Difficulty Indices			
	Maximum Values	Minimum Values	Mean Values	Standard Deviation
Predetermined Control Group	0.943	0.315	0.6745	0.1718
Combined Experimental Groups A & B	0.731	0.058	0.4372	0.2117
Experimental Group A	0.917	0.000	0.5469	0.2486
Experimental Group B	0.893	0.000	0.3971	0.2121

This data indicates that the experimental groups experienced more difficulty with the items than did the control groups used to predetermine the item difficulty indices.

To test the hypothesis that the item difficulty index of individual test items is replicable over time when the test items are administered to different groups of examinees, the item difficulty indices obtained from both experimental groups was correlated with the predetermined item difficulty indices obtained from the test bank.

The null hypothesis - that the correlation between the item difficulty indices of the experimental groups and the predetermined item difficulty indices is not significantly different from 0.0 would be rejected at the 0.005 level of probability. The correlation results are summarized in table VI.

Table VI

Pearson Correlation Coefficients for the Item Difficulty Indices

	Experimental Item Difficulty Index		
	Combined Groups A and B	Group A Only	Group B Only
Predetermined Item Difficulty	0.6341 (P=0.000)	0.7241 (P=0.000)	0.5698 (P=0.000)

The data indicates a high correlation between the predetermined item difficulty indices and the experimental indices, and the null hypothesis was rejected. The item difficulty indices are replicable over time.

Readability Analysis

The data obtained from the reading difficulty analysis determined from the BASIC programme written for the Compucolor II microcomputer was examined to ascertain if each of the three methods of determining the index of readability for each item produced consistent measures of readability. The null hypothesis - that the correlation between the three readability indices is not significantly different from 0.0 would be rejected at the 0.005 level of probability. Table VII summarizes the Pearson Correlation Coefficients for the three readability formulae used.

Table VII

Pearson Correlation Coefficients for the readability indices

	Reading Grade Level	Gunning Fog Level
Automated Readability Index	0.9998 (P=0.000)	0.9103 (P=0.000)
Gunning Fog Index	0.9072 (P=0.000)	

The high correlation coefficient resulted in the null hypothesis being rejected. The three readability formulae produced readability indices data that correlated very highly.

The Reading Grade Level, Gunning Fog Index and the Automated Readability Index were investigated to determine any significant trends. The data is summarized in table VIII.

Table VIII

Summary of Readability Indices Data

	Mean Value	Standard Deviation	Minimum Value	Maximum Value
Automated Readability Index	53.8577	6.8577	41.500	71.130
Reading Grade Level	6.4943	3.4922	0.148	15.258
Gunning Fog Index	10.5034	3.5537	5.713	17.896

The range of readability indices was investigated, and all items having a readability index greater than 10.0 measured by the Gunning Fog Index were studied. The data shown in table IX was developed.

Test items found to have a readability index greater than 10.0 on the Gunning Fog Index were found to be the longer items, containing on average 95 words, and with more sentences containing above average word length. The mean item difficulty index for these items was below the average for the total examination.

The Gunning Fog Index is known to inflate the readability index (Clark, 1977), however the mean grade reading level as measured by the Gunning Fog Index is at the education level required for entry into the trade (Alberta Advanced Education and Manpower, 1980).

Table IX

Summary of Readability Data for the Test Items with Gunning Fog Index Greater than 10.0

	Test Items with Gunning Fog Index Over 10.0	All Test Items
Mean Gunning Fog Index	13.405	8.939
Mean Number of words	95.14	71.55
Mean Sentence Length	14.052	10.799
Mean Word Length	4.99	4.77
Mean Item Difficulty Index	0.435	0.458

Item Difficulty Index and Readability Index

To determine if the item difficulty indices could be predicted from the item reliability indices, the readability indices data obtained from the BASIC computer programme was correlated with the predetermined item difficulty indices. The null hypothesis - that the correlation between the predetermined item difficulty indices and the readability indices is significantly different from 0.0 would be rejected at the 0.05 level of probability. The correlation data is summarized in table X.

Table X

Pearson Correlation Coefficient for Item Readability Indices and the Item Difficulty Indices

	Reading Grade Level	Automated Readability Index	Gunning Fog Index
Predetermined Item Difficulty Index	-0.1634 (P=0.314)	-0.1633 (P=0.314)	-0.0124 (P=0.939)

Because the correlation between the data was so low, the null hypothesis could not be rejected. The readability indices could not be used to predict the item difficulty indices.

Summary

The experimental data was analysed using the Pearson Correlation Coefficients programme avaliable from the Statistical Package for the Social Sciences (Nie, et al., 1975). The correlations indicate that the two experimental groups produced results on the test instrument that show the item difficulty indices were replicable between groups. Further analysis of the experimental item difficulty indices show that the item difficulties are replicable over time.

Analysis of the readability data showed a wide variation in the readability levels of the items, but the correlation between the different methods of determining the readability indices indicated stability of measurements.

The data indicates that although the item difficulty indices are replicable over time and between experimental groups, the item difficulty indices cannot be predicted from the readability indices.

V. Conclusion and Recommendations

This research has considered three major questions. First are the test item difficulty indices replicable when the items are administered to different groups of students with similar training backgrounds? Second, can an index of readability of multiple choice test items be established using a microcomputer? And third, can the index of readability be used to predict the item difficulty index of individual test items?

Conclusions

Although there are some group differences, the data has supported the thesis that the item difficulty indices are replicable when the test items are administered to different groups of examinees with similar training backgrounds.

This research has demonstrated that a microcomputer can be used to determine an item readability index for multiple choice test items, and that the three methods used to determine the readability indices produced consistent results when analyzing the same data.

This research demonstrated that under the conditions used the item readability indices cannot be used to predict the item difficulty indices.

Discussion and Implications

Although the item difficulty indices for content referenced multiple choice test items have been shown statistically to be replicable, there are group differences as evidenced by variations in both the range, and the maximum and minimum values of the item difficulty indices.

Two experimental groups were used in this research, both classes were taught by different instructors at different times of the academic year, therefore a portion of the group differences could be attributed to the different emphasis placed on the training objectives by the different instructors involved.

Some variation in the mean and in the range of the item difficulty indices between classes may be accounted for in variations in the students' different reading abilities, and in the different knowledge levels of the subject being tested, which may be compounded by the response set or cognitive style of the examinees.

Although technical jargon may be considered to inflate the index of readability, this programme demonstrated that the mean readability index of the test items used was at the grade 10 education level expected of apprentices entering the trade. The range of the indices of readability provides justification to use the BASIC programme to predetermine the reading difficulty level of items.

Test items selected for use in this research were chosen from 2000 items in the gasfitting item bank maintained by the Pipetrades section at the Northern Alberta Institute of Technology, and the predetermined item difficulty index used as part of the selection criteria was determined from the use of the items in examinations used over the period starting in 1972. These examinations have been used repeatedly during that time, and the item difficulty indices are based on the same use of the test item in the same examination used repeatedly. The use of the same test item in a different context likely changes the response set the students would be in when they attempt the test items in this research, this may also account for some variation in the group differences.

For this research the test items were selected using the criteria that the final test instrument represented the full range of item difficulty from 0.30 to 0.95 with a stratified distribution throughout this range. This is not a normal procedure to follow when designing an examination instrument, and it is possible that this deliberately stratified distribution of item difficulty indices has contributed to examinee instability and introduced some bias into the results.

The mean predetermined item difficulty index was 0.6745, slightly higher than the ideal item difficulty index of 0.5 recommended by most authorities, and slightly lower then the 0.74 recommended by Sax (1974). This item difficulty index of 0.6745 should have produced a test that would optimally discriminate between the academically good and poor students, and minimized guessing.

Investigation of the item difficulty indices achieved by the experimental groups indicate that both groups achieved mean item difficulty indices below the predetermined item difficulty indices. This indicates that the experimental groups both experienced more difficulty answering the items than did the groups used to obtain the predetermined item difficulty indices. The lower achievement results could have contributed to an increase in the amount of guessing the experimental groups performed in obtaining the answers to the items.

Analysis of the Gunning Fog Index and Reading Grade Level indices of the test instrument indicates that the mean readability level of the test instrument was at than the grade 10 education level required of all students entering the trade.

Using the Gunning Fog Index, which is known to inflate the reading grade levels (Clark, 1977), the range of readability indices for the test items is from a low of grade 5, to a high of grade 17. When all test items having a grade reading level greater than grade 10 are isolated, their mean readability level is 13.405 on the Gunning Fog Index scale, and their mean item difficulty index is 0.435. In comparison with the test means, this implies that a higher proportion of students scored correct answers on the items that were rated as more difficult to read.

Many researchers indicate that the test items' difficulty is a complex interaction of many variables, of which readability is only one contributory factor. This research has demonstrated that where the mean readability level of the test items is below the education level that can be legitimately expected of the examinees, readability of the test item contributes insignificantly to the test items' difficulty.

The reduced experimental groups item difficulty indices must be attributed to other components of item difficulty such as: students' lack of knowledge of the subject matter being tested; students' misconceptions of the material examined; or the homogeniety of the alternative distractors used in the items.

The BASIC programme used to determine the item reliability indices for the test items worked well, but is tediously slow in execution.

To be more effective the programme needs to be optimized for speed of execution, and should be modified to read the test items from disk, rather than have the teacher or instructor type in the items.

Utility of the programme could be further enhanced by merging it into the programme used to initially compose and write the original test items onto the disk. By doing so, a teacher or instructor would be able to determine the readability index of a test item before it is stored onto the disk. Additionally this procedure would allow the readability index to be stored on the disk with the test item and would thus allow test items to be selected at a future date with consideration of the readability index.

Suggestions for further research

This study has been based on a small proportion of the test items in the gasfitting test item bank maintained in the pipetreades department at the Northern Alberta Institute of Technology. It would be useful to repeat this study using other groups of items evaluating different objectives of the course.

Because this research did not select test items from the different levels or domains of the taxonomy of educational objectives, it would be valuable to repeat this study and to select the test items by the classifications in the taxonomies of educational objectives recall through synthesis and evaluation.

To determine if the stratified distribution of item difficulties used in this study did bias the test results, it would be helpful to replicate this study using a different distribution of item difficulty indices in the test instrument.

This research did not consider the relationship between indices of reliability and the reading difficulty indices, it would be useful to repeat this study to determine if the readability index can be used to predict the item reliability index of the individual test items.

This study has focused on the item difficulty indices and the readability of multiple chice test items which include the technical jargon of the trade. This study should be replicated and the technical jargon of the test items controlled for.

Summary

This research has answered the three questions posed in the introduction. It has demonstrated that the test item difficulty indices are replicable between groups with similar training and expertize, and that a microcomputer can be used to determine indices of readability of individual test items. The research has also demonstrated that where the mean readability level of the test items is below the reading ability level that can be legitimately expected of the students being tested, the readability of the test item contribute to the item difficulty index.

The BASIC computer programme used in this research has been demonstrated to be effective, although slow in execution. With optimization for speed of operation, and modification to permit items to be read from the computer disk, the BASIC programme could be used by classroom teachers to obtain accurate determination of the reading grade level of written materials and test items they produce.

REFERENCES AND BIBLIOGRAPHY

Alberta Advanced Education and Manpower. Alberta's Apprenticeship Programme 1980. Alberta Advanced Education and Manpower, 1980

Aiken, L. R. Relationship between the item difficulty and discrimination index. Educational and Psychological Measurement, 1979, 39(4).

Baker, A. What's your reading level. Interface Age, May 1981, 6(5), 27-29.

Benson, J., & Crocker, L. The effect of item format and reading ability on objective test performance: A question of validity. Educational and Psychological Measurement, 1979, 39(2), 381-387.

Berk, R. A. (Editor), Criterion Referenced Measurement The state of the Art. Baltimore: The John Hopkins University Press, 1980

Bloom, B. S., Hastings, J. T., & Madaus, G. F. Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill Book Co., 1971.

Bormuth, J. R. New developments in readability research. in J. R. Bormuth, Readability in 1968, Chicago: National Council of Teachers of English, 1968, 40-47.

Bormuth, J. R. Readability: a new approach. Reading Research Quarterly, 1966, 1(3), 79-103.

Burtt, H. E. Typography and readability. In E. Dale (editor) Readability. National Conference on Research in English, 1949.

Canadian Standards Association. Installation Code for Natural Gas Burning Appliances and Equipment, B 149.1, Ottawa: Canadian Standards Association, 1976

Canadian Standards Association. Installation Code for Propane Gas Burning Appliances and Equipment, B 149.2, Ottawa: Canadian Standards Association, 1976

Carlson, R. Reading level difficulty. Creative Computing, April 1980, 60-61.

Cawley, R. W. V., Miller, S. A., & Milligan, J. N. Cognitive style and the adult learner. Adult Education. 1976, 26(2), 101-116.

Chall, J. S. Interpretation of the results of standardized reading tests. In R. Farr. Measurement and Evaluation of Reading. New York; Harcourt Brace and World, Inc., 1970.

Clark, A. K. Readability of Recommended Textbooks in Industrial Education. Unpublished research report for the Alberta Advisory Committee for Educational Studies, 1977.

Clark, A. K. Readability in technical writing - principles and procedures. IEEE Transactions on Professional Communications, June 1975, PC-18(2).

Dale, E., & Chall, J. S. The concept of readability. In E. Dale Readability. Illinois: The National Council of Teachers of English, 1949, 1-7.

Denova, C. C. Test Construction for Training Evaluation. New York: Von Nostrand Reinhold Co., 1979.

Ebel, R. L. Essentials of Educational Measurement. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1972.

Ebel, R. L. Measuring Educational Achievement. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1965.

Ebel, R. L. Practical Problems in Educational Measurement. Massachusetts: Heath and Company, 1980.

Farr, R. Reading: What Can be Measured. Delaware: International Reading Association Research Fund, 1969.

Finley, C. J., & Berdie, F. S. The National Assessment Approach to Exercise Development. National Assessment of Educational Progress, 1970.

Fry, A. E. A readability formula that saves time. Journal of Reading, 1968, 11, 513-516.

Garrison, K. C., & Hales, L. W. Educational Psychology: An Intergration of Psychology and Educational Practices. Chas. E. Merril Publishing Co., 1972.

Gilliland, J. Readability. London: University of London Press Ltd., 1972.

Goodman, D., & Schwab, S. Computerized testing for readability. Creative Computing, April 1980, 46-51.

Gronlund, N. E. Measurement and Evaluation in Teaching. 4th. Edn. New York: Macmillan Publishing Co., Inc., 1981.

Gunderson, D. V. Reading Problems: Glossary of Terminology. Reading Research Quarterly, 1969, 9(4), 534-547.

Harvey, W. L. Learning systems for the future. Journal of Educational Technology Systems, 1978, 6(1), 3-6.

Hadlai Hull, C., & Nie, N. H. SPSS Update 7-9 New Procedures and Facilities for Release 7-9. New York: McGraw Hill Book Company, 1981.

Hopkins, K. D., & Stanley, J. C. Educational and Psychological Measurement and Evaluation. Englewood Cliffs. New Jersey: Prentice Hall, Inc., 1981.

Huck, S. W. Test Performance under the conditions of known item difficulty. Journal of Educational Measurement. 1978, 15(1), 53-58.

Hughes, P. Word Count. Interface Age. 6(8), 112; 156.

Knapp, T. R. Statistics for Educational Measurement. Scranton: Intext Educational Publishers, 1971.

Klare, G. R., & Campbell, L. R. Measuring the Readability of Highschool Newspapers. Iowa: Quill and Scroll Foundation, 1967.

Klare, G. R., Rowe, P. P., St John, M. G., & Stolurow, L. M. Automation of the Flesch reading ease readability formula with various options. Reading Research Quarterly. Summer 1969, 4, 550-559.

Klecka, W. R., Nie, N. H., & Hadlai Hull, C. SPSS Primer. New York: McGraw-Hill Book Company, Inc., 1975.

Kwolek, W. F. A readability survey of technical and popular literature. Journalism Quarterly. Summer 1973, 50, 455-264.

Lennon, R. T. What can be measured? In R. Farr, Measurement and Evaluation of Reading. New York: Harcourt Brace and World, Inc., 1970.

Laubach, R. S., & Koschnick, K. Using Readability Formulas for Easy Adult Materials. New York: New Readers Press, 1977.

Lindvall, C. M., & Nilko, A. J. Measuring Pupil Achievement and Aptitude. New York: Harcourt Brace Jovanovich Inc., 1975.

Lorge, I. Readability formulae - an evaluation. In, E. Dale, Readability. Illinois: The national Council of Teachers of English, 1949, 8-16.

Lorge, I. The Lorge Formula for Estimating Difficulty of reading Materials. New York: 1959.

Lyman, H. B. Test Scores and What They Mean. Englewood Cliffs, New Jersey: Prentice Hall Inc., 1978.

Marshall, J. C., & Hales, L. W. Classroom Test Construction. Reading, Massachusetts: Addison-Wesley Publishing Company, 1971.

Marshall, J. C., & Hales, L. W. Essentials of Testing. Reading, Massachusetts: Addison-Wesley Publishing Company, 1972.

Mehrens, W. A., & Ebel, R. L. Principles of Educational and Psychological Measurement. Chicago: Rand McNally and Company, 1967.

Mehrens, W. A., & Lehman, I. J. Measurement and Evaluation in Education and Psychology. New York: Holt Rinehart and Winston, Inc., 1973.

Milagros, R. A. The validity of the Miller-Coleman readability scale. Reading Research Quarterly. Spring 1969, 4, 342-357.

Miller, D. M. Interpreting Test Scores. New York: John Wiley and Sons, 1972.

Miller, W. R. Readability versus reading ability. The Journal of Educational Research. December 1962, 56(4), 205-208.

Murtuza, V. R. Applying Norm Referenced and Criterion Referenced Measurement in Education. Boston: Allyn and Bacon Inc., 1977.

Myers, C. T. The relationship between item difficulty and test validity and reliability. Educational and Psychological Measurement. 1962, 22, 565-571.

Nie, N. H., Hadlai Hull, c., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. SPSS Statistical Package for the Social Sciences. New York: McGraw Hill Book Company, 1975.

Noll, V. H., Scannell, D. P., & Craig, R. C. Introduction to Educational Measurement. Boston: Houghton Mifflin Company, 1979.

Noonan, L. Reading level: determination and evaluation. Creative Computing, March 1981. 166-173.

Norris, R. G. Characteristics of Adults that Facilitate and/or Interfere with Learning. Florida: The Florida State University. 1977. (ERIC Document Reproduction Service No. ED 150 442)

Oosterhof, A. C. Similarity of various item discrimination indices. Journal of Educational Measurement. 1976, 13(2), 145-149.

Parry, R. R. Mind Your P's and Q's. Kilobaud Microcomputing, June 1981, 5(6), 58-67.

Popham, W. J. Criterion Referenced Measurement. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1978.

Popham W. J. Criterion Referenced Measurement, An Introduction. Englewood Cliffs, New Jersey: Educational Technology Publications, 1971.

Popham, W. J. Modern Educational Measurement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1981.

Powers, R. D., Summer, W. A. & Kearl, B. E. A recalculation of four adult readability formulas. Journal of Educational Psychology, April 1958, 49, 99-105.

Pyrezak, F. Validity of discrimination index as a measure of item quality. Journal of Educational Measurement, 1973, 10(3), 227-231.

Quereshi, M. Y., & Fisher, T. L. Logical Versus empirical estimates of item difficulty. Educational and Psychological Measurement, 1977, 37(1), 91-100.

Rankin, E. F. The Cloze procedure - its validity and utility. In, R. Farr, Measurement and Evaluation of Reading. New York: Harcourt Brace and World, Inc., 1970, 237-253.

Rindler, S. E. Pitfalls in assessing test speededness. Journal of Educational Measurement, 1979, 16(4), 261-270.

Rowley. G. L. Which examinees are most favoured by the use of multiple choice tests. Journal of Educational Measurement, 1974, 11(1)

Sax, G. Principles of Educational Measurement and Evaluation. California: Wadsworth Publishing Company Inc., 1974.

Sherrifs, A. C. & Boomer, D. S. Who is penalized by the penalty for guessing. Journal of Educational Psychology, February 1954. 81-90.

Shields. M. E. An Item Analysis Method for Evaluating Competency Based Test Items. Masters of Education Thesis, University of Alberta. 1980.

Smith, F. M., & Adams, S. Educational Measurement for the Classroom Teacher. New York: Harper and Row, 1972.

Smith, E. A. & Senter, R. J. Automated Readability Index. Aerospace Medical and Research Laboratories, 1967.

Stodala, Q., & Stordahl, K. Basic Educational Tests and Measurement. Chicago: Science Research Associates, Inc., 1967.

Storey, A. G. The Measurement of Classroom Learning. Chicago: Science Research Associates, Inc., 1970.

Swain, S. L. & Bell, D. A. Applications of indices of readability to highschool vocational material. Canadian Vocational Journal, April 1976, 12(1), 23-25.

Thorndike, R. L. (Editor) Educational Measurement. Washington: American Council on Education, 1971.

Thorndike, R. L., & Hagen, E. Measurement and Evaluation in Psychology and Education. New York: John Wiley and Sons, Inc., 1977.

Torgerson, T. L., & Adams, G. S. Measurement and Evaluation. New York: Holt Rinehart and Winston, 1960.

Tyler, L. E. Tests and Measurements. Englewood Cliffs, New Jersey: Prentice Hall, 1971.

Williams, R. L. Sentence construction with a computer. Creative Computing, April 1980, 52-58.

Woods, D. A. Test Construction. Columbus, Ohio: Charles E. Merrill Books, Inc., 1960.

APPENDICES

APPENDIX 1

TYPICAL TEST ITEMS ANALYSED

IN THIS RESEARCH

5) A GAS RANGE MAY BE CONNECTED TO THE GAS PIPING BY MEANS OF A FLEXIBLE METAL CONNECTOR PROVIDED THE FLEXIBLE METAL CONNECTOR IS NOT LONGER THAN.

A) 2 FEET.
B) 3 FEET.
C) 4 FEET.
D) 5 FEET.
E) 6 FEET.

PASSAGE DIFFICULTY ANALYSIS:-

GRADE LEVEL....................................	.148333
AUTOMATED READABILITY INDEX....................	41.5
RUDOLPH FLESCH INDEX...........................	87.94
MURMUR DIPHTHONGS (AR,IR,ER,OR)................	3
VOWEL DIGRAPHS (AI,AY,EE,EA,OA)................	7
CONSONANT DIGRAPHS (SH,TH,WR,WH,CH,CK,KN,NG)...	6
R-BLENDS (BR,CR,DR,FR,GR,PR,TR)................	1
L-BLENDS (BL,CL,FL,GL,PL)......................	4
NUMBER OF 9 OR MORE LETTER WORDS...............	3
NUMBER OF 3 OR MORE VOWEL WORDS................	6
NUMBER OF SENTENCES............................	6
NUMBER OF WORDS................................	36
NUMBER OF LETTERS..............................	142
AVERAGE WORDS PER SENTENCE.....................	6
AVERAGE WORD LENGTH............................	3.94444
READING LEVEL - GUNNING FOG INDEX..............	7.4
TOTAL VOWELS (INCLUDING DOUBLES)...............	55
VOWELS (DOUBLES COUNT AS 1)....................	48

NUMBER OF WORDS		WORD LENGTH
7	....	1
5	....	2
6	....	3
6	....	4
4	....	5
2	....	6
3	....	8
3	....	9

WHEN INSTALLING VENT CONNECTORS WE AVOID UNNECESSARY TURNS AND OTHER CONSTRUCTION FEATURES BECAUSE WE USUALLY.

A) WISH TO MINIMIZE CONDENSATION IN THE VENT CONNECTOR.

B) WANT TO MAKE THE INSTALLATION LOOK GOOD.

C) MUST KEEP FRICTIONAL RESISTANCE TO A MAXIMUM.

D) WANT TO ENTRAIN SUFFICIENT AIR.

E) HAVE A LIMITED STACK ACTION IN THE VENT TO OVERCOME OBSTRUCTIVE FORCES.

PASSAGE DIFFICULTY ANALYSIS:-

GRADE LEVEL....................................	8.80056
AUTOMATED READABILITY INDEX....................	58.1667
RUDOLPH FLESCH INDEX...........................	33.2
REGULAR DIPHTHONGS (OU,OW,OI,OY,EW)............	1
MURMUR DIPHTHONGS (AR,IR,ER,OR)................	7
VOWEL DIGRAPHS (AI,AY,EE,EA,OA)................	4
CONSONANT DIGRAPHS (SH,TH,WR,WH,CH,CK,KN,NG)...	8
R-BLENDS (BR,CR,DR,FR,GR,PR,TR)................	4
S-BLENDS (SC,SK,SL,SM,SN,SP,ST,SW).............	7
OTHER BLENDS (PT,DW,TW,MP,FT,NK,NT)............	7
3 LETTER BLENDS (SQU,SHR,SPR,STR,THR,SPL,SCR)..	2
NUMBER OF 9 OR MORE LETTER WORDS...............	11
NUMBER OF 3 OR MORE VOWEL WORDS................	17
NUMBER OF SENTENCES............................	6
NUMBER OF WORDS................................	54
NUMBER OF LETTERS..............................	295
AVERAGE WORDS PER SENTENCE.....................	9
AVERAGE WORD LENGTH............................	5.46296
READING LEVEL - GUNNING FOG INDEX..............	13.9704
TOTAL VOWELS (INCLUDING DOUBLES)...............	120
VOWELS (DOUBLES COUNT AS 1)....................	105

NUMBER OF WORDS		WORD LENGTH
2	....	1
9	....	2
5	....	3
13	....	4
4	....	5
2	....	6
5	....	7
3	....	8
1	....	9
5	....	10
2	....	11
3	....	12

APPENDIX 2

COMPUCOLOR II BASIC PROGRAMME LISTING

FOR DETERMINING READABILITY

```
0 GOTO  65000
10 REM  HERE WE START THE MAIN PROGRAMME
100 CLEAR  10000:DIM  V(30),QL$(100),LG(30)
110 PLOT  12:Y$= "*":GOSUB  10000
120 PLOT  3,21,10,6,3,14:PRINT "R E A D I N G   L E V E L"
130 PLOT  3,13,15:PRINT "D I F F I C U L T Y   E V A L U A T I O N"
140 PLOT  3,32,19,6,1,15:PRINT "BY":PLOT  3,27,22:PRINT "A. E. DOUG
HTY"
150 Y$= "PLEASE PRESS SPACEBAR TO CONTINUE":Z$= " ":GOSUB  10100
160 PRINT :PLOT  12,3,4,0,14,6,5:PRINT "READING LEVEL DIFFICULTY EV
ALUATION  BY  A. E. DOUGHTY.
165 PLOT  15,3,9,3,6,1:PRINT "PLEASE FOLLOW THESE INSTRUCTIONS EXAC
TLY !!!"
170 PLOT  6,3,2,0,0,242,0,107,127,107,127,0,0,0,255
180 X= 1:Y= 6:H= 22:W= 62:C= 1
190 GOSUB  64000:PLOT  27,24
200 PLOT  6,6:Y$= "
            ":IF  ZX= 100 GOTO  350
210 GOSUB  9000:PRINT "TYPE YOUR DATA TO BE ANALYSED - ONE LINE AT
A TIME         ":GOSUB  9000:PRINT Y$
220 GOSUB  9000:PRINT "DON'T ALLOW YOUR TYPED LINES TO EXTEND PAST
THIS BORDER --->":PLOT  6,6:GOSUB  9000:PRINT Y$
230 GOSUB  9000:PRINT "DON'T SPLIT OR HYPHENATE WORDS ON THE SCREEN
            ":GOSUB  9000:PRINT Y$
240 GOSUB  9000:PRINT "TERMINATE EACH LINE BY DEPRESSING THE ENTER
KEY            ":GOSUB  9000:PRINT Y$
250 GOSUB  9000:PRINT "END THE STEM OF QUESTIONS WITH EITHER  .  ;
 ?  OR  !       ":GOSUB  9000:PRINT Y$
260 GOSUB  9000:PRINT "DISTRACTORS SHOULD BE IDENTIFIED WITH A RIGH
T ELLIPSE  )  ":GOSUB  9000:PRINT Y$
270 GOSUB  9000:PRINT "NOT WITH ANY OTHER PUNCTUATION.
            ":GOSUB  9000:PRINT Y$
280 GOSUB  9000:PRINT "DELETE ALL OTHER PUNCTUATION I.E.  ";CHR$ (3
4);"  '  ,  :  ETC.       ":GOSUB  9000:PRINT Y$
290 GOSUB  9000:PRINT "USE  /  AS A DECIMAL POINT, NOT THE  .  I.E.
 2/5 MEANS 2.5 ":GOSUB  9000:PRINT Y$
300 GOSUB  9000:PRINT "END EACH SENTENCE AND DISTRACTOR WITH A  .
AND A RETURN  ":GOSUB  9000:PRINT Y$
310 GOSUB  9000:PRINT "TYPE A LINE CONTAINING ONLY  ][  TO TERMIN
ATE DATA INPUT  ":GOSUB 9000:PRINT Y$
320 H= 24:Y$= "PLEASE PRESS SPACEBAR TO CONTINUE":GOSUB  10100
330 PRINT :PLOT  12,3,4,0,14,6,5:PRINT "READING LEVEL DIFFICULTY EV
ALUATION  BY  A. E. DOUGHTY.
335 PLOT  6,3,2,0,0,242,0,107,127,107,127,0,0,0,255
340 PLOT 15
350 PRINT :PLOT  3,3,3,6,3:PRINT "TYPE YOUR DATA TO BE ANALYSED  -
 ONE LINE AT A TIME NOW:-"
360 H= 22:PLOT  6,2:GOSUB  9000:INPUT "  ";QL$(0):PLOT  3,2,27:PRIN
T "0":GOSUB  9000
362 REM  *** CONTINUED ***
```

```
362 REM   *** CONTINUED ***
365 IF  RIGHT$ (QL$(0),1)< > "?"AND RIGHT$ (QL$(0),1)< > "!"AND
RIGHT$ (QL$(0),1)< > ";"AND RIGHT$ (QL$(0),1)< > "." THEN  QL$(0
)= QL$(0)+ " "
370 FOR K= 1TO 100:PRINT :PLOT  3,1,3,6,3,6,7* RND (1)+ 1:PRINT
"REMEMBER TO INPUT A LINE ENDING WITH ][ TO END YOUR DATA INPUT"
380 PLOT 6,2:GOSUB  9000:INPUT "  ";QL$(K):IF  QL$(K)= "][" THEN
 V(25)= K- 1:K= 150:GOTO  400
390 PLOT  3,1,27:PRINT K:GOSUB  9000
395 IF  RIGHT$ (QL$(K),1)< > "?"AND RIGHT$ (QL$(K),1)< > "!"AND
RIGHT$ (QL$(K),1)< > ";"AND RIGHT$ (QL$(K),1)< > "." THEN  QL$(K
)= QL$(K)+ " "
400 NEXT :PLOT  12,3,5,1,14,6,3:PRINT "P A S S A G E   D I F F I
 C U L T Y   A N A L Y S I S":PLOT  15
410 PLOT  2,0,115,242,127,115,255
420 PLOT  3,24,6,31,6,5:PRINT "CURRENTLY ANALYSING:-":PLOT  15,6
,6
430 PLOT  3,29,9:PRINT "- DIPHTHONGS":PLOT  3,27,11:PRINT "- MUR
MUR DIPHTHONGS":PLOT  3,27,13:PRINT "- VOWEL DIGRAPHS"
440 PLOT  3,24,15:PRINT "- CONSONANT DIGRAPHS":PLOT  3,29,17:PRI
NT "R - BLENDS":PLOT  3,29,19:PRINT "S - BLENDS"
450 PLOT  3,29,21:PRINT "L - BLENDS":PLOT  3,28,23:PRINT "- OTHE
R BLENDS"
460 PLOT  3,14,25:PRINT "- 3 LETTER BLENDS":PLOT  3,42,25:PRINT
"- TRIGRAPHS"
470 PLOT  3,7,27:PRINT "- WORDS,        - SENTENCES,       - 9 OR
MORE LETTER WORDS"
500 PLOT  6,1:J= 0:L= 1:V= 0
510 FOR  K= 1TO (LEN (QL$(J))- 1):A$= MID$ (QL$(J),K,2):PLOT  3,
47,6:PRINT A$:GOSUB  20000
520 NEXT  K:M$=  RIGHT$ (QL$(J),1)
550 J= J+ 1:L= 1:V= 0:IF  J= <  V(25) GOTO  510
600 FOR  J= 0TO V(25):FOR  K= 1TO (LEN (QL$(J))- 2):A$= MID$ (QL
$(J),K,3):PLOT  3,47,6:PRINT A$:GOSUB  30000
610 NEXT  K,J
1400 REM  HERE WE CALCULATE THE DIFFICULTY LEVEL
1410 XX= ((V(17)+ V(18))/ 2):X= (XX/ V(20))* 100
1420 PLOT  3,2,6,14,6,3:PRINT "DIFFICULTY INDEX FOR THIS PASSAGE
 IS APPROXIMATELY -":PLOT  15
1430 V(24)= .4* (X+ (V(20)/ V(19))):PLOT  3,55,6,14:PRINT V(24):
PLOT  15:GOSUB 10100
1500 PRINT :PLOT  12,3,5,1,14,6,3:PRINT "P A S S A G E   D I F F
 I C U L T Y   A N A L Y S I S":PLOT  15
1510 PLOT  2,0,115,242,127,115,255:V(23)= V(21)/ V(20):V(22)= V(
20)/ V(19)
1520 PLOT  3,21,6,6,3:PRINT "# OF WORDS  # OF LETTERS":PLOT 6,6
1530 FOR  K= 1TO 18:PLOT 3,25,7+ K:PRINT LG(K);SPC( 11);K:NEXT K
1540 PLOT  3,1,27,6,1:PRINT "AVERAGE LETTERS/WORD";TAB( 24);V(23
);TAB( 35);"AVERAGE WORDS/SENTENCE";TAB( 58);V(22)
1550 GOSUB  10100
1555 REM   *** CONTINUED ***
```

```
1555 REM   *** CONTINUED ***
1560 V(2)= V(22)+ (9* V(23))
1570 V(1)= (0.5* V(22))+ (4.71* V(23))- 21.43
1580 V(3)= 206.835- ((V(22)* 1.015)+ (((V(27)/ V(20))* 100)* 0.846)
)::V(3)= (INT ((V(3)* 100)+ 0.5))/ 100
2000 PLOT 12:PRINT :PLOT  3,5,15:PRINT "SET THE PRINTER - PRESS RET
URN WHEN READY...";:INPUT  ZX$
2005 POKE  33289,115:TMP= PEEK (33265):PLOT  27,18,3:PLOT 27,13
2010 FOR K= 0TO V(25):PRINT QL$(K):NEXT K
2020 PRINT
2030 PRINT "PASSAGE DIFFICULTY ANALYSIS:-":PRINT
2040 PRINT "GRADE LEVEL....................................";V(1)
2050 PRINT "AUTOMATED READABILITY INDEX...................";V(2)
2060 PRINT "RUDOLPH FLESCH INDEX..........................";V(3)
2080 IF  V(5)< > 0 THEN  PRINT "REGULAR DIPHTHONGS (OU,OW,OI,OY,EW)
............";V(5)
2090 IF  V(6)< >  0 THEN  PRINT "MURMUR DIPHTHONGS (AR,IR,ER,OR)...
.............";V(6)
2100 IF  V(7)< > 0 THEN  PRINT "VOWEL DIGRAPHS (AI,AY,EE,EA,OA)....
............";V(7)
2110 IF  V(8)< >  0 THEN  PRINT "CONSONANT DIGRAPHS (SH,TH,WR,WH,CH
,CK,KN,NG)...";V(8)
2120 IF  V(9)< >  0 THEN  PRINT "R-BLENDS (BR,CR,DR,FR,GR,PR,TR)...
.............";V(9)
2130 IF  V(10)< >  0 THEN  PRINT "S-BLENDS (SC,SK,SL,SM,SN,SP,ST,SW
)............";V(10)
2140 IF  V(11)< > 0 THEN  PRINT "L-BLENDS (BL,CL,FL,GL,PL).........
.............";V(11)
2150 IF  V(12)< > 0 THEN  PRINT "OTHER BLENDS (PT,DW,TW,MP,FT,NK,NT
)...........";V(12)
2160 IF  V(13) < >  0 THEN  PRINT "3 LETTER BLENDS (SQU,SHR,SPR,STR
,THR,SPL,SCR)..";V(13)
2170 IF  V(14)< > 0 THEN  PRINT "NUMBER OF TRIGRAPHS (EAU,IGH).....
.............";V(14)
2180 PRINT "NUMBER OF 9 OR MORE LETTER WORDS................";V(17)
2190 PRINT "NUMBER OF 3 OR MORE VOWEL WORDS................";V(18)
2200 PRINT "NUMBER OF SENTENCES...........................";V(19)
2210 PRINT "NUMBER OF WORDS...............................";V(20)
2220 PRINT "NUMBER OF LETTERS.............................";V(21)
2230 PRINT "AVERAGE WORDS PER SENTENCE....................";V(22)
2240 PRINT "AVERAGE WORD LENGTH...........................";V(23)
2250 PRINT "READING LEVEL - GUNNING FOG INDEX..............";V(24)
2252 PRINT "TOTAL VOWELS (INCLUDING DOUBLES)...............";V(26)
2255 PRINT "VOWELS (DOUBLES COUNT AS 1)...................";V(27)
2260 PRINT :PRINT "NUMBER OF WORDS          WORD LENGTH"
2270 FOR K= 1TO 20:IF  LG(K)< > 0 THEN  PRINT "    ";LG(K);TAB( 7);
"    ...............    ";K
2280 NEXT K:PLOT 12
7000 POKE  33265,TMP
7500 PLOT  12
7505 REM   *** CONTINUED ***
```

```
7505 REM    *** CONTINUED ***
7510 PLOT  6,7* RND (1)+ 1:PLOT  3,10,5:PRINT "DEPRESS   A <RETURN>
   FOR ANOTHER PRINTOUT"
7520 PLOT  6,7* RND (1)+ 1:PLOT  3,10,8:PRINT "OR DEPRESS   I <RETU
RN>  FOR ANOTHER INPUT"
7530 INPUT  VW$
7540 IF  VW$= "A" GOTO  2000
7550 IF  VW$= "I" GOTO  8998
7560 GOTO  7500
8998 ZX= 100:FOR  K= OTO 30:V(K)= 0:LG(K)= 0:NEXT K:GOTO  330
8999 END
9000 X= CALL (0)
9010 PLOT  3,1,Y+ H- 1:PRINT "
                                "
9020 PLOT  3,X- 1,Y+ H- 1
9030 RETURN
9040 END
10000 REM  HERE WE DRAW A BORDER
10010 PLOT  12,6,7* RND (1)+ 1:FOR  K= 1TO  7:Y$= Y$+ Y$:NEXT  K:Y$
= LEFT$ (Y$,64)
10020 PLOT  3,0,2:FOR  K= 1TO  31:PLOT  28:PRINT  TAB( 32- K)LEFT$
(Y$,2* K):NEXT K
10030 X$= LEFT$ (Y$,1):FOR  K= 1TO  29:PRINT " "X$SPC( 60)X$:NEXT K
10040 FOR  K= 1TO 31:PLOT  28:PRINT  " "LEFT$ (Y$,K)SPC( 62- K* 2)L
EFT$ (Y$,K):NEXT K
10050 PLOT  3,2,2,18:RETURN
10060 END
10100 REM  HERE WE HAVE A SPACEBAR SUBROUTINE
10110 PLOT  6,7* RND (1)+ 1:PLOT  3,(64- LEN (Y$))/ 2,29,6,5:PRINT
Y$;
10120 POKE  33278,0:POKE  33265,14:OUT  8,255:POKE  33279,1
10130 IF  PEEK (33278)< > ASC (Z$) THEN  GOTO  10120
10140 POKE  33265,0:PLOT  3,0,0:RETURN
10150 END
20000 REM  HERE WE COUNT WORDS ETC
20100 IF A$= ") " THEN  L= 0:V= 0:RETURN
20500 IF  LEFT$ (A$,1)= " " THEN  L= 1:V= 0:GOTO  20600
20510 IF  RIGHT$ (A$,1)= ")" THEN  L= 1:V= 0:RETURN
20530 IF  RIGHT$ (A$,1)= "?"THEN  20590
20540 IF  RIGHT$ (A$,1)= "!"THEN  20590
20550 IF  RIGHT$ (A$,1)= ";"THEN  20590
20560 IF  RIGHT$ (A$,1)= "."THEN  20590
20570 V(21)= V(21)+ 1:IF  RIGHT$ (A$,1)= "." THEN  V(20)= V(20)+ 1:
PLOT  3,2,27:PRINT V(20):LG(L)= LG(L)+ 1:L= 0:GOTO  20600
20580 L= L+ 1:IF  L= 9 THEN  PLOT  3,36,27:V(17)= V(17)+ 1:PRINT V(
17)
20585 GOTO 20600
20590 V(21)= V(21)+ 1:V(19)= V(19)+ 1:V(20)= V(20)+ 1:PLOT 3,2,27:P
RINT V(20):PLOT 3,18,27:PRINT V(19):LG(L)= LG(L)+ 1:L= 0:V= 0:GOTO
 20600
20595 REM   *** CONTINUED ***
```

```
20595 REM   *** CONTINUED ***
20600 REM  HERE WE CHECK FOR VOWELS
20610 IF  LEFT$ (A$,1)= "A" THEN V(26)= V(26)+ 1:V(27)= V(27)+ 1:GO
TO 21300
20620 IF  LEFT$ (A$,1)= "E" THEN  V(26)= V(26)+ 1:V(27)= V(27)+ 1:G
OTO 21300
20630 IF  LEFT$ (A$,1)= "I" THEN  V(26)= V(26)+ 1:V(27)= V(27)+ 1:G
OTO 21300
20640 IF  LEFT$ (A$,1)= "O" THEN  V(26)= V(26)+ 1:V(27)= V(27)+ 1:G
OTO 21300
20650 IF  LEFT$ (A$,1)= "U" THEN  V(26)= V(26)+ 1:V(27)= V(27)+ 1:G
OTO 21300
20655 IF  LEFT$ (A$,1)= "Y" THEN  V(26)= V(26)+ 1:V(27)= V(27)+ 1:G
OTO 21300
20660 REM  HERE WE SEARCH FOR REGULAR DIPHTHONGS
20670 IF  A$ =  "OU" THEN  V(5)= V(5)+ 1:V(4)= V(4)+ 1:PLOT  3,25,9
:PRINT  V(5):RETURN
20680 IF  A$ =  "OW" THEN  V(5)= V(5)+ 1:V(4)= V(4)+ 1:PLOT  3,25,9
:PRINT  V(5):RETURN
20690 IF  A$ =  "OI" THEN  V(5)= V(5)+ 1:V(4)= V(4)+ 1:PLOT  3,25,9
:PRINT  V(5):RETURN
20700 IF  A$ =  "OY" THEN  V(5)= V(5)+ 1:V(4)= V(4)+ 1:PLOT  3,25,9
:PRINT  V(5):RETURN
20710 IF  A$ =  "EW" THEN  V(5)= V(5)+ 1:V(4)= V(4)+ 1:PLOT  3,25,9
:PRINT  V(5):RETURN
20730 REM  HERE WE SEARCH FOR MURMUR DIPHTHONGS
20740 IF  A$ = "AR" THEN  V(6)= V(6)+ 1:V(4)= V(4)+ 1:PLOT  3,23,11
:PRINT V(6):RETURN
20750 IF  A$ = "IR" THEN  V(6)= V(6)+ 1:V(4)= V(4)+ 1:PLOT  3,23,11
:PRINT V(6):RETURN
20760 IF  A$ = "ER" THEN  V(6)= V(6)+ 1:V(4)= V(4)+ 1:PLOT  3,23,11
:PRINT V(6):RETURN
20770 IF  A$ = "OR" THEN  V(6)= V(6)+ 1:V(4)= V(4)+ 1:PLOT  3,23,11
:PRINT V(6):RETURN
20780 REM  HERE WE SEARCH FOR VOWEL DIPHTHONGS
20790 IF  A$ = "AI" THEN  V(7)= V(7)+ 1:V(4)= V(4)+ 1:PLOT  3,23,13
:PRINT V(7):RETURN
20800 IF  A$ = "AY" THEN  V(7)= V(7)+ 1:V(4)= V(4)+ 1:PLOT  3,23,13
:PRINT V(7):RETURN
20810 IF  A$ = "EE" THEN  V(7)= V(7)+ 1:V(4)= V(4)+ 1:PLOT  3,23,13
:PRINT V(7):RETURN
20820 IF  A$ = "EA" THEN  V(7)= V(7)+ 1:V(4)= V(4)+ 1:PLOT  3,23,13
:PRINT V(7):RETURN
20830 IF  A$ = "OA" THEN  V(7)= V(7)+ 1:V(4)= V(4)+ 1:PLOT  3,23,13
:PRINT V(7):RETURN
20840 REM  HERE WE SEARCH FOR CONSONANT DIPHTHONGS
20850 IF  A$ = "SH" THEN  V(8)= V(8)+ 1:V(4)= V(4)+ 1:PLOT  3,20,15
:PRINT V(8):RETURN
20860 IF  A$ = "TH" THEN  V(8)= V(8)+ 1:V(4)= V(4)+ 1:PLOT  3,20,15
:PRINT V(8):RETURN
20865 REM   *** CONTINUED ***
```

```
20865 REM   *** CONTINUED ***
20870 IF  A$ = "WR" THEN  V(8)= V(8)+ 1:V(4)= V(4)+ 1:PLOT  3,20,15
:PRINT V(8):RETURN
20880 IF  A$ = "WH" THEN  V(8)= V(8)+ 1:V(4)= V(4)+ 1:PLOT  3,20,15
:PRINT V(8):RETURN
20890 IF  A$ = "CH" THEN  V(8)= V(8)+ 1:V(4)= V(4)+ 1:PLOT  3,20,15
:PRINT V(8):RETURN
20900 IF  A$ = "CK" THEN  V(8)= V(8)+ 1:V(4)= V(4)+ 1:PLOT  3,20,15
:PRINT V(8):RETURN
20910 IF  A$ = "KN" THEN  V(8)= V(8)+ 1:V(4)= V(4)+ 1:PLOT  3,20,15
:PRINT V(8):RETURN
20920 IF  A$ = "NG" THEN  V(8)= V(8)+ 1:V(4)= V(4)+ 1:PLOT  3,20,15
:PRINT V(8):RETURN
20930 REM  HERE WE SEARCH FOR R - BLENDS
20940 IF  A$ = "BR" THEN  V(9)= V(9)+ 1:PLOT  3,25,17:PRINT V(9):RE
TURN
20950 IF  A$ = "CR" THEN  V(9)= V(9)+ 1:PLOT  3,25,17:PRINT V(9):RE
TURN
20960 IF  A$ = "DR" THEN  V(9)= V(9)+ 1:PLOT  3,25,17:PRINT V(9):RE
TURN
20970 IF  A$ = "FR" THEN  V(9)= V(9)+ 1:PLOT  3,25,17:PRINT V(9):RE
TURN
20980 IF  A$ = "GR" THEN  V(9)= V(9)+ 1:PLOT  3,25,17:PRINT V(9):RE
TURN
20990 IF  A$ = "PR" THEN  V(9)= V(9)+ 1:PLOT  3,25,17:PRINT V(9):RE
TURN
21000 IF  A$ = "TR" THEN  V(9)= V(9)+ 1:PLOT  3,25,17:PRINT V(9):RE
TURN
21010 REM  HERE WE SEARCH FOR S-BLENDS
21020 IF  A$ = "SC" THEN  V(10)= V(10)+ 1:PLOT  3,25,19:PRINT V(10)
:RETURN
21030 IF  A$ = "SK" THEN  V(10)= V(10)+ 1:PLOT  3,25,19:PRINT V(10)
:RETURN
21040 IF  A$ = "SL" THEN  V(10)= V(10)+ 1:PLOT  3,25,19:PRINT V(10)
:RETURN
21050 IF  A$ = "SM" THEN  V(10)= V(10)+ 1:PLOT  3,25,19:PRINT V(10)
:RETURN
21060 IF  A$ = "SN" THEN  V(10)= V(10)+ 1:PLOT  3,25,19:PRINT V(10)
:RETURN
21070 IF  A$ = "SP" THEN  V(10)= V(10)+ 1:PLOT  3,25,19:PRINT V(10)
:RETURN
21080 IF  A$ = "ST" THEN  V(10)= V(10)+ 1:PLOT  3,25,19:PRINT V(10)
:RETURN
21090 IF  A$ = "SW" THEN  V(10)= V(10)+ 1:PLOT  3,25,19:PRINT V(10)
:RETURN
21100 REM  HERE WE SEARCH FOR L-BLENDS
21110 IF  A$ = "BL" THEN  V(11)= V(11)+ 1:PLOT  3,25,21:PRINT V(11)
:RETURN
21120 IF  A$ = "CL" THEN  V(11)= V(11)+ 1:PLOT  3,25,21:PRINT V(11)
:RETURN
21125 REM   *** CONTINUED ***
```

```
21125 REM   *** CONTINUED ***
21130 IF  A$ = "FL" THEN  V(11)= V(11)+ 1:PLOT  3,25,21:PRINT V(11)
:RETURN
21140 IF  A$ = "GL" THEN  V(11)= V(11)+ 1:PLOT  3,25,21:PRINT V(11)
:RETURN
21150 IF  A$ = "PL" THEN  V(11)= V(11)+ 1:PLOT  3,25,21:PRINT V(11)
:RETURN
21160 REM  HERE WE SEARCH FOR OTHER - BLENDS
21170 IF  A$ = "PT" THEN  V(12)= V(12)+ 1:PLOT  3,24,23:PRINT V(12)
:RETURN
21180 IF  A$ = "DW" THEN  V(12)= V(12)+ 1:PLOT  3,24,23:PRINT V(12)
:RETURN
21190 IF  A$ = "TW" THEN  V(12)= V(12)+ 1:PLOT  3,24,23:PRINT V(12)
:RETURN
21200 IF  A$ = "MP" THEN  V(12)= V(12)+ 1:PLOT  3,24,23:PRINT V(12)
:RETURN
21210 IF  A$ = "FT" THEN  V(12)= V(12)+ 1:PLOT  3,24,23:PRINT V(12)
:RETURN
21220 IF  A$ = "NK" THEN  V(12)= V(12)+ 1:PLOT  3,24,23:PRINT V(12)
:RETURN
21230 IF  A$ = "NT" THEN  V(12)= V(12)+ 1:PLOT  3,24,23:PRINT V(12)
:RETURN
21240 RETURN
21300 REM  HERE WE SEARCH FOR DOUBLE VOWELS
21310 IF  RIGHT$ (A$,1)= "A" THEN  V(27)= V(27)- 1:GOTO  20670
21321 IF  RIGHT$ (A$,1)= "E" THEN  V(27)= V(27)- 1:GOTO  20670
21330 IF  RIGHT$ (A$,1)= "I" THEN  V(27)= V(27)- 1:GOTO 20670
21340 IF  RIGHT$ (A$,1)= "O" THEN  V(27)= V(27)- 1:GOTO 20670
21350 IF  RIGHT$ (A$,1)= "U" THEN  V(27)= V(27)- 1:GOTO 20670
21360 IF  RIGHT$ (A$,1)= "Y" THEN  V(27)= V(27)- 1:GOTO 20670
21370 V= V+ 1:IF  V= 3 THEN  V(18)= V(18)+ 1:V= 4
21380 GOTO  20670
30000 REM  HER WE LOOK FOR TRIGRAPHS
31250 IF  A$ = "SQU" THEN  V(13)= V(13)+ 1:PLOT  3,10,25:PRINT V(13
):RETURN
31260 IF  A$ = "SHR" THEN  V(13)= V(13)+ 1:PLOT  3,10,25:PRINT ,V(13
):RETURN
31270 IF  A$ = "SPR" THEN  V(13)= V(13)+ 1:PLOT  3,10,25:PRINT V(13
):RETURN
31280 IF  A$ = "STR" THEN  V(13)= V(13)+ 1:PLOT  3,10,25:PRINT V(13
):RETURN
31290 IF  A$ = "THR" THEN  V(13)= V(13)+ 1:PLOT  3,10,25:PRINT V(13
):RETURN
31300 IF  A$ = "SPL" THEN  V(13)= V(13)+ 1:PLOT  3,10,25:PRINT V(13
):RETURN
31310 IF  A$ = "SCR" THEN  V(13)= V(13)+ 1:PLOT  3,10,25:PRINT V(13
):RETURN
31320 IF  A$ = "EAU" THEN  V(14)= V(14)+ 1:PLOT  3,39,25:PRINT V(14
):RETURN
31330 IF  A$ = "IGH" THEN  V(14)= V(14)+ 1:PLOT  3,39,25:PRINT V(14
):RETURN
31340 RETURN
31345 REM   *** CONTINUED ***
```

```
31345 REM   *** CONTINUED ***
64000 GOSUB  65410
64010 Z= 28672+ 128* Y+ X+ X::AD= TM+ 2:GOSUB  65400
64020 Z= 128- W- W:AD= TM+ 25:GOSUB  65400
64030 POKE  TM+ 5,H- 1:POKE  TM+ 7,W* (C+ 1):POKE  TM+ 19,35* (1- C
)
64040 RETURN
65000 GOSUB  65410:RESTORE  65010
65010 DATA  33,-1,-1,6,-1,14,-1,17,128,0,25,126
65020 DATA  17,128,255,25,119,35,-1,13,194,-1,-1
65030 DATA  17,-1,-1,25,5,194,-1,-1,201
65040 IF  TM> 65503 THEN  TM= TM- 32:GOTO 65080
65050 FOR  I =  1 TO  32:READ  A
65060 IF  A> = 0 AND  A< >  PEEK (TM+ 1) THEN  I = 32:TM= TM- 32
65070 NEXT
65080 RESTORE  65010
65090 FOR  I =  1 TO  32:READ  A:POKE  TM+ I,A- (A< 0):NEXT
65100 Z= TM+ 1:AD= 33283:GOSUB  65400
65110 Z= TM:AD= ER:GOSUB  65400
65120 Z= TM+ 6:AD= TM+ 30:GOSUB  65400
65130 Z= TM+ 8:AD= TM+ 22:GOSUB  65400
65140 CLEAR  25:GOTO  10
65400 ZZ= INT (Z/ 256):POKE  AD,Z- 256* ZZ:POKE  AD+ 1,ZZ:RETURN
65410 ER= 32940:TM= 256* PEEK (ER+ 1)+ PEEK (ER):RETURN
65535 REM   *** CONCLUDED ***
```

APPENDIX 3

TYPICAL PASSAGES USED FOR VALIDATING THE BASIC COMPUTER PROGRAMME

UNCLE SAM IS THE MOST EXTENSIVE LAND OWNER IN THE COUNTRY. HE HAS UNDER HIS CONTROL ABOUT TWO HUNDRED MILLION ACRES OF VACANT LAND.
THESE VAST TRACTS ARE LARGELY DESERT LAND IT IS TRUE BUT SOME SECTIONS ARE MOUNTAINOUS SOME ARE FORESTED AND OTHER PORTIONS ARE SUITABLE FOR PASTURE LANDS.
ALL OF THIS GOVERNMENT LAND LIES OUTSIDE THE ORIGINAL THIRTEEN COLONIES AND OUTSIDE THE STATES OF OHIO INDIANA ILLINOIS TENNESSEE AND KENTUCKY.
UNCLE SAM IS DESIROUS OF HAVING THIS LAND KNOWN AS THE PUBLIC DOMAIN MADE PRODUCTIVE.
THE TASK OF PREPARING IT FOR AGRICULTURE IS GIVEN TO THE UNITED STATES RECLAMATION SERVICE.

PASSAGE DIFFICULTY ANALYSIS:-

GRADE LEVEL....................................	10.2506
AUTOMATED READABILITY INDEX....................	61.2941
RUDOLPH FLESCH INDEX...........................	37.8
REGULAR DIPHTHONGS (OU,OW,OI,OY,EW)............	10
MURMUR DIPHTHONGS (AR,IR,ER,OR)................	19
VOWEL DIGRAPHS (AI,AY,EE,EA,OA)................	4
CONSONANT DIGRAPHS (SH,TH,WR,WH,CH,CK,KN,NG)...	16
R-BLENDS (BR,CR,DR,FR,GR,PR,TR)................	9
S-BLENDS (SC,SK,SL,SM,SN,SP,ST,SW).............	7
L-BLENDS (BL,CL,FL,GL,PL)......................	5
OTHER BLENDS (PT,DW,TW,MP,FT,NK,NT)............	7
NUMBER OF 9 OR MORE LETTER WORDS...............	8
NUMBER OF 3 OR MORE VOWEL WORDS................	20
NUMBER OF SENTENCES............................	6
NUMBER OF WORDS................................	102
NUMBER OF LETTERS..............................	502
AVERAGE WORDS PER SENTENCE.....................	17
AVERAGE WORD LENGTH............................	4.92157
READING LEVEL - GUNNING FOG INDEX..............	12.2902
TOTAL VOWELS (INCLUDING DOUBLES)...............	205
VOWELS (DOUBLES COUNT AS 1)....................	183

NUMBER OF WORDS		WORD LENGTH
15	....	2
23	....	3
16	....	4
11	....	5
9	....	6
10	....	7
10	....	8
3	....	9
2	....	10
3	....	11

ONCE UPON A TIME THERE WAS A MAN NAMED CHOU WHO AFTER COMPETING FOR SEVERAL OFFICIAL APPOINTMENTS WITHOUT SUCCESS NOTICED ONE DAY THAT AS THE YEARS ADVANCED HIS HAIR WAS TURNING GREY.
WHILE WEEPING OVER HIS MISFORTUNE IN THE STREET HE WAS ASKED BY A PASSERBY TO TELL THE CAUSE OF HIS SORROW.
I HAVE NEVER ONCE SUCCEEDED IN MY OFFICIAL CAREER REPLIED HE AND NOW I AM GRIEVED TO THINK OF MY OLD AGE AND THE LOST OPPORTUNITIES.
THAT IS WHY I AM CRYING.
NEVER ONCE SUCCEEDED RETURNED THE STRANGER.
WELL WHEN A YOUTH I DEVOTED MYSELF TO LITERARY STUDIES.

PASSAGE DIFFICULTY ANALYSIS:-

GRADE LEVEL	7.64538
AUTOMATED READABILITY INDEX	56.3086
RUDOLPH FLESCH INDEX	53.22
REGULAR DIPHTHONGS (OU,OW,OI,OY,EW)	6
MURMUR DIPHTHONGS (AR,IR,ER,OR)	18
VOWEL DIGRAPHS (AI,AY,EE,EA,OA)	8
CONSONANT DIGRAPHS (SH,TH,WR,WH,CH,CK,KN,NG)	21
R-BLENDS (BR,CR,DR,FR,GR,PR,TR)	5
S-BLENDS (SC,SK,SL,SM,SN,SP,ST,SW)	5
L-BLENDS (BL,CL,FL,GL,PL)	1
OTHER BLENDS (PT,DW,TW,MP,FT,NK,NT)	5
3 LETTER BLENDS (SQU,SHR,SPR,STR,THR,SPL,SCR)	2
NUMBER OF 9 OR MORE LETTER WORDS	6
NUMBER OF 3 OR MORE VOWEL WORDS	15
NUMBER OF SENTENCES	6
NUMBER OF WORDS	101
NUMBER OF LETTERS	443
AVERAGE WORDS PER SENTENCE	16.8333
AVERAGE WORD LENGTH	4.38614
READING LEVEL - GUNNING FOG INDEX	10.8917
TOTAL VOWELS (INCLUDING DOUBLES)	186
VOWELS (DOUBLES COUNT AS 1)	163

NUMBER OF WORDS	WORD LENGTH
8	1
16	2
22	3
16	4
11	5
5	6
10	7
7	8
3	9
1	10
1	12
1	13