The use of readability formulas in health care

Article *in* Psychology Health and Medicine · February 1996

DOI: 10.1080/13548509608400003

CITATIONS
313

READS
2,511

2 authors:

ProfessorPhilip Ley
The University of Sydney
46 PUBLICATIONS **3,268** CITATIONS
SEE PROFILE

Tony Florio
UNSW Sydney
26 PUBLICATIONS **1,302** CITATIONS
SEE PROFILE

Psychology, Health & Medicine
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713441652

The use of readability formulas in health care

P. Ley [a]; T. Florio [b]
[a] University of Sydney, Australia [b] University of New South Wales, Australia

To cite this Article Ley, P. and Florio, T.(1996) 'The use of readability formulas in health care', Psychology, Health & Medicine, 1: 1, 7 — 28
To link to this Article: DOI: 10.1080/13548509608400003
URL: http://dx.doi.org/10.1080/13548509608400003

The use of readability formulas in health care

P. LEY[1] & T. FLORIO[2]
[1]*University of Sydney, &* [2]*University of New South Wales, Australia*

Abstract *Readability formulas are being increasingly used to measure the understandability of written information in clinical and health settings. This paper examines the most commonly used formulas (Dale-Chall Formula, Flesch Reading Ease, Flesch-Kincaid Formula, Fog Index, Fry Readability Graph, and SMOG Grading). Their reliability and validity when used in health-related areas are discussed, and findings resulting from their use are described. These findings show that much of the material written for patients and clients, in the areas of informed consent, illnesses and their investigation and treatment, and lifestyle advice, is too difficult for many of them to understand. It is also concluded that increasing readability usually leads to improvement in understanding and occasionally in co-operation with treatment. Finally, methods for supplementing the information gained from readability formulas are described.*

Introduction

The term 'readability' refers to the understandability of written text. This paper presents a review of the use of readability formulas in the measurement of the readability of written health-related communications, and of the findings which have emerged from their use. Readability formulas have not been the only methods used in assessing the understandability of such materials. Other methods include: (a) analysis of the vocabulary used; (b) the Cloze procedure; and (c) pre-testing of materials with *ad hoc* comprehension tests. These other methods will be briefly discussed later in this review as it is desirable, wherever possible, to use all of them in the assessment of written information.

Readability formulas have been increasingly applied to the measurement of written material in the areas of illness, treatment, investigative procedures, informed consent, behaviour therapy manuals, self-help manuals, and the wording of psychological tests and personality inventories. Many investigations have been purely descriptive. Others have assessed the effects of increased readability on understanding, recall, and compliance; and on the validity and reliability of psychological tests. There will be no attempt to summarize the literature on the readability of psychological assessment devices. To do this properly would at least double the length of this paper. However, occasional reference will be made to this literature as appropriate. Recent exemplars of the various approaches to the readability of psychological tests include the studies of Prout & Chizik (1988); Lubin *et al.* (1990); Caspi *et al.* (1992); Kaufman *et al.* (1992); and Schinka & Borum (1993).

Address for correspondence: Tony Florio, School of Psychiatry, University of New South Wales, NSW 2031, Australia. Tel: + 61 2 382 1566, Fax: + 61 2 399 6337. Email: t.florio@unsw.edu.au.

1354-8506/96/010007-22 © Journals Oxford Ltd

Construction of formulas

Readability formulas are multiple regression equations which predict the reading ability required to understand a given piece of text. The equations usually involve one or more of the following:

- average word length in syllables;
- average sentence length in words;
- proportion of common words used;
- proportion of words with three or more syllables in them;
- proportion of words which are monosyllabic.

Other factors have also been used. For example, Bormuth (1969) investigated 169 possible starters. Included in these were several sophisticated psycholinguistic variables. In general the addition of these seems to lead to little improvement over the use of the variables listed above (e.g. Wagenaar *et al.*, 1987). Indeed some very complicated psycholinguistic variables, such as the sum of Yngve word depths per sentence, turn out to correlate very highly with simpler variables, such as number of words per sentence. For example, Bormuth (1969) reported the correlation between these two particular variables to be 0.99. However, more research is needed before any definite conclusion about the usefulness of theory-based psycholinguistic variables can be drawn.

The criterion variable against which the formulas are validated is nearly always a set of passages. The passages have usually been assigned to Reading Grade Levels in one of two ways. In the case of students, the performance on a comprehension test is the normal criterion. Thus a passage will be assigned to the mean Grade at which students can correctly answer 50% (Dale-Chall); 75% (Flesch); 90% (Fog); or 100% (SMOG) of questions on a comprehension test, or obtain the equivalent score on a Cloze measure (Taylor, 1953; Bormuth, 1969). The Cloze measure involves presenting subjects with the passage to be assessed with every fifth word replaced by a blank. The score is the percentage of missing words that the subject can guess absolutely correctly (paraphrase and synonym will not do).

A commonly used validation criterion was the set of data known as the McCall-Crabbs Passages (McCall & Crabbs, 1925; 1950; 1961). The passages comprised pieces of writing about 150 words in length, each followed by 8 to 10 multiple choice questions. Performance of children in different reading grades in answering these questions was also provided. The McCall-Crabbs passages have been criticized as a defective criterion by several authors (e.g. Dale & Chall, 1948a; McLaughlin, 1969; Jacobson *et al.*, 1978; Olson, 1989). Nevertheless, formulas based on this admittedly shaky foundation correlate highly with other criteria used in formula construction. For example, in US armed forces studies, the Flesch Formula, initially validated against the McCall-Crabbs Passages, showed correlations between 0.78 and 0.92 with a criterion based on a sophisticated scaling system (Caylor *et al.*, 1973). In addition, readability formulas based on the McCall-Crabbs Passages are valid predictors of a variety of relevant criteria (see Tables 4 and 9 below).

Adult samples can be given a standardized reading test and divided into groups according to their reading ability grade, e.g. those reading at the Grade 8 level of reading ability, those reading at the Grade 9 level, and so on. Passages are presented to these groups who then complete some measure of passage comprehension. A passage is assigned to the Grade level at which the selected criterion percentage of those with a given grade of reading ability achieve the required score. Thus in the case of the Flesch-Kincaid Formula a passage was assigned to a Reading Grade Level by the following rule: 'A passage is at Reading Grade x if 50% of those who read at the Grade x level obtain a score at or above 35% correct

on a Cloze test of the passage, and this is the lowest Grade at which this is true'. (A Cloze score of 35% is approximately equal to getting 75% correct on a multiple choice comprehension test; a Cloze score of 55% is approximately equal to 90% correct on a conventional test.)

Another method of validation is to show that a formula correlates highly with established formulas. This method is open to the objection that a high correlation with another readability formula is quite compatible with much lower correlation with a comprehension criterion. For example, many established readability formulas correlate about 0.7 with the McCall-Crabbs criterion. If we have a new formula which correlates 0.9 with one of these established formulas, the correlation of the new formula with the comprehension criterion could be anywhere in the range from 0.32, which would be useless for most purposes, to 0.94, which would be excellent (Jensen, 1980).

The formulas most commonly used in health care settings

Several of the formulas to be described are included in various computer word-processing programs. If using these formulas by hand, read Klare (1974) or the original sources for more detailed directions. With the exception of the Dale-Chall formula all can be fairly easily calculated, with the SMOG Grading being easiest. The calculation of the Dale-Chall formula requires use of Dale's list of 3,000 words (Dale & Chall, 1948b). There is also a slight complication with the Fog Index in that certain long words are not counted (Gunning, 1973).

With shorter pieces of text assess the whole of the text and prorate where necessary. For example the SMOG formula uses the number of long words (words of three or more syllables) in 30 sentences as its measure. If text has only 20 sentences simply multiply the number of long words by 1.5 and enter the answer into the formula. If there are 40 sentences multiply the number of long words by 0.75 and so on. The formulas, in order of frequency of usage in the literature to be summarized in Table 8, are as follows.

Flesch Reading Ease (Flesch, 1948; 1962)

This formula was validated against the McCall-Crabbs Passages. For any passage it predicted the mean Grade Level of those who correctly answered 75% of the comprehension questions about that passage. The formula is:

Reading Ease (RE) = 206.835 − 0.846 wl − 1.015 sl

where:

wl = number of syllables per 100 words

sl = average number of words in a sentence.

RE scores can be converted to Grade equivalents by use of a table—see Table 1 below.

McLaughlin's SMOG Grading (McLaughlin, 1969)

This formula was also validated against the McCall Crabbs Passages. This time the formula predicts, for any passage, the mean Grade Level of those who correctly answer 100% of the questions. The formula is:

Reading Grade Level = $3 + \sqrt{lw_{30}}$

where:

$\sqrt{lw_{30}}$ = square root of the number of words of three or more syllables per 30

sentences (i.e. three sets of 10 consecutive sentences, one set from the beginning, one from the middle, and one from the end of the document).

McLaughlin's paper gives details of other formulas he considered. SMOG Grade was the formula recommended for the readability measurement of cancer leaflets by the National Cancer Institute (1979).

Fry Readability Graph (Fry, 1968)

The graph gives approximate reading grade level. The user has to calculate average number of sentences per 100 words, and average number of syllables per 100 words. Using these values, the Reading Grade Level is then read from the graph. The formula behind the graph was validated against text graded by comprehension tests, and by showing that the formula correlated highly with other formulas (Botel, Dale-Chall, Flesch, SRA). The correlation with comprehension test gradings was 0.93, and with other formulas ranged from 0.78 to 0.98.

Gunning Fog Index (Gunning, 1973)

The formula was partly validated against the McCall-Crabbs Passages. It predicted for any passage the average Grade of those who correctly answered 90% of questions about the passage. The formula is:

Reading Grade Level = 0.4 (sl + % lw)

where:

sl = average number of words per sentence

% lw = percentage of words which are three or more syllables in length (with some exceptions)

Flesch-Kincaid Formula (Kincaid *et al.*, 1975)

This formula was validated for US armed forces use. The passages against which the formula was validated were drawn from various armed forces manuals. The passages were scaled by use of the Cloze procedure. Personnel of known reading grade ability (as measured by a standardized psychological test) were presented with the passages with every fifth word deleted. Their task was to guess what the missing word was. The score assigned to performance was the percentage of blank spaces filled in with the exactly correct word. Passages were then assigned to a Reading Grade Level by the rule: 'A passage is at Reading Grade x if 50% of those who read at the Grade x level obtain a score at or above 35% correct on a Cloze test of the passage, and this is the lowest Grade at which this is true'. It will be recalled that a Cloze score of 35% correct is approximately equal to correctly answering 75% of the questions on a comprehension test. The formula is:

Reading Grade Level = 0.39 sl + 11.8 spw − 15.59

where:

sl = average number of words per sentence

spw = average number of syllables per word

Table 1. *The interpretation of readability formulas*

Type of writing	Reading Ease	Reading grade Level	Verbal description	Fog Index Grade estimate	% of people likely to understand
Comics	90–100	5	very easy		93
Pulp fiction	80–90	6	easy	7	91
Slick fiction	70–80	7	fairly easy	9	88
Digests	60–70	8–9	standard	10	83
Quality	50–60	10–12	fairly difficult	12	54
Academic	30–50	13–16	difficult		33
Scientific	0–30	college	very difficult graduate		4.5

The Dale-Chall Formula (Dale & Chall, 1948a; 1948b)

Once again the McCall-Crabbs Passages were used. This time the formula predicts the mean Grade Level of those who could correctly answer 50% of questions about the passage. The formula is:

Reading Grade Level = 0.1579 D + 0.496 sl + 3.6365
where
D = % of words not in Dale list of 3,000 words
sl = average number of words in a sentence.
(N.B. A table has to be used to apply a correction)

The interpretation of formula-derived readability estimates

Table 1 summarizes the guidelines (columns 1–4 and 6) provided by Flesch (1962) and (column 5) by Gunning (1973) for the interpretation of Reading Ease Scores and Fog Index Grades respectively. Note that the Fog Index Grades assigned to the different types of text are higher than those suggested for Reading Ease. This is because Flesch assigned text to the Grade at which 75% of questions were answered correctly, while Gunning used the criterion of 90% correct.

As Flesch makes clear, the percentage of the population likely to understand (shown in the final column) should be interpreted with extreme caution. It is based simply on USA Census data concerning the percentage of the adult population who have completed the number of years of education needed to bring them up to the indicated Grade level. It will be apparent that this use of years of schooling presents at least two difficulties. Firstly, the figures are now out of date. For example, the 1990 Census data show that only 10% have less than 9 years at school, and only 24% fail to complete Grade 12, while 20% obtain a college degree of some sort. The second problem (well recognized by Flesch) lies in the assumption that adults' reading ability is at the level of the highest Grade they completed at school. Detailed evidence of a gap between last school grade completed and reading ability will be presented later. This gap causes considerable problems of interpretation. For example, using data on years of schooling we might well assume that about 90% of adults nowadays can read at the 'Digests' Level (Grade 8–9). However, Kozol (1986) reports that 40% of American adults have reading ability below the Grade 9 level. Thus Flesch's estimate, given in Table 2,

Table 2. *Intra-class correlation coefficients for various combinations of formulas*

Investigation	Formulas used	Average estimated reliability of single formula	Estimated reliablity of combination of all the formulas
Morris *et al.* (1980) medication leaflets	Dale-Chall, Flesch Fog and SMOG	0.97	0.99
Kanouse *et al.* (1981) medication leaflets	Flesch, Fog and SMOG	0.74	0.89
Meade & Byrd (1989) anti-smoking pamphlets	Dale-Chall Flesch, Fog, Fry and SMOG	0.79	0.95
Klingbeil *et al.* (1994) paediatric leaflets	Fog, Fry and SMOG	0.74	0.89
All of the above	Flesch, Fog and SMOG	0.93	0.97
All of the above	Fog, Fry and SMOG	0.73	0.89
Ley (1995) individual warning, safety direction and first aid statements	Flesch, Flesch-Kincaid, Fog and SMOG	0.76	0.93
Ley (1995)	Flesch, Fog and SMOG	0.77	0.91

of 83% understanding at 9th Grade level is probably an overestimate, despite the increase in years of education since he prepared the table.

Reliability of formulas

Scorer reliability

Klare (1963) reviewed a number of studies which reported satisfactory inter-scorer reliability. More recent investigations support the earlier findings. For example, O'Farrell and Keuthen (1983) reported inter-scorer reliability of 0.99 for behaviour therapy self-help manuals, and Wells (1994) found inter-scorer reliability of 0.92 for HIV/AIDS materials. It would not really be expected that inter-scorer reliability would be low as the procedures involved are essentially mechanical in nature.

Internal consistency reliability estimates

Estimates of reliability can be obtained from calculating intra-class correlation coefficients where this is possible. The results of these calculations for a number of investigations are shown in Table 2. It can be seen that reliability estimates for the use of single formula range from 0.74 to 0.97, while the use of combinations of formulas leads to estimated reliabilities of 0.89 to 0.99.

There is also some evidence from health-related studies that the readability formulas described above give approximately equal means when applied to the same text. Table 3 summarizes evidence from studies which allow means of different readability estimates to be calculated for the same written materials. The Flesch, Fog and SMOG formulas give almost identical results when applied to the same set of texts, but the Fry formula gives a lower

Table 3. *Comparability of estimates of mean Reading Grade Level from different formulas*

	Morris *et al.* (1980); Meade & Byrd (1989) (9 pieces of text)	Morris *et al.* (1980); Kanouse *et al.* (1981); Meade & Byrd (1989) (15 pieces of text)	Morris *et al.* (1980); Kanouse *et al.* (1981); Meade & Byrd (1989); Kingbeil *et al.* (1995) (48 pieces of text)	Meade & Byrd (1989); Klingbeil *et al.* (1995) (38 pieces of text)
Dale-Chall	10.0			
Flesch	10.0	11.0		
Fog	10.4	11.3	10.6	10.4
SMOG	10.1	10.7	10.2	10.1
Fry				8.7

estimate than Fog and SMOG. However, note that samples of materials are often small, and that there is overlap between columns. Note also that these results only apply to these formulas. Large differences have sometimes been found between formulas (Stokes, 1978: Morris *et al.*, 1980).

Using a sample of 216 component statements of warnings, first aid advice, and safety directions statements, Ley (1995) reported mean Grade Level estimates of 8.7, 7.1, 8.5, and 8.0, respectively, for the Flesch, Flesch-Kincaid, Fog, and SMOG measures. The greater differences for these component statements than for the complete brochure texts above probably reflect the lower reliability and validity of the use of readability formulas at the single statement level.

Intra-text variability

Most readability measurement has involved taking a number of samples of the text being assessed. Thus, Flesch Reading Ease requires several samples of 100 words, and SMOG requires three samples of 10 sentences. From the analysis of these samples the readability of the whole text is estimated. The reliability of these estimates will depend on the adequacy of the sampling scheme used. It will also depend on the homogeneity of the text. It is not uncommon for different samples of a text to be written at different levels of difficulty. For example, in an assessment of 33 paediatric leaflets using three readability formulas (Fog, Fry, and SMOG) Klingbeil *et al.* (1994) found that, depending on the formula used, 48 to 84% of the pamphlets had intra-sample variability of three or more grades, and 9 to 43% had intra-sample variability of five or more grades. A consequence of this is that sometimes readers might not be able to understand certain parts of a text even though the text as a whole achieves an acceptable readability value. More attention needs to be paid to this problem.

Validity of formulas

Original criterion validity

Four of the formulas most often used in health-related studies were validated, at least in part, against the McCall-Crabbs Reading Passages which provided data for each passage about scores obtained by children in different school grades on tests of comprehension of that passage. The correlation between this criterion and Dale-Chall score and Flesch score was

Table 4. *Summary of validity studies reported by Klare (1963; 1976)*

Criterion	Positive	Mixed	Negative
Klare (1963)			
Comprehension/ memory	11	2	8
Increased probability of article being read	5	3	1
Judgemental	12	3	2
Reading speed	6	0	0
Klare (1976)			
Experimental raising of readability leads to increased comprehension	19	6	11

0.7 in each case, 0.59 with a slightly modified version of Gunning's Fog Index, and 0.985 with McLaughlin's SMOG. The Fry measure correlated 0.93 with passage grading based on a different comprehension test. The Cloze procedure has also been used in the assignment of criterion passages to reading grade levels, especially in military settings (e.g. Caylor *et al.*, 1973; Kincaid *et al.*, 1975). As mentioned earlier, the Flesch-Kincaid formula was validated in this way, and was found to correlate 0.7 with the Cloze-derived criterion.

Other evidence of validity from non-health-related investigations

Formulas have been validated against a variety of other criteria. The results of reviews of Klare (1963; 1976) are summarized in Table 4, and results in health-related settings will be reviewed later in this paper. Klare's 1963 review classified results into 'positive', 'negative', and 'indeterminate'. The criteria used are not fully spelled out except in the case of correlation coefficients. Thus if the correlation between a formula and criterion was 0.5 or higher the relationship was said to be positive. Below 0.5 it was said to be negative. In the 1976 review the criterion of 'positive' was a statistically significant result, while non-significant results were classed as 'negative'.

Inter-correlations amongst formulas in health-related studies

Formulas show high inter-correlations. In an investigation of the readability of four medication package inserts, Morris *et al.* (1980) found that each of 13 formulas ranked the inserts in the same order. Across a set of health-related information materials Meade and Smith (1991) reported correlations of 0.95, 0.79, 0.96, and 0.95 between the Flesch and Dale-Chall, Fog, Fry, and SMOG respectively. Dale-Chall correlated 0.74 with Fog, 0.79 with Fry, and 0.75 with SMOG. Fog correlated 0.93 with Fry and 0.99 with SMOG, while the correlation between Fry and SMOG was 0.93. Rivera *et al.* (1992) reported that for a series of sets of contraceptive instructions, the Dale-Chall correlated 0.67 with Fry, and 0.78 with SMOG, and that the correlation between Fry and SMOG was 0.84. Wells (1994) reported that scored SMOG estimates correlated 0.97 with computer estimates of the Fog Index, and 0.96 with Flesch Reading Ease, Fry, and Flesch-Kincaid. Klingbeil *et al.* (1995) found that the Fog correlated 0.92 with Fry, and 0.99 with SMOG, while the correlation between Fry

Table 5. *Distribution of absolute differences between estimates of Reading Grade Level as assessed by different formulas*

Absolute difference	Percentage of comparisons between formulas showing a difference of stated size	
	Brochures, pamphlets and leaflets (% of 171 comparisons)	Individual statements (% of 1296 comparisons)
0	40	15
1	39	28
2	17	19
3	2	15
4	1	12
5	0	6
6	1	3
7		1
8		0
9		1

and SMOG was 0.99. In an analysis of the readability of individual warnings, safety directions, and first aid instructions, Ley (1995) reported that Flesch-Kincaid correlated 0.78 with Fog, and 0.70 with SMOG, while the correlation between FOG and SMOG was = 0.92. In addition this study found that the correlation between hand-scored Flesch Reading Ease and computer-based Flesch-Kincaid was 0.91.

Differences in Grade Levels assigned by the different formulas in health-related studies

It is obviously possible that, although formulas show high inter-correlations, they assign different Grade values to a piece of text. Two data sets were used to investigate this problem. The first was that of Ley (1995). This consisted of individual statements making up health and danger warnings, safety directions and first aid instructions. Each was assessed by the Flesch, Flesch-Kincaid, Fog and SMOG formulas. The distribution of the 1,296 absolute differences between formulas was obtained. The second set was derived from those studies in the literature which gave sufficient detail for absolute differences to be calculated. This set comprised brochures, leaflets and pamphlets (Morris *et al.*, 1980; Kanouse *et al.*, 1981; Meade & Byrd, 1989; Klingbeil *et al.*, 1995). Formulas involved were Flesch, Fog, Fry, and SMOG. This analysis yielded 171 absolute differences. The results of these analyses are summarized in Table 5. As would be expected, the differences for documents are smaller than for individual statements.

Given that readability formulas can produce differing estimates of grade level, which estimate should be used if more than one is available? As most health-related materials are intended for general use, one alternative is to take the highest estimate of grade level. If this is in fact an overestimate of difficulty, little damage will be done. The other alternative is to take the average of the results of more than one formula. This will produce a more reliable estimate than a single formula (see Table 2).

Criteria of acceptable difficulty

Several authorities have suggested that text for general consumption should be written at the 6th to 8th Grade Level rather than the more usual 10th to 11th at which they are now written (e.g. Dale & O'Rourke, 1982), while some have even suggested that health educational materials should be written at the 3rd to 5th Grade level (Wells, 1994). Evidence relevant to the problem of selecting suitable criteria comes from studies of the reading ability of clinic samples, and from studies of the Flesch Reading Ease Scores of a variety of newspapers, magazines and journals.

Literacy of clinical samples

It might be thought that a person's reading ability could be assessed from number of years of education. To some extent this is true, but a gap between years of schooling and reading ability (as measured by standardized tests) has been noted in several studies of clinic and hospital samples. These are summarized in Table 6, where it can be seen that a gap of 2 to 4 years is commonly found. Attempts to estimate reading ability from years of education will overestimate reading ability.

The results summarized above for small samples might well be more generally true. Thus Doak *et al.* (1985) suggest a gap of 4 or so years between years of schooling and reading ability for the USA as a whole. It would be prudent to assume that this finding would apply to other English-speaking countries also. Thus, in estimating likely average reading ability, it would be wise to subtract 4 years from years of education. Even then, of course, a large proportion will read below this average level.

Written information for patients should be made as easy as possible. The average reading ability of the samples referred to in Table 6 ranges between 3 and 11. Obviously material for general use has to cater, as far as possible, for the lower part of this range. Other findings of these investigations show that 36 to 40% can only read at or below the 4th Grade level (Doak & Doak, 1980; Davis *et al.*, 1990); 54% below Grade 6; 77% below Grade 9 (Meade & Byrd, 1989); and 85% below Grade 10 (Doak & Doak, 1980). These data are based largely on public hospital and clinic patients, and to some extent overestimate the percentage of poor readers in the population at large. Kozol (1986) estimated that 25 million American adults (about 15%) can only read at 4th Grade level or below, and a further 40 million read between the 5th and 8th Grade levels. This is a total of about 40% of American adults reading at or below the Grade 8 level. It is unlikely that low reading ability is limited to the USA. So, either leaflets should be written at a level low enough to cater for as many of these poor readers as possible, or different clinics, hospitals, and health care professionals will need to write their own material at a suitable level for their clients.

A criterion based on Reading Ease scores—the 'tabloid' line

It has been suggested that a Reading Ease score of 70 might be a reasonable criterion to set as the absolute upper limit of acceptable difficulty for written materials for general use, and that any text with Reading Ease below 70 should be re-written (Nicoll & Harrison, 1984). These authors called this the 'tabloid line', because they noted that the content of mass

Table 6. *Some data on literacy levels and years of education*

Sample studied	Mean years at school	Mean Reading Grade on standardized test	Gap (Reading Grade minus years of education)
Welfare claimants	(a) 8	5.3	−2.7
(Bendick & Cantu, 1978)	(b) 12	7.5	−4.5
Smokers (Meade, 1988)	11	7	−4
Primary care patients (Hardie *et al.*, 1979)	10	7	−3
Primary care patients— all smokers (Meade & Byrd 1989)	10	6	−3.8
Primary care (Davis *et al.*, 1990)			
(a) university clinic	(a) 10.5	(a) 6.8	(a) −3.7
(b) community clinic	(b) 10.2	(b) 5.4	(b) −4.8
(c) private practice	(c) 13.4	(c) 10.8	(c) −2.6
Hospital patients—mostly naval and coastguard families (Doak & Doak 1980)	—	—	−4.5
American Indians—diabetics (Hosey *et al.*, 1990)	—	—	−3.0
Out-patients (Jackson *et al.*, 1994)			
(a) aged 60 +	(a) 7.3	(a) 2.9	(a) −4.4
(b) aged below 60	(b) 10.2	(b) 5.4	(b) −4.8
Cancer patients (Jubelirer *et al.*, 1994)	12.5	10.5	−2

circulation tabloid newspapers, which accounted for over 70% of newspapers sold in the UK, had Reading Ease scores of 70 or higher. This suggests that material at this level of difficulty is preferred by most people. A Reading Ease Score of 70 corresponds to a Reading Grade Level of 7. In the light of what has been said in the previous section this might be too high for many readers. However, if the average number of years of education is 12 or so, subtracting the expected 4-year gap leaves us with an estimated average population reading ability of about 8. Thus the tabloid line criterion is probably at a level below the average reading ability of the population. However, note that about 15% of the USA population cannot cope adequately with material with a Reading Ease score of 100.

In an attempt to give some feel for the meaning of Reading Ease scores, Table 7 lists some typical values drawn from Morrow (1980), Nicoll & Harrison (1984), Hammerschmidt & Keane (1992), Grossman *et al.* (1994), and Murphy *et al.* (1994). Table 7 also provides estimates of the percentage of people who would be expected to understand material at the various levels. These percentages are based on the table given in Flesch (1962) and, where available, the estimates of Kozol (1986).

Table 7. *The 'Tabloid Line' and the readability of various texts*

Reading Ease Score	Newspapers, periodicals and fiction etc.	Medical journals	Flesch's 1962 estimate of percentage likely to understand, and estimates based on Kozol's 1986 report
90–100	Spiderman Comic		93 (Kozol 85)
80–90	Hemingway's 'Snows of Kilimanjaro'		91
70–80	P. G. Wodehouse 'The Inimitable Jeeves'		88
	UK Daily Mirror		
	UK Sun		
	Better Homes and Gardens		
The 'Tabloid line'			
60–70	Newsweek		83
	New Zealand 'popular' magazines		(Kozol 60% above
	Reader's Digest		Reading Ease of 65)
	UK Daily Mail		
	Lincoln's Gettysburg Address		
50–60	Conan Doyle 'The Naval Treaty' (Sherlock Holmes)		54
	New Yorker		
	UK Guardian		
	UK Daily Telegraph		
40–50	New Zealand newspaper editorials		33
	UK Financial Times		
	Life insurance policies		
30–40		*Journal of the American Medical Association*	
30 or less	UK Times	*New England Journal of Medicine*	5
		Cancer	

The comprehensibility of health care materials as assessed by formulas

Ley (1982) and Ley & Morris (1984) reviewed investigations of the readability of written information. These reviews included studies of leaflets for welfare clients (Bendick & Cantu, 1978); leaflets concerning diabetes (Thrush & Lanese, 1962); X-ray procedures (Ley *et al.*, 1972); dental procedures (Ley, 1974); non-prescription drugs (Pyrczak & Roth, 1976); optometry and eye care (French *et al.*, 1978); prescription drugs (Liguori, 1978); health educational topics (Cole, 1979); cancer (National Cancer Institute, 1979); and consent to

surgery (Grundner, 1980). The results were analysed in terms of Reading Ease, and of the percentage of leaflets which would be estimated to be understood by various percentages of the population. Of the 170 leaflets involved in these studies, they estimated that only 15% would be understood by 75% or more of the population; 33% would be understood by 40 to 74% of people; and 68% would be understood by less than 40% of people. Further details of these studies can be found in the works cited.

Table 8 summarizes the results of several further investigations which have analysed the readability of collections of written materials intended for general use. These investigations have covered a wide range of topics. Their results reinforce the conclusions of Ley (1982) and Ley & Morris (1984) that much of the literature produced for patients, clients and the general public is too difficult. The mean of the reported Reading Grade Level means is 11.9 (s.d. 2.3). Ninety per cent of the means are above 9, and 50% above 11.

Effects of increasing readability

Table 9 summarizes results of experiments in which readability was manipulated as an independent variable. An important methodological point is worth making here. It is unrealistic to expect changes in comprehension or memory or other variables in experimental studies if either (a) the highest level of difficulty used is below the level at which subjects can read, or (b) the lowest level of difficulty used is above the level at which people can read. Ideally, therefore, there should be direct measurement of the reading ability of the subjects. This is virtually never done.

It can be seen from this table that while increased readability led to significant improvement in 22 out of the 36 comparisons listed, in no case did it lead to significant worsening of understanding, recall, or compliance.

Other methods of readability assessment

Although readability formulas have been the commonest method for assessing written materials, other methods have also been used in assessing the understandability of health information. These include:

- vocabulary analysis;
- the Cloze procedure;
- pre-testing by using *ad hoc* comprehension tests.

Vocabulary analysis

Vocabulary analysis attempts to ensure that none of the words used is too difficult for the intended audience of the written material. The most direct way of doing this is to use a vocabulary data base which provides information about the school grade level at which a given word is likely to be understood. For general use, words above the Grade 6 level might be replaced with easier words. The most commonly used data base is *The Living Word Vocabulary* (Dale & O'Rourke, 1982). This reports, for 44,000 words, the school grade at which US children and young persons were able to choose the correct definition of a given word in a multiple choice test.

Another strategy is to use words which occur with high frequency in written English text. There are several sources of such data available which are still reasonably current including

Table 8. *Studies of the readability of written information for patients concerning informed consent, illness and lifestyle*

Investigation	Topic area	Formula used	Mean Reading Grade Level (RGL) or Reading Ease Score (RE)
Informed consent			
Morrow (1980)	Informed consent—treatment	Flesch RE	RE = 41
Handelsman *et al.* (1986)	Informed consent—treatment & assessment	(a) Flesch RE (b) Fry	(a) RE = 46 (b) RGL = 12.5
Ogloff & Otto (1991)	Informed consent—health-related research		
	Adults	(a) Flesch RE (b) Fry	(a) RE = 46 (b) RGL = 15.6
	Teenagers aged 13–17	(a) Flesch RE (b) Fry	(a) RE = 50 (b) RGL = 15.2
	Children aged 6–12	(a) Flesch RE (b) Fry	(a) RE = 67 (b) RGL = 11.4
Hammerschmidt & Keane (1992)	Informed consent—research	Fry	RGL = 15
Meade & Howser (1992)	Informed consent—cancer	SMOG	RGL = 14.3
Eisenstaedt *et al.* (1993)	Informed consent—blood transfusion	SMOG	RGL = 14.6
Hopper *et al.* (1993)	Informed consent—X-ray procedures	(a) Flesch Formula (b) Flesch-Kincaid (c) Fog	Average RGL = 12.4
Ross *et al.* (1994)	Informed consent—AIDS medication	(a) Flesch RE (b) Fry	(a) RE = 17 (b) Fry = 16 +
Grossman *et al.* (1994)	Informed consent—clinical oncology	(a) Flesch RE (b) Flesch-Kincaid (c) Fog	(a) RE = 53 (b) RGL = 11.1 (c) RGL = 14.1
Murphy *et al.* (1994)	Informed consent—research	(a) Flesch RE (b) Flesch-Kincaid (c) Fog	(a) RE = 48 *Median* RGL (b) 12.1 (c) = 10.6
Hopper *et al.* (1995)	Informed consent—X-ray procedures	(a) Flesch Formula (b) Flesch-Kincaid (c) Fry	Average RGL (a) clinical = 14.5 (b) research = 12.3
Illnesses			
Doak & Doak (1980)	Various	SMOG	RGL = 10
Le Bas (1989)	Psychiatric leaflets	Flesch RE	RE = 52
Zion & Aiman (1989)	Obstetric & gynaecological leaflets	SMOG	RGL = 11.7
Davis *et al.* (1990)	Various—mainly illness & treatment	Fog Index	RGL = 13.3
Meade *et al.* (1992)	Cancer leaflets	SMOG	RGL = 11.9
Klingbeil *et al.* (1995)	Paediatric leaflets	(a) Fog (b) Fry (c) SMOG	(a) RGL = 10.5 (b) RGL = 8.8 (c) RGL = 9.8

Table 8. *Cont.*

Investigation	Topic area	Formula used	Mean Reading Grade Level (RGL) or Reading Ease Score (RE)
Health education, self-help, lifestyle			
O'Farrell & Keuthen (1983)	Behaviour therapy self-help manuals	Flesch Formula	RGL = 11.7
Nicoll & Harrison (1984)	Various	Flesch RE	RE = 55
Meade (1988)	Anti-smoking leaflets	SMOG	RGL = 10.6
Meade & Byrd (1989)	Anti-smoking leaflets	(a) Dale-Chall (b) SMOG (c) Fry (d) Fog Index (e) Flesch	(a) RGL = 10 (b) RGL = 10.8 (c) RGL = 9.0 (d) RGL = 10.9 (e) RGL = 10.1
Glanz & Rudd (1990)	Cholesterol leaflets	(a) SMOG (b) Fog Index	(a) RGL = 10.8 (b) RGL = 10.9
Ledbetter *et al.* (1990)	Condom directions	(a) Dale-Chall (b) Flesch (c) Flesch-Kincaid (d) FOG (e) Fry (f) SMOG	Average RGLs commercial = 10.3 prepared by health care providers = 8.7
Primas *et al.* (1992)	Prenatal care	SMOG	RGL = 10.2
Rivera *et al.* (1992)	Contraceptive instructions	(a) Dale-Chall (b) Fry (c) SMOG	(a) RGL = 8.9 (b) RGL = 8.1 (c) RGL = 7.5
Bradley *et al.* (1994)	Non-prescription drugs	(a) Flesch (b) Fog Index (c) SMOG	(a) RGL = 15.4 (b) RGL = 14.8 (c) RGL = 14.8
Wells (1994)	HIV/AIDS educational materials	SMOG	*Median* RGL (a) comic books = 8.2 (b) cards/inserts = 10.2 (c) brochures = 11.0 (d) pamphlets = 11.8 (e) books/monographs = 12.3
Ley (1995)	First aid instructions	(a) Flesch RE (b) Flesch-Kincaid (c) Fog index (d) SMOG	(a) RE = 40 (b) RGL = 10 (c) RGL = 13 (d) RGL = 11

the Brown Corpus (a corpus in this context is a large database of examples of written text derived from a large collection of different sources—newspapers, novels, government forms, plays, and so on), the COBUILD Corpus, and the Lancaster-Oslo-Bergen Corpus. For more details see Hofland & Johansson (1982), Sinclair (1987), Johansson & Hofland (1989), and for general information see McArthur (1992). There is also a large corpus of over 18 million words of Australian text associated with the Macquarie Dictionary, but there is as yet no

Table 9. *Effects of increasing the readability of written health care information*

Investigation	Topic	Percentage improvement	Difference in Grade level	*p*
Studies of memory—comprehension				
Ley *et al.* (1972)	(1) Barium meal X-ray	−6	3	n.s.
	(2) Cholecystogram	+34	4	<0.01
Bradshaw *et al.* (1975)	Weight reducing diet instructions	(a) +29	8	n.s.
		(b) +72	8	<0.05
Ley *et al.* (1979)	Menopause information	+22	2	<0.05
Eaton & Holloway (1980)	Warfarin medication	—	5	<0.001
Kanouse *et al.* (1981)	Medication information			
	(1) Erythromycin	+4	3	n.s.
	(2) Estrogen	−1	2	n.s.
	(3) Flurazepam	+1	2	n.s.
Ley (1982)	Glaucoma information	+34	2	<0.001
Ley *et al.* (1985)	Warning about danger of inhaling volatile substances			
	(1) Grades 5 & 6	+126	4	<0.001
	(2) Grades 9–11	+45	4	<0.001
	(3) Parents	+14	4	n.s.
	(4) Secretaries	+48	4	<0.01
Taub *et al.* (1986)	Informed consent information	+4	6	n.s.
Ley (1988)	Antibiotic information recall	(a) +116	3	<0.01
		(b) +128	6	<0.01
Meade (1988)	Anti-smoking literature	—	5	<0.05
McGraw & Sturmey (1989)	Portage material (mental retardation)	+33	3	<0.01
Handelsman & Martin (1992)	Informed consent	(a) +27	6	n.s.
		(b) +61	6	<0.05
		(c) +3	6	n.s.
Overland *et al.* (1993)	Diabetes information	(a) +30	3	<0.001
		(b) +27	5	<0.001
Studies of compliance & behavioural outcomes				
Ley *et al.* (1975)	(1) Tranquillizing medication errors	(a) +47	8	<0.05
		(b) +61	12	<0.05
	Antidepressant medication errors	(a) +44	6	<0.05
		(b) +82	12	<0.05
Ley (1978)	Weight loss	(a) +60	—	<0.001
		(b) −10	—	n.s.
		(c) −4	—	n.s.
Kanouse *et al.* (1981)	(1) Erythromycin—missed or late dose	+2	3	n.s.
	(2) Estrogen—missed or late dose	+64	2	?
	(3) Flurazepam—unnecessary dose taken	+35	2	?
Ley (1988)	Antibiotic information reading speed	(a) +29	3	<0.01
		(b) +24	6	<0.01

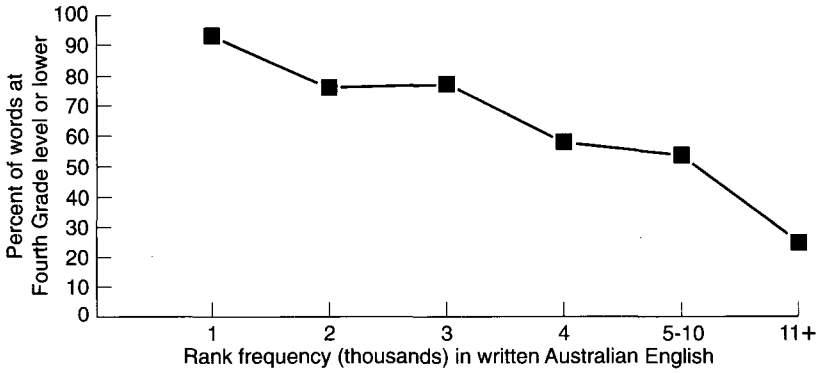

FIG. 1. *The relationship between rank frequency in English and difficulty, in words used in first aid and safety instructions, and warnings.*

published account of this. In addition some dictionaries based on the COBUILD Corpus contain frequency data on the words used in the definitions in the dictionary (Collins COBUILD, 1990).

It is well established that the frequency of occurrence of words is correlated with the percentage of people likely to understand them (Klare, 1963). Ley (1995) has reported a recent example of this relationship for a sample of words used in warning statements concerned with potentially dangerous chemical products and medicines. The results are summarized in Figure 1. The words used in the warnings were classified as easy if they were at Grade 4 or lower in the Living Word Vocabulary database. Each word was also given a rank frequency in terms of the frequency of occurrence of that word in the Macquarie Corpus of written Australian English. It can be seen that the words occurring most frequently in the language as a whole are likely to be easy words, and that as the frequency with which words occur in written language goes down, difficulty rises. Thus, of the words in the warnings which were amongst the thousand most frequently used Australian words, 93% were easy words. In contrast, of 'warnings' words which were not amongst the 11,000 commonest words, only 25% were easy.

It will be recalled that one of the predictor variables used in the Dale-Chall readability formula is the Dale list of 3,000 words which can be understood by students in the Fourth Grade. Further, a few investigators have used the Wide Range Readability Formula (Harris & Jacobson, 1982). This formula also uses a vocabulary measure, and it has been used in assessments of the readability of psychological tests (e.g. Caspi *et al.*, 1992). Thus a vocabulary measure has been implicit in some research. However, other investigations have examined vocabulary in its own right in assessing the readability of health-related materials. Amongst these studies are assessments of literature for diabetics (Thrush & Lanese, 1966); health education pamphlets (Cole, 1979); medication labels (Wilson & Hogan, 1983); words used in psychological tests (e.g. Prout & Chizik, 1988; Lubin *et al.*, 1990; Schinka & Borum, 1993); and product safety warnings (Fletcher & Abood, 1988; Ley, 1995). This research supports the conclusion that health-related literature will often be too difficult for its intended audience in that some of the words used are too difficult.

The Cloze procedure

This has already been mentioned several times. In the Cloze Procedure (Taylor, 1953) every fifth word in a piece of written text is omitted. The reader has to guess the missing word. The score is the percentage of guessed words which are exactly correct. Investigations have

established a reasonably stable relationship between Cloze scores and the percentage of multiple choice questions likely to be correctly answered. Thus a score of 35% correct on a Cloze test is approximately equivalent to correctly answering 75% of multiple choice questions about the passage, and a score of 50% correct on the Cloze test is approximately equivalent to correctly answering 90% of multiple choice questions. The starting point for counting off the fifth word will vary from subject to subject so that all words are covered. This review has already cited examples of the indirect use of Cloze procedure, and it has occasionally been used in its own right (e.g. Holcomb, 1983). Because it is a simple objective procedure it has great potential value in the pre-testing of written health materials with target and local samples.

It is desirable wherever possible to use vocabulary analysis and Cloze procedures as well as readability formulas in assessing the understandability of health materials. We suggest how this might be done in the next section.

A suggested procedure for the improvement of written information

The procedure is as follows.

(1) If you have access to a word-processing package which uses one or more of the readability formulas described above, use the formula(s). If more than one formula reports a reading grade level either take the mean of those grades, or take the highest estimate. There are other good formulas available, but they have not been used much with health-related information. Be aware also that some formulas were validated for relatively restricted ranges. For example the Spache Formula was intended for early primary school grades (Spache, 1953; 1974), while the Coleman and Liau-Coleman Formulas was originally intended for use with college level readers (Liau *et al.*, 1976). If you do not have such analyses at your disposal use either the Flesch Reading Ease Formula, and/or the Flesch-Kincaid Formula and/or the SMOG Grading. They are all relatively easy to calculate. Details of these are given earlier in this paper.

(2) In addition (if possible) list all of the words used in the text, and assess their difficulty level using a database such as *'The Living Word Vocabulary'* and substitute alternatives for words at or above the Grade 8 level. Use words at Grade 4 or Grade 6 level for this purpose. Alternatively, try to re-write the text using, as far as possible, only the words in Dale's list of 3,000 words (Dale & Chall, 1948b). If no such databases are available to you, use instead an appropriate language corpus and assesses the rank frequency of the words used. Wherever possible try to ensure that the words used are amongst the 5,000 most frequently used words. If a rarer word has to be used, then provide simple definitions and, where possible, paraphrases.

(3) Revise the text in the light of the above analyses. Aim for a Reading Grade Level of 4 to 5. This might be difficult to achieve, and will usually lead to a longer document. However, set an upper difficulty limit of 6th to 7th Grade, or a Reading Ease Score of 70 to 80.

(4) Examine the text (subjectively) to eliminate any obvious ambiguities and other causes of misunderstanding. Remember that, while a readability formula result indicating a high level of difficulty is virtually always right, a readability formula result indicating easy text can sometimes be wrong. For example, the following sentence, derived from Blumenthal (1966) via Wright (1977), obtains a Reading Ease score of 100. The words used in it are all easy words. But it is incomprehensible to most intelligent adults at first (or even second) reading:

> The rat that the cat that the dog chased killed ate the malt.

(5) Wherever possible, test the text on samples of its intended audience using the Cloze procedure. If possible, continue revisions until nearly all of the sample obtain a score of 55% or higher in the Cloze test. At the very least, try to ensure that nearly all obtain a score of 35% or more.

References

BENDICK, M. & CANTU, M.C. (1978) The literacy of welfare clients, *Social Services Review*, March, 56–68.

BLUMENTHAL, A.L. (1966) Observations with self embedded sentences, *Psychonomic Science*, 6, 453–454.

BORMUTH, J.R. (1969) *Development of readability analyses* (Project No. 7–0052). (Washington, DC, US Office of Education).

BRADLEY, B., SINGLETON, M. & PO, A.L. (1994) Readability of patient information on over-the-counter (OTC) medicines, *Journal of Clinical Pharmacy and Therapeutics*, 19, 7–15.

BRADSHAW, P.W., LEY, P., KINCEY, J.A. & BRADSHAW, J. (1975) Recall of medical advice: comprehensibility and specificity, *British Journal of Social and Clinical Psychology*, 14, 55–62.

CASPI, A., BLOCK, J., BLOCK, J.H., KLOPP, B., LYNAM, D., MOFFIT, D.E. & STOUTHAMER-LOEBER, M. (1992) A common language version of the California Child Q-set for personality assessment, *Psychological Assessment*, 4, 512–523.

CAYLOR, J.S., STICHT, T.S., FOX, L.C. & FORD, J.P. (1973) *Methodologies for Determining Reading Requirements of Military Occupational Specialties* (HumRRO Technical Report 73–5) (Presidio of Monterey, California, Human Resources Research Organization).

COLE, R. (1979) The understanding of medical terminology used in printed health education materials, *Health Education Journal*, 38, 111–121.

COLLINS COBUILD (1990) *Students' dictionary: Helping students with real English* (London, Harper Collins Publishers).

DALE, E. & CHALL, J.S. (1948a) A formula for predicting readability, *Educational Research Bulletin*, 27, 11–20.

DALE, E. & CHALL, J.S. (1948b) A formula for predicting readability: instructions, *Educational Research Bulletin*, 27, 27–54.

DALE, E. & O'ROURKE, J. (1982) *The Living Word Vocabulary: a national vocabulary inventory* (Chicago, World Book—Childcraft International).

DAVIS, T.C., CROUCH, M.A., MILLER, S. & ABDEHOU, D.M. (1990) The gap between patient reading comprehension and the readability of patient education materials, *Journal of Family Practice*, 31, 533–538.

DOAK, L.G. & DOAK, C.C. (1980) Patient comprehension profiles: recent findings and strategies, *Patient Counselling and Health Education*, 1, 101–106.

DOAK, C.C., DOAK, L.G. & ROOT, J.H. (1985) *Teaching patients with low literacy skills* (New York, J.B. Lippincott).

EISENSTAEDT, R.S., GLANZ, K., SMITH, D.G. & DERSTINE, T. (1993) Informed consent for blood transfusion: a regional hospital survey, *Transfusion*, 33, 558–561.

FLESCH, R. (1948) A new readability yardstick, *Journal of Applied Psychology*, 32, 221–233.

FLESCH, R. (1962) *The art of plain talk (revised edition)* (New York, Harper and Row).

FLETCHER, D. & ABOOD, D. (1988) An analysis of the readability of product warning labels: implications for curriculum development for persons with moderate and severe mental retardation, *Education and Training in Mental Retardation*, 12, 224–227.

FRENCH, C., MELLOR, M. & PARRY, L. (1978) Patient's view of the ophthalmic optician. Part 1: Communication between practitioners and patients, *The Ophthalmic Optician*, 28, 784–786.

FRY, E. (1968) A readability formula that saves time, *Journal of Reading*, 11, 513–516; 575–578.

GLANZ, K. & RUDD, J. (1990) Readability and content analysis of print cholesterol materials, *Patient Education and Counselling*, 16, 109–118.

GROSSMAN, S.A., PANTADOSI, S. & COVAHEY, C. (1994) Are informed consent forms that describe clinical oncology research protocols readable by most patients and their families? *Journal of Clinical Oncology*, 12, 2211–2215.

GRUNDNER, T.M. (1980) On the readability of surgical consent forms, *New England Journal of Medicine*, 302, 900–902.

GUNNING, R. (1973) *The art of clear writing (revised edition)* (New York, McGraw Hill).

HAMMERSCHNIDT, D.E. & KEANE, M.A. (1992) Institutional Review Board (IRB) review lacks impact on the readability of consent forms for research, *American Journal of Medical Sciences*, 304, 348–351.

HANDELSMAN, M.M., KEMPER, M.B., KESSON-CRAIG, P., MCLAIN, J. & JOHNSRUD, C. (1986) Use, content and readability of informed consent forms for treatment, *Professional Psychology*, 17, 514–518.

HANDELSMAN, M.M. & MARTIN, W.L. (1992) Effects of readability on the impact and recall of written informed consent material, *Professional Psychology: Research and Practice*, 23, 500–503.

HARDIE, N.R., GAGNON, J.P. & ECKEL, F.M. (1979) Feasibility of symbolic directions on prescription labels, *Drug Intelligence and Clinical Pharmacy*, 13, 588–95.

HARRIS, A.J. & JACOBSON, M.D. (1982) *Basic reading vocabularies* (New York, MacMillan).

HOFLAND, K. & JOHANSSON, S. (1982) *Word frequency in British and American English* (Bergen, Norwegian Computing Centre for the Humanities).

HOLCOMB, C.A. (1983) The Cloze procedure and readability of patient oriented drug information, *Journal of Drug Education*, 13, 347–357.

HOPPER, K.D., LAMBE, H.A. & SHIRK, S.J. (1993) Readability of informed consent forms for use with iodinated contrast media, *Radiology*, 187, 279–283.

HOPPER, K.D., TENHAVE, T.R. & HERTZEL, J. (1995) Informed consent forms for clinical and research imaging procedures: how much do patients understand? *American Journal of Roentgenology*, 164, 493–496.

HOSEY, G.M., FREEMAN, W.L. & STRACQUALURSI, F. (1990) Designing and evaluating diabetes education materials for American Indians, *The Diabetes Educator*, 16, 407–414.

JACKSON, R.H., DAVIS, T.C., MURPHY, P., BAIRNSFATHER, L.E. & GEORGE, R.B. (1994) Reading deficiencies in older patients, *American Journal of Medical Sciences*, 308, 79–82.

JACOBSON, M.D., KIRKLAND, C.E. & SELDEN, R.W. (1978) An examination of the McCall-Crabbs Standard Test Lessons in Reading, *Journal of Reading*, 21, 224–230.

JENSEN, A.R. (1980) *Bias in mental testing* (London, Methuen).

JOHANSSON, S. & HOFLAND, K. (1989) *Frequency analysis of English vocabulary and grammar* (Cambridge, Cambridge University Press).

JUBELIRER, S.J., LINTON, J.C. & MAGNETTI, S.M. (1994) Reading versus comprehension: implications for patient education and consent in an outpatient oncology clinic, *Journal of Cancer Education*, 9, 26–29.

KANOUSE, D.E., BERRY, S.H., HAYES-ROTH, B., ROGERS, W.H. & WINKLER, J.D. (1981) *Informing patients about drugs: Summary Report* (Santa Monica, Ca, Rand Corporation).

KAUFMAN, K.L., TARNOWSKI, K.J., SIMONIAN, S.J. & GRAVES, K. (1992) Assessing the readability of family self assessment self-report inventories, *Psychological Assessment*, 3, 697–700.

KINCAID, J.P., FISHBURNE, R.P., ROGERS, R.L. & CHISSOM, B.S. (1975) *Derivation of new readability formula for navy enlisted personnel* (Millington, Tennessee, Navy Research Branch).

KLARE, G.R. (1963) *The measurement of readability* (Iowa, Iowa State University Press).

KLARE, G.R. (1974) Assessing readability, *Reading Research Quarterly*, 10, 62–102.

KLARE, G.R. (1976) A second look at the validity of readability formulas, *Journal of Reading Behavior*, 8, 129–152.

KLINGBEIL, C., SPEECE, M.W. & SCHUBINER, H. (1995) Readability of pediatric patient education materials, *Clinical Pediatrics*, 34, 96–102.

KOZOL, J. (1986) *Where stands the Republic? Illiteracy: a warning and a challenge to the Nation's Press* (Atlanta, Cox Enterprises, Inc).

LE BAS, J. (1989) Comprehensibility of patient education literature, *Australian and New Zealand Journal of Psychiatry*, 23, 542–546.

LEDBETTER, C., HALL, S., SWANSON, J.M. & FORREST, K. (1990) Readability of commercial versus generic health instructions for condoms, *Health Care for Women International*, 11, 295–304.

LEY, P. (1974) Communications in the clinical setting, *British Journal of Orthodontics*, 1, 173–177.

LEY, P. (1978) Psychological and behavioural factors in weight loss, in: G.A. BRAY (Ed.) *Recent advances in obesity research, 2* (London, Newman Publishing).

LEY, P. (1982) Satisfaction, compliance and communication, *British Journal of Clinical Psychology*, 21, 241–254.

LEY, P. (1988) *Communicating with patients* (London, Chapman and Hall).

LEY, P. (1995) *Effectiveness of label statements for drugs and poisons* (Canberra, Australian Government Publishing Service).

LEY, P. & MORRIS, L.A. (1984) Psychological aspects of written information for patients, in: S. RACHMAN (Ed.) *Contributions to Medical Psychology 3* (Oxford, Pergamon Press).

LEY, P., FLAHERTY, B., SMITH, F., MARTIN, J. & RENNER, P. (1985) *A comparative study of the effects of two warning messages about volatile substances* (Sydney, New South Wales, Drug and Alcohol Authority).

LEY, P., GOLDMAN, M., BRADSHAW, P.W., KINCEY, J.A. & WALKER, C. (1972) The comprehensibility of some X-ray leaflets, *Journal of the Institute of Health Education*, 10, 47–53.

LEY, P., JAIN, V.K. & SKILBECK, C.E. (1975) A method for decreasing patients' medication errors, *Psychological Medicine*, 6, 599–601.

LEY, P., PIKE, L.A., WHITWORTH, M.A. & WOODWARD, R. (1979) Effects of source, context of communication, and difficulty level on the success of health education communications, *Health Education Journal*, 38, 47–52.

LIAU, T.L., BASSIN, C.R., MARTIN, C.J. & COLEMAN, E.B. (1976) Modification of the Coleman readability formulas, *Journal of Reading Behavior*, 8, 381–386.

LIGUORI, S. (1978) A quantitative assessment of the readability of PPIs, *Drug Intelligence and Clinical Pharmacy*, 12, 712–716.

LUBIN, B., COLLINS, J.E., SEEVER, M., VAN WHITLOCK, R. & DENNIS, A.J. (1990) Relationships among readability, reliability and validity in a self-report adjective check list, *Psychological Assessment*, 2, 256–261.

MCARTHUR, T. (1992) *The Oxford companion to the English language* (Oxford, Oxford University Press).

MCCALL, W.A. & CRABBS, L.M. (1925; 1950; 1961) *Standard test lessons in reading* (New York, Teachers' College, Columbia University).

MCLAUGHLIN, G.H. (1969) SMOG Grading—a new readability formula, *Journal of Reading*, 12, 639–646.

MEADE, C.D. (1988) *Effect of document simplification on patients' comprehension* (Unpublished PhD thesis, University of Illinois).

MEADE, C.D. & BYRD, J.C. (1989) Patient literacy and the readability of smoking literature, *American Journal of Public Health*, 79, 204–206.

MEADE, C.D. & HOWSER, D.M. (1992) Consent forms: how to determine and improve their readability, *Oncology Nursing Forum*, 19, 1523–1528.

MEADE, C.D. & SMITH, C.F. (1991) Readability formulas: cautions and criteria, *Patient Education and Counselling*, 17, 153–158.

MEADE, C.D., DIEKMANN, J. & THORNHILL, D.G. (1992) Readability of American Cancer Society patient education literature, *Oncology Nursing Forum*, 19, 52–55.

MORRIS, L.A., MYERS, A. & THILMAN, D.G. (1980) Application of the readability concept to patient-oriented drug information, *American Journal of Hospital Pharmacy*, 37, 1504–1509.

MORROW, G. (1980) How readable are surgical consent forms? *Journal of the American Medical Association*, 244, 56–58.

MURPHY, L., GAMBLE, G. & SHARPE, N. (1994) Readability of subject information leaflets for medical research, *New Zealand Medical Journal*, 107, 509–510.

NATIONAL CANCER INSTITUTE (1979) *Readability testing in cancer communications* (Washington, DC, Department of Health Education and Welfare) (DHEW Publication Number (NIH) 79–1689).

NICOLL, A. & HARRISON, C. (1984) The readability of health care literature, *Developmental Medicine and Child Neurology*, 26, 596–600.

O'FARRELL, T.J. & KEUTHEN, N.J. (1983) Readability of behavior self-help manuals, *Behavior Therapy*, 14, 449–454.

OGLOFF, J.R.P. & OTTO, R.K. (1991) Are research participants truly informed? Readability of informed consent forms used in research, *Ethics and Behaviour*, 1, 239–252.

OLSON, A.V. (1989) A question of reading validity, *Journal of Research and Development in Education*, 19, 33–40.

OVERLAND, J.E., HOSKINS, P.L. & YUE, D.K. (1993) Low literacy: a problem in diabetes education, *Diabetic Medicine*, 10, 857–850.

PRIMAS, P., LEFOR, N., JOHNSON, J., HELMS, S.M., COATES, L. & COE, M.K. (1992) Prenatal literature testing: a pilot project, *Journal of Community Health*, 17, 61–67.

PROUT, H.T. & CHIZIK, R. (1988) Readability of child and adolescent self report inventories, *Journal of Consulting and Clinical Psychology*, 54, 152–154.

PYRCZAK, F. & ROTH, D.H. (1976) The readability of directions on non-prescription drugs, *Journal of the American Pharmaceutical Association*, 16, 242–245.

RIVERA, R., REED, J.S. & MENIUS, D. (1992) Evaluating the readability of informed consent forms used in contraceptive clinical trials, *International Journal of Gynaecology & Obstetrics*, 38, 227–230.

ROSS, M.W., JEFFORDS, K. & GOLD, J. (1994) Reasons for entry into and understanding of HIV/AIDS clinical trials: a preliminary study, *AIDS Care*, 6, 77–82.

SCHINKA, J.A. & BORUM, R. (1993) Readability of adult psychopathology inventories, *Psychological Assessment*, 5, 384–386.

SINCLAIR, J.M. (ED.) (1987) *Looking up: An account of the COBUILD Project in lexical computing* (London, Collins ELT).

SPACHE, G. (1953) A new readability formula for primary grade reading materials, *Elementary School Journal*, 53, 410–413.

SPACHE, G.D. (1974) *Good reading for poor readers* (Champaign, Illinois, Garrard).

STOKES, A. (1978) The reliability of readability formulas, *Journal of Research in Reading*, 1, 21–34.

TAUB, H.A., BAKER, M.T. & STURR, J.F. (1986) Informed consent research: effects of readability, patient age, and education, *Journal of the American Geriatric Society*, 34, 601–606.

TAYLOR, W.L. (1953) Cloze procedure: a new tool for measuring readability, *Journalism Quarterly*, 30, 415–433.

THRUSH, R.S. & LANESE, R.R. (1962) The use of printed materials in diabetes education, *Diabetes*, 11, 132–137.

WAGENAAR, W.A., STIRRED, R. & WIJLHUIZEN, G.J. (1987) Readability of instructional text written for the general public, *Applied Cognitive Psychology*, 1, 155–167.

WELLS, J.A. (1994) Readability of AIDS/HIV educational materials: the role of the medium of communication, target audience, and producer characteristics, *Patient Education and Counselling*, 24, 249–259.

WILSON, J. & HOGAN, L. (1983) Readability testing of auxiliary labels, *Drug Intelligence and Clinical Pharmacy*, 17, 54–55.

WRIGHT, P. (1977) Presenting technical information: a survey of research findings, *Instructional Science*, 6, 93–134.

ZION, A.B. & AIMAN, J. (1989) Level of reading difficulty in the American College of Obstetricians and Gynecologists patient education materials, *Obstetrics and Gynecology*, 7, 955–961.