

A Systematic Review of Readability and Comprehension Instruments Used for Print and Web-Based Cancer Information

Daniela B. Friedman, MSc
Laurie Hoffman-Goetz, PhD, MPH

Adequate functional literacy skills positively influence individuals' ability to take control of their health. Print and Web-based cancer information is often written at difficult reading levels. This systematic review evaluates readability instruments (FRE, F-K, Fog, SMOG, Fry) used to assess print and Web-based cancer information and word recognition and comprehension tests (Cloze, REALM, TOFHLA, WRAT) that measure people's health literacy. Articles on readability and comprehension instruments explicitly used for cancer information were assembled by searching MEDLINE and PsycINFO from 1993 to 2003. In all, 23 studies were included; 16 on readability, 6 on comprehension, and 1 on readability and comprehension. Of the readability investigations, 14 focused on print materials, and 2 assessed Internet information. Comprehension and word recognition measures were not applied to Web-based information. None of the formulas were designed to determine the effects of visuals or design factors that could influence readability and comprehension of cancer education information.

Keywords: *health literacy; cancer; readability; comprehension; word recognition; patient education information; print material; Internet*

LITERATURE REVIEW

Health Literacy

The term *health literacy* refers to "the capacity to obtain, interpret, and understand basic health information and services and the competence to use such information and services to enhance health" (U.S. Department of Health and Human Services, 2000) as well as "cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand, and use information in ways which promote and maintain good health" (Nutbeam, 1998, p. 264). Health literacy is a critical factor in accessing health information and empowering people to take charge of their health (World Health Organization, 2000).

Daniela B. Friedman and Laurie Hoffman-Goetz, Department of Health Studies and Gerontology, Faculty of Applied Health Sciences, University of Waterloo, Ontario, Canada.

Address reprint requests to Laurie Hoffman-Goetz, PhD, MPH, Professor, Department of Health Studies and Gerontology, Faculty of Applied Health Sciences, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1; phone: (519) 885-1211 ext. 3098; fax: (519) 746-2510; e-mail: lhgoetz@healthy.uwaterloo.ca.

This work was supported in part by a grant from the Social Sciences and Humanities Research Council of Canada.

Health Education & Behavior, Vol. 33 (3): 352-373 (June 2006)

DOI: 10.1177/1090198105277329

© 2006 by SOPHE

Individuals with basic literacy skills tend to take words literally, read at a slow pace, skip over unfamiliar words, have problems identifying key concepts, focus on insignificant details, and overlook the context of the writing (C. C. Doak, Doak, Friedell, & Meade, 1998). Limited health literacy is associated with reduced understanding of screening procedures for cancer (Davis, Williams, Marin, Parker, & Glass, 2002). Individuals with limited literacy have higher distress levels about developing cancer (Sharp, Zurawski, Roland, O'Toole, & Hines, 2002), have little knowledge about cancer prevention (Lindau et al., 2002), and are diagnosed at later stages of cancer (Bennett et al., 1998).

Readability and Comprehension

Dale and Chall (1949) defined readability as "the sum total (including the interactions) of all those elements within a given piece of printed material that affects the success a group of readers have with it" (p. 23). The importance of readability results from the following three factors: (1) the increase in the amount of reading required for success, (2) the decline in reading skills of high school graduates, and (3) the simplification of text as a partial solution for the combination of increased reading and decreased reading skills (Klare, 1984).

Readability formulas are predictive devices providing quantitative estimates of the reading difficulty of printed health information by determining the complexity of vocabulary using word length counts. They do not however take into account readers' prior knowledge, motivation, or comprehension (Bailin & Grafstein, 2001). Independent of readability measures, comprehension instruments are used to evaluate individuals' understanding of written medical text. Results of such tests are significant for the development of printed health information that is easily understood (Davis, Michielutte, Askov, Williams, & Weiss, 1998). Active reading comprehension requires "encoding of words read into working memory and the processing of word groups into a useful meaning" (Gausman Benson & Forman, 2002, p. 96). Patients' ability to understand cancer information can be affected by their anxiety levels, emotional distress, vulnerability following diagnosis, or unfamiliarity with the surroundings (Estey, Musseau, & Keehn, 1994).

Both readability and comprehension are important concepts in the study of health literacy. Health educators and researchers need to be able to match the readability levels of print and Web-based cancer information with the literacy (comprehension) levels of their target patients and groups.

REVIEW OBJECTIVES

The objectives of this review are twofold: first, to provide a systematic review of readability tools used to assess written cancer information and of comprehension tools used to evaluate individuals' health literacy skills; and second, to assess the usefulness of these instruments in studies evaluating print and Web-based cancer information.

This review is the first to assess multiple literacy instruments employed for both print and Web-based cancer information. Other studies have summarized and compared readability (Ley & Florio, 1996) or comprehension (Foltz & Sullivan, 1998) instruments for health or cancer information, discussed limitations of readability tools for the identification of patients with low literacy (Parker, Williams, Baker, & Nurss, 1996), and typically

evaluated information in one medium—print (Chung, Horowitz, Cantos, & Siriphanth, 2000; Slaten, Parrott, & Steiner, 1999) or Internet (Berland et al., 2001; Wilson, Baker, Brown-Syed, & Gollop, 2000).

METHODOLOGY

Literature Search Terms

A list of articles on instruments explicitly used to test the readability and comprehension of cancer information was assembled by searching the computerized databases of MEDLINE and PsycINFO. These sources were extensively searched for papers written over a 10-year period (1993-2003) using the terms *cancer* and *health* along with the following search terms alone or in combination: *literacy*, *readability*, *reading ability*, *reading assessment tool*, *readability formula*, *Internet*, *World Wide Web*, *comprehension*, *comprehension tool*, *word recognition*, *reliability*, and *validity*.

Inclusion Criteria of Cancer Literacy Studies

For inclusion in this review, articles had to be in English, published from 1993 to 2003, available through the university library system, and discussed at least one literacy tool used to evaluate print or Web-based cancer information.

Inclusion Criteria of Literacy Tools

For inclusion in the review, the following inclusion criteria were used: availability of references through the selected databases, tool in existence for at least 5 years prior to 2003, used for English text, and applicable to health information for adult populations. Five readability formulas (Flesch Reading Ease [FRE], Flesch-Kincaid [F-K], Fog Index, SMOG, and Fry Readability Graph), two comprehension instruments (Cloze and Test of Functional Health Literacy in Adults [TOFHLA/S-TOFHLA]), and two word recognition tests (Wide Range Achievement Test [WRAT-R] and Rapid Estimate of Adult Literacy in Medicine [REALM/REALM-R]) met the inclusion criteria. Once these tools were selected for review, their specific test names were entered into MEDLINE to ensure that a comprehensive list of references on the readability and comprehension of cancer information was obtained.

Coding of Literacy Tools

For each readability, comprehension, and word recognition instrument, the following characteristics were noted: test procedure, applicability to cancer studies, reliability (interrater or test-retest reliability), validity (concurrent validity/interformula correlations), ease of use (administration time, staff training), reference to visuals (illustrations, diagrams, graphs, charts), and typographic or design variables (font and font size, use of white space on the page, headings, margins). These formatting factors are considered important for the readability and comprehension of cancer messages (L. G. Doak, Doak, & Meade, 1996).

RESULTS

A total of 416 citations was obtained using the identified search terms. Of these, 183 were not applicable to the review, 29 were on general literacy principles or literacy instruments, 158 were on health literacy, and 46 were on literacy and cancer. Of the 46 articles on literacy and cancer, 25 were empirical studies. Using the inclusion criteria, 23 articles were selected for this systematic review—16 studies were on readability, 6 studies were on comprehension, and 1 study evaluated both readability and comprehension of cancer information. Articles on health literacy were read for background information on literacy instruments.

Review of Readability Formulas Used to Assess Print and Web Cancer Information

Five readability formulas were selected for review based on frequent use in studies on health literacy (Ley & Florio, 1996). Table 1 summarizes the commonly used instruments in the assessment of written health materials; background information, procedures, and formulas (if applicable) are presented for each instrument. These instruments are described next to provide a context for their application to cancer information.

Flesch Reading Ease Scale

$$\text{Flesch Reading Ease score} = 206.835 - 0.846 \text{ wl} - 1.015 \text{ sl}$$

where wl = number of syllables per 100 words (average word length) and sl = number of words per sentence (average sentence length).

The FRE (Flesch, 1948) measures the readability of text written between Grade 5 and college-graduate level. The formula is based on two variables—average sentence length (based on number of words) and average word length (based on number of syllables). The FRE has been validated against the McCall-Crabbs Standard Test Lessons in Reading (McCall & Crabbs, 1961). FRE results correlate highly with other readability formulas, including the Fry and SMOG at .96 and .95, respectively (Meade & Smith, 1991).

The Flesch instrument tests at least three passages of text consisting of 100 words each. Titles, subtitles, headings, captions, date lines, and signature information are excluded from the passage selection. The number of words in each passage is added up and divided by the number of selected sentences. The total number of syllables per word is determined, divided by the total number of words in the sample, and multiplied by 100. The score derived from the FRE formula is referred to as the Flesch Reading Ease Scale score, which ranges from 100 (very easy to read) to 0 (unreadable).

Flesch-Kincaid Readability Formula

$$\text{Flesch-Kincaid Reading Grade Level} = 0.39 \text{ sl} + 11.8 \text{ spw} - 15.59$$

where sl = average number of words per sentence and spw = average number of syllables per word.

The Flesch-Kincaid (Kincaid, Fishburne, Rogers, & Chissom, 1975) formula is a modified version of the FRE that generates a reading grade level (RGL) for print informa-

(text continues on p. 358)

Table 1. Characteristics of Literacy Tools

Literacy Tools	Variables and Computation	Passage Length Required	Administration Time	Interpretation	Strengths	Weaknesses
Readability Flesch Reading Ease (Flesch, 1948)	Sentence length (number of words) and word length (number of syllables)	Three 100-word passages	Quick: Computer software	Reading Ease Scale (0 = unreadable; 100 = very easy to read)	Quick administration with software; highly correlated with other instruments	Computer scores underestimate difficulty; longer manual administration
Flesch-Kincaid (Kincaid, Fishburne, & Rogers, & Chissom, 1975)	Sentence length (number of words) and word length (number of syllables)	Three 100-word passages	Quick: computer software	Reading grade level	Quick administration with software; highly correlated with other instruments	Computer scores underestimate difficulty; longer manual administration
Fog Index (Gunning, 1968)	Sentence length (number of words) and word length (polysyllabic words)	100 consecutive words	4 to 5 minutes	Score is the number of formal years of education needed to read information	Highly correlated with other instruments	Longer manual administration
SMOG (McLaughlin, 1969)	Polysyllabic words	30 sentences (10 near the beginning, 10 in the middle, 10 at the end)	Time-consuming: manual computation	Reading grade level	Most frequently used readability test; highly correlated with other instruments	Score based on strict criterion of 100% comprehension; time-consuming

Fry (Fry, 1968)	Sentence length (number of words) and word length (number of syllables)	Three 100-word passages	Time-consuming: manual computation	Reading grade level	Validated in Spanish; ease of administration	Longer administra- tion time com- pared to Flesch Reading Ease and F-K
Comprehension Cloze (Taylor, 1953)	Score is ease of filling deleted words into sentences within paragraph	Variable number of sentences and random number of blanks	20 to 50 minutes	Adequate, challenging, and problematic comprehension	Context of information provided as opposed to word lists	Time-consuming
Test of Functional Health Literacy in Adults (Parker, Baker, Williams, & Nurss, 1995)	Modified Cloze procedure with deletion of random words	Two subtests: 50-item reading comprehension and 17-item numeracy test	22 minutes	Inadequate, marginal, or functional health literacy	Measures functional health literacy	Time-consuming; inapplicable content
Shortened Test of Functional Health Literacy in Adults (Baker, Williams, Parker, Gazmararian, & Nurss, 1999)	Modified Cloze procedure with deletion of words	Two shortened subtests: 36-item reading comprehension and 4-item numeracy test	12 minutes	Inadequate, marginal, or functional health literacy	Quick measure of functional health literacy	Inapplicable content

(continued)

Table 1 (continued)

Literacy Tools	Variables and Computation	Passage Length Required	Administration Time	Interpretation	Strengths	Weaknesses
Word recognition						
Rapid Estimate of Adult Literacy in Medicine (Davis et al., 1991)	Raw score converted to grade range	66 words	2 to 3 minutes	Four grade ranges: < Grade 3; Grades 4 to 6; Grades 7 to 8; Grade 9+	Quick screening test; not intimidating; available in Spanish	Does not discriminate above Grade 9 level; can over- and underestimate reading skills
Rapid Estimate of Adult Literacy in Medicine-Revised (Bass, Wilson, & Griffith, 2001)	Raw score converted to grade range	8 words	< 2 minutes	Score below 6 implies \leq Grade 6 level	Quick screening test; not intimidating; available in Spanish	Does not discriminate above a Grade 9 level; can over- and underestimate reading skills
Wide Range Achievement Test Revised (Jastak & Wilkinson, 1987)	Raw score converted to specific grades	~ 40 words	3 to 5 minutes	Score \geq Grade 7 implies good reader	Quick screening test for children and adults	Low face validity; does not discriminate above Grade 7 level; no medical terms

tion. The F-K formula is the standard readability test of the U.S. Department of Defense and is used to determine the difficulty of text written between Grade 5 and college level.

Computer software programs (e.g., Microsoft Word) calculate the F-K and the FRE by estimating the number of syllables per passage. The correlation between the computer-calculated score of the F-K and the manually computed score of FRE is excellent at .91 (Ley & Florio, 1996). However, because computer software programs recognize each period as the end of a sentence, abbreviations, numbers with decimals, and bullets may lower the RGL and underestimate text difficulty. Follow-up editing of text by health educators would help generate more useful and accurate computer-based scores.

Fog Index

$$\text{Fog Index} = 0.4 (\text{sl} + \%lw)$$

where sl = average number of words per sentence (number of words/number of sentences) and %lw = percentage of words with three or more syllables.

Gunning's (1968) Fog Index assesses the readability level of text ranging from Grade 4 to collegiate level. This tool counts syllables for polysyllabic words only, requiring less time to administer manually than the Flesch formulas. Fog was validated against the McCall-Crabbs Test Lessons with a correlation of .59 and created to predict the average grade of individuals who could answer 90% of the passages' comprehension questions correctly (Ley & Florio, 1996). Concurrent validity is high, with Fog scores correlating extremely well with both Fry ($r = .93$) and SMOG ($r = .99$) (Meade & Smith, 1991). The Fog formula produces a score representing the number of formal years of education required to be able to read the written information.

SMOG Readability Formula

$$\text{SMOG RGL} = 3 + \sqrt{lw_{30}}$$

where $\sqrt{lw_{30}}$ is the square root of the number of words with three or more syllables per 30 sentences.

The SMOG readability formula (McLaughlin, 1969) was designed to predict readability based on total number of polysyllabic words. Polysyllabic words in 30 sentences of text are counted, and the square root of the nearest perfect square to this number is taken. The RGL is determined by adding 3 to the resulting number (McLaughlin, 1969).

The formula was validated against the McCall-Crabbs Standard Test Lessons ($r = .99$) and intended to produce average RGLs for individuals with a 100% comprehension of the passages (Ley & Florio, 1996). Interrater reliability of the SMOG procedure has shown to be 100% for the assessment of Web-based information on mental health (King, Winton, & Adkins, 2003). SMOG correlates highly with the Fog Index ($r = .97$ to $.99$), FRE ($r = .95$ to $.96$), F-K ($r = .93$), and the Fry Readability Graph ($r = .93$ to $.96$) (Meade & Smith, 1991).

SMOG is an accurate and convenient method of analyzing the readability of health education literature and was recommended by the U.S. National Cancer Institute (1979) for the assessment of cancer pamphlets. However, SMOG results in scores one to two grade levels higher because it is based on stricter criteria for readability and classifies RGLs based on 100% comprehension ability (Meade & Smith, 1991).

Fry Readability Graph

The Fry (1968) formula measures sentences per 100 words and syllables per 100 words, plotting these results on a graph and obtaining an approximate RGL of the information. Fry results correlate highly with other readability formulas such as FRE ($r = .96$) and SMOG ($r = .93$) and Fry is recommended due to its ease of administration (Fry, 1968; Meade & Smith, 1991).

The average number of sentences and the average number of syllables per 100-word passage are plotted on the Fry Readability Graph to determine an RGL of the material as approximated by the grade areas on the graph.

The Fry formula has been validated in both English and Spanish and takes into account that Spanish text tends to have a greater number of polysyllabic words than English text at an identical reading level (Gilliam, Pena, & Moutain, 1980).

Applicability of Readability Tools to Print and Web-Based Cancer Information

All five readability formulas (FRE, F-K, Fog Index, SMOG, and Fry) were typically used to assess print and Web-based cancer information. Table 2 provides summaries of the 16 studies using these instruments for readability of cancer information. Of these investigations, 14 focused on print information and 2 on Internet-based information. Overall, mean reading grade levels ranged from Grade 6 to 14.1 despite the recommended score of Grade 5 or 6 for medical print information (Meade, McKinney, & Barnas, 1994). FRE scores represented fairly difficult reading styles with scores of 65 or lower.

Research specifically conducted on the readability of cancer information on the Internet found that current Web sites are written at a very difficult level for individuals with only marginal literacy skills (Berland et al., 2001; Wilson et al., 2000). An examination of 100 English and 6 Spanish Web sites found reading levels to be at a Grade 9 level or higher using the Fry Readability Graph (Berland et al., 2001). Studies examining print cancer information used SMOG alone most often to assess text difficulty (Butow, Brindle, McConnell, Boakes, & Tattersall, 1998; Chung et al., 2000; Friedman & Hoffman-Goetz, 2003; Guidry, Fagan, & Walker, 1998; Price & Everett, 1996; Singh, 2003) despite concerns about single-instrument accuracy of reading levels (Meade & Smith, 1991). A recent study on the readability of cancer Web sites used multiple evaluation tools for this reason (Friedman, Hoffman-Goetz, & Arocha, 2004).

Slaten and colleagues (1999) used both SMOG and the Fog Index to examine skin cancer prevention brochures intended for parents of young children, whereas Beaver and Luker (1997) used SMOG and FRE to evaluate breast cancer information. One study on printed oral cancer information employed FRE alone (Humphris, Duncalf, Holt, & Field, 1999).

Two studies (Dollahite, Thompson, & McNew, 1996; Glazer, Kirk, & Bosler, 1996) measured concurrent validity or interformula correlations, which ranged from .70 to .97, even though five other studies used multiple tools to assess readability (Beaver & Luker, 1997; Gribble, 1999; Grossman, Piantadosi, & Covahey, 1994; Mohrmann et al., 2000; Slaten et al., 1999). Only one study reported numerical results for interrater reliability (100% agreement) of administering SMOG on printed cancer information (Singh, 2003). Little attention to reliability testing of the readability procedures may have resulted from

Table 2. Readability Studies on Print and Internet Cancer Information

Investigators	Cancer Topic(s)	Medium	Number of Documents Tested/Created	Readability Formula Used ^a	Mean Reading Grade Level or Reading Ease Score
Friedman and Hoffman-Goetz (2003)	General	Print	45	SMOG	12.5
Singh (2003)	General	Print	10	SMOG	12.1
Berland et al. (2001)	Multiple conditions (including breast cancer)	Internet	18 English, 7 Spanish	Fry	13.2, 9.9
Chung, Horowitz, Cantos, and Siriphaht (2000)	Oral cancer	Print	19	SMOG	6 to 13
Mohrmann et al. (2000)	Breast cancer	Print	61	FRE, F-K	65.0, 7.5
Wilson, Baker, Brown-Syed, and Gollop (2000)	Therapy, sites and types, treatment, prevention, screening, ethnic groups	Internet	49	F-K	12.0
Gribble (1999)	Informed consent for genetic testing	Print	12	FRE, F-K	45.4, 13.2
Humphris, Duncalf, Holt, and Field (1999)	Oral cancer	Print	1	FRE	80.0
Slaten, Parrott, and Steiner (1999)	Skin cancer	Print	8	SMOG, Fog	8.5, 10.6
Butow, Brindle, McConnell, Boakes, and Tattersall (1998)	Treatment	Print	5	SMOG	10.8
Guidry, Fagan, and Walker (1998)	Breast and prostate cancer	Print	46	SMOG	9.0
Beaver and Luker (1997)	Breast cancer	Print	50	SMOG, FRE	10, 58.7
Dollahite, Thompson, and McNew (1996)	Multiple conditions (including cancer)	Print	20 (cancer)	FRE, Fry	50.4, 11.0
Glazer, Kirk, and Bosler (1996)	Breast cancer	Print	19	Fog, FRE, F-K	11.2, 59.2, 9.2
Price and Everett (1996)	Multiple cancers	Print	4	SMOG	6.0 (based on one pamphlet)
Grossman, Piantadosi, and Covahey (1994)	Informed consent forms	Print	137	FRE, F-K, Fog	52.6, 11.1, 14.1

NOTE: FRE = Flesch Reading Ease; F-K = Flesch-Kincaid.

a. Only formulas selected for review are included in this table.

the availability of word processing software and computer programs that are able to calculate readability scores automatically. None of the studies evaluated design and formatting factors of the documents.

Review of Comprehension and Word Recognition Instruments Commonly Used to Assess Health Literacy

Testing individual literacy skills informs health educators about individuals with poor reading skills who are in need of plain language health education information (Davis et al., 1998). Two comprehension instruments (Cloze and TOFHLA/S-TOFHLA) and two reading assessment/word recognition tests (REALM/REALM-R and WRAT-R) were selected for review based on frequency of use in the assessment of functional health literacy in adults (Foltz & Sullivan, 1998). Table 1 provides a summary of these instruments.

Word recognition tests measure the ability of an individual to recognize and pronounce isolated words. It is assumed that if people have difficulty with basic reading skills (pronunciation), they will also have trouble with comprehension (Davis et al., 1998).

Cloze Procedure

Taylor's (1953) Cloze procedure evaluates comprehension through testing deleted words from written passages at set intervals chosen randomly, based on a table of random numbers or by counting out every *n*th word, and having participants fill in the blanks. If enough words are deleted at random, the blanks are intended to represent an adequate percentage of words of varying complexity. The Cloze test assumes that individuals with superior reading skills will comprehend the context of the information and be able to complete the sentences with little difficulty (Davis et al., 1998).

The number of correct guesses is presented as a proportion of the total number of blanks. The resulting Cloze score is an index of predictability or ease of filling in the blanks. So, 100% correct guesses yields a score of 1.0. Scoring is as follows: (1) $> 56\%$ (or $0.56/1.0$) = adequate comprehension; (2) 44% to 56% = text found to be challenging; (3) $< 44\%$ = comprehension problems. Cloze is a valid and reliable measure of patient comprehension that can be used to pretest written health information (Estey et al., 1994; Hafner, 1966; Ley & Florio, 1996; Taylor, 1953), though it has a heavy time burden for administrators and participants.

TOFHLA and S-TOFHLA

The TOFHLA (Parker, Baker, Williams, & Nurss, 1995) measures functional health literacy in adult populations. The test consists of two subtests: (1) reading comprehension: 50 test items in a modified Cloze procedure where four multiple-choice options are provided (topics of test passages include preparation for an upper gastrointestinal series, a patients' rights and responsibilities section of a Medicaid application, and a standard hospital consent form); and (2) numeracy subtest: 17 phrases that test individuals' understanding of prescription directions and appointment information from authentic health care materials.

Both the comprehension and numeracy subtest scores are multiplied by 2.941 to generate a score ranging from 0 to 50. The total score ranges from 0 to 100: (1) 0 to 59 = inadequate literacy; (2) 60 to 74 = marginal literacy; and (3) 75 to 100 = adequate literacy (Baker, Parker, Williams, & Scott Clark, 1998). A Spanish-language version of the test is available. The correlation between the reading comprehension and numeracy subtests is moderate at .79 for the English and .70 for the Spanish version. The English TOFHLA correlates well with word recognition tests such as the REALM ($r = .84$) and WRAT-R ($r = .74$). It has high internal consistency of test items with a Cronbach alpha of .98 (for both English and Spanish) and excellent test-retest reliability at .92 and .84 in English and Spanish, respectively (Parker et al., 1995).

A shortened version of the TOFHLA (S-TOFHLA; Baker, Williams, Parker, Gazmararian, & Nurss, 1999) has excellent internal consistency of test items (.98) and significantly high correlations ($p < .001$) with the original TOFHLA ($r = .91$; Parker et al., 1995) and the REALM ($r = .80$; Baker et al., 1999).

REALM and REALM-R

The REALM (Davis et al., 1991) is a reading recognition test of medical terminology. It does not measure literal comprehension of written information. Instead, the instrument evaluates individuals' ability to read and pronounce 66 common medical terms presented in order of increasing syllable number and pronunciation difficulty. The test can be administered in 2 to 3 minutes with only minimal training (Davis et al., 1993). Raw scores of 0 to 66 are categorized into four RGLs: (1) 0 to 18 = Grade 3 or below; (2) 19 to 44 = Grades 4 through 6; (3) 45 to 60 = Grades 7 and 8; and (4) 61 to 66 = Grade 9+.

REALM correlates significantly ($p < .001$) with other reading tools including the WRAT-R ($r = .88$; Davis et al., 1993). Test-retest reliability is excellent at .99 (Bass, Wilson, & Griffith, 2001). REALM is considered to have high face validity because it measures relevant health-related terms (Davis et al., 1994). However, there are significant limitations of the REALM for evaluation of the comprehension of health information. First, it is not a valid indicator of reading ability in Spanish (Nurss, Baker, Davis, Parker, & Williams, 1995). Second, it overestimates people's reading skills of medical terminology due to excellent pronunciation despite poor comprehension (Bennett et al., 1998). Individuals may be able to pronounce words well on the REALM but perform poorly when comprehension is assessed with the S-TOFHLA. Also, they may find terms difficult to pronounce on the REALM but generate better results on the S-TOFHLA when they are provided with context.

The REALM-R (Rapid Estimate of Adult Literacy in Medicine-Revised; Bass et al., 2001) allows for a quick screen of potential literacy problems in medical settings. This instrument uses a series of eight medical words from the terms on the original REALM (*fatigue, jaundice, directed, allergic, colitis, constipation, anemia, and osteoporosis*). Scores of 6 or lower imply a reading ability at a Grade 6 level or below. Test scores of the REALM-R correlate moderately with the WRAT-R ($r = .78$) and REALM ($r = .72$). Internal consistency or reliability of test items within the REALM-R is excellent at .91 (Bass, Wilson, Griffith, & Barnett, 2002).

WRAT-R

The WRAT-R (Jastak & Wilkinson, 1987) is the most standardized word recognition and pronunciation test. This tool, originally designed to assess the basic skills of school-aged children, consists of three subtests: reading, spelling, and arithmetic. The reading portion of the instrument can be administered in 3 to 5 minutes, evaluating individuals' pronunciation of approximately 40 items. Item consistency of the WRAT-R is extremely high at .94. From a raw score, grade scores can be generated. Individuals with scores of \geq Grade 7 are considered to have high reading skills, whereas those with scores $<$ Grade 7 are considered poorer readers (Meade et al., 1994). Once an individual has trouble pronouncing three consecutive words, the test is stopped (Estey, Kemp, Allison, & Lamb, 1993).

Limitations of the WRAT-R are its availability and validity in English only and its use in health settings even though it consists of nonmedical test words. In addition, patients rate it as having poor face validity. They score lower on the WRAT-R compared to other tests such as the REALM because words become difficult more rapidly and one third of test words are at a Grade 9 level or higher (Davis et al., 1994).

Applicability of Comprehension and Word Recognition Tools to Written Cancer Information

Research on cancer patients' comprehension of cancer information has been conducted on print material only. Empirical studies on the comprehension of Internet-based information were not evident in the literature. Of the four comprehension and word recognition instruments reviewed (Cloze, TOFHLA/S-TOFHLA, REALM/REALM-R, WRAT-R), the REALM was used most often to evaluate print cancer information (Bennett et al., 1998; Davis et al., 1996; Foltz & Sullivan, 1996; Lindau et al., 2002; Sharp et al., 2002). The WRAT-R was employed in one study (Meade et al., 1994). Average reading levels ranged from Grade 4 to 9+ according to the REALM and WRAT-R. The Cloze procedure and TOFHLA, a modified version of Cloze, were not used in any of the cancer studies. The TOFHLA instrument was likely not used because of its very specific content that does not include cancer information. The Cloze and TOFHLA/S-TOFHLA are also more time-consuming than the REALM/REALM-R and WRAT-R.

None of the studies reported numerical test-retest results. Foltz and Sullivan (1996) provided limited test-retest procedures showing test stability of the REALM, although numerical results were not provided. Table 3 identifies the studies that measured patients' comprehension or pronunciation of written cancer information.

One additional study tested both the readability level and comprehension of cancer education materials (Cooley et al., 1995). The difficulty of 30 printed cancer pamphlets showed an average readability level of Grade 9 (F-K) and reading scores of Grade 9 (WRAT-R). Nevertheless, 27% of participants possessed reading skills below a Grade 6 level.

DISCUSSION

This review is the first to describe and evaluate multiple readability and comprehension instruments used for both print and Internet cancer information. Five readability formulas (Flesch Reading Ease, Flesch-Kincaid, Fog Index, SMOG, and Fry Readability

Table 3. Comprehension Studies on Print Cancer Information

Investigators	Population Tested (N)	Comprehension Tool Used ^a	Mean Comprehension Score
Lindau et al. (2002)	Women in ambulatory clinics (529)	REALM	63.0 = Grade 9+
Sharp, Zurawski, Roland, O'Toole, and Hines (2002)	African American women in colposcopy clinics (130)	REALM	45% at or below Grade 9
Bennett et al. (1998)	Low-income men in prostate cancer clinics (212)	REALM	< Grade 6 for 39.6% with Stage D cancers and for 24.8% with Stages A to C cancers
Davis et al. (1996)	Women aged 40+ in outpatient clinics (417)	REALM	40 = Grades 4 to 6
Foltz and Sullivan (1998)	Cancer patients (63)	REALM	50.8 = Grades 7 and 8
Meade, McKinney, and Barnas (1994)	Clinic patients (1,100)	WRAT-R	Grade 7

NOTE: REALM = Rapid Estimate of Adult Literacy in Medicine; WRAT-R = Wide Range Achievement Test-Revised.

a. Only formulas selected for review are included in this table.

Graph), two comprehension instruments (Cloze and TOFHLA/S-TOHFLA), and two word recognition tools (REALM/REALM-R and WRAT-R) were examined. Although this review represents a systematic investigation of the literature, certain references may not have been identified due to availability constraints.

The readability formulas were found to correlate reasonably well with each other, having high concurrent validity (.934 to .976; Dollahite et al., 1996; Glazer et al., 1996). Only one readability study presented interrater agreement results (Singh, 2003).

Although the Fog Index and SMOG formula require less time to calculate than other formulas (e.g., FRE, F-K, and Fry) because only polysyllabic words are counted, a number of the formulas (FRE, F-K) can be calculated quickly and automatically using computer software. SMOG tests 30 sentences (typically 600 words), which is a more robust measure than formulas using smaller samples of 100 words such as the Fog Index and Fry Readability Graph.

None of the formulas were designed to determine the effects of visuals or design factors, which are recognized as having influence on both the readability and comprehension of patient education information (L. G. Doak et al., 1996). Calculations accompanying original readability formulas that determine number of words per line versus amount of white space are recommended for the evaluation of print and Web-based information. Appropriate font size is especially important for Internet users who may not feel comfortable reading small print on a computer screen. Illustrations should clearly depict the concept being conveyed with informative, legible, and correctly labeled captions. Visual components should also reflect the subpopulations they are meant to represent (e.g., seniors' Web site should portray illustrations of older persons). It would be useful to create an assessment tool that would evaluate the appeal of diagrams and quantify changes in readability according to graphic and typographic considerations. Separate readability scores would also need to be calculated for captions of key diagrams that refer to and often clarify a particular piece of text. Despite the limitations of readability formulas, readability testing is quick, efficient, and involves a process that encourages writers to choose their words carefully and respectfully based on the reading level of the general population.

Critique of Readability Formulas

Although readability formulas are predictive measures of text difficulty, they have been criticized for relying on word and sentence factors and for ignoring possible effects of reader motivation, design, and graphics on readability and comprehension. Next, we identify several limitations of readability formulas applied to print materials in health (cancer) education.

Criterion Used in Formula Development

The readability formulas reviewed here were validated using passages from the McCall-Crabbs standard test lessons (McCall & Crabbs, 1961), which were not intended for the validation of readability formulas. The test lesson passages are too short in length to reveal higher-level characteristics of text organization that are known to help readers understand written text. Formulas may yield different reading grade levels depending on the passages used and the criterion of comprehension employed. For instance, SMOG is estimated to test for 100% complete comprehension and overrates word difficulty,

whereas Fog and Flesch estimate RGLs for lower 90% and 75% comprehension, respectively (Ley & Florio, 1996).

Word and Sentence Length Variables

A limitation of readability formulas is their atheoretical approach, relying solely on text properties such as word and sentence length (Bailin & Grafstein, 2001). Readability formulas may underestimate the difficulty of medical information for complex, monosyllabic words that are rated as easy to read according to the formulas (D'Alessandro, Kingsley, & Johnson-West, 2001).

Readability formulas are based on the assumption that there is a linear relationship between increased word and sentence length and increased reading ability requirements. It is assumed that problems arising from word difficulty compared with sentence difficulty remain constant and that skills needed for increased word difficulty rise at an equal proportion to skills needed for sentence difficulty. However, poor readers may have more trouble with word recognition than competent readers, and competent readers may find sentence structure more difficult than terminology compared to poor readers. Moreover, readability formulas cannot measure text structure and do not differentiate between ordered and scrambled sentences even though comprehension is influenced by word order.

Prior Knowledge and Motivation

Readability formulas overlook effects of reader motivation and prior knowledge on the readability of written information (Klare, 1984). The knowledge processes involved in deriving meaning from information are significant determinants of reading ability (Bailin & Grafstein, 2001; Klare, 1984). Interactive features of reading including previous experiences, motivation, and enthusiasm for the information are all important factors to consider when creating or providing cancer education materials to patients.

Interformula Scoring

An additional weakness is that RGL scores of text often differ between readability formulas, making it difficult to determine which formula yields the most accurate result. For example, F-K has produced reading levels four to five grade levels lower than those of Fry and SMOG (D'Alessandro et al., 2001). A solution to these discrepancies is to assess readability using multiple formulas, taking the score that produces the highest estimated RGL as overestimating difficulty will lead to less damage than underestimating information and considering it readable. An alternative solution is to use an average score or a range of scores from several formulas applied to the same passage. This procedure may generate a more reliable estimate than using a single formula (Meade & Smith, 1991). The use of multiple readability formulas improves the reliability of readability scores.

Visuals and Design Factors

None of the reviewed formulas measured the potential effects of illustrations, diagrams, or design, which are thought to improve the readability of cancer information (L. G. Doak et al., 1996). Text accompanied by realistic and tailored visuals can be more

than 40% more influential than text alone (Rehe, 1984). Culturally acceptable pictorial representations of women undergoing mammograms have been successful in promoting breast screening practices among African American women with limited literacy (Paskett, Tatum, Wilson, Dignan, & Velez, 1996).

Design factors influence the readability of health information. Font sizes of 12 point or larger assist with readability as does black print on a white background (Bradley, Singleton, & Li Wan Po, 1994; L. G. Doak et al., 1996). The FRE, F-K, SMOG, Fog Index, and Fry formulas test text difficulty without attention to these additional formatting variables that may affect readability (e.g., layout, font, captions, reduced amount of text due to inclusion of illustrations).

Critique of Comprehension and Word Recognition Tools

Test Administration

An advantage of using word recognition tools (REALM, REALM-R, WRAT-R) in a medical setting is that they can be administered in less than 5 minutes, are easy to score, and do not require intensive training on the part of the administrator. Alternatively, comprehension tests (Cloze, TOFHLA) are more time-consuming and require administrators to have adequate training in test delivery (Paskett et al., 1996). The S-TOFHLA was created to identify individuals with low functional health literacy in less time than the original TOFHLA. However, patients' ability to understand health information can be affected by anxiety and emotional distress.

The studies in this review used mainly clinic patients (Bennett et al., 1998; Cooley et al., 1995; Davis et al., 1996; Foltz & Sullivan, 1996; Gilliam et al., 1980; Lindau et al., 2002; Sharp et al., 2002). It would be important to determine the comprehension of community participants searching for print or Web-based cancer prevention information. Although research has shown that Internet users report difficulty in understanding and assessing the quality of health information on the Web (Chi-Lum, 1999), systematic studies of cancer comprehension on the Web have not been conducted.

Shame of Low Literacy

Although the majority of comprehension and word recognition tests are established as valid and reliable tools in the literature, the WRAT-R has low face validity because individuals unable to pronounce its difficult words have found the test to be overwhelming (Davis et al., 1994). Low literacy skills can be associated with shame and embarrassment, and individuals with basic literacy attempt to hide reading difficulties to avoid stigmatization by others. Individuals with low functional health literacy as measured by the TOFHLA have admitted to having feelings of shame because of their poor reading abilities (Parikh, Parker, Nurss, Baker, & Williams, 1996).

Inadequate literacy may not be only a reading problem. Health educators need to be aware that poor functional health literacy can also imply problems in understanding oral communication (Baker et al., 1998, 1999). There is a need to assess the benefits of audio, video, and computer-assisted education for individuals with basic literacy skills (Davis et al., 2002). Suggestions for improving the readability and comprehension of cancer information include choosing terminology carefully for print information and verbal explanations and not assuming that written words will be understood simply because they are recognized when spoken.

Applicability of Test Content

A significant weakness of comprehension instruments is that the content being tested may not be applicable to the health of the test population. For instance, the comprehension subtest of the TOFHLA and S-TOFHLA measures understanding of upper gastrointestinal series (Baker et al., 1999; Parker et al., 1995). However, the studies reviewed tested cancer patients or individuals at risk for cancer.

In addition, the WRAT-R was not designed to test people on medical terminology but rather to evaluate their pronunciation using a standard word list. The REALM/REALM-R is recommended by health literacy researchers and considered more suitable than the WRAT-R because it tests individuals on relevant medical terms (C. C. Doak et al., 1998).

Visuals and Design Factors

Similar to readability formulas, none of the comprehension or word recognition tests incorporated an evaluation of the effects of visuals and illustrations on the meaningfulness of comprehension test scores. Relevant visual aids that are simple, uncluttered, and easily identifiable have been seen to improve the retention of information and compliance with medical instructions (Davis et al., 2002; L. G. Doak et al., 1996).

Recommendations and Implications for Practice

Measuring the ability of people to read and understand cancer information is an important skill for health educators. Word recognition tests have been applied to individuals' understanding of print cancer information only. The use of specific readability and comprehension tools by health educators will depend on the individuals being tested and on the testing context. Table 4 provides recommendations on specific literacy tools for use with diverse target populations. For quick readability testing, the FRE or F-K formulas should be used because scores can be calculated automatically by computer software. However, if time permits, it is preferable to use multiple tools (i.e., SMOG and F-K) to ensure validity of the readability scores. The use of multiple assessment tools can also help improve the readability of a passage by documenting the range of the scores and revising text until readability is reduced to a suitable readability level. For readability assessment in Spanish, the Fry test is appropriate. Comprehension testing by health educators will depend on the intended populations. For brief screening for patient comprehension of printed text, a word recognition test (i.e., REALM/REALM-R) followed by a quick evaluation of functional health literacy (S-TOFHLA) could be used. Word recognition tests such as the REALM/REALM-R are a quick method of screening for reading ability and can provide an initial measure of an individual's literacy comprehension. Tests such as the S-TOFHLA offer a more detailed analysis of patients' understanding of authentic health care materials. The REALM used on its own may not provide an accurate portrayal of literacy because it can both over- and underestimate an individual's reading ability. The WRAT-R is not recommended for assessing health literacy because it is not based on medical terminology.

In neighborhood settings, such as a library or a local community center, a word recognition test along with a longer Cloze test would be appropriate to consider by health educators. However, individuals with basic literacy skills may be intimidated or uncomfortable by lengthy testing sessions; in this case, either a short Cloze test or semistructured interview questions may be useful for probing comprehension of cancer information. A

Table 4. Recommended Literacy Assessment Instruments for Different Populations

Target Group	Word Recognition Tool	Comprehension Tool
Patients	REALM/REALM-R depending on time available	TOFHLA/S-TOFHLA depending on time available
Community groups	REALM	Cloze, TOFHLA
Groups with basic literacy skills	REALM-R	Interview Questions or 1-paragraph Cloze test
Multicultural groups	Tools are not valid in languages other than English	Spanish TOFHLA/S-TOFHLA or interview questions

NOTE: REALM = Rapid Estimate of Adult Literacy in Medicine; REALM-R = Rapid Estimate of Adult Literacy in Medicine-Revised; TOFHLA = Test of Functional Health Literacy in Adults; S-TOFHLA = Shortened Test of Functional Health Literacy in Adults.

word recognition test would be inappropriate for individuals with basic literacy because it could have negative consequences for self-efficacy and self-esteem. Word recognition tools have not been validated in languages other than English and therefore would not be suitable for Spanish language applications. The TOFHLA/S-TOFHLA have been developed for testing in Spanish-speaking populations.

Effective communication of cancer information in print and Web-based formats needs to be tailored to the abilities of various subpopulations. Cancer education materials should be pretested for readability, comprehension, relevance, interest, and credibility. During the content development stages of cancer Web sites intended for individuals with basic literacy skills, cancer organizations, Web site developers, health writers, and graphic designers need to work in partnership and focus on clarity, plain language, and relevant illustrations. For individuals who require more detailed information, links to scientific information should also be provided. Clear communication by health care providers and health educators, Web site operators, editors, and writers of mass print publications is necessary to enable better understanding of cancer information disseminated to the public.

References

- Bailin, A., & Grafstein, A. (2001). The linguistic assumptions underlying readability formulae: A critique. *Language & Communication, 21*, 285-301.
- Baker, D. W., Parker, R. M., Williams, M. V., & Scott Clark, W. (1998). Health literacy and the risk of hospital admission. *Journal of General Internal Medicine, 13*, 791-798.
- Baker, D. W., Williams, M. V., Parker, R. M., Gazmararian, J. A., & Nurss, J. (1999). Development of a brief test to measure functional health literacy. *Patient Education and Counseling, 38*, 33-42.
- Bass, P. F., Wilson, J. F., & Griffith, C. H. (2001). A shortened instrument for literacy screening. *Journal of General Internal Medicine, 16*, 117.
- Bass, P. F., Wilson, J. F., Griffith, C. H., & Barnett, D. R. (2002). Residents' ability to identify patients with poor literacy skills. *Academic Medicine, 77*, 1039-1041.
- Beaver, K., & Luker, K. (1997). Readability of patient information booklets for women with breast cancer. *Patient Education and Counseling, 31*, 95-102.

- Bennett, C. L., Ferreira, M. R., Davis, T. C., Kaplan, J., Weinberger, M., Kuzel, T., et al. (1998). Relation between literacy, race, and stage of presentation among low-income patients with prostate cancer. *Journal of Clinical Oncology*, 16, 3101-3104.
- Berland, G. K., Elliott, M. N., Morales, L. S., Algazy, J. I., Kravitz, R. L., Broder, M. S., et al. (2001). Health information on the Internet: Accessibility, quality, and readability in English and Spanish. *Journal of the American Medical Association*, 285, 2612-2621.
- Bradley, M., Singleton, M., & Li Wan Po, A. (1994). Readability of patient information leaflets on over-the-counter (OTC) medicines. *Journal of Clinical Pharmacy and Therapeutics*, 19(1), 7-15.
- Butow, P., Brindle, E., McConnell, D., Boakes, R., & Tattersall, M. (1998). Information booklets about cancer: Factors influencing patient satisfaction and utilisation. *Patient Education and Counseling*, 33, 129-141.
- Chi-Lum, B. (1999). Friend or foe? Consumers using the Internet for medical information. *Journal of Medical Practice Management*, 14, 196-198.
- Chung, V., Horowitz, A. M., Cantos, M. T., & Siriphand, P. (2000). Oral cancer educational materials for the general public: 1998. *Journal of Public Health Dentistry*, 60(1), 49-52.
- Cooley, M. E., Moriarty, H., Berger, M. S., Selm-Orr, D., Coyle, B., & Short, T. (1995). Patient literacy and the readability of written cancer educational materials. *Oncology Nursing Forum*, 22, 1345-1351.
- Dale, E., & Chall, J. S. (1949). The concept of readability. *Elementary English*, 26, 23.
- D'Alessandro, D. M., Kingsley, P., & Johnson-West, J. (2001). The readability of pediatric patient education materials on the World Wide Web. *Archives of Pediatrics & Adolescent Medicine*, 155, 807-812.
- Davis, T. C., Arnold, C., Berkel, H. J., Nandy, I., Jackson, R. H., & Glass, J. (1996). Knowledge and attitude on screening mammography among low-literacy, low income women. *Cancer*, 78, 1912-1920.
- Davis, T. C., Crouch, M. A., Long, S. W., Jackson, R. H., Bates, P., & George, R. B. (1991). Rapid assessment of literacy levels of adult primary care patients. *Family Medicine*, 23, 433-435.
- Davis, T. C., Long, S. W., Jackson, R. H., Mayeaux, E. J., George, R. B., Murphy, P. W., et al. (1993). Rapid estimate of adult literacy in medicine: A shortened screening instrument. *Family Medicine*, 25, 391-395.
- Davis, T. C., Mayeaux, E. J., Fredrickson, D., Bocchini, J. A., Jr., Jackson, R. H., & Murphy, P. W. (1994). Reading ability of parents compared with reading level of pediatric patient education materials. *Pediatrics*, 93, 460-468.
- Davis, T. C., Michielutte, R., Askov, E. N., Williams, M. V., & Weiss, B. D. (1998). Practical assessment of adult literacy in healthcare. *Health Education & Behavior*, 25, 613-624.
- Davis, T. C., Williams, M. V., Marin, E., Parker, R. M., & Glass, J. (2002). Health literacy and cancer communication. *CA: A Cancer Journal for Clinicians*, 52, 134-149.
- Doak, C. C., Doak, L. G., Friedell, G. H., & Meade, C. D. (1998). Improving comprehension for cancer patients with low literacy skills: Strategies for clinicians. *CA: A Cancer Journal for Clinicians*, 48, 151-162.
- Doak, L. G., Doak, C. C., & Meade, C. D. (1996). Strategies to improve cancer education materials. *Oncology Nursing Forum*, 23, 1305-1312.
- Dollahite, J., Thompson, C., & McNew, R. (1996). Readability of printed sources of diet and health information. *Patient Education and Counseling*, 27, 123-134.
- Estey, A., Kemp, M., Allison, S., & Lamb, C. (1993). Evaluation of a patient information booklet. *Journal of Nursing Staff Development*, 9, 276-282.
- Estey, A., Musseau, A., & Keehn, L. (1994). Patient's understanding of health information: A multihospital comparison. *Patient Education and Counseling*, 24, 73-78.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233.
- Foltz, A., & Sullivan, J. (1996). Reading level, learning presentation preference, and desire for information among cancer patients. *Journal of Cancer Education*, 11, 32-38.
- Foltz, A., & Sullivan, J. (1998). Get real: Clinical testing of patient reading abilities. *Cancer Nursing*, 21, 162-166.

- Friedman, D. B., & Hoffman-Goetz, L. (2003). Cancer coverage in North American publications targeting seniors. *Journal of Cancer Education*, 18, 43-47.
- Friedman, D. B., Hoffman-Goetz, L., & Arocha, J. F. (2004). Readability of cancer information on the Internet. *Journal of Cancer Education*, 19, 117-122.
- Fry, E. (1968). A readability formula that saves time. *Journal of Reading*, 11, 513-516.
- Gausman Benson, J., & Forman, W. B. (2002). Comprehension of written health care information in an affluent geriatric retirement community: Use of the test of functional health literacy. *Gerontology*, 48, 93-97.
- Gilliam, B., Pena, S., & Moutain, L. (1980). The Fry graph applied to Spanish readability. *Reading Teacher*, 426-430.
- Glazer, H. R., Kirk, L. M., & Bosler, F. E. (1996). Patient education pamphlets about prevention, detection, and treatment of breast cancer for low literacy women. *Patient Education and Counseling*, 27, 185-189.
- Gribble, J. N. (1999). Informed consent documents for BRCA1 and BRCA2 screening: How large is the readability gap? *Patient Education and Counseling*, 38, 175-183.
- Grossman, S. A., Piantadosi, 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.
- Guidry, J. J., Fagan, P., & Walker, V. (1998). Cultural sensitivity and readability of breast and prostate printed cancer education materials targeting African Americans. *Journal of the National Medical Association*, 90, 165-169.
- Gunning, R. (1968). *The technique of clear writing*. New York: McGraw-Hill.
- Hafner, L. (1966). Cloze procedure. *Journal of Reading*, 9, 415-421.
- Humphris, G. M., Duncalf, M., Holt, D., & Field, E. A. (1999). The experimental evaluation of an oral cancer information leaflet. *Oral Oncology*, 35, 575-582.
- Jastak, S., & Wilkinson, G. S. (1987). *Wide range achievement test-revised*. Wilmington, DE: Jastak Associates.
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formula for navy enlisted personnel*. Millington, TN: Navy Research Branch.
- King, M. M., Winton, A. S. W., & Adkins, A. D. (2003). Assessing the readability of mental health Internet brochures for children and adolescents. *Journal of Child and Family Studies*, 12, 91-99.
- Klare, G. R. (1984). Readability. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 681-744). White Plains, NY: Longman.
- Ley, P., & Florio, T. (1996). The use of readability formulas in health care. *Psychology, Health & Medicine*, 1, 7-28.
- Lindau, S. T., Tomori, C., Lyons, T., Langseth, L., Bennett, C. L., & Garcia, P. (2002). The association of health literacy with cervical cancer prevention knowledge and health behaviors in a multiethnic cohort of women. *American Journal of Obstetrics and Gynecology*, 186, 938-943.
- McCall, W. A., & Crabbs, L. M. (1961). *Standard test lessons in reading*. New York: Teachers College Press.
- McLaughlin, G. H. (1969). SMOG grading—A new readability formula. *Journal of Reading*, 12, 639-646.
- Meade, C. D., McKinney, P., & Barnas, G. P. (1994). Educating patients with limited literacy skills: The effectiveness of printed and videotaped materials about colon cancer. *American Journal of Public Health*, 84, 119-121.
- Meade, C. D., & Smith, C. F. (1991). Readability formulas: Cautions and criterias. *Patient Education and Counseling*, 17, 153-158.
- Mohrmann, C. C., Coleman, E. A., Coon, S. K., Lord, J. E., Heard, J. K., Cantrell, M. J., et al. (2000). An analysis of printed breast cancer information for African American women. *Journal of Cancer Education*, 15, 23-27.
- Nurss, J. R., Baker, D. W., Davis, T. C., Parker, R. M., & Williams, M. V. (1995). Difficulties in functional health literacy screening in Spanish-speaking adults. *Journal of Reading*, 38, 632-637.

- Nutbeam, D. (1998). Health promotion glossary. *Health Promotion International*, 13, 349-364.
- Parikh, N. S., Parker, R. M., Nurss, J. R., Baker, D. W., & Williams, M. V. (1996). Shame and health literacy: The unspoken connection. *Patient Education and Counseling*, 27, 33-39.
- Parker, R. M., Baker, D. W., Williams, M. V., & Nurss, J. R. (1995). The test of functional health literacy in adults: A new instrument for measuring patients' literacy skills. *Journal of General Internal Medicine*, 10, 537-541.
- Parker, R. M., Williams, M. V., Baker, D., & Nurss, J. R. (1996). Literacy and contraception: Exploring the link. *Obstetrics and Gynecology*, 88(3), 72S-77S.
- Paskett, E. D., Tatum, C., Wilson, A., Dignan, M., & Velez, R. (1996). Use of a photoessay to teach low-income African American women about mammography. *Journal of Cancer Education*, 11, 216-220.
- Price, J. H., & Everett, S. A. (1996). Developing cancer pamphlets for economically disadvantaged African Americans. *Patient Education and Counseling*, 28, 159-167.
- Rehe, R. F. (1984). *Typography: How to make it more legible* (5th ed.). Carmel, CA: Design Research International.
- Sharp, L. K., Zurawski, J. M., Roland, P. Y., O'Toole, C., & Hines, J. (2002). Health literacy, cervical cancer risk factors, and distress in low-income African-American women seeking colposcopy. *Ethnicity & Disease*, 12, 541-546.
- Singh, J. (2003). Reading grade level and readability of printed cancer education materials. *Oncology Nursing Forum*, 30, 867-870.
- Slaten, D., Parrott, R., & Steiner, C. (1999). Readability of skin cancer prevention brochures targeting parents of young children. *Journal of the American Academy of Dermatology*, 40, 997-999.
- Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*, 30, 415-433.
- U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. (2000). *Healthy People 2010*. Available January 15, 2002, from <http://www.health.gov/healthypeople>
- U.S. National Cancer Institute. (1979). *Readability testing in cancer communications* (DHEW Publication No. 79-1689). Washington, DC: Department of Health Education and Welfare.
- Wilson, F. L., Baker, L. M., Brown-Syed, C., & Gollop, C. (2000). An analysis of the readability and cultural sensitivity of information on the National Cancer Institute's Web site: CancerNet. *Oncology Nursing Forum*, 27, 1403-1409.
- World Health Organization. (2000). *Health promotion—Provisional agenda item 3.2, EB107/4-December 1, 2000*. Retrieved January 15, 2002, from http://www.who.int/gb/EB_WHA/PDF/EB107/ee4.pdf.