Published in final edited form as:

Educ Gerontol. 2009 ; 35(7): 653–668. doi:10.1080/03601270902885504.

Comprehension of Health-related Written Materials by Older Adults

Chiung-ju Liu,
Boston University

Susan Kemper, and
University of Kansas

James A. Bovaird
University of Nebraska-Lincoln

Abstract

This study examined how Flesch Reading Ease and text cohesion affect older adults' comprehension of common health texts. All older adults benefited when high Flesch Reading Ease was combined with high cohesion. Older adults with small working memories had more difficulty understanding texts high in Flesch Reading Ease. Additionally, older adults with low verbal ability or older than 77 years of age had difficulty understanding texts high in text cohesion but low in Flesch Reading Ease. These results imply that writers must increase Flesch Reading Ease without disrupting text cohesion to ensure comprehension of health-related texts.

Understanding health-related written materials, such as medication leaflets or health education brochures, is critical for older adults. Unfortunately, older adults often have difficulty comprehending written health information (Benson & Forman, 2002; Gazmararian et al., 1999; Williams et al., 1995). Comprehension problems are associated with poor health status and more hospital admissions (Baker et al., 2002; Baker, Parker, Williams, & Clark, 1998; Baker, Parker, Williams, Clark, & Nurss 1997; Schillinger et al., 2002). This issue has become a concern to the health care system because of the consequences of high health costs (Marwick, 1997). Improving comprehension by increasing the readability of health-related written materials is commonly suggested (Adams, Smith, Metts, & Ross, 1979; Cotugna, Vickery, & Carpenter-Haeefe, 2005; Eaton & Holloway, 1980; Johnson & Stern, 2004; Singh, 2003). Readability is typically assessed in terms of grade level and recommendations are often specified relative to an eighth grade reading standard. However, this approach may be limited by the formulas used to assess readability.

Readability formulas are generally a function of word length and sentence length, such as the Flesch Reading Ease index (Flesch, 1948). The number of syllables per word and the number of words per sentence capture only the shallow features of a text and do not directly reflect content and organization. Comprehension is a complex mental process that involves not only decoding words and sentences but also making connections of ideas and deriving meaning from a written text. Readability formulas may not, therefore, measure actual determinates of reading difficulty. Kintsch and Vipond (1979) suggested that researchers must look for other predictor variables to supplement readability formulas in order to determine the real readability of a text. One such variable may be text cohesion.

Text comprehension involves three hierarchical levels of processing and representation. The basic level is the surface form which represents the exact words and syntactic structures of the text. The second level is that of propositional abstraction which represents the idea units of the text in a textbase; propositions consist of predicates, corresponding to verbs, states, or properties, and arguments, corresponding to nouns. The textbase is an interconnected representation of the propositions, linked through shared arguments. The highest level is the situation model (Radvansky, 1999). In the situation model, information from the first two levels is integrated with prior knowledge. According to Kintsch's construction-integration (CI) model (1998), comprehension is a series of cycles of construction and integration as processing cycles extract information at the basic level and building a textbase and an integrated situation model at the highest level. In the construction phase, propositions are extracted from the surface form and linked to each other and to earlier ones based on overlapping or shared arguments. In the integration phase, those propositions sharing many arguments with other propositions are kept in working memory and linked to new propositions extracted from the text and old propositions derived from prior knowledge.

The CI model suggests that a critical determinate of reading comprehension is text cohesion. Text cohesion arises from argument overlap and from the use of explicit features (e.g., words, phrases, or sentences) to guide the reader through the text, by making connections among ideas and sentences (Halliday & Hasan, 1976). Britton and Gülgöz (1991) have shown that increasing argument overlap can improve comprehension. Similarly, Meyer (2003) suggested that readers rely on causal and temporal connectives such as "before" or "because" in order to identify the structure of a text. Additionally, cohesion can also be increased based on implicit world knowledge, such as using arguments located close together in semantic space (Kintsch, 1998). Increasing text cohesion through argument overlap, causal and temporal connectives, and semantic and conceptual similarity may aid readers by reducing processing demands on working memory during either the construction or integration phase of text processing.

Working memory allows people to process and retrieve information simultaneously during reading. Prior studies have shown that working memory limitations affect older adults' ability to process texts with complex syntactic structures or texts high in propositional density (Norman, Kemper, & Kynette, 1992; Stine & Wingfield, 1988; Stine-Morrow, Ryan, & Leonard, 2000; Stine-Morrow, Milinder, Pullara, & Herman, 2001). When text cohesion is disrupted, readers must make inferences which require additional working memory to retrieve relevant information from long-term memory, leading to age-related decrements in inferential reasoning (Cohen, 1979, 1981; Zacks, Hasher, Doren, Hamm, & Attig, 1987).

However, knowledge that accumulated over life, especially domain specific knowledge, may help older adults process text information (Beier & Ackerman, 2005). McNamara (2001; McNamara & Kintsch, 1996) demonstrated background knowledge and text cohesion interact: for high-knowledge readers, low cohesion facilitates comprehension because the readers can actively integrate the text with their prior knowledge to construct a coherent representation; for low-knowledge readers, however, high cohesion facilitates comprehension because less inferential processing is required.

Prior research has also identified some factors that are related to older adults' comprehension of health information, such as age and years of education (Baker, Gazmararian, Sudano, & Patterson, 2000; Baker, Williams, Parker, Gazmararian, & Nurss, 1999; Gazmararian et al., 1999; Morrow et al., 2006; Williams et al., 1995). These studies have found that increasing age hampers older adults' comprehension and early life education, which is an indicator of general literacy, is positively related to context-specific literacy skills, such as reading health-related text, in later life. On the other hand, there are many reasons to assume that older adults have more knowledge of health and medicine than young adults since they have more chronic

illnesses than other age groups. Interestingly, Brown and Park (2002) found that prior knowledge of a disease did not enhance older adults' learning new information about that disease.

In the research reported here, we were particularly interested in how well older adults could comprehend short health-related texts. We contrasted texts differing on two factors previously shown to affect comprehension: (a) Flesch Reading Ease, indicated by word length and sentence length, and (b) text cohesion. We also investigated how age, educational level, health-literacy, verbal ability, prior knowledge, and working memory affected older adults' comprehension of health-related texts. Moreover, we included the interaction between Flesch Reading Ease and text cohesion in the study. Although the two concepts may seem orthogonal, an interaction may arise when readers' characteristics are considered. Our long-term goal is to provide suggestions for writing health-related materials in order to improve older adults' comprehension.

Method

Health texts

There were two stages involved in selecting the health texts. The first stage was to obtain a sizeable sample of short health-related information written for older adults. The second stage was to select 16 experimental texts that varied in word and sentence length as well as text cohesion from the sample.

In the first stage, texts from websites, health education brochures, and magazines were collected. The first author accessed the websites such as that maintained by the National Institute of Health (NIH) and MedlinePlus, as well as links suggested by these websites to locate texts. For example, the National Eye Institute website was suggested by the NIH website on the topic of "Age-related macular degeneration." These websites were visited from April to June 2005. Health education brochures were obtained from the patient education center at a local hospital in a college town at the Midwest in May 2005. Texts from magazines were obtained from a local public library. The first author consulted a librarian (May 2005) about which magazines that seniors often read. To obtain the most up-to-date health information, only issues from January to May of 2005 were chosen.

As a result, 200 health texts were selected (websites, $n = 157$; health education brochures, $n = 35$; and magazines, $n = 8$). These texts were related to senior health issues in four domains: (1) medication knowledge, e.g. how to take or how to store medications; (2) disease knowledge, e.g., explanation of a disease and related symptoms; (3) treatment procedure, e.g., how a test is performed or how a disease is treated; and (4) health policy, e.g., Medicare. Three criteria were used to select the texts. First, text length was limited to less than 300 words because we were interested in short texts so that several could be read in a single testing session and so that the text could be easily printed on a letter-sized page with the font size at 14. Second, we selected texts that did not include illustrations so that the text could be read alone without referring to the illustrations. Third, texts with a similar topic were allowed as long as they were from different publishers since the purpose of this stage was to sample health text with different written styles.

Each text was analyzed with the Coh-Metrix online software (Graesser, McNamara, Louwerse, & Cai, 2004; McNamara, Louwerse, Cai, & Graesser, 2005). The Coh-Metrix program provides several indices including measures of word frequency, grammatical complexity, and cohesion. We first selected Flesch Reading Ease, one of the most commonly used readability formulas (Cramer, 1978), from the Coh-Metrix analysis. Flesch Reading Ease measures text difficulty by computing word length and sentence length (See Equation 1). The score ranges

from 0 to 100 with a higher score indicating lower reading grade level and the use of more short words and short sentences; for this sample of texts, Flesch Reading Ease had a mean of 54.99 ($SD = 17.32$).

$$\begin{aligned} \text{Flesch Reading Ease} &= 206.835 - (1.015 \times \text{ASL}) - (84.6 \times \text{ASW}) \\ \text{ASL} &= \text{Average sentence length.} \\ \text{ASW} &= \text{Average number of syllables per word.} \end{aligned} \quad (1)$$

We also derived a text cohesion index from Coh-Metrix according to the following procedure. First, a factor analysis was conducted to obtain a single factor that accounted for the most variance in text cohesion among the 200 health texts. The eigenvalue of the first principal factor was 5.38 and it accounted for 25.60% of the variance. Six indices that loaded highly on the factor ($\lambda > .75$) were then selected: adjacent argument overlap ($\lambda = .90$), argument overlap for all sentences ($\lambda = .91$), adjacent stem overlap ($\lambda = .90$), all stem overlap ($\lambda = .89$), and two Latent Semantic Analysis (LSA) measures- LSA “sentence adjacent” ($\lambda = .81$), and LSA “sentence all” ($\lambda = .76$). These six indexes represent either referential cohesion or LSA aspects of cohesion.

Referential cohesion arises when an argument such as a noun, pronoun, or noun phrase refers to another argument in the text. “Adjacent argument overlap” estimates the proportion of adjacent sentences that share one or more referring expressions. “Argument overlap for all sentences” estimates the proportion of all sentence pairs in a paragraph that share one or more arguments. “Adjacent stem overlap” estimates the proportion of adjacent sentences that share one or more word stems. A stem is a lexical unit of a word, which carries significant semantic content, to which morphemes can be added. For example, *urinate* and *urination* share the same stem. “All Stem overlap” estimates the proportion of all sentence pairs in a paragraph that share one or more word stems. LSA is a statistical representation of word knowledge based on a large corpus of texts. The large corpus of text was decomposed into hundreds of concepts or topics in dimensions. The conceptual similarity between any two texts (e.g., words, clauses, sentences, texts) was evaluated. In Coh-Metrix, the LSA “sentence adjacent” measure captures how conceptually similar each sentence is to the next sentence. Likewise, the LSA “sentence all” measure captures how conceptually similar each sentence is to every other sentence in the text.

The text cohesion index for the current study was created by averaging the standardized scores of these six indices and then multiplying by 100 with a higher score indicating more cohesion (Range: $-205 \sim 214$, $M = 0.00$; $SD = 86.77$). The text cohesion index represents the degree of argument repetition and semantic and conceptual similarity in a text.

In the second stage, the 200 sample health texts were divided along two dimensions defined by quartile scores on Flesch Reading Ease and text cohesion, yielding 16 cells. Then 2 sets of 16 texts were selected by choosing 2 texts from each cell to assure the texts covered the range of Flesch Reading Ease and text cohesion of the entire sample of 200 texts and to eliminate the relationship between the two text measures. Table 1 provides four example texts that varied in Flesch Reading Ease and text cohesion. Additionally, the texts were assigned to 2 lists of 16 texts so that similar health topics were not repeated within a list to prevent readers learning from earlier texts.

Six yes/no comprehension questions were developed for each text. Each question explicitly tested factual information that was directly stated in the text. Table 1 shows some example questions. The two lists were pre-tested on young adults to identify very difficult questions in

order to avoid ceiling effects when the older adults were tested. We assumed that any questions young adults had difficulty answering would also be difficult for older adults.

Young adults were recruited from a university at the Midwest by posted signs and from the introductory psychology research pool. The students were paid \$10.00 per hour or with class credit. Each student read 16 texts. The texts were presented in a random order for each student. The title of each text was printed at the top of the page. Each text was printed with 14-point font. No bold, italic, or bullet formatting was applied. After reading a text, the student answered six questions printed on the back of paper by circling YES, NO, or DON'T KNOW. They were instructed to read at their normal reading pace and not to look back at the text once they turned to the questions. Additionally, the student rated each question for subjective difficulty with a 4-point Likert scaling (1 = very easy, 2 = easy, 3 = hard, and 4 = very hard).

Forty-two students (17 males), ranging in age from 19 to 28 years ($M = 21.30$, $SD = 2.01$) with 12 to 20 years of education ($M = 15.23$, $SD = 1.73$), read texts. None reported undergraduate majors related to health care. Accuracy rates and subjective difficulty rating were computed for each comprehension question. Accuracy rates ranged from 14.29% to 100.00% for the first selection of texts ($M = 70.54\%$, $SD = 22.97\%$) and 24.00% to 100.00% for the second selection ($M = 76.98\%$, $SD = 19.14\%$). Subjective difficulty ratings ranged from 1.24 to 2.81 ($M = 1.87$, $SD = .32$) and 1.10 to 2.62 ($M = 1.94$, $SD = .31$) for the two selections of texts. There was a significant negative relationship between accuracy rates and subjective difficulty rating ($r = -.73$ and $-.64$, $p < .01$); higher accuracy rates were associated with subjectively easier questions.

One text of the first selection list was replaced with a comparable text from the second list since the young adults had difficulty answering questions about that text. Among the final list of 16 texts, 2 texts were related to medication knowledge, 11 were related to disease knowledge, and 3 were related to treatment procedure. For this set of 16 texts, Flesch Reading Ease ($M = 55.41$, $SD = 18.37$) and text cohesion ($M = 3.52$, $SD = 95.63$), were not correlated, $r(16) = .08$, $p = .75$. Both indices were rescaled with z-scores for data analysis.

The final step was to compare the 16 texts on 53 other measures obtained from the Coh-Metrix analysis including various type-token ratios, measures of grammatical complexity, and word frequency. In general, these measures did not correlate with either Flesch Reading Ease or cohesion, all $r(16) < .12$, $p > .05$. The exception was that the relationship between the word frequency and Flesch Reading Ease was significant, $r(16) = .65$, $p < .01$, using the log frequency of content words obtained from the Coh-Metrix analysis ($M = 2.09$, $SD = .24$), reflecting Zipf's law (Zipf, 1935) that word length varies with word frequency. ANOVA was used to compare texts classified as high versus low in Flesch Reading Ease and high versus low in cohesion on these measures. With the exception of word frequency, all main effects and interactions were non-significant indicating that the Reading Ease and cohesion were not confounded with other text variables. Texts high in Reading Ease were composed of more frequent words ($M = 3.57$, $SD = .23$) compared to texts low in Reading Ease ($M = 0.67$, $SD = .47$), $F(1,15) = 11.637$, $p = .004$.

Participants

124 community dwelling older adults (32 males) between ages of 63 and 95 years ($M = 77.60$; $SD = 6.93$) were recruited from a local community by accessing previous research participant pools from the second author's research lab; by signs posted in local churches, grocery stores, or senior centers; or by word of mouth. The appropriate sample size of older adults was pre-determined by using the Optimal Design software (Raudenbush, Spybrook, Liu, & Congdon, 2005) based on the number of parameters of the study and a desired statistical power level of .80. All the older adults but one were Caucasian and most had at least 12 years of education

(Range: 8~24; $M = 15.08$; $SD = 2.94$). None of older adults reported severe central nervous system diseases or visual or cognitive impairments. The participants were either tested at the university research site or at the participant's own home. They received \$10.00 per hour for their time and \$5.00 for travel to the study site.

Working memory

The Digits Forward and Digits Backward tests (Wechsler, 1958), as well as the Daneman and Carpenter reading span test (1980) were used to test older adults' working memory. In the Digits Forward or the Digits Backward test, participants were required to recall a series of digits varying in length. The score was the total number of trials that the participant recalled correctly. In the Daneman and Carpenter reading span test, participants read a series of sentences and recall the last word of each sentence. The number of sentences in each set increased progressively. The score was the set size that the participant could successfully recall for two out of three trials. A compound latent factor score of working memory was calculated from these three measures. A higher compound latent factor score indicates higher working memory capacity.

Prior knowledge

As an index of prior knowledge of health information, we asked participants whether they had or have physical conditions or medical treatments that were mentioned in the texts because we assumed that older adults may have acquired some health information from personal experience. A list of physical conditions and treatments regarding text topics were given to the participants in the beginning of the study and they were asked to check all those they had or have experienced. We did not assess detailed knowledge of these topics to reducing priming effects.

Verbal ability

The American Version of the Nelson Adult Reading Test (AMNART) (Grober & Sliwinski, 1991) was used to measure participants' vocabulary knowledge. It is a 45-item test that requires participants to pronounce rare words one at a time. Correct pronunciation cannot be based on spelling-to-sound corresponding rules but must be retrieved from a memory representation. The full score is 45 with a higher score indicating better vocabulary knowledge. The test takes less than 5 minutes to administer and has been shown to correlate with the WAIS-R Verbal subtests (Lastine-Sobecks, Jackson, & Paolo, 1998; See also Wiens, Bryan, & Crossen, 1993).

Health literacy

The short version of the Test of Functional Health Literacy in Adults (S-TOFHLA) (Nurss, Parker, Williams, & Baker, 2001) was used to measure older adults' health literacy skills. S-TOFHLA includes two health texts, X-ray preparation and Medicaid rights and responsibility. The test is a 36-item Cloze test. Participants chose a word from four possible answers to complete sentences. For example, "*Your doctor has sent you to have a _____ X-ray,*" is accompanied with four possible answers: *a. stomach, b. diabetes, c. stitches, or d. germs*. The time limit of the test is 7 minutes. A score from 23 to 36 is considered an adequate level of health literacy, from 17 to 22 is marginal health literacy, and 0 to 16 is inadequate health literacy.

Procedure

First, each older adult was given a consent form to read and sign. Demographic information and prior knowledge of health information were then collected. Next, the participant took the health literacy test followed by the verbal ability test. The verbal ability test was audio recorded

for later scoring. Then, the participant read the 16 health texts and answered the comprehension questions. The format and procedure for administering the 16 texts was identical to that used for young adults except that older adults were asked to provide a subjective difficulty rating for each text rather than for each question. Lastly, the working memory tests were administered in the following order: Digits Forward, Digits Backward, and the Daneman and Carpenter reading span test. The study usually took about 70 minutes to finish and participants could take a break between tests.

Data analysis

The analysis can be viewed as an example of an items analysis from traditional psycholinguistics (Clark, 1973; Raaijmakers, Schrinemakers, & Gremmen, 1999) where variability in responses can be attributed to both text and person characteristics and both the texts administered and the persons tested can be considered as samples from larger populations of like texts and persons. In the current context, responses can be viewed as jointly nested, or cross-nested, within both persons and texts. Conversely, both persons and texts can be viewed as random rather than fixed sources of variability in individual responses. Multilevel modeling was used because it allows researchers to estimate variances and predictors at multiple levels, such as the response and experimental stimuli (texts) and person levels (older adults). For more details on the utility of multilevel modeling in this and similar contexts, the interested reader is directed towards Hoffman and Rovine (2007) and Locker, Hoffman, and Bovaird (in press).

In the current study, predictors related to text difficulty were specified in the level-one model,

$$Y_{ji} = \beta_{0i} + \beta_{1i}(F_{ji}) + \beta_{2i}(C_{ji}) + \beta_{3i}(F_{ji} \times C_{ji}) + U_{0i} + W_{0j} + \varepsilon_{ji} \quad (2)$$

where F denotes Flesch Reading Ease, C denotes text cohesion, and $F \times C$ denotes the interaction between the two indices. The dependent variable, Y_{ji} , is the comprehension score of older adult i on text j , ranging from 0 to 6 counts. Predictors related to older adults' characteristics were specified in the level-two model,

$$\beta_{0i} = \gamma_{00} + \gamma_{01}(M_i) + \gamma_{02}(H_i) + \gamma_{03}(K_i) + \gamma_{04}(V_i) + \gamma_{05}(E_i) + \gamma_{06}(A_i) \quad (3)$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}(M_i) + \gamma_{12}(H_i) + \gamma_{13}(K_i) + \gamma_{14}(V_i) + \gamma_{15}(E_i) + \gamma_{16}(A_i) \quad (4)$$

$$\beta_{2i} = \gamma_{20} + \gamma_{21}(M_i) + \gamma_{22}(H_i) + \gamma_{23}(K_i) + \gamma_{24}(V_i) + \gamma_{25}(E_i) + \gamma_{26}(A_i) \quad (5)$$

$$\beta_{3i} = \gamma_{30} + \gamma_{31}(M_i) + \gamma_{32}(H_i) + \gamma_{33}(K_i) + \gamma_{34}(V_i) + \gamma_{35}(E_i) + \gamma_{36}(A_i) \quad (6)$$

where M denotes working memory, H denotes health literacy, K denotes prior knowledge, V denotes verbal ability, E denotes education, and A denotes age. These effects, along with F , C , and $F \times C$ above, represent the fixed effects in the model. There are three random sources of variance in the model: W_{0j} is the random effect of health texts, U_{0i} is the individual random intercept of older adults, and ε_{ji} was the residual. Furthermore, multilevel modeling allows researchers to estimate interactions between predictors at different levels. If one substitutes coefficients from the level-two model into the level-one model, the comprehension score of a

text is not only predicted by the Flesch Reading Ease and text cohesion but also by the older adults' characteristics and interactions between both. All multilevel modeling was conducted with SAS PROC MIXED.

Results

Older adults who participated in the study on average took 4.41 ($SD = 3.13$) kinds of medication per day and spent 19.58 ($SD = 10.96$) hours reading per week. They all had corrected vision. Two older adults did not complete reading all 16 texts; one omitted 1 text and the other skipped 14 texts. Because multilevel modeling estimates model parameters with full-information maximum likelihood, the two participants were retained in the data analysis. For the measure of prior knowledge, two older adults had 4 physical conditions or medical treatments mentioned in the texts, 2 had 3, 10 had 2, and 37 had 1. Although the data of prior knowledge was slightly positive skewed, the results were the same regardless of the application of data transformations. There were thirteen older adults whose former occupation was related to health care, so former occupation was added in the level-two models (γ_{07}) as a control variable. Additionally, most older adults had adequate health literacy skills but three had inadequate or marginal health literacy skills ($S\text{-TOFHLA} = 15, 18, \text{ and } 22$, respectively). The $S\text{-TOFHLA}$ scores were extremely negatively skewed despite the use of data transformations to normalize the distribution. Consequently, the three older adults with inadequate to marginal health literacy skills were excluded from the analysis and health literacy was not estimated in the multilevel model because the remaining 121 older adults showed a ceiling effect on the test. Table 2 shows means and standard deviations for the measures of working memory, verbal ability, and health literacy for all older adults and for the 121 older adults used in subsequent analyses.

Overall, the mean difficulty rating of the texts by older adults was 2.31 ($SD = 0.59$). The subjective difficulty rating was not related to the Flesch Reading Ease index ($r = .003, p = .91$), nor to the text cohesion index ($r = -.04, p = .11$).

A modified *forward stepwise* procedure for model specification was used based on Snijders & Bosker (1999, p. 94). The first model is the empty model (Model 1) which did not include any predictors. The second model included main effects (Model 2) related to text and older adults. Model 2 represented a significant improvement from the empty model, $\Delta\text{-2LL}(9) = 44, p < .05$, and provided a better fit as indicated by smaller absolute fit indices such as the AIC and BIC, relative to the empty model. Model 3 included cross-level interactions. Non-significant interactions were removed one at a time to improve the parsimony of the model, and this final model was the best fitting model. Model 3 represented a significant improvement from Model 2, $\Delta\text{-2LL}(7) = 23, p < .05$, also with smaller AIC and BIC than Model 2. See Table 3 for multilevel modeling results.

Older adults had good comprehension of the health texts. On average, they had 4.65 correct answers out of 6 questions per text ($\gamma_{00} = 4.65, p < .001$). Older adults with larger working memories had better comprehension than those with smaller working memories ($\gamma_{01} = 0.18, p < .001$), especially as Flesch Reading Ease increased ($\gamma_{11} = 0.06, p < .05$). Figure 1 shows comprehension scores for a typical older adult with high working memory (latent working memory score is 1 SD above the mean) and another older adult with low working memory (latent working memory score is 1 SD below the mean). The figure suggests that Flesch Reading Ease did not affect comprehension of older adults with larger working memories while older adults with smaller working memories had more difficulty understanding the texts as Flesch Reading Ease increased. This is an unexpected finding: it indicates that increasing Flesch Reading Ease, by using shorter sentences and words, impairs the comprehension of older adults with smaller working memories. Increasing Flesch Reading Ease is intended to improve comprehension but may in fact impair comprehension by readers with smaller working

memories who are unable to overcome the limitations on the expression of causal and temporal information imposed by the use of short words and sentences.

Older adults' education and their prior knowledge were not significantly related to their comprehension ($\gamma_{05} = 0.02, p = .14$; $\gamma_{03} = 0.03, p = .59$; respectively). However, if their former occupation was related to health care, older adults did have higher comprehension scores ($\gamma_{07} = 0.28, p < .05$). Verbal ability was positively related to comprehension ($\gamma_{04} = 0.02, p < .05$). Older adults with high verbal ability generally had better comprehension than older adults with low verbal ability. Moreover, participants' age was significantly related to comprehension ($\gamma_{06} = -.01, p < .05$) which showed that increasing age may impair comprehension.

The main effects of Flesch Reading Ease and text cohesion were not significant ($\gamma_{10} = -0.07, p = .52$; $\gamma_{20} = -0.14, p = .21$; respectively). However, the interaction was marginally significant ($\gamma_{30} = 0.21, p = .08$) and the 3-way interaction of verbal ability, Flesch Reading Ease, and text cohesion was significant ($\gamma_{34} = 0.007, p < .05$). Figure 2 shows comprehension scores that are given for health texts with high text cohesion (Index score = 150) and low text cohesion (Index score = -150) and for a typical older adult with high verbal ability (AMNART score = 38, 1 SD above the mean) and an older adult with low verbal ability (AMNART score = 24, 1 SD below the mean). The figure suggests that when Flesch Reading Ease is low, increasing text cohesion can lead to comprehension problems, especially for older adults with limited verbal ability. This finding implies that the use of long sentences and long words, e.g., low Flesch Reading Ease, combined with a high degree of argument overlap and semantic and conceptual similarity, e.g., high text cohesion, can produce texts that are difficult to understand, especially for older adults with limited verbal ability. Further, when Flesch Reading Ease is high, increasing text cohesion will benefit all older adults, suggesting that the use of short sentences and words can enable older readers to utilize argument overlap and semantic and conceptual similarity to construct an integrated textbase and situation model.

Additionally, the 3-way interaction of age, Flesch Reading Ease, and text cohesion was also significant ($\gamma_{36} = -0.006, p < .05$). Figure 3 shows comprehension scores of a typical young-old adult at age of 70 (1 SD below the mean age) and an old-old adult at age of 85 (1 SD above the mean age) on health texts with high text cohesion (Index score = 150) and low text cohesion (Index score = -150). The figure suggests that when text cohesion is low, increasing Flesch Reading Ease can lead to comprehension problems. However, when text cohesion is high, increasing Flesch Reading Ease can benefit all older adults, especially those less than 77 years of age.

Discussion

This study evaluated the ability of older adults to comprehend common health-related texts that varied in along two dimensions, reading grade level and text cohesion. Our primary findings were that increasing Flesch Reading Ease by using shorter words or sentences led to worse comprehension by older adults with smaller working memories. This finding is unexpected as increasing Flesch Reading Ease is generally assumed to lead to improved comprehension (cf: Adams, et al, 1979; Cotuga et al, 2005). Our findings suggest that older adults with limited working memories do not benefit from the use of short sentences and short words to increase Flesch Reading Ease. Increasing Flesch Reading Ease, may, in fact, lead to the omission of critical connections among ideas as sentences by deleting causal and temporal subordinate clauses or phrases. Other strategies for increasing Flesch Reading Ease, such as using short words such as pronouns, e.g., *it*, may also disrupt text comprehension by requiring readers to make more inferences. These strategies, while designed to improve comprehension, may, in fact, challenge older adults with smaller working memories.

We also found that Flesch Reading Ease interacts with text cohesion: when Flesch Reading Ease is high, all older adults, especially those with high verbal ability or those less than 77 years of age, benefit from increasing text cohesion. The ability to fully process texts high in argument overlap and conceptual similarity into an integrated textbase and situation model may be contingent on the ability to fully process a text at lexical and syntactic levels. Hence, the use of short words and short sentences to increase Flesch Reading Ease may enable readers to take full advantage of increased text cohesion. However, when Flesch Reading Ease is low, all older adults, especially those with limited verbal ability or those greater than 77 years of age, may be overwhelmed by increased coherence as they struggle to process long words and long sentences.

Our study demonstrates that the comprehension of health-related text is a joint function of multiple text factors and reading ability. It also suggests two strategies for improving older adults' comprehension of written health-related texts. One strategy is to use short words and short sentences to increase Flesch Reading Ease. The use of this strategy carries with it some risk for older adults, especially those with limited working memories, as it may lead to comprehension problems if key words and ideas are abridged. The second strategy is to repeat important words, phrases, and ideas in order to increase text cohesion through argument overlap and semantic and conceptual similarity. This strategy must be applied in tandem with the first as our results indicate that all older adults will benefit from increasing text cohesion when Flesch Reading Ease is high but will be hurt from increasing text cohesion when Flesch Reading Ease is low. Older adults may be able to exploit argument overlap, semantic and conceptual similarity, and temporal and causal connectives as they construct a textbase and situation model only when the texts are easy to process lexically and syntactically because they were written using short words and short sentences.

Limitations and directions for future research

This study found no evidence that older adults' education level affected their comprehension of health-related texts, probably because most of the participants were well educated. Moreover, 98% of older adults in the present study had adequate health literacy skills. According to S-TOFHLA criteria, these older adults should be able to read and interpret most health texts. Therefore, it is not surprising that the intercept of comprehension in the multilevel model was high and the average rating of subjective difficulty was "neutral." Nonetheless, even these well-educated older adults had difficulty comprehending some texts, especially those low in Reading Ease and high in cohesion. Extrapolating from our findings would suggest that older adults with more limited educations or limited health literacy would tend to have even more difficulty understanding health-related texts at low Flesch Reading Ease decreases but high text cohesion.

In order to approximate daily life situations, the study used real world materials that varied along many dimensions, such as content and text organization; further, we imposed few controls on the testing situation. Our results may be an artifact of these choices. Additional studies, comparing different types of texts, varying text properties, and probing different experimental manipulations will help confirm and clarify our findings. An important consideration may be reading format. Furthermore, older adults may read and re-read texts with few constraints on reading time in daily life. Stine-Morrow et al. (2001) reported that older adults allocate more time than young adults as texts become more resource demanding. Our older participants, especially the young-old adults, may have adopted a similar reading strategy by taking more time to read some texts, such as those low in cohesion and Flesch Reading Ease, in order to maintain their good overall comprehension accuracy of 78% (compared to the 74% overall accuracy of the young adult pre-testers). Future studies should examine additional measures of reading behavior, such as monitoring reading times or eye

fixations. In addition, we assessed comprehension using yes/no questions; it would be appropriate to also assess inferential reasoning and recall.

Conclusions

The characteristics of texts and the characteristics of older readers interact to determine the comprehension of health-related written materials. Increasing Flesch Reading Ease by using short words and short sentences may not be the best solution for improving older adults' comprehension since increasing Flesch Reading Ease often means leaving out key information, such as causal and temporal connections among ideas. Readers must therefore engage in additional processing to construct these text linkages in order to build an accurate situation model. On the other hand, increasing text cohesion may benefit older adults since repeating similar words, phrases, and ideas may reduce processing demands associated with constructing a situation model from the text. However, our results suggest that the interaction of low-level processing demands due to lexical and syntactic analysis may interact with high-level processing demands due to the integration of propositional and conceptual information. Thus, the optimal strategy may be to both increase both Flesch Reading Ease and text cohesion. Although generalization of our findings is limited, these results should guide writers who are developing health-related written materials to consider not only the use of short sentences and short words to increase Flesch Reading Ease as well as the use of argument repetition and conceptual similarity to increase text cohesion.

Acknowledgements

This research was supported by grants AG 018892, P30 HD-002528, and P30 DC-005803 from the National Institute of Health and undertaken by the first author in partial fulfillment of the requirements for doctoral degree in Gerontology from the University of Kansas. Correspondence concerning this article may be addressed to Chiung-ju Liu, Health and Disability Research Institute, School of Public Health, Boston University, 580 Harrison Ave. 4th FL., Boston, MA 02118. Electronic mail may be addressed to liu_chiung_ju@hotmail.com.

References

- Adams RC, Smith TP, Metts JK, Ross JW. Readability-its applicability to education of patients by pharmacy. *Hospital Pharmacy* 1979;14:654–662. [PubMed: 10244957]
- Baker DW, Gazmararian JA, Sudano J, Patterson M. The association between age and health literacy among elderly persons. *Journal of Gerontology: Social Sciences* 2000;55B:S368–S374.
- Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D, Ren J, Peel J. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. *American Journal of Public Health* 2002;92:1278–1283. [PubMed: 12144984]
- Baker DW, Parker RM, Williams MV, Clark WS. Health literacy and the risk of hospital admission. *Journal of General Internal Medicine* 1998;13:791–798. [PubMed: 9844076]
- Baker DW, Parker RM, Williams MV, Clark WS, Nurss J. The relationship of patient reading ability to self-reported health and use of health services. *American Journal of Public Health* 1997;87:1027–1030. [PubMed: 9224190]
- Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. *Patient Education and Counseling* 1999;38:33–42. [PubMed: 14528569]
- Beier ME, Ackerman PL. Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge: Promising results in a real-world learning environment. *Psychology and Aging* 2005;20(2):340–355.
- Benson JG, Forman WB. Comprehension of written health care information in an affluent geriatric retirement community: Use of the Test of Functional Health Literacy. *Gerontology* 2002;48:93–97. [PubMed: 11867931]

- Britton BK, Gülgöz S. Using Kintsch's computational model to improve instrumental text: Effects of repairing inference calls on recall and cognitive structure. *Journal of Educational Psychology* 1991;83:329–345.
- Brown SC, Park DC. Roles of age and familiarity in learning health information. *Educational Gerontology* 2002;28:695–710.
- Clark HH. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 1973;12:335–359.
- Cohen G. Language comprehension in old age. *Cognitive Psychology* 1979;11:412–429. [PubMed: 487746]
- Cohen G. Inferential reasoning in old age. *Cognition* 1981;9:59–72. [PubMed: 7196819]
- Cotugna N, Vickery CE, Carpenter-Haeefe KM. Evaluation of literacy level of patient education pages in health-related journals. *Journal of Community Health* 2005;30:213–219. [PubMed: 15847246]
- Cramer E. A quick guide to readability formulas. *Curriculum Review* 1978;17:416–417.
- Daneman M, Carpenter PA. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 1980;19:450–466.
- Eaton ML, Holloway RL. Patient comprehension of written drug information. *American Journal of Hospital Pharmacy* 1980;37:240–243. [PubMed: 7361799]
- Flesch R. A new readability yardstick. *Journal of Applied Psychology* 1948;32:221–233. [PubMed: 18867058]
- Gazmararian JA, Baker DW, Williams MV, Parker RM, Scott TL, Green DC, Fehrenbach SN, Ren J, Koplan JP. Health literacy among Medicare enrollees in a managed care organization. *Journal of American Medical Association* 1999;281:545–551.
- Graesser AC, McNamara DS, Louwerse MM, Cai Z. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers* 2004;36:193–202.
- Grober E, Sliwinski M. Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology* 1991;13:933–949. [PubMed: 1779032]
- Halliday, MAK.; Hasan, R. *Cohesion in English*. Longman; Hong Kong: 1976.
- Hoffman L, Rovine MJ. Multilevel models for experimental psychologists: Foundations and illustrative examples. *Behavior Research Methods* 2007;39:101–117. [PubMed: 17552476]
- Johnson JL, Stern EB. Readability of patient education materials: A comparison of rural and urban cardiac rehabilitation sites in Minnesota. *Journal of Cardiopulmonary Rehabilitation* 2004;24:121–127. [PubMed: 15052116]
- Kintsch, W. *Comprehension: A paradigm for cognition*. Cambridge University Press; New York: 1998.
- Kintsch, W.; Vipond, D. Reading comprehension and readability in educational practice and psychological theory. In: Nilsson, L-G., editor. *Perspectives on memory research: Essay's in honor of Uppsala University's 500th Anniversary*. Lawrence Erlbaum Associates; Hillsdale, NJ: 1979.
- Lastine-Sobecks JL, Jackson ST, Paolo AM. Identifying the pronunciation of irregularly spelled words: Relation to verbal IQ. *The Clinical Neuropsychologist* 1998;12:189–192.
- Locker L Jr, Hoffman L, Bovaird JA. On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods*. in press
- Marwick C. Patients' lack of literacy may contribute to billions of dollars in higher hospital costs. *Journal of American Medical Association* 1997;278:971–972.
- McNamara DS. Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology* 2001;55:51–62. [PubMed: 11301728]
- McNamara DS, Kintsch W. Learning from texts: Effects of prior knowledge and text coherence. *Discourse Process* 1996;22:247–288.
- McNamara, DS.; Louwerse, MM.; Cai, Z.; Graesser, A. Coh-Metrix version 1.4. Jan 1. 2005 Retrieved [May 2005], from <http://cohmetrix.memphis.edu>
- Meyer BJF. Text coherence and readability. *Topics in Language Disorders* 2003;23:204–224.
- Morrow D, Clark D, Tu W, Wu J, Weiner M, Steinley D, Murray MD. Correlates of Health Literacy in Patients with Chronic Heart Failure. *The Gerontologist* 2006;46:669–676. [PubMed: 17050758]

- Norman S, Kemper S, Kynette D. Adults' reading comprehension: Effects of syntactic complexity and working memory. *Journal of Gerontology: Psychological Sciences* 1992;47:258–265.
- Nurss, JR.; Parker, R.; Williams, M.; Baker, D. *Test of Functional Health Literacy in Adults*. Peppercom Books & Press; NC: Snow Camp: 2001.
- Raaijmakers JGW, Schrijnemakers JMC, Gremmen F. How to deal with the “Language-As-Fixed-Effect Fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language* 1999;41:416–426.
- Radvansky GA. Aging, memory, and comprehension. *Current Directions in Psychological Science* 1999;8:49–53.
- Raudenbush, SW.; Spybrook, J.; Liu, X-F.; Congdon, R. Optimal design for longitudinal and multilevel research: Documentation for the “Optimal Design” software. 2005.
http://sitemaker.umich.edu/group-based/optimal_design_software, retrieved at June 2005
- Schillinger D, Grumbach K, Piette J, Wang F, Osmond D, Daher C, Palacios J, Sullivan GD, Bindman AB. Association of health literacy with diabetes outcomes. *Journal of American Medical Association* 2002;288:475–482.
- Singh J. Reading grade level and readability of printed cancer education materials. *Oncology Nursing Forum* 2003;30:867–870. [PubMed: 12949600]
- Snijders, TAB.; Bosker, R. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage; Thousand Oaks, CA: 1999.
- Stine EAL, Wingfield A. Memorability functions as an indicator of qualitative age differences in text recall. *Psychology and Aging* 1988;3:179–183. [PubMed: 3268257]
- Stine-Morrow EAL, Milinder L, Pullara O, Herman B. Patterns of resource allocation are reliable among younger and older readers. *Psychology and Aging* 2001;16:69–84. [PubMed: 11302369]
- Stine-Morrow EAL, Ryan S, Leonard JS. Age differences in on-line syntactic processing. *Experimental Aging Research* 2000;26:315–322. [PubMed: 11091938]
- Wechsler, D. *The measurement and appraisal of adult intelligence*. Williams and Wilkins; Baltimore: 1958.
- Wiens AN, Bryan JE, Crossen JR. Estimating WAIS-R FSIQ from the National Adults Reading Test-Revised in normal subjects. *The Clinical Neuropsychologist* 1993;7:70–84.
- Williams MV, Parker RM, Baker DW, Parikh NS, Pitkin K, Coates WC, Nurss JR. Inadequate functional health literacy among patients at two public hospitals. *Journal of American Medical Association* 1995;274:1677–1682.
- Zacks RT, Hasher L, Doren B, Hamm V, Attig MS. Encoding and memory for explicit and implicit information. *Journal of Gerontology* 1987;42:418–422. [PubMed: 3598090]
- Zipf, GK. *The psychobiology of language*. Houghton-Millflin; New York: 1935.

Figure 1.
The 2-Way Interaction of Flesch Reading Ease and Working Memory on Comprehension.

Figure 2.
The 3-Way Interaction of Verbal Ability, Flesch Reading Ease, and Text Cohesion on Comprehension.

Figure 3.
The 3-Way Interaction of Age, Flesch Reading Ease, and Text Cohesion on Comprehension.

Table 1
Examples of Four Health Texts.

A. How can I find out what my average blood glucose is?

High Flesch Reading Ease (Score = 79).

High text cohesion (Score = 47).

Ask your health care provider to do an A1C test. This blood test shows the average amount of glucose in your blood during the past 2 to 3 months. Have this test done at least twice a year. If your A1C result is not as good as it should be, your health care provider will do this test more often to see if it is improving as your treatment changes. Your A1C result plus your blood glucose meter results can show whether your blood glucose is under control. Aim for a result below 7 percent. If your A1C test result is below 7 percent, then your blood glucose is in a desirable range and your diabetes treatment plan is working. The lower your A1C is, the lower your chance of getting eye, nerve, and kidney damage. If your test result is more than 8 percent, you need a change in your diabetes plan. Your health care team can help you decide what part of your plan to change. You may need to change your meal plan, your diabetes medicines, or your exercise plan.

Question: Is A1C test a blood glucose test?

B. Hyperthermia-Too hot for your health.

High Flesch Reading Ease (Score = 82).

Low text cohesion (Score = -109).

Almost every summer there is a deadly heat wave in some part of the country. Too much heat is not safe for anyone. It is even riskier if you are older or if you have health problems. It is important to get relief from the heat quickly. If not, you might begin to feel confused or faint. Your heart could become stressed, and sometimes this causes death. Your body is always working to keep a balance between how much heat it makes and how much it loses. Your brain is the thermostat. It sends and receives signals to and from parts of your body that affect temperature, such as the spinal cord, muscles, blood vessels, skin, and glands that make substances known as hormones. Too much heat causes sweating. When the sweat dries from your skin, the surface of your body cools and your temperature goes down.

Question: Does the surface of the body cool down when the sweat dries from the skin?

C. Urinary retention.

Low Flesch Reading Ease (Score = 15).

High text cohesion (Score = 76).

Urinary retention, or bladder-emptying problems, is a common urological problem with many possible causes. Normally, urination can be initiated voluntarily and the bladder empties completely. Urinary retention is the abnormal holding of urine in the bladder. Acute urinary retention is the sudden inability to urinate, causing pain and discomfort. Causes can include an obstruction in the urinary system, stress, or neurologic problems. Chronic urinary retention refers to the persistent presence of urine left in the bladder after incomplete emptying. Common causes of chronic urinary retention are bladder muscle failure, nerve damage, or obstructions in the urinary tract. Treatment for urinary retention depends on the cause.

Question: Can obstruction in the urinary tract cause chronic urinary retention?

D. Multi-infarct dementia.

Low Flesch Reading Ease (Score = 32).

Low text cohesion (Score = -120).

Multi-infarct dementia (MID), a common cause of dementia in the elderly, occurs when blood clots block small blood vessels in the brain and destroy brain tissue. Probable risk factors are high blood pressure and advanced age. CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy) is an inherited form of MID. This disease can cause stroke, dementia, migraine-like headaches, and psychiatric disturbances. Symptoms of MID, which often develop in a stepwise manner, include confusion, problems with recent

memory, wandering or getting lost in familiar places, loss of bladder or bowel control (incontinence), emotional problems such as laughing or crying inappropriately, difficulty following instructions, and problems handling money. Usually the damage is so slight that the change is noticeable only as a series of small steps. However over time, as more small vessels are blocked, there is a gradual mental decline. MID, which typically begins between the ages of 60 and 75, affects men more often than women.

Question: Does multi-infarct dementia happen when blood clots block vessels in the brain and destroy brain tissue?

Table 2
Means and Standard Deviations of Working Memory, Verbal Ability, and Health Literacy.

Measures	Working memory			Verbal ability	Health literacy
	Digits Forward	Digits Backward	Reading span		
n ^a = 124	Mean (SD)	7.83 (2.34)	6.47 (2.00)	2.84 (0.82)	30.98 (6.99)
n ^b = 121	Mean (SD)	7.87 (2.34)	6.53 (1.98)	2.86 (0.82)	31.26 (6.79)
					34.16 (3.04)
					34.55 (1.69)

Note. AMNART = The American Version of the Nelson Adult Reading Test. S-TOFHLA = The Short Version of the Test of Functional Health Literacy in Adults.

^aThe number of total participants.

^bThe number of participants with adequate health literacy only.

Table 3

Multilevel Model Parameters.

Liu et al.

Page 20

Parameters	Model 1		Model 2		Model 3	
	Estimate	SE	Estimate	SE	Estimate	SE
<u>Fixed Effects:</u>						
Intercept (γ_{00})	4.717 [‡]	0.122	4.646 [‡]	0.113	4.647 [‡]	0.113
Flesch Reading Ease (γ_{10})			-0.067	0.105	-0.068	0.105
Text cohesion (γ_{20})			-0.143	0.111	-0.143	0.111
Working memory (γ_{01})			0.175 [‡]	0.051	0.175 [‡]	0.051
Health literacy (γ_{02})			-	-	-	-
Prior knowledge (γ_{03})			0.029	0.055	0.029	0.055
Verbal ability (γ_{04})			0.016 [*]	0.007	0.015 [*]	0.007
Education (γ_{05})			0.023	0.016	0.024	0.016
Age (γ_{06})			-0.014 [*]	0.006	-0.013 [*]	0.006
Occupation (γ_{07})			0.282 [*]	0.139	0.281 [*]	0.140
Flesch Reading Ease by Text cohesion (γ_{30})			0.206	0.112	0.205	0.112
Working memory by Flesch Reading Ease (γ_{11})					0.060 [*]	0.024
Verbal ability by Flesch Reading Ease (γ_{14})					-0.008 [*]	0.003
Verbal ability by Text cohesion (γ_{24})					0.002	0.003
Age by Flesch Reading Ease (γ_{16})					0.000 ^c	0.003
Age by Text cohesion (γ_{26})					-0.002	0.003
Verbal ability by Flesch Reading Ease by Text cohesion (γ_{34})					0.007 [*]	0.003
Age by Flesch Reading Ease by Text cohesion (γ_{36})					-0.006 [*]	0.003
<u>Variance Components:</u>						
Random Text Variance (W_{0j})	0.197 ^{**}	0.072	0.155 ^{**}	0.057	0.155 ^{**}	0.057
Random Intercept Variance (U_{0j})	0.262 [‡]	0.040	0.172 [‡]	0.028	0.173 [‡]	0.028
Residual Variance (ϵ_{ij})	0.778 [‡]	0.026	0.788 [‡]	0.026	0.788 [‡]	0.026
<u>Fit Statistics:</u>						
ML Deviance ^d (number of parameters)	5269 (4)		5225 (13)		5202 (20)	

Parameters	Model 1		Model 2		Model 3	
	Estimate	SE	Estimate	SE	Estimate	SE
AIC, BIC ^e	5277, 5269		5251, 5225		5242, 5202	

Note.

– : Not estimated. SE = Standard Error.

[‡] $p < .001$.

^{**} $p < .01$.

^{*} $p < .05$.

^c Non-zero estimate, 0.0003.

^d $-2(\log \text{likelihood})$ based on Maximum Likelihood (ML) estimation.

^e Absolute fit indices - a smaller value indicates a better fitting and/or more parsimonious model.