© International Reading Association 0034-0553/93 USS1.25+.00

Bonnie J.F. Meyer
The Pennsylvania State University, University Park, PA, USA

Michael Marsiske
Max Planck Institute, Berlin, Germany

Sherry L. Willis
The Pennsylvania State University, University Park, PA, USA

Text processing variables predict the readability of everyday documents read by older adults

An eight-step model is presented to predict the difficulty older adults will have in reading documents. Components of this model come from theoretical and empirical work on discourse processing and include such factors as emphasis, structure, and position of the answer in a propositional analysis of everyday documents. The documents examined include charts, schedules, tables, labels, and forms. One aspect of this investigation is an examination of whether factors found to influence learning and memory from expository and narrative texts also influence comprehension from documents. Another important aspect of this work is the assessment of how well such a model predicts the readability of documents by older adults. A unique feature of this study is that it brings together the research domains of reading comprehension, aging, and document readability and literacy.

Discourse processing and aging

In the literature on discourse processing and aging (for reviews see Cohen, 1988; Hartley, 1989; Hultsch & Dixon, 1984; Meyer, 1987; Meyer & Rice, 1989; Zelinski & Gilewski, 1988), the majority of studies report age deficits in processing narrative and expository texts. The magnitude of age deficits ranges from minimal to large. For example, in self-paced leisure reading of expository text, older persons read at a 20% slower rate and remembered 10% fewer critical main ideas than young adults (Meyer & Rice, 1989). However, when the reading is paced, older adults are definitely disadvantaged, and the magnitude of age difference has been found to be dependent on the verbal ability of the individual: 53% fewer ideas remembered by elderly of average verbal ability and 20% fewer for high-verbal individuals (Vincent, 1985). These observed age differences may be tolerable in leisure reading, but when the elderly must adequately comprehend documents, missing some critical main ideas and failing to adjust strategies to take the extra time needed for adequate comprehension could be quite harmful or at least inconvenient. As a practical example, 80% of the elderly take prescription drugs (Murray, Darnell, Weinberger, & Martz, 1986); potential consequences from failure to adequately comprehend drug labels point to the importance of document literacy for older adults.

Working memory limitations, strategy differences, and the intrusion of irrelevant information are all reader variables that can lead to the poorer reading comprehension of older adults. Working memory appears to decline with aging: this decline has been attributed to loss of processing capacity and also to a decline in processing speed (Babcock & Salthouse, 1990; Salthouse, 1991; Salthouse, Mitchell, Skovronek, & Babcock, 1989). Age group differences in text recall are attenuated by educational level and verbal ability (Dixon, Hultsch, Simon, & von Eye, 1984; Hultsch, Hertzog, & Dixon, 1984; Meyer & Rice, 1981, 1983). Education may lead to the development and use of text-processing strategies which can in part compensate for some age-related declines in reading comprehension. Meyer, Young, and Bartlett (1989) taught older adults deficient in effective text processing to use a strategy that doubled their recall from texts. Effective use of such a strategy can help older adults to focus on the task and away from irrelevant thoughts; deficits in inhibitory mechanisms have been proposed as a source of age deficits in performance (Hasher & Zacks, 1988).

In addition to these reader variables, the literature on reading comprehension and aging shows the influence of a number of text variables. Text variables studied include syntactic complexity (e.g., left-branching sentences, Kemper, 1987), sensitivity to text structure (e.g., Dixon et al., 1984; Meyer & Rice, 1981, 1989), and signaling. Signaling is information in text that does not add new content about a topic but gives emphasis to certain aspects of the semantic content or points out aspects of the structure of the content (Meyer, 1975). Examples include words like *however*, but also include boldface or underlining, or salient positions in a chart or schedule for a document; in our model we call this factor emphasis. Signals are a way of cutting down the amount of processing required because high-level relationships are clearly delineated instead of having to be inferred by the reader or listener. In a study with young adults, Britton, Glynn, Meyer, and Penland (1982) found signaling to reduce the demands on cognitive capacity, as measured by reaction time to a secondary task (reactions to random clicks while reading text with and without signaling). Other studies with young learners have pointed to the facilitative effects of emphasis devices such as headings and typographical cues (e.g., Glynn, Britton, & Tillman, 1985; Hartley & Jonassen, 1985).

Meyer and Rice (1983) used multiple regression techniques to analyze data from 146 young, 127 middle, and 159 older adults. In these analyses, presence of signals had significant beta weights in the explanatory equations for total recall, recall of logical relations, structure of recall, and on the levels effect (main ideas remembered better than details). Equations were also calculated separately for each age group: Comparing across age groups, signals had a relatively larger effect on recall of logical relations and the levels effect (high-level main ideas remembered better than low-level details) for the older age group than it did for the middle and young groups.

Meyer and Rice (1989) noted that when no signals were supplied, the amount of high-level information recalled by elderly people varied inversely with the amount of specific detail in the text (e.g., specific detail = "1820," general detail = "early last century"). Data suggest that older adults can identify and store logical relationships in text when these are not explicitly signaled; that is, they are capable of making these inferences during self-paced reading. However, when specific details are present and signals are removed, older adults appear to process details at the expense of logical relationships and main ideas. The greater effects of signaling on these older adults than on young adults may result from reduced cognitive capacity or processing speed for older adults where effort spent in processing details reduces that available for figuring out implicit logical relationships among major propositions in a text.

Signals (Meyer & Rice, 1989) appear to improve recall of superordinate propositions accompanied by a slight drop in recall of details. Overall, Meyer and Rice (1989) found that the amount of information remembered was increased by signaling for young and older adults. Signaling appears to free processing resources so that they can be allocated more effectively.

Research indicates that aging deficits in discourse processing do not result from the lower-order processes involved in lexical access and comprehending sentences (e.g., Burke & Yee, 1984; Howard, 1980), nor from drawing inferences when the facts for these inferences are available (e.g., Light, Zelinski, & Moore, 1982). However, age deficits appear to occur whenever memory is taxed (e.g., Light & Burke, 1988). For example, Light et al. (1982) found large age deficits when the task required the element of the inference to be carried forward through the text by working memory to assure availability when the inference question was posed. Light and Albertson (1988) explain that older adults are just as likely to draw inferences as younger adults except when they cannot remember relevant facts. Prior propositions that cannot be remembered cannot be integrated with current information; the inability to retrieve relevant information from long-term memory into the working memory buffer in text processing would lead to a breakdown in the ability to form a coherent memory representation and in the ability to reason from document information.

Our conceptualization of prose processing integrates the work of a number of investigators in the domain of discourse processing (Fletcher & Bloom, 1988; Kintsch & van Dijk, 1978; Mayer, 1985; Meyer et al., 1989; Trabasso & van den Broek, 1985; van den Broek, 1989). Text is processed in cycles which roughly correspond to simple sentences. Information being processed takes capacity or space in working memory, as does a buffer of ideas (propositions) from previously processed sentences. The purpose of the buffer is to carry forward through the text those propositions which are central to the gist of the text. High-level propositions (main ideas) are hypothesized to be more frequently retrieved from long-term memory and kept in the working memory buffer as old propositions to be related to incoming new propositions in the reading process. This extra processing of high-level ideas leads to their better recall by readers throughout the adult life span.

Although the buffer size for a person is probably fixed, the contents of that buffer can be made more or less efficient by the strategy used by the reader. For example, a reader using the structure strategy (Meyer et al., 1989) would keep the overall organizational structure and ideas bound by it frequently in the buffer to be related to incoming ideas. This would result in efficient processing as the new information was simply connected to the buffer information to create a coherent text representation. However, a reader using a less systematic strategy would be less likely to find information in the buffer to connect to incoming ideas and would need to conduct time-consuming reinstatement searches into long-term memory. Signals or emphasis in texts can restate previous text information currently not in the working memory buffer so that a memory reinstatement search is not required. A heading in a chart should function in a similar manner. Most of the factors in our proposed model to predict document readability are based on this conceptualization of text processing.

Document literacy and aging

Document literacy (Kirsch & Jungeblut, 1986) involves reading skills necessary to understand and use printed materials occurring in a variety of nonprose formats. These formats include charts, schedules, tables, labels, and forms (Kirsch & Mosenthal, 1990). Older adults encounter such documents in their everyday lives (e.g., medicine bottle labels, directions for utilizing products, public transportation schedules, financial documents). The opportunity to function independently in our society may rest partly on the ability to comprehend everyday documents. Level of functioning on these tasks (e.g., comprehension of medication information, managing one's finances) has been shown to be predictive of timing of institutionalization and mortality (Fillenbaum, 1985). The elderly report that the ability to care for themselves and manage their own affairs is one of their major concerns. Maintenance of independent functioning in the elderly is also of concern for society as the number of elderly continues to increase.

Although an adequate level of functioning in everyday activities is critical for the elderly and for society, prior research indicates that many older adults perform poorly on everyday tasks involving printed material. Errors in self-administered medications are made by 40-60% of the elderly (Ouslander, 1981). Inability to comprehend prescription labels is a significant predictor of patient noncompliance; 24% of the elderly fail to understand the primary prescription label, and 39% misinterpret the auxiliary label (Murray et al., 1986). Many elderly have difficulty comprehending commonly encountered forms. In a national study, 33% of the adults could not comprehend a Medicare form and 10% could not interpret a personal loan form (Robeck & Wilson, 1977).

Research on the pervasiveness and significance of the elderly's problems with tasks of daily living has been limited due to previous studies focusing on a single domain of tasks (e.g., medications) and by the lack of adequate assessment instruments. Epidemiological studies have traditionally depended on the elderly's self-report of whether they can perform tasks of daily living without assistance (Fillenbaum, 1985; Ford, Haug, Jones, Roy, & Folmar, 1990). Likewise, social service providers have traditionally used self-report items in intake procedures (Fillenbaum, 1988). However, the few studies examining the relationship between the elderly's self-report and clinical assessment of functioning have consistently found that the elderly overestimate their level of competence (Fillenbaum, 1988; Ford et al., 1988).

Document literacy and the readability of documents are important issues for older adults of today who want to maintain independent functioning. This issue will become even more important over the next few decades; the number of Americans over the age of 65 is expected to double to 65 million by the year 2030.

The Educational Testing Service (ETS) Test of Basic Skills, 1977 edition, was the source of documents used in this investigation; this test requires readers to answer questions about charts (e.g., bus schedules), labels (e.g., plant spray labels and prescriptions), and forms (e.g., tax forms). Many of the items on this test represent instrumental activities required to remain independent in the community (Lawton & Brody, 1969).

Readability

Readability formulas yield a quick and easy way to predict readability, and as a result they have become a practical solution to the problem of estimating readability (Anderson & Davison, 1988; Klare, 1963, 1974-1975).

Currently there are over 30 readability formulas that can be used to estimate textbook difficulty. The formulas typically involve measures of sentence length and word difficulty to determine a grade level for the reading materials. The Dale-Chall (1948) formula and the Fry Readability Graph (1968) are two of the more frequently used measures. The former is based on average sentence length and percentage of words not on the Dale List of 3,000 words, while the latter involves counting syllables and number of sentences and then plotting them on a graph. Klare's (1974-1975) exhaustive review of readability formulas established that the formulas utilizing word length or difficulty and sentence length are sufficient to make relatively good predictions about readability. Criticisms of readability formulas focus on their incompleteness in characterizing texts of varying difficulty and a lack inherent in the simple formula for providing productive guidelines for creating readable texts (Davison & Kantor, 1982; Pearson, 1974-1975).

We developed our model to predict readability of everyday documents because most of the documents do not meet the requirements for applying the readability procedures. In applying the Fry readability formula, which is based on sentence length and word length, we found that the documents did not contain the recommended number of words. Fry recommends three 100-word samples; none of the documents contained 300 words. Less than a third of the documents even contained 100 words. In addition, many did not have sentences. Thus, readability formulas useful in predicting readability of text materials do not appear to be suited to appraising the readability of most documents. Due to the lack of an established approach to readability of documents, we devised the proposed model to predict readability of documents by older adults.

There have been many positive steps to simplify documents in recent years (e.g., Atwood, Baker, & Duffy, 1985; Chapanis, 1965; Waller, 1984; Wright, 1980, 1987). Many attempts to improve document readability have followed the "redesign-test-redesign-and-test" approach (Wright, 1979). This approach has been useful in improving individual documents, but lacks generalizability (Kirsch & Mosenthal, 1990). The model posited does not follow this approach either.

Instead, the model focuses on theoretical and empirical work on discourse processing and applies it to the reading comprehension of documents. Although the model was initially developed independently, it is similar to the approach of Kirsch and Mosenthal (1990) and ultimately has included some of their components. Kirsch and Mosenthal (1988) developed a propositional grammar to describe documents. They explained that this was necessary because existing grammars (e.g., Frederikson, 1975; Kieras & Dechert, 1985; Kintsch, 1977; Meyer, 1975) tended to apply to a particular type of discourse structure, not documents. The grammar utilized by Kirsch and Mosenthal (1990) is similar to the Meyer (1975, 1985a) approach. The model and grammar used in the present study are derived from Meyer (1975, 1985a) and the work of Kirsch and Mosenthal (1990).

Kirsch and Mosenthal (1990) identified a set of critical variables that underlie the performance of a national sample of young adults (ages 21 to 25 years) on the National Educational Progress (NAEP) study of young adult literacy. The variables identified accounted for 89% of the variance for the total population of 3,618 young adults. Among differing education levels, the variables accounted for 56% of variance for young adults with 0 to 8 years of schooling, 81% for young adults with 9 to 12 years of schooling, 88% for young adults with high school degrees, and 84% for young adults with post-high school degrees.

The variables in the Kirsch and Mosenthal study included six materials variables based on their grammar, three materials-by-task variables, and three response variables. The only materials variable that predicted performance dealt with the amount of information in the document, a measure of the number of arguments in the document. Kirsch and Mosenthal found two of their materials-by-task variables to predict performance. These variables related to the variable in our model called paraphrases.

The Kirsch and Mosenthal work is an important study and provides a foundation for a theoretical model of document processing. Many of the 12 variables identified in their study are complex, and considerable work is required to ascertain them from documents. The current model presented in our study is compatible with the orientation of Kirsch and Mosenthal, but offers a somewhat simpler approach. In addition, we have looked at a different population of document users, older adults rather than young adults.

Figure 1 displays the model developed to analyze the readability of everyday documents over a wide range of formats (e.g., charts, labels) and to derive an item readability score, based on this analysis. The model includes eight steps.

Step 1 (document size/number of propositions) is a measure of the amount of material to be learned and has been shown in previous literature to relate to readability (Kintsch & Vipond, 1979; Kirsch & Mosenthal, 1990).

Step 2 (position in content structure) in this scheme deals with the location of the answer to the question in the propositional analysis (content structure) of the document. Position in the content structure attributes greater difficulty and thus gives more points for a

Figure 1 Text-based model for assessing difficulty involved in processing everyday documents (readability level)

1. Document size/number of propositions
Count the number of organizing categories in the text

↓

2. Position in content structure
The level at which the answer to the question is embedded in the text

↓

3. Paraphrases

0. Literal match between answer and question
1. Simple word substitution (e.g., *drug, medication*)
2. Low text-based inference of phrase (e.g., *finish all, how long take*)
3. High text-based inference (e.g., *height* with *cm*): Answer and question at different levels of generalization
4. Match requiring prior specialized knowledge

↓

4. Mismatched relationships
Add 1 point for each relationship not identical between analyses of question and document

↓

5. Discourse structure
Discourse structure organizing the answer:
-1 = sequence, 0 = description, 1 = comparison
2 = causation, problem/solution

↓

6. Emphasis
Subtract one point if answer is emphasized (italics, boldface, capitals, initial item in text)

↓

7. Search complexity
Add 1 point for each place in content structure that must be searched to arrive at answer

↓

8. Plausibility of multiple choice foils

↓

Readability level

Note. Higher score indicates the material is more difficult to read.

question whose answer comes from the lower levels in the content structure. The rationale for this step comes from the large body of literature showing greater memory for information higher in the content structure than for that lower in the structure (e.g., Kintsch & van Dijk, 1978; Meyer, 1975; Meyer & McConkie, 1973). Also, Walker and Meyer (1980) found that information in sentences higher in the structure is more easily integrated than information in lower-level sentences.

Step 3 (paraphrases) involves adding points to the accumulating readability score for each content word in the question that is a paraphrase of a word in the document. The rationale for Step 3 comes from the work on comprehension questions of Anderson (1972).

Step 4 (mismatched relationships) adds one point for each relationship among the ideas in the question and those in the document that are not identical. The types of relations among the ideas in both the document and question are analyzed. Those relationships among the ideas that are not constant between the analyses of the document and of the question are counted and added to the readability score. Facilitation is expected when the relational structure of the question and the document are the same (e.g., Meyer, 1975; Thorndyke, 1977).

Step 5 (discourse structure) relates to the types of structures used to organize the answer in the content structure of the document. One point is added for each comparison relationship, and two points are added to the difficulty score for each causation or problem/solution relationship (Meyer, 1985a). No points are given for the easier structure of description (usually with enumeration) that appears to be very common in documents. One point is subtracted for the easiest structure, sequence. A number of investigations indicate that sequences are easier for learners to use than causation and problem/solution (Englert & Hiebert, 1984; Hidi & McLaren, 1988; Meyer et al., 1989; Raphael, Englert, & Kirschner, 1986, Richgels, McGee, Lomax, & Sheard, 1987). Other studies (Meyer & Freedle, 1984) have shown greater recall from structures with more inherent, structural links (e.g., comparison has more structural links than description) (Meyer, 1985b), if learners were able to use these more complex structures to organize their recall. However, as Horowitz (1987) points out, these facilitative effects on recall appear to interact with individual differences factors (Ellis, 1982; Rickards & Slife, 1987; Vincent, 1985), task, and topic. In light of the investigations pointing to ease of use of sequences, discourse structure was included as one factor in our model.

Step 6 (emphasis) deals with the highlighting of the answer in the document. If the answer is emphasized through initial placement in the document, signaling, boldface, capitalization, or some other technique, then less difficulty is anticipated for the task. Such emphasis is thought to ease the processing load (Britton et al., 1982). As mentioned in the review of discourse processing and aging above, facilitation from such emphasis devices has been reported (Brown & Smiley, 1977; Kieras, 1985; Marshall & Glock, 1978-79; Meyer, 1985a, 1985b; Meyer, Brandt, & Bluth, 1980).

Step 7 (search complexity) involves how many places in the content structure a subject would have to examine to find the answer. One point is added for each place in the content structure that it is necessary to examine in order to arrive at an answer. Charts and tables often require a search of two or more places in the content structure. This step is similar to both rational task analysis (e.g., Gagne, 1985) and componential analysis (e.g., Sternberg, 1977).

The final step 8 (plausibility of multiple choice foils) was added for use with multiple choice test items. As the number and relationships among alternative answers increase, the difficulty level of the item increases (Brown, 1986; Drum, Calfee, & Cook, 1981; Kirsch & Mosenthal, 1990).

Documents and questions that yield low scores on these steps in the model should be easier for older adults to answer. Our sample of older adults was average to slightly below average in terms of education. Seventy percent had 0-12 years of schooling. High school-educated older adults show greater deficiencies in use of effective prose learning strategies and recall than comparable young adults and college-educated older adults (Meyer & Rice, 1983, 1989). Since the older adults may have ineffective strategies and limitations in short-term memory, documents and questions with low values on the variables from the model should be particularly helpful. That is, easier reading should result from increased saliency of information (position in the content structure and emphasis), literal questions posed in a manner similar in structure to the document, and fewer propositions to search and compare in memory.

The study examines the efficacy of this proposed model for predicting the reading comprehension of everyday documents by older adults.

Method

Sample

The sample was composed of 482 community-dwelling Caucasian Pennsylvania elders (405 females, 77 males), ranging in age from 52 to 93 years (M = 73.39

years, $SD = 6.60$ years). Of the subjects, 232 were participants in a 7-year follow-up of the Adult Development and Enrichment Project (ADEPT) conducted in 1986. The rest of the sample ($n = 250$) were subjects from the same geographical locations as the original sample, but tested for the first time in 1986. Educational level ranged from 3 to 22 years, ($M = 11.59$ years, $SD = 2.89$ years). Mean annual income reported ranged from under US$2,000 to over $28,000 (*median* = $9,876.46, SD = $10,153.13). Married persons constituted 32.8% of the sample, and the remaining 67.2% of the sample included widowed, single, divorced, and separated persons. Only 6.7% of the sample was employed full or part time, with the remainder of the subjects reporting themselves as homemakers or retired. Self-reported health, hearing, and vision ranged from 1 (very good) to 6 (very poor). Mean health self-rating was 2.11 ($SD = 0.88$). Mean hearing self-rating was 2.53 ($SD = 1.08$). Mean vision self-rating was 2.50 ($SD = 0.92$). Subjects reported a mean life satisfaction rating of 2.88 ($SD = 1.10$), with a range of 1 (extremely happy) to 6 (very unhappy).

Subjects were paid for their participation at a rate of US$5.00 per hour. Participation involved 5-10 hours of testing (over a 2-week period), plus a take-home questionnaire packet.

Table 1 The ADEPT ability battery: Second-order dimensions, first-order primary mental abilities, and selected marker tests

General dimension	Primary ability	Test	Source
Gf	CFR	Culture Fair Test (Scale 2, Form A) and Power Matrices (Scale 3, Form A, 1936 ed., and Form B, 1961 ed.)	Cattell & Cattell (1957, 1961, 1963)
Gf	I	ADEPT Induction Diagnostic Test (Form A)	Blieszner, Willis, & Baltes (1981)
Gf/Gc	CMR	Word matrix	Guilford (1969)
Gc	EMS	Social translations	Horn (1967)
Gc	V	Verbal meaning	Thurstone (1962)
Ms	Ms	Auditory number span—delay	After Ekstrom et al. (1976)
Ps	Ps	Number comparison	Ekstrom et al. (1976)

Measures

Ability battery. The psychometric ability battery used in the present study was developed within the fluid (Gf) and crystallized (Gc) model of intelligence (see Cattell, 1971). While multiple-marker tests of each of seven primary mental abilities were used, the present study reports results using only one marker for each ability. These seven tests were previously identified as the purest markers (i.e., highest factor loadings) of the abilities they represent (Baltes, Cornelius, Nesselroade, Spiro, & Willis, 1980). Table 1 presents the ability battery, including the specific measures used, the primary mental abilities they represent, and the broad second-order dimensions on which the primary abilities load.

The primary abilities of figural relations (CFR) and induction (I) were selected to represent fluid intelligence. Subjects must observe patterns of figures or letters in marker tests of these abilities and educe the pattern of relationships contained within them. Crystallized intelligence (Gc) was represented by marker tests of experiential evaluation (EMS), which required subjects to manipulate problems of a social nature, and verbal comprehension (V); the measure of verbal comprehension was a recognition vocabulary test. Semantic relations (CMR) measures, which require subjects to complete verbal analogies, were used to represent both fluid and crystallized intelligences. Two other primary abilities were represented in the present study: memory span (Ms), which was assessed using a test of digit span; and perceptual speed (Ps). Perceptual speed measures assess the speed with which subjects make simple visual discriminations.

All but one of the measures were adapted versions of published psychometric ability tests. Test adaptations consisted mainly of enlargement of test stimuli, modification of response format to simplify test taking for older adults, and reduction of the number of test items. All reduced measures had alpha reliabilities above 0.65. The ADEPT induction test was not previously published and was first developed in the initial phase of ADEPT (Blieszner, Willis, & Baltes, 1981; Willis, Blieszner, & Baltes, 1981).

Everyday task measure. The Educational Testing Service's Test of Basic Skills was the measure of everyday task performance used in the present study. The measure contains 65 items and measures subjects' ability to comprehend printed materials, including charts and forms (e.g., bus schedule), labels (e.g., plant spray label), and technical documents (e.g., warranty).

Procedure

Subjects were assessed in small groups (3-12) by a tester and a proctor. All testers and proctors ($n = 6$) ranged from young to middle adulthood. A small number of subjects required individual testing sessions, due to transportation or sensory difficulties. Testing sessions were usually held over a 2-day to 3-week interval. Tests were administered in community settings, usually the senior citizen centers to which participants belonged.

Derivation of scores

Standardization of ability data. Ability data for all subjects in 1986 were standardized to their 1979 base (M = 50, SD = 10). This procedure was done to preserve change in test performance levels from 1979 to 1986.

Score derivations for everyday task measure. To examine the utility of the readability scoring procedure, a subset of 62 items was selected from the 65-item everyday task measure. These items were representative of a variety of item formats (charts, forms, labels, everyday prose). Three items were eliminated because they involved the interpretation of graphic and pictorial information (e.g., cartoons, maps) and did not fit well into a linguistic parsing scheme.

Individual item responses were recoded into a dichotomous (1 = right, 0 = wrong) format. The maximum score on the 62 items was therefore 62.

The proportion of subjects ($N = 483$) who incorrectly answered each of the 62 selected items was computed. This proportion incorrect ranged from 2.4% to 70.5%, with a mean of 24.12% ($SD = 16.63\%$). The sample proportion answering an item incorrectly served as a second criterion for determining item difficulty (a higher proportion of subjects would be expected to incorrectly answer a hard item relative to an easy one), against which the readability score components could be compared.

Derivation of readability scores. Figure 1 depicts the model, based on research in discourse processing, that was utilized to derive readability scores. The model includes eight kinds of variables that were expected to influence the difficulty of comprehending the items on the test of everyday memory. Appendix A presents an example of a document and two questions that are similar to those found on the everyday memory test; due to security concerns for the Test of Basic Skills, similar rather than actual materials are displayed. Appendix B also lists the scores for the two questions on the factors from the model. The first variable dealt with the number of propositions in the document (as indexed by the number of organizing categories; see Kirsch & Mosenthal, 1990). The second variable dealt with the position of the answer to the question in the content structure of the analyzed document. The second element of the difficulty model for an item was, therefore, the number of points equal to level of answer in content structure of the document.

The third variable focused on paraphrases. Points were added for paraphrased words/phrases in the question and answer that must be matched with related words/phrases in the document. As seen in Figure 1, from 0 to 4 points could be given for a paraphrase depending on the variance between the words to be matched. No points were given when the match was literal. One point was given for a paraphrase that involved a simple word substitution (e.g., *drug* for *medication*, *address* for *where one lives*, *twice a day* for *two times a day*, *a gallon* for *one gallon*). Two points were added to the difficulty score for a paraphrase that involved a string of words or required prior knowledge of common English language usage (e.g., *in combination* for *at same time*, *years of service* for *after working x years*, *other study* for *learned how to repair radios in the Navy*). Three points were given when phrases in the text and in the question were at different levels of generalization and required inferences to match the concepts (e.g., height with a specific measured height, such as cm). Four points were added for paraphrases that required the reader to have prior specialized knowledge in order to match text material and answer (e.g., *250 mg* for *.250 g*, *gallon* for *4 quarts*, knowing both the generic and brand name of a drug). For Question 1 about the application for employment, no paraphrases were found (score = 0), while 3 points were given to Question 2: *address* matched to *location of school* for 1 point and *learned how to repair radios in the Navy* matched to *other study* for 2 points.

The fourth variable was the number of mismatched relationships between the document and the question. First, a propositional analysis (Meyer, 1975, 1985a) was completed to determine the relationship structure of the document and the question. Then, the relationships among ideas (verbatim or paraphrased) were examined to see if they were the same in the analysis of the question as they were for the analysis of the document. Each difference noted in this comparison added one point to the difficulty scores. Care was taken to be sure that the document and the question had been analyzed to the same degree of specificity so that any difference did not simply reflect a difference in the level of analysis.

The fifth variable related to the types of structures used to organize the answer in the content structure of the document. One point was subtracted for each sequence relationship. As seen in Appendixes A and B, spaces labeled 1 through 9 in the application qualified this document as organized by a sequence, and one point was subtracted. One point was added for each comparison relationship, and 2 points were added to the difficulty score for each causation or problem/solution relationship (Meyer, 1985a).

The sixth variable seen in the model relates to emphasis. One point was subtracted from the readability score each time the answer was emphasized in the document (e.g., an answer could be emphasized by appearing as the first item in the document, by underlining, boldface print, large type, or other types of signaling; see Meyer, 1985b). For Question 1, address is emphasized in the document by capitalization, and 1 point was subtracted.

The seventh variable involved search complexity. One point was added for each place in the content structure that it was necessary to examine in order to arrive at an answer. Charts and tables often required a search of two or more places in the content structure. Question 1 in Appendix A required search in only one place (*address*) in the structure, while Question 2 required looking at two places (*location of school* and *other study*).

The eighth and final variable indexed the plausibility of multiple choice foils. When none of the multiple choice distractors appeared in the text, a score of 1 was assigned. As the number of distractors increased, as the critical features shared by the answer and the distractors increased, and as the proximity of the distractor to the answer in the content structure increased, a higher plausibility score was assigned, to a maximum of 5. This model element was adapted from one presented by Kirsch and Mosenthal (1990), although the direction of scoring was reversed. The distractors for the questions in the Appendixes shared the critical feature of belonging to the class of numbers with the answer, but they occurred under different organizing categories than the answer and thus were not very difficult.

As will be described below, the individual components of this readability model for each item were simultaneously entered into a multiple regression, with the sample proportion incorrect for each item as the dependent variable. The obtained beta weights provided a formula for weighting and summing model elements into a total "readability score." Higher readability scores meant harder items.

Results

Two questions were addressed: (a) What is the relationship between an empirical estimate of item difficulty (i.e., sample proportion incorrect on each item) and the theory-based estimate of item difficulty (i.e., the model components contained in the readability score), and (b) What is the relationship between ability performance and item difficulty, as determined by the item readability scores? That is, is there a stronger relationship with ability performance for items having higher (more difficult) readability scores than for items having lower readability scores?

Table 2 Correlation between sample proportion incorrect and readability model components for 62 ETS basic skills items

Model component	Percentage incorrect
Document size/number of propositions	-0.194[a]
Position in content structure	0.229[a]
Paraphrases	0.295*
Mismatched relationships	0.288*
Discourse structure	0.017
Emphasis	0.101
Search complexity	0.094
Plausibility of multiple choice foils	0.216[a]

[a] $p < .20$.
* $p < .05$.

Table 3 Full model: Prediction of item difficulty (sample proportion incorrect) from readability model components for 62 ETS basic skills items

Independent variable	Parameter estimate	*T*-value for parameter
Intercept	-8.24	-9.92***
Document size/number of propositions	-0.75	-2.14*
Position in content structure	3.16	1.28[b]
Paraphrases	2.43	2.05*
Mismatched relationships	2.58	1.87[a]
Discourse structure	-1.72	-0.81
Emphasis	-2.62	-0.38
Search complexity	-0.51	-1.48[b]
Plausibility of multiple choice foils	7.04	2.37*

Notes. Dependent variable: sample proportion incorrect. Model $R^2 = 0.32$, $p < .01$.
[a] $p < .10$. [b] $p < .20$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Relationship between text-based model components and item difficulty. With regard to the first question, the proportion of subjects answering each item incorrectly was correlated with each of the eight readability score components for that item. The average correlations across the 62 items are shown in Table 2. In addition, the sample proportion incorrect was regressed on the readability model components. Table 3 presents a summary of the multiple regression results.

Based on the regression results reported above, a reduced regression model was estimated containing only significant and near significant ($p < .20$) parameter estimates from the original full-model regression. These reduced regression results are reported in Table 4. Using the results of the reduced regression model, *a readabili-*

Table 4 Reduced model: Prediction of item difficulty (sample proportion incorrect) from readability model components for 62 ETS basic skills items

Independent variable	Parameter estimate	*T*-value for parameter
Intercept	-9.95	-11.00***
Document size/number of propositions	-0.49	-1.90[a]
Position in content structure	2.91	1.32[b]
Paraphrases	2.30	2.01*
Mismatched relationships	2.55	1.92[a]
Search complexity	-0.55	-1.73[a]
Plausibility of multiple choice foils	7.42	2.56**

Notes. Dependent variable: sample proportion incorrect. Model $R^2 = 0.29$, $p < .003$.
[a]$p < .10$. [b]$p < .20$.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 5 Correlation of easy and hard everyday task measure item subsets with primary mental ability marker tests (sample size for each correlation in parentheses)

	Hard item subset	Easy item subset	*T*-value for difference between hard and easy correlations[a]
Figural relations (fluid)	0.67 (n = 482)	0.63 (n = 482)	2.33**
Induction (fluid)	0.61 (n = 482)	0.55 (n = 482)	3.00**
Perceptual speed (speed)	0.52 (n = 445)[b]	0.50 (n = 445)	0.52
Verbal ability (crystallized)	0.69 (n = 353)[b]	0.67 (n = 353)	1.15
Experiential evaluation (crystallized)	0.59 (n = 353)[b]	0.51 (n = 353)	-3.13**
Semantic relations (fluid/crystallized)	0.43 (n = 353)[b]	0.40 (n = 353)	-1.08
Memory span (memory)	0.53 (n = 353)[b]	0.47 (n = 353)	-2.00*

[a]*T*-ratio was adjusted for changes in *df* and correlation between easy and hard item subsets.
[b]Ability data for this measure was available for a subset of individuals.
*$p < .05$. **$p < .01$.
Note. All correlation coefficients were significant ($p < .0001$). Difference between the two correlations is significant; hard subset score is more highly correlated with ability than easy subset score.

ty score was generated for each ETS item. The readability score was defined as the *estimated item difficulty* produced when the model-based difficulty components (i.e., number of propositions, content level of answer in text, paraphrases, number of mismatched relationships, number of search points, and plausibility of multiple choice foils) were weighted by their regression estimates and summed.

Relationship between readability scores and psychometric ability performance. The model-based readability scores were correlated with the sample proportion incorrect for the 62 ETS items; the obtained Pearson product-moment correlation coefficient was 0.54.

In order to investigate the relationship between item readability and psychometric ability performance, the 31 dichotomously scored (0 = wrong, 1 = right) ETS items that received the highest readability scores were summed into a "hard" item subset (maximum possible score = 31). The 31 ETS items with the lowest readability scores were summed into an "easy" item subset (maximum possible score = 31). Subjects' scores on these two item subsets were then separately correlated with their scores on the standardized marker tests of each of the seven primary mental abilities. Table 5 presents the resulting correlation matrix.[1]

Easy and hard item subsets differed significantly in their correlations with the psychometric abilities of figural relations (CFR), induction (I), experiential evaluation (EMS) and memory span (Ms) (p < .05). In all cases, the hard item subset was more highly correlated with psychometric ability performance than the easy item subset. These significant differences were obtained for markers of fluid (I, CFR) and crystallized (V, EMS) intelligences, as well as for a marker of general memory (Ms).

Discussion

The results obtained in this study present some interesting findings concerning the utility of applying discourse processing techniques to paper-and-pencil measures of everyday task performance. The significant correlation coefficients obtained when readability scores (obtained through the discourse-processing analysis model outlined above) were correlated with sample proportion incorrect on ETS basic skills items suggest that a text-processing analysis of test materials may be useful in predicting and accounting for sample response patterns. This has both theoretical and practical significance. Theoretically, it suggests that the dimensions of text readability identified in prior research *are* salient for predicting actual comprehension in *everyday* task materials for older adults. Consequently, the present findings are an extension of prior findings, using more ecologically valid materials and text formats than may have been used in prior research.

Practically, the text-processing model outlined in the present paper presents a useful algorithm for the selection of test items. Generation of readability scores provides an a priori method for varying item difficulty in a text-based test. Traditionally, items could only be excluded following the results of pilot testing on standardization samples. Unfortunately, such an approach could be very sample dependent; a particular sample may have different response patterns than other samples. Since pilot tests are often conducted on convenience samples, inferences about response patterns in a broader population from a pilot sample may be dangerous. The present analyses suggest that it may be possible to evaluate item properties even before pilot testing, providing an initial estimate of item quality that is less sample dependent. Further, the present results imply that exclusion of items from such a test may now be guided by theoretical considerations; the present item difficulty analysis system provides *reasons* for excluding particular items. Traditional exclusion heuristics, based solely on sample response patterns, are largely atheoretical.

Two important caveats to the above interpretation should be stated. First, the findings are based on only a limited set of everyday task materials. A larger, broader set of everyday texts and problems is needed to replicate the predictive utility of our readability scores. Second, the findings are based on a sample of 482 older adults residing in Pennsylvania; generalization of findings to a larger sample of older adults and to other age groups needs to be examined. At the same time, it should be noted that the mean educational level of our sample (11.59) was close to the national average (12) for those over 65. Kirsch and Mosenthal's (1990) findings that the ability to predict item difficulty was reduced with lower-education subjects is relevant to the present study. Kirsch and Mosenthal could account for 56% of the variance for adults ages 21 to 25 with 0 to 8 years of schooling, while they account for 89% of the variance with their entire sample of 3,618 young adults. Future research needs to examine the usefulness of our text-processing-based item-difficulty model for understanding the everyday printed task performance of other groups of older adults, as well as other age groups. Despite these limitations, our findings, taken together with those of Kirsch and Mosenthal (1990) with young adults, present support for the validity of this technique for predicting the readability of documents and tests querying such documents.

Since we wanted to explore and evaluate a theoretically based model of readability, we chose to simultaneously evaluate a number of conceptually important readability components. We adopted a lenient alpha criterion in our multiple regression models because our interest was in evaluating as comprehensive a model as possible. The overall multiple regression model accounted for a significant proportion of the variance in item difficulty: $R^2 = .29, p < .01$.

The findings of differential magnitude of correlation with abilities for hard and easy item subsets (as defined by readability scores) suggest an additional way of understanding differential item readability. Hard items correlate significantly more than easy items with fluid intelligence abilities (figural relations and induction), with crystallized intelligence abilities (experiential evaluation), and with memory span. Although the magnitude of these correlation differences is small, the hard items are consistently more strongly related to the mental abilities than the easy items. This pattern is compatible with the notion that more difficult items may require a larger crystallized knowledge base on which to draw. They may also require a greater application of processing strategies and reasoning to text materials (i.e., greater fluid intelligence). The memory demands that difficult items present may be greater (this last point agrees well with Meyer's previous research, which suggests that the elements that make an item more "difficult" also tend to make text less memorable). The present findings therefore suggest that text elements that make reading materials more difficult may do so because they increase the intellectual demands of text processing. Future research will need to more systematically explore how performance under varying conditions of the readability model relates to intellectual ability.

The present analyses focused specifically on the usefulness of a discourse-processing approach for understanding item difficulty on a measure of everyday task performance. The text materials used in the present study were similar to materials encountered in the everyday world: bus schedules, technical documents, charts, forms, directions, product labels. Several intervention implications emerge from our findings. First, it may be possible to analyze the printed materials encountered in everyday life for readability. Second, it may be possible to suggest a strategy for increasing the readability of everyday text. Third, it may be possible to design intervention strategies that help individuals to attend to salient aspects of everyday text and to improve their comprehension of these materials.

Since the present study was conducted with a sample of older adults, the intervention implications of these findings become increasingly important. The opportunity to remain independent in the community may rest largely on the ability to comprehend everyday text materials (e.g., medicine bottle labels, public transportation schedules, financial documents) (Walmsley & Allington, 1982). A heuristic that is helpful in understanding the locus of age-related deficits (if any) in dealing with everyday printed materials can also help in the design of intervention strategies for these older adults. It follows that an effort to improve older adults' abilities to deal with everyday texts may actually help to improve their fitness for an independent life in the noninstitutional community. Further research needs to examine these issues.

The present study, focusing on text factors, must be embedded within a larger model of literacy considering the complex social context in which older adults use documents in everyday life. The current model does not incorporate social or motivational factors. The social context of reading a bus schedule or a prescription label was not examined; there was no attempt to describe how the social or motivational context in everyday life would help to define the form and function of authentic literate actions. Document literacy involves more than simply performing well on written questions about documents; it involves comprehending documents for specific purposes and specific contexts of use. Future research will need to incorporate social and motivational factors.

In summary, the present findings may be helpful in understanding four related issues. First, they provide further evidence that text factors help to determine the readability and comprehensibility of text. Second, they suggest that ecologically valid texts, encountered in the everyday world, may also be more or less difficult due to text factors intrinsic in them. Third, the results suggest that more difficult texts may be more difficult because they impose higher intellectual/processing demands, particularly in the domains of fluid and crystallized intelligences, and in working memory. Fourth, the results suggest that it may be possible to construct highly comprehensible everyday printed materials that reduce the intellectual demands imposed on the reader. A large-scale revision of texts found in the everyday world seems like an unlikely short-term goal. Consequently, the results also suggest that it may be possible to instruct older adults in reading strategies geared to comprehending documents. Strategy training has proved helpful for older adults in both reading expository texts (Meyer et al., 1989) and performing on intelligence tests (Willis, 1990). Thus, strategy training focused on reading and using documents successfully may improve older adults' abilities to deal with the documents they encounter in everyday life.

REFERENCES

ANDERSON, R.C. (1972). How to construct achievement tests to assess comprehension. *Review of Educational Research, 42*, 145-170.

ANDERSON, R.C., & DAVISON, A. (1988). Conceptual and empirical bases of readability formulas. In A. Davison & G. Green (Eds.), *Critical approaches to readability* (pp. 23-54). Hillsdale, NJ: Erlbaum.

ATWOOD, N., BAKER, E., & DUFFY, T.M. (1985). *Cognitive approaches to the analysis of technical understanding.* (CDC Tech. Rep. No. 24). Pittsburgh, PA: Carnegie Mellon University, Communications Design Center.

BABCOCK, R., & SALTHOUSE, T.A. (1990). Effects of increased processing demands on age differences in working. *Psychology and Aging, 5*(3), 421-428.

BALTES, P.B., CORNELIUS, S., NESSELROADE, J.R., SPIRO, A., & WILLIS, S.L. (1980). Integration versus differentiation of fluid/crystallized intelligence in old age. *Developmental Psychology, 16*, 625-635.

BLIESZNER, R., WILLIS, S.L., & BALTES, P.B. (1981). Training research in aging on the fluid ability of inductive reasoning. *Journal of Applied Developmental Psychology, 2*, 247-265.

BRITTON, B.K., GLYNN, S.M., MEYER, B.J.F., & PENLAND, M.J. (1982). Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology, 74*, 51-61.

BROWN, A.L., & SMILEY, S.S. (1977). Rating the importance of structural units of prose passages: A problem in metacognitive development. *Child Development, 48*, 1-8.

BROWN, J. (1986). Similarity relations in recognition: A model to explain Tulving's data. *Journal of Memory and Language, 25*, 197-201.

BURKE, D.M., & YEE, P.L. (1984). Semantic priming during sentence processing by young and older adults. *Developmental Psychology, 20*, 903-910.

CATTELL, R.B. (1971). *Abilities: Their structure, growth and action.* Boston, MA: Houghton-Mifflin.

CHAPANIS, A. (1965). Words, words, words. *Human Factors, 7*, 1-17.

COHEN, G. (1988). Age differences in memory for texts: Production deficiency or processing limitations? In L.L. Light & D.M. Burke (Eds.), *Language, memory, and aging* (pp. 171-190). New York: Cambridge University Press.

DALE, E., & CHALL, J.S. (1948). A formula for predicting readability. *Educational Research Bulletin, 27*, 11-20, 37-54.

DAVISON, A., & KANTOR, R.N. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading Research Quarterly, 18*, 187-209.

DIXON, R.A., HULTSCH, D.F., SIMON, E.W., & VON EYE, A. (1984). Verbal ability and text structure effects on adult age difference recall. *Journal of Verbal Learning and Verbal Behavior, 23*, 569-578.

DRUM, P.A., CALFEE, R.C., & COOK, L.K. (1981). The effects of surface structure variables on performance in reading comprehension tests. *Reading Research Quarterly, 16*, 486-514.

EKSTROM, R.B., FRENCH, J.W., HARMAN, H., & DERMAN, D. (1976). *Kit of factor referenced cognitive tests* (revised). Princeton, NJ: Educational Testing Service.

ELLIS, J.P. (1982). *Hemispheric specialization for the processing of prose and maps.* Unpublished doctoral dissertation, Arizona State University, Tempe.

ENGLERT, C.S., & HIEBERT, E.H. (1984). Children's developing awareness of text structures in expository materials. *Journal of Educational Psychology, 76*, 65-74.

FERGUSON, G.A. (1953). *Statistical analysis in psychology and education.* New York: McGraw-Hill.

FILLENBAUM, G. (1985). Screening the elderly: A brief instrumental activities of daily living measure. *Journal of the American Geriatrics Society, 33*, 698-706.

FILLENBAUM, G. (1988). *Multidimensional functional assessment of* *older adults. The Duke older Americans resources and services procedures.* Hillsdale, NJ: Erlbaum.

FLETCHER, C.R., & BLOOM, C.P. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language, 27,* 235-244.

FORD, A., FOLMAR, S., SALMON, R., MEDALIE, J., ROY, A., & GALAZKA, S. (1988). Health and function in the old and very old. *Journal of the American Geriatrics Society, 36,* 187-197.

FORD, A., HAUG, M., JONES, P., ROY, A., & FOLMAR, S. (1990). Race-related differences among elderly urban residents: A cohort study, 1975-1984. *Journal of Gerontology, 45,* 163-171.

FREDERIKSON, C. (1975). Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology, 7,* 371-458.

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

GAGNE, R.M. (1985). *The conditions of learning.* New York: Holt, Rinehart, & Winston.

GLYNN, S.M., BRITTON, B.K., & TILLMAN, M.H. (1985). Typographical cues in text: Management of the reader's attention. In D.H. Jonassen (Ed.), *The technology of text* (Vol. 2, pp. 192-209). Englewood Cliffs, NJ: Educational Technology.

HARTLEY, J.T. (1989). Memory for prose: Perspective on the reader. In L.W. Poon, D.C. Rubin, & B.A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 135-156). Cambridge, England: Cambridge University Press.

HARTLEY, J.T., & JONASSEN, D.H. (1985). The role of headings in printed and electronic text. In D.H. Jonassen (Ed.), *The technology of text* (Vol. 2, pp. 237-263). Englewood Cliffs, NJ: Educational Technology.

HASHER, L., & ZACKS, R.T. (1988). Working memory, comprehension, and aging: A review and a new view. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193-226). New York: Academic Press.

HIDI, S., & MCLAREN, J. (1988). *The development of text structures in children's written expositions.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

HOROWITZ, R. (1987). Rhetorical structure in discourse processing. In R. Horowitz & J. Samuels (Eds.), *Comprehending oral and written language* (pp. 117-160). New York: Academic Press.

HOWARD, D.V. (1980). Category norms: A comparison of the Battig and Montague (1968) norms with the responses of adults between the ages of 20 and 80. *Journal of Gerontology, 35,* 225-231.

HULTSCH, D.F., & DIXON, R.A. (1984). Memory for text materials in adulthood. In P. Baltes & O.G. Brim, Jr. (Eds.), *Life-span development and behavior* (Vol. 6, pp. 77-108). New York: Academic.

HULTSCH, D.F., HERTZOG, C., & DIXON, R.A. (1984). Text recall in adulthood: The role of intellectual abilities. *Developmental Psychology, 20,* 1192-1209.

KEMPER, S. (1987). Syntactic complexity and elderly adults' prose recall. *Experimental Aging Research, 13,* 47-52.

KIERAS, D.E. (1985). Thematic processes in the comprehension of technical prose. In B.K. Britton & J. Black (Eds.), *Analyzing and understanding expository text* (pp. 89-108). Hillsdale, NJ: Erlbaum.

KIERAS, D.E., & DECHERT, C. (1985). *Rules for comprehensible technical prose: A survey of the psycholinguistic literature* (Tech. Rep. No. 21, TR-85-ONR-21). Ann Arbor: University of Michigan.

KINTSCH, W. (1977). *Memory and cognition.* New York: Wiley.

KINTSCH, W., & VAN DIJK, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363-394.

KINTSCH, W., & VIPOND, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L.G. Nilsson (Ed.). *Perspectives on memory research* (pp. 329-365). Hillsdale, NJ: Erlbaum.

KIRSCH, I.S., & JUNGEBLUT, A. (1986). *Literacy: Profiles of America's Young Adults.* (Final Report No. 16-PL-02). Princeton, NJ: NAEP and Educational Testing Service.

KIRSCH, I.S., & MOSENTHAL, P.B. (1988). *Understanding document literacy: Variables underlying the performance of young adults* (Research Report No. RR-88-62). Princeton, NJ: Educational Testing Service.

KIRSCH, I.S., & MOSENTHAL, P.B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly, 25,* 5-30.

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

KLARE, G.R. (1974-1975). Assessing readability. *Reading Research Quarterly, 10,* 62-102.

LAWTON, M.P., & BRODY, E. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist, 9,* 179-185.

LIGHT, L.L., & ALBERTSON, S.A. (1988). Comprehension of pragmatic implications in young and old adults. *Language, memory, and aging* (pp. 133-153). Cambridge, England: Cambridge University Press

LIGHT, L.L., & BURKE, D. (1988). Patterns of language and memory in old age. In L. Light & D.M. Burke (Eds.), *Language, memory, and aging* (pp. 244-272). Cambridge, England: Cambridge University Press.

LIGHT, L.L., ZELINSKI, E.M., & MOORE, M. (1982). Adult age differences in reasoning from new information. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8,* 435-447.

MARSHALL, N., & GLOCK, M.D. (1978-1979). Comprehension of connected discourse: A study into the relationship between the structure of text and information recalled. *Reading Research Quarterly, 14,* 10-57.

MAYER, R.E. (1985). Structural analysis of science prose: Can we increase problem-solving performance? In B.K. Britton & J. Black (Eds.), *Analyzing and understanding expository text* (pp. 65-88). Hillsdale, NJ: Erlbaum.

MEYER, B.J.F. (1975). *The organization of prose and its effects on memory.* Amsterdam: North Holland.

MEYER, B.J.F. (1985a). Prose analysis: Purposes, procedures and problems. In B.K. Britton & J. Black (Eds.), *Analyzing and understanding expository text* (pp. 11-64, 269-304). Hillsdale, NJ: Erlbaum.

MEYER, B.J.F. (1985b). Signaling the structure of text. In D.H. Jonassen (Ed.), *The technology of text* (pp. 64-89). Englewood Cliffs, NJ: Educational Technology.

MEYER, B.J.F. (1987). Reading comprehension and aging. In K.W. Schaie (Ed.), *Annual review of gerontology and geriatrics* (Vol. 7, pp. 93-116). New York: Springer.

MEYER, B.J.F., BRANDT, D.M., & BLUTH, G.J. (1980). Use of the top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly, 16,* 72-103.

MEYER, B.J.F., & FREEDLE, R.O. (1984). Effects of discourse type on recall. *American Educational Research Journal, 21,* 121-143.

MEYER, B.J.F., & MCCONKIE, G.W. (1973). What is recalled after hearing a passage? *Journal of Educational Psychology, 65,* 109-117.

MEYER, B.J.F., & RICE, G.E. (1981). Information recalled from prose by young, middle, and old adult readers. *Experimental Aging Research, 7,* 253-268.

MEYER, B.J.F., & RICE, G.E. (1983). Learning and memory from text across the adult life span. In J. Fine & R.O. Freedle (Eds.), *Developmental studies in discourse* (pp. 291-306). Norwood, NJ: Erlbaum.

MEYER, B.J.F., & RICE, G.E. (1989). Prose processing in adulthood: The text, the reader and the task. In L.W. Poon, D.C. Rubin, & B.A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 157-194). Cambridge, England: Cambridge University Press.

MEYER, B.J.F., YOUNG, C.J., & BARTLETT, B.J. (1989). *Memory improved: Enhanced reading comprehension and memory across the life span through strategic text structure.* Hillsdale, NJ: Erlbaum.

MURRAY, M., DARNELL, J., WEINBERGER, M., & MARTZ, B. (1986). Factors contributing to medication noncompliance in early public housing tenants. *Drug Intelligence and Clinical Pharmacy, 20,* 146-151.

OUSLANDER, J. (1981). Drug therapy in the elderly. *Annals of Internal Medicine, 95,* 711-722.

PEARSON, D. (1974-1975). The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations. *Reading Research Quarterly, 10,* 155-192.

RAPHAEL, T.E., ENGLERT, C.S., & KIRSCHNER, B.M. (1986). *The impact of text structure instruction and social context on students' comprehension and production of expository text.* (Research Series No. 177). East Lansing: Michigan State University, Institute for Research on Teaching.

RICHGELS, D.J., MCGEE, L.M., LOMAX, R.G., & SHEARD, C. (1987). Awareness of four text structures: Effects of recall on expository text. *Reading Research Quarterly, 22,* 177-196.

RICKARDS, J.P., & SLIFE, B.D. (1987). Interaction of dogmatism and rhetorical structure in text recall. *American Educational Research Journal, 24,* 635-641.

ROBECK, M., & WILSON, J. (1977). *Psychology of reading: Foundations of instruction.* New York: Wiley.

SALTHOUSE, T.A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science, 2*(3), 179-183.

SALTHOUSE, T.A., MITCHELL, D.R.D., SKOVRONEK, E., & BABCOCK, R.L. (1989). Effects of adult age and working memory on reasoning and spatial abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 507-516.

STERNBERG, R.J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities.* Hillsdale, NJ: Erlbaum.

THORNDYKE, P.W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology, 9,* 77-110.

TRABASSO, T., & VAN DEN BROEK, P.W. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24,* 612-630.

VAN DEN BROEK, P.W. (1989). Causal reasoning and inference making in judging the importance of story statements. *Child Development, 60,* 286-297.

VINCENT, J.P. (1985). *Effects of discourse types on memory of prose by young, middle-age, and old adults with average vocabularies.* Unpublished doctoral dissertation, Arizona State University, Tempe.

WALKER, C.H., & MEYER, B.J.F. (1980). Integrating different types of information in text. *Journal of Verbal Learning and Verbal Behavior, 19,* 263-275.

WALLER, R. (1984). Designing a government form: A case study. *Information Design Journal, 4,* 36-57.

WALMSLEY, S., & ALLINGTON, R. (1982). Reading abilities of elderly persons in relation to the difficulty of essential documents. *Gerontologist, 22,* 36-38.

WILLIS, S.L. (1990). Current issues in cognitive training research. In E. Lovelace (Ed.), *Aging and cognition.* Amsterdam: North-Holland.

WILLIS, S.L., BLIESZNER, R., & BALTES, P.B. (1981). Intellectual training research in aging: Modification of performance on the fluid ability of figural relations. *Journal of Educational Psychology, 73,* 41-50.

WRIGHT, P. (1979). The quality control of document design. *Information Design Journal, 1,* 33-42.

WRIGHT, P. (1980). Strategy and tactics in the design of forms. *Visible Language, 14,* 151-193.

WRIGHT, P. (1987). Writing technical information. In E.A. Rothkopf (Ed.), *Review of Research in Education* (pp. 327-385). Washington, DC: American Educational Research Association.

ZELINSKI, E.M., & GILEWSKI, M.J. (1988). Memory for prose and aging: A meta-analysis. In M.L. Howe & C.J. Brainerd (Eds.), *Cognitive development in adulthood* (pp. 133-160). New York: Springer-Verlag.

FOOTNOTES

[1]The *t* test for the difference between correlation coefficients with correlated samples must be adjusted for correlations between the criterion variables being correlated with a common predictor. After Ferguson (1953), the following formula was used:

$$t = \frac{(r_{12} - r_{13})\,[(N-3)(1 + r_{23})]^{1/2}}{\{2[1 - (r_{12})^2 - (r_{13})^2 - (r_{23})^2 + 2r_{12}r_{13}r_{23}]\}^{1/2}}$$

$$df = N - 3$$

Walker and Lev (cited in Ferguson, 1953) note that, while the above formula for the *t* ratio makes no assumption as to the distribution of *X* and *Y* (here, "easy" and "hard" subset scores), generalization is only possible for all samples for which *X* and *Y* have exactly the same set of values as those in the observed sample. This stringent assumption makes use of the above formula less than ideal for the present purposes, and caveats to the generalizability of the *t* test results obtained must be noted.