HOW DO TASK DEMANDS AND AGING AFFECT LEXICAL PREDICTION DURING ONLINE READING OF NATURAL TEXTS?

Sally Andrews[1], Aaron Veldre[1,2], Roslyn Wong[1], Lili Yu[2], & Erik D. Reichle[2]

The University of Sydney, Australia[1]

Macquarie University, Australia[2]

Author Note

Sally Andrews passed away in May, 2022, shortly before the revised version of this paper was submitted. Sally made outstanding contributions to reading research and psychology education throughout her career. She was a beloved colleague, mentor, and friend. She is greatly missed.

The data, analysis code, and materials from the present study are publicly available at the Open Science Framework website: https://osf.io/8qfta/. Portions of these data were presented at the 61st Annual Meeting of the Psychonomic Society. We thank Justine Greenaway and Joshua Vinson for their assistance with data collection.

Funding: This research was supported by the Australian Research Council's *Discovery Projects* funding scheme (grant numbers DP18102705 and DP190100719).

Correspondence concerning this article should be addressed to Aaron Veldre, School of Psychology, The University of Sydney, NSW 2006, Australia. Email: aaron.veldre@sydney.edu.au

Abstract

Facilitated identification of predictable words during online reading has been attributed to the generation of predictions about upcoming words. But highly predictable words are relatively infrequent in natural texts, raising questions about the utility and ubiquity of anticipatory prediction strategies. This study investigated the contribution of task demands and aging to predictability effects for short natural texts from the Provo corpus. The eye movements of 49 undergraduate students (mean age 21.2) and 46 healthy older adults (mean age 70.8) were recorded while they read these passages in two conditions: (i) '*reading for meaning*' to answer occasional comprehension questions; (ii) '*proofreading*' to detect 'transposed letter' lexical errors (e.g., *clam* instead of *calm*) in intermixed filler passages. The results suggested that the young adults, but not the older adults, engaged anticipatory prediction strategies to detect semantic errors in the proofreading condition, but neither age group showed any evidence of costs of prediction failures. Rather, both groups showed facilitated reading times for unexpected words that appeared in a high constraint within-sentence position. These findings suggest that predictability effects for natural texts reflect partial, probabilistic expectancies rather than anticipatory prediction of specific words.

Keywords

aging, eye movements, predictability, reading, task demands

Generating predictions has been proposed to be a universal principle of human information processing (e.g., Bar, 2009; Clark, 2013; Friston, 2010). Consistent with this view, a substantial body of evidence from a range of different methodologies and paradigms clearly demonstrates that more predictable words are processed more quickly and efficiently than less predictable words (e.g., DeLong et al., 2014; Kuperberg & Jaeger, 2016; Rayner et al., 2004). However, debate persists about when the benefits of predictability arise, and exactly what is predicted (e.g., Huettig, 2015; Staub, 2015). As discussed below, there are at least three possible sources of these effects.

The Role of Predictability in Reading

Many current theories of language processing assume that comprehension depends on an active meaning-construction process in which prior knowledge and experience are used to generate expectancies about how the discourse will unfold (e.g., Altmann & Mirkovic, 2009; Kuperberg, 2013). Such *anticipatory prediction* is argued to play a central role in the speed and efficiency of comprehension by giving the language processor a 'head-start' in resolving ambiguities and reducing the processing load imposed by noisy bottom-up perceptual information (e.g., Christiansen & Chater, 2016; Kutas et al., 2011; Van Petten & Luka, 2013). But questions have been raised about how often anticipatory prediction would benefit natural language processing (Jackendoff, 2007).

Estimates of how likely words in texts are to be predicted are typically derived from responses in the *cloze task* (Taylor, 1953) in which readers are asked to guess upcoming words from their preceding sentence contexts. For example, a sample of participants would be asked to generate their predictions about what the next word will be in the sentence: "The cat chased the …." The predictability metric of *cloze probability* is then the proportion of cloze task responses that correspond to a particular word. Thus, a word can be the most expected completion in a given context ("mouse" in the preceding example), which would be reflected by a high cloze probability (i.e., the *modal* response in the cloze task). An unpredictable word (i.e., non-modal cloze response) in the same position (e.g., “centipede”) may have a much lower cloze probability. Applications of the cloze task to natural texts have shown that very few words are highly predictable: The average cloze probability of content words is typically no more than .3, suggesting that readers’ predictions may often be incorrect (Gough et al., 1981; Rubenstein & Abori, 1958). The impact of such misprediction depends on what is assumed to be predicted (see DeLong et al., 2014; Staub et al., 2015, for discussion). If prediction is defined as “the all-or-none process of activating … a word in advance of perceptual input” (DeLong et al., 2014, p. 632) – a process that we will refer to as *lexical prediction* (Luke & Christianson, 2016) – then incorrect predictions might be expected to yield a processing cost associated with suppressing the disconfirmed prediction and/or adapting to unexpected alternative input (e.g., DeLong et al., 2014; Kutas et al., 2011). As elaborated below, the evidence for such prediction costs is inconsistent.

Alternative theoretical accounts attribute predictability effects to broad, probabilistic activation of features at multiple linguistic levels – visual, phonological, lexical, syntactic, and/or discourse (e.g., Gibson et al., 2013; Hale, 2011; Levy, 2008). Returning to our previous example involving a cat, such a graded prediction might correspond to any noun referring to a small animal or toy that a cat would chase. In contrast to the active, resource-demanding processes implicated in lexical prediction accounts, these *graded-prediction* accounts typically assume a passive, diffuse activation process that yields benefit due to pre-activation without causing cost when predictions are not confirmed (e.g., Neely, 1977; Posner & Snyder, 1975). Moreover, although specific lexical predictions are only possible for relatively highly constrained sentence contexts, graded prediction accounts allow for partial predictions about at least some attributes of upcoming words in conditions that preclude full lexical prediction.

The benefits of graded prediction should therefore extend across sentences that vary across the full range of *contextual constraint*, where the latter term can be operationally defined in terms of the probability of guessing the modal cloze response for a given within-sentence position. To illustrate this point, our previous example involved a sentence that provides high contextual constraint for the word "mouse," thereby affording a precise predication for that specific word. In constrast, a sentence beginning with "The dog ate the …" provides very little contextual constraint because a large number of non-modal responses would be expected, resulting in both the modal cloze response having a low cloze probability and many graded predictions of nouns referring to the various food and non-food items that dogs enthusiastically ingest.

Despite the main theoretical difference between the lexical-prediction and graded-prediction accounts, both accounts assume that the mental processes that generate predictions operate before the upcoming word has been fully processed. A third logical possibility is that predictability benefits arise after words have been identified, during a late stage of meaning construction that reflects the greater ease of integrating predictable words into their discourse representation (e.g., Brown & Hagoort, 1993; Reichle, 2021; Traxler & Foss, 2000).

Of course, these three different explanations of predictability effects are not mutually exclusive. Their contributions to the benefits and costs of predictability may also depend on characteristics of reading materials, task demands, and individual differences between readers. The broad aim of the present research was to evaluate the relative contributions of these factors to the impact of predictability on the reading of short naturalistic texts.

Although the role of predictive processes in reading has been a major focus of recent research (see Kutas et al., 2011; Staub, 2015; for reviews), most studies have, at least implicitly, focused on lexical prediction by defining predictability as a cloze probability above .65, a level that requires the use of highly constraining sentence materials (Huettig & Mani, 2016; Luke & Christianson, 2016). The majority of studies have also used single sentence materials which limit both the richness of the contextual cues available to support prediction and the incentive for readers to engage in integrative processing. Research using electrophysiological *event-related potential* (*ERP*) methods to investigate predictability effects in reading also typically involves *rapid serial visual presentation* (*RSVP*) paradigms that display words sequentially for 400-500 ms, which is substantially longer than the typical fixation duration (e.g., 200-300 ms; Rayner, 1998) during natural reading. The RSVP format may encourage predictive processing by increasing the time spent processing each word, and by depriving readers of both parafoveal preview of upcoming words and the opportunity to regress in response to processing difficulties. It is therefore unclear whether the conclusions drawn from existing research on predictive processing using ERP generalize to natural reading.

An important contribution to addressing these limitations is provided by Luke and Christianson's (2016, 2017) recent investigations of the impact of predictability on readers' eye movements during reading of short naturalistic texts. To allow assessment of effects of both lexical and graded prediction, they collected cloze responses for every word in the passages from a sample of young adult university students and computed the proportion of cloze responses that matched the orthographic form of the target word in a particular position, and measures of both the semantic and syntactic similarity of the cloze responses to the target word. The outcomes of this labor-intensive data collation have been made publicly available as the *Provo corpus* (Luke & Christianson, 2017) to encourage other researchers to build upon their efforts.

Luke and Christianson (2016; hereafter *L&C16*) reported a number of important findings that provide the foundation for the present research. First, they confirmed that the likelihood of generating the actual upcoming target words in their texts was relatively low: the mean cloze probability of content words was only .13, being slightly higher for nouns and verbs (.17) than adjectives and adverbs (.10), but only .30 even for function words. The percentage of words exceeding the typical .67 cloze probability criterion that has often been used in defining "highly predictable" words (e.g., see Staub, 2015) was very low: only 5% of content words and 19% of function words. Consequently, the likelihood of a correct lexical prediction (i.e., the percentage of modal cloze responses corresponding to target words) was also low: only 21% of content words and 40% of function words. However, the probability of correctly predicting the syntactic class of the word was substantially higher, exceeding .7 for nouns and verbs. Additionally, the average semantic similarity between the cloze responses and target words (estimated from *Latent Semantic Analysis*, *LSA*; Landauer, Foltz, & Laham, 1998) was significantly higher than expected by chance, particularly for nouns and verbs, suggesting that graded predictions about the upcoming targets were possible even though precise lexical forms were rarely generated.

Secondly, L&C16 evaluated whether these different predictability metrics influenced reading behavior by recording the eye movements of an independent sample of 84 university students who read the passages for meaning but were not directly tested for comprehension. Analyses of the eye-movement measures showed that target words that were associated with higher cloze probabilities were also associated with significant facilitative effects on virtually all measures, not only for targets corresponding to modal responses, but also for 'unexpected' words that mismatched the modal responses. To illustrate this finding using our previous example involving the cat sentence, this result would be equivalent to observing shorter fixations on a higher cloze non-modal continuation of the sentence, such as "yarn," than on a lower cloze non-modal continuation, like "centipede." Contrary to lexical-prediction accounts, the strength of the facilitatory relationship with cloze probability was equivalent for modal and non-modal words for virtually all eye-movement measures. The only two exceptions were the probability of skipping and of regressing out of content words: Content words that corresponded to the modal response were more likely to be skipped than those corresponding to non-modal responses, and surprisingly, were also *more* likely to trigger immediate regressions back to earlier in the sentence. These findings indicate that predictability effects are not restricted to the most probable word in a particular context:

Although few words in these natural passages were highly predictable, cloze probability significantly facilitated reading times for words that differed from the most probable continuation. Direct evidence for the contribution of graded predictions was provided by analyses showing that, even when cloze probability was controlled, both the LSA similarity and the part-of-speech (*PoS*) match of the cloze responses to the actual target words significantly facilitated all fixation-time measures on content words, and that reading times for function words were significantly facilitated by PoS match.

The third important issue addressed by L&C16 was whether their eye-movement data provided evidence of the cost of prediction failure that logically follows from anticipatory lexical-prediction accounts. This was assessed by analyzing whether processing of unexpected, non-modal words was slower when they appeared in highly constraining sentence contexts (thereby replacing words having higher cloze probabilities), as would be expected if a strong prediction for a specific word led to greater processing cost when the prediction was violated. The data provided minimal support for this possibility. To the degree that a strong, contextually-induced prediction for a specific word significantly influenced the processing of an unexpected word, such effects were generally facilitative rather than inhibitory, particularly for function words. The only indication of processing cost was a small increase in immediate regressions out of unexpected words in contextually constrained sentences.

L&C16 interpreted their findings as indicating that full lexical prediction occurs infrequently during the reading of natural texts, and that predictability effects in such contexts are more compatible with graded-prediction accounts in which readers' predictions about upcoming text are "partial and even sparse" (p. 47) but yield reliable processing benefits, even for words at the low end of the cloze probability range. L&C16 contrasted their findings with evidence suggesting that "the language processor can and does generate fine-grained details about upcoming material, including prediction of specific lexical items" (p. 46) in the ‘prediction-friendly’ environments that have dominated research on predictive language processing. Such environments are characterized by a much higher incidence of highly constraining sentence contexts than is typical of natural texts like those in the Provo corpus (Luke & Christianson, 2017). Predictive processing may be further promoted by unnatural presentation conditions like the slow, word-by-word presentation method used in the self-paced reading paradigm and many ERP studies, and by the pre-exposure of visual stimuli in many visual world eye-tracking studies (Huettig & Mani, 2016). The dominance of such methodologies may have led researchers to overestimate the extent to which readers generate specific lexical predictions during normal reading. Rather than the generation of specific lexical predictions being a ubiquitous component of language processing (as assumed by strong anticipatory prediction accounts; e.g., Altmann & Mirkovic, 2009; Christiansen & Chater, 2016), L&C16 suggest that effective language processing may depend on learning *when* to implement predictive strategies. This flexible predictive processing perspective (Brothers et al., 2017; Wlotko & Federmeier, 2015) highlights the importance of specifying the factors determining when lexical predictions are (and are not) generated. The present research addressed this goal by extending L&C16’s methodology to investigate two factors that have been shown to modulate predictability effects: task demands and cognitive changes associated with older age.

The Impact of Task Demands on Predictability Effects

Although the Provo materials consist of naturally occurring texts that are more extended than the single sentences used in most previous research, they are relatively short (an average of 50 words and 2.5 sentences per paragraph) and the different passages are unconnected. Moreover, L&C16’s eye-movement data were collected under conditions in which comprehension was not monitored by the occasional comprehension questions used in most studies. In combination with the lack of continuity across texts, the absence of any test of comprehension may have reduced readers’ incentive to engage anticipatory prediction processes. Even when comprehension is assessed, the use of easy, lexical-verification questions is associated with a 'less deliberate' reading strategy characterized by reduced regressions and rereading than when comprehension is assessed with more frequent and/or difficult questions (Andrews & Veldre, 2021; Radach et al., 2008). Thus, the limited evidence of lexical prediction in L&C16's data may, in part, reflect the use of a shallow reading strategy that readers deemed to be 'good enough' for the limited comprehension demands of the task (Ferreira et al., 2002).

To determine whether task demands influence the extent to which readers engage lexical prediction during reading of natural texts, the present study compared eye movements for the Provo corpus passages under standard comprehension demands with a proofreading task adapted from Schotter et al. (2014) that encouraged contextual prediction. Schotter et al. investigated how task demands modulated the impact of word frequency and predictability on eye-movement measures of sentence reading by comparing two proofreading conditions that differed only in the nature of the errors that readers were required to detect. In Experiment 1, participants were instructed either to read for meaning and respond to yes/no comprehension questions presented after one-third of the sentences, or to detect nonwords created by transposing two letters of a word (e.g., *trcak* from *track*) that occurred in one-third of the sentences. Experiment 2 was identical except that the errors created by letter transposition yielded a different word (e.g., *trial* instead of *trail*) and participants were instructed to detect "misspelled words that spell check cannot catch [because they] produce an actual word but not the word that the writer intended" (p. 11). This clever design therefore manipulated task demands while maintaining identical presentation format, eye-tracking methods, and critical stimulus materials.

Both proofreading conditions yielded longer total reading time and average fixation durations than the comprehension conditions, demonstrating that readers adopted a more cautious reading strategy, particularly when required to detect word errors. However, rather than reflecting generalized slowing in response to task difficulty, the eye-movement data suggested that reading strategies differed between the word- and nonword-proofreading tasks. The critical evidence for these ‘task-specific’ modulations was provided by comparisons of the effects of word frequency and predictability for the same ‘filler’ sentences, which never contained errors, under standard reading for meaning conditions relative to each of the proofreading tasks. Frequency effects for the filler sentences increased in both proofreading conditions, suggesting that the process of determining lexical status is a “first step in checking for spelling errors” (Schotter et al., 2014, p. 19). However, predictability effects only significantly increased in the word-proofreading task, presumably because the semantic fit of words in the sentence context was crucial to error detection. Schotter et al. interpreted these data as demonstrating “a qualitative readjustment of different subcomponents of overall reading” (p. 19) to optimize detection of word versus nonword errors. Notably, the increased predictability effects observed in the word-proofreading task relative to comprehension was limited to the late measure of total-fixation duration. However, the assessment of predictability was based on high-constraint sentence contexts in which the target word was either high in cloze probability (mean = .64) or a plausible but unexpected word that was virtually never produced in the cloze task (mean = .01). Consistent with previous literature using the same materials (Balota et al., 1985; Rayner & Well, 1996), significant main effects of predictability were observed on both early and late reading measures in all conditions, suggesting that high-constraint contexts may have encouraged lexical prediction. Thus, the low contextual constraints of the texts used in the present experiment might be expected to reduce readers’ tendency to predict upcoming words in the comprehension task, but to exaggerate the task-specific differences in the effect of predictability on eye movements between the comprehension versus proofreading tasks.

Cognitive Aging and Predictability

There is a growing body of literature comparing predictability effects in younger and older adults, but the outcomes are often contradictory. Behavioral investigations of off-line tests of reasoning and comprehension have found no evidence of age differences in the encoding and use of context to make predictive inferences from text (e.g., Valencia-Laver & Light, 2000), even in conditions that revealed global age-related differences in overall memory performance (e.g., Light et al., 1991) or speed-accuracy trade-offs (McKoon & Ratcliff, 2018). Older adults' cloze responses (Lahar et al., 2004) and judgements of semantic fit and plausibility (Little et al., 2004) are also very similar to those of young adults, suggesting that the structure of semantic memory and the capacity to generate conscious predictions is preserved in older age (Payne & Silcox, 2019).

However, age differences in contextual prediction *have* been observed in studies of online processing, but the nature of these differences varies with task demands and methodology. In tasks requiring comprehension and memory of rapidly presented or perceptually degraded speech, older adults typically perform more poorly for low-constraint materials, but equivalently to young adults for high-constraint contexts (see Payne & Silcox, 2019, for a review). Similar age-related enhancement of the benefit of contextual constraint has been found during reading under conditions of visual degradation (e.g., Madden, 1988). Such findings converge with the evidence from off-line tasks in showing that older adults *can* apply contextual prediction strategies to compensate for limitations in perceptual input, but these findings do not address the question of whether and when prediction is applied during online reading (i.e., under more natural conditions).

This issue was directly addressed by eye-movement research comparing the effects of word frequency and predictability on sentence reading in younger and older adults (Rayner et al., 2006; Rayner et al., 2010; Rayner et al., 2013). This research was interpreted as indicating that older readers adopt a 'risky reading' strategy characterized by increased reliance on contextually-based predictions. This increased reliance upon prediction is posited to compensate for an overall slower rate of lexical processing and/or age-related declines in visual acuity that reduce parafoveal processing. The major eye-movement signatures claimed to index this strategy were slower reading times due to longer fixations and more regressions (attributed to slower lexical processing), and increased skipping, longer forward saccades, and enhanced word frequency and predictability effects (which were attributed to a compensatory increased reliance upon prediction to “guess” upcoming words; Rayner et al., 2006). Although the risky reading hypothesis has motivated further research (e.g., Paterson et al., 2020) and computational modelling (McGowan & Reichle, 2018; see also Laubrock et al., 2006), questions have been raised about the strength and consistency of the evidence underpinning it (Payne & Silcox, 2019; Veldre et al., 2021, 2022).

For example, a recent meta-analysis confirmed that older adults show a trade-off between longer fixations, on one hand, and increased skipping rates and longer forward saccades, on the other, but only with readers of alphabetic languages: Although older readers of Chinese read more slowly, they did not show higher skipping rates or longer forward saccades than young adults (Zhang et al., 2022). However, the meta-analysis did not support the central claim of the risky reading hypothesis that older adults are more likely than younger adults to use lexical and contextual knowledge to predict upcoming words because there were no significant differences in older adults’ word frequency or predictability effects in the reading of either alphabetic or non-alphabet scripts. However, the authors noted that these findings need to be interpreted cautiously because the impact of age on word frequency and predictability effects had only been investigated in seven and three studies, respectively. There is also evidence that the frequency and difficulty of comprehension questions influence the strategies revealed by young readers’ eye movements (Andrews & Veldre, 2021; Radach et al., 2008), and such manipulations have been found to modulate the extent of parafoveal processing in older adults (Wotschak & Kliegl, 2013). Thus, older adults’ reliance on prediction may depend on task demands (e.g., Paterson et al., 2020; Wotschak & Kliegl, 2013).

Evidence for the contribution of task-specific factors is provided by research using ERP methods to investigate age differences in predictive processing. Comparisons of younger and older readers' sensitivity to contextual constraint and cloze probability have consistently shown that the N400 component (an ERP signature of a word's contextual fit; see Federmeier & Laszlo, 2009; Kutas & Federmeier, 2011, for reviews) is reduced in amplitude and delayed in older adults, even in strongly constraining contexts (see Wlotko et al., 2010, for a review). Older adults' ERP waveforms, at least at the group level, also fail to show a post-N400 frontally distributed positivity that has been consistently observed in young adults for plausible but unexpected words in high-constraint contexts (e.g., DeLong et al., 2014; Wlotko et al., 2012) and attributed to processes triggered by the violation of predictions (e.g., Kuperberg & Jaeger, 2016; Van Petten & Luka, 2012). These age differences have been interpreted as evidence that, unlike young readers, "older adults do not routinely utilize context in an anticipatory manner to predict likely upcoming input" (Payne & Silcox, 2019, p. 33). This "shift away from the predictive use of sentential context" (Wlotko et al., 2012, p. 986) is attributed to age-related deterioration in cortical connectivity that limits engagement of frontally mediated, top-down processes that support anticipatory prediction in young adults (e.g., Federmeier et al., 2010). This conclusion is essentially opposite to that drawn from eye-movement studies of online reading – that the capacity to make contextual predictions is preserved, or even enhanced, in older adults.

Determining the source of these discrepant conclusions is complicated by differences between the experimental procedures and task demands associated with different methodologies. The RSVP procedures typically used in ERP studies constrain readers' control over the rate of input and their ability to access information from the parafovea, as well as their ability to use strategies like 'risky reading', which may play an important role in natural reading. Adapting to these unnatural conditions may also selectively disrupt older relative to younger readers (Choi et al., 2017; Veldre et al., 2022). The constraints of ERP recording also mean that few studies assess online comprehension (Payne & Silcox, 2019). One notable study was reported by Dave et al. (2018) which demonstrated enhanced effects of both the benefits and costs of lexical prediction in older adults' ERP waveforms in a task that explicitly required prediction, suggesting that task demands may modulate older readers' engagement of predictive processes.

Given the inherent limitations of the reviewed ERP studies, one might argue that eye-movement studies of sentence reading more closely approximate natural reading because the individual words are displayed simultaneously, and because these studies usually include comprehension questions to encourage reading for meaning. However, comprehension questions are typically presented for only 25-33% of sentences and often assess little more than the presence/absence of particular words (Andrews & Veldre, 2021). Moreover, like the eye-movement studies of young adults, most of the eye-movement evidence for the risky reading hypothesis has been collected in prediction-friendly environments dominated by high-constraint sentence contexts.

The Present Study

The aim of the present study was to investigate the impact of task demands and age on predictability effects during online reading of the Provo corpus of natural texts. Samples of younger, university students and healthy older (60+) adults were compared under (1) *standard comprehension* conditions in which comprehension questions were presented after half of the passages, and (2) *proofreading* conditions in which they were required to detect transposed letter (*TL*) word errors that occurred in half of the passages. Our analyses used the methods developed by L&C16 to investigate the relative contribution of cloze probability and indices of graded prediction to eye-movement measures of the benefits and costs of predictability, and how these differed as a function of both task demands and age.

If L&C16's failure to find evidence of reliance on lexical prediction was due to the absence of any monitoring of comprehension rather than the low contextual constraint of their naturalistic texts, then the present comprehension conditions may reveal enhanced benefits and costs of contextual constraints that might be expected if readers generate specific lexical predictions about upcoming words. These effects may also be magnified in the proofreading condition where semantic fit is critical to identifying incorrect TL words. Alternatively, if readers normally generate automatic, graded predictions, then processing benefits may be observed for both high cloze probability words (i.e., words corresponding to modal responses) and words that are compatible with their semantic and syntactic context (i.e., words corresponding to non-modal responses). Any costs associated with prediction violation should be restricted to the proofreading condition which, due to the nature of the task, would be expected to encourage more frequent specific lexical predictions. Finally, if the predictability effects are due solely to late integration processes, then any costs observed in the comprehension conditions will largely manifest in regressions triggered by integration failures, which will in turn enhance the accuracy of proofreading performance by facilitating identification of semantically incongruent words.

The results of the present study will also add to the sparse evidence about how aging modulates predictability effects. If older readers rely more on context to compensate for slow lexical processing, then they should exhibit patterns of eye movements consistent with the risky reading hypothesis. That is, their compensatory strategies may reflect increased reliance on partial visual and/or orthographic information extracted from upcoming words (Choi et al., 2017), consistent with McGowan and Reichle's (2018) simulations using the E-Z Reader model. This strategy will yield benefits when predictions are correct, but result in regressions (to repair comprehension) when they are not. Older readers may also be able to apply their superior vocabulary and reading experience to construct the gist of sentence meaning from the partial information extracted from a skimming strategy (Zhang et al., 2022). One final implication is that evidence for larger predictability effects in older than younger readers would suggest that the absence of such effects in ERP studies is due to the unnatural presentation conditions that they typically employ.

Method

Participants

The final sample consisted of 49 younger adults recruited from The University of Sydney and Macquarie University (*M* = 21.8 years, range: 18-30 years; 33 females) who received course credit for their participation, and 46 community-dwelling, older adults (*M* = 70.8 years, range: 60-88 years; 30 females) who received financial reimbursement. All participants reported that they were native English speakers. Young adults reported that they had normal or corrected-to-normal vision. For the older adults, corrected vision was assessed to be within the normal range (at least 20/40) using a modified Snellen test at the experimental viewing distance. The data of a further three older adults were excluded due to eye-tracker calibration difficulty and/or self-reported visual impairments (e.g., cataracts, macular degeneration). To assess the cognitive capacity of the older adults, both groups completed the Nelson Denny Reading Test (Brown et al., 1993) under the reduced time constraints that Andrews et al. (2020) demonstrated to yield more normal score distributions for samples of skilled readers. As summarized in the upper section of Table 1, the two groups' performance did not significantly differ on the Reading Comprehension subtest, but the older adults' Vocabulary scores were significantly higher than those of the young adults. The study was approved by The University of Sydney Human Research Ethics Committee (protocol 2018/605) and the Faculty Ethics Subcommittee at Macquarie University (protocol 52020526715992). All participants provided informed consent.

Materials and Design

All participants read two blocks of 45 passages. In the first, *comprehension* block, they were instructed to read the passages for meaning and told that they would be presented with questions assessing their comprehension after some of the sentences. In the second, *proofreading* block, they were instructed that their task was to detect whether the passage contained any errors in which a word had been incorrectly typed to form a different word. Following Schotter et al. (2014), the proofreading condition was always presented as the second block to avoid contaminating normal reading behavior in the comprehension block by inducing sensitivity to TL words.

The critical materials comprised 50 2-3 sentence passages from the Provo corpus which were always presented in a format parallel to L&C16 (i.e., without any proofreading errors or subsequent comprehension questions). A further 40 ‘filler’ passages were constructed to match the length, structure, and general style of the Provo passages. Each of these passages contained a critical target word in which two adjacent letters could be transposed to form another word that was implausible in the sentence context (e.g., *wrap/warp; coats/coast*). These words were distributed approximately equally across the initial, medial, and final sections of the passage to prevent participants from developing a selective search strategy, but were never one of the first or last five words of the passage. A two-alternative multiple-choice question that required at least moderate comprehension of the passage was created for each filler passage. Four counterbalanced lists were constructed. In the comprehension block of each list, the filler passages contained the contextually appropriate TL word (i.e., the semantically appropriate word from a given pair of TL word neighbors; e.g., *calm* in Table 2). In the proofreading block, the filler passages contained the implausible TL neighbor word (i.e., the semantically inappropriate word from a given pair of TL word neighbors; e.g., *clam* in Table 2). The presentation of the critical Provo passages was thus identical in both blocks. This arrangement ensured that all participants only read each passage once, but across lists, all passages were presented under both comprehension and proofreading instructions. Examples of the stimulus materials are presented in Table 2.

---TABLES 1 AND 2 ABOUT HERE ---

Apparatus and Procedure

The majority of young adults (*n* = 46) were tested at Macquarie University. The remaining young adults (*n* = 3) and all older adults were tested at The University of Sydney. At both testing sites, participants' eye movements were recorded at 2,000 Hz using an EyeLink 1000 Plus tracker. Chin and forehead rests were used to minimize head movements. Viewing was binocular but fixation position was monitored from the right eye. The texts were presented over two or three lines in black monospaced font on a grey background. At the University of Sydney testing site, passages were displayed on a 21-in. CRT monitor (refresh rate 140 Hz) and participants were seated 60 cm from the monitor. At the Macquarie University site, passages were displayed on a 24.5-in. LCD monitor (refresh rate 240 Hz) and participants were seated 95 cm from the monitor. At these distances, approximately 3 characters subtended 1 degree of visual angle for all participants.

Participants read two blocks of 45 passages, each consisting of 25 Provo passages and 20 filler passages. In the first comprehension block, the filler passages were followed by a comprehension question that required a binary key-press response. Two practice passages followed by comprehension questions were presented before the individually randomized sequence of 45 experimental passages. Participants then received instructions for the proofreading block which emphasized that the errors they were required to detect were mistyped words that did not fit the context. Four practice passages were then presented before the 45 randomly ordered experimental passages. Each experimental passage was followed by the question '*Was there an error?*' to which participants made a binary key press response. A 9-point calibration procedure initiated each block. Each passage was presented when the participant made a stable fixation on a fixation point presented in the location of the first letter of the passage, or a new calibration procedure was performed if the calibration error was more than 0.5 degrees of visual angle.

The data, analysis code, and materials from the present study are publicly available at the Open Science Framework website (osf.io/8qfta).

Results and Discussion

Mean comprehension accuracy ranged from 70-100% and mean proofreading accuracy was between 60 and 100%. As summarized in Table 1, comprehension accuracy was equivalently high in both groups, but older adults performed the proofreading task significantly more accurately than younger adults. Both groups made fewer false alarms to correct Provo passages than misses to filler passages containing TL word errors.

All analyses of the eye-movement data were conducted on the Provo passages alone, which were presented in identical format in both task conditions. Analyses were conducted on both global measures of passage reading and local measures of word-level processing. Before calculating reading measures, those fixations less than 80 ms were merged with fixations within one letter space, and the remaining fixations less than 80 ms along with trials that were prematurely terminated or timed out were excluded. All analyses were conducted using the *lme4* package (Version 1.1-21; Bates, Maechler, Bolker & Walker, 2015) in *R* (Version 36.3; R Core Team, 2020) to test linear or logit mixed-effect models that included fixed effects of task and age, assessed by effect coded contrasts, and their interaction. To control for the significant difference in total Nelson Denny test score between the older and younger samples, this continuous variable was included as a covariate in all analyses to allow effects due to age group to be separated from effects of verbal proficiency.

Global Reading Measures

Initial analyses assessed the effects of task demands, age, and verbal proficiency on five global eye movement measures: (1) total passage reading time; (2) total number of fixations; (3) average fixation duration; (4) average forward saccade length; and (5) total number of regressions (see Table 3 for summary data and Table 4 for LMM output). The random-effects structure for each model included participant and item random intercepts, participant random slopes for task, and item random slopes for the effects of age, task, verbal proficiency, and the Age × Task interaction. Models that failed to converge with this structure were trimmed to retain random slopes for which there were significant corresponding fixed effects.

---TABLES 3 & 4 ABOUT HERE ---

The effect of task was significant on all measures reflecting longer reading times, more fixations, shorter saccades, and more regressions in the proofreading than the comprehension task. Higher verbal proficiency was associated with significantly shorter reading times, fewer and shorter fixations, and longer saccades. The main effect of age group, and the interactions between age and task, were significant on forward saccade length and average fixation duration. Older adults made longer average saccades and longer average fixations than young adults. Older readers also showed a larger increase in fixation duration and decrease in saccade length in the proofreading relative to the comprehension block than did younger readers.

Thus, both younger and older adults implemented a more cautious reading strategy in the proofreading than comprehension task, and although there were few overall differences between the age groups' reading behavior when controlling for verbal proficiency, older adults showed the trade-off between longer reading times and longer forward saccades observed in Zhang et al.'s (2022) meta-analysis and posited by the risky reading strategy (Rayner et al., 2006), and they also adjusted their reading to task demands more than young adults.

Predictability Effects on Word-Level Processing

To evaluate whether task demands and age modulated predictability effects, we next conducted analyses of eye-movement measures of word-level processing based on those conducted by L&C16. To assess the relative contributions of both lexical predictions and graded predictions of semantic and syntactic features, we tested models including target word cloze probability as well as the LSA similarity and PoS match of each target word to the corresponding cloze responses from the Provo corpus (Luke & Christianson, 2017).

Eye-movement data were collated separately for content words (n = 1,444) and function words (n = 946). We computed reading time measures of *first fixation* (i.e., duration of the initial first-pass fixation), *gaze duration* (i.e., sum of all first-pass fixations), *total duration* (i.e., sum of all fixations, including regressions back to the word), and the probability of skipping, making a regression out of the word to earlier in the sentence, and making a regression into the word from later in the sentence. We only report analyses of content words because they have dominated previous investigations of predictability effects. Table 5 presents the mean content word data for both age groups in each task. The mean data for function words are presented in Appendix Table A1 to allow comparison with L&C16's data.

---TABLE 5 ABOUT HERE ---

Following L&C16, analyses were conducted on log-transformed duration measures. To control for the effects of return sweeps from the end of one line of text to the beginning of the next, which affects up to 20% of fixations in multi-line texts, we followed Slattery and Parker's (2019) recommendation to include line initial and final position as fixed factors in the model. Similarly, sentence final and passage final positions were coded as fixed factors to control for the impact of structure building and wrap-up processes (Andrews & Veldre, 2021; Perfetti & Helder, 2021).[1] The models tested the fixed effects of task, age, and verbal proficiency along with continuous measures of log-transformed target cloze probability, LSA similarity, and (log-transformed) PoS match[2], as well as their interactions with task and age

[1] These factors were not controlled in L&C16's analyses of the Provo eye-movement data. The minor differences in analysis outcomes when they were excluded from our models are noted below.

[2] L&C16's decision to log-transform predictability scores was motivated by evidence that the relationship between predictability and response time is logarithmic (Smith & Levy, 2013), and because log-transformed scores explained more variance in eye-movement measures than raw scores. More recent research, including meta-analysis of eye-movement data, has suggested that the relationship between predictability and response time is linear (Brothers & Kuperberg, 2021). We report models in which cloze probability and PoS match were

group. All models contained random intercepts for subjects and words. Maximal random-effects structures corresponding to random slopes for all main effects and interactions resulted in non-convergence. Following L&C16, we trimmed the random-effects structures until convergence and, wherever possible, retained random slopes for the predictors that yielded significant fixed effects. The overall patterns of results are described below, and the full model outputs are presented in Table 6.

---TABLE 6 ABOUT HERE---

Proofreading was associated with significantly longer reading times, less skipping, and more regressions-out compared to reading for comprehension. Higher proficiency was associated with significantly shorter reading times and more skipping. Age group yielded a significant main effect on first-fixation duration, and there were significant interactions between age group and task on all fixation-duration measures and on skipping and regressions-out because, relative to younger readers, older readers showed a larger increase in reading time and regressions and a larger decrease in skipping in the proofreading task (see Figure 1). These findings parallel the global eye-movement measures in showing that task exerted a stronger influence than age group, but that the older adults' reading strategy was more sensitive to task demands than that of young adults.

---FIGURE 1 ABOUT HERE ----

There were significant facilitatory effects of target word cloze probability and LSA similarity on all fixation-duration measures, while POS match yielded a significant facilitatory effect on gaze duration but not first fixation or total duration. Higher LSA similarity and POS match were also associated with significantly higher skipping, and the main effect of cloze probability was also significant in the model for regressions-in. Higher

---

log transformed to allow comparison with L&C16's results. Parallel analyses including raw scores yielded slightly stronger effects of cloze probability but did not change the pattern of significant interactions.

LSA similarity was associated with significantly lower regressions-out but *higher* regressions-in.

Cloze probability did not significantly interact with task on any fixation-duration measure, but the effect of cloze probability on skipping and regressions-in was significantly stronger in the proofreading task. LSA similarity significantly interacted with task on gaze and total duration and on regressions-out, while POS match showed a significant task interaction on regressions-in.

Age group significantly interacted with the LSA similarity effect on gaze duration and with cloze probability on total duration due to weaker effects in older readers. Skipping rates also showed significant interactions between age group and both cloze probability and LSA similarity because older readers' word skipping was less sensitive to either predictability metric than younger readers. Regressions-in also showed significant 3-way interactions between age, task, and both cloze probability and POS match because older adults showed more marked differences between tasks in the impact of both predictability metrics on regressions-in: Relative to young adults, older adults made more regressions to words that were high in cloze probability, and to words low in POS match in the proofreading than the comprehension task.

In summary, the results of this set of analyses converge with L&C16's evidence that metrics assessing graded semantic and syntactic similarity make significant contributions over and above the effects of cloze probability, suggesting that partial semantic and syntactic information influenced eye movements over and above the impact of specific lexical predictions. The task interactions with the predictability effects generally confirm that the proofreading task achieved the intended goal of increasing reliance on prediction. Relative to young adults, older adults showed weaker effects of target cloze probability on fixation durations and skipping, but stronger effects of target cloze probability and PoS match on regressions-in. This pattern may reflect a trade-off between a reduced reliance on generating specific lexical predictions in the early processes (as indexed by skipping) and increased benefits of semantic fit on late rereading and integration processes (as indexed by regressions-in and total reading time).

Predictability Effects for Expected and Unexpected Words

As reviewed earlier, only about 30% of the words in the natural passages of the Provo corpus corresponded to the most expected word for that position (i.e., the modal cloze response), so specific lexical predictions are often violated. L&C16 proposed that reliance on such predictions could be diagnosed by comparing the direction of the relationship between the cloze probability and reading measures of expected (i.e., modal) and unexpected (i.e., non-modal) words. If readers make anticipatory lexical predictions, then there should be a stronger facilitatory effect of cloze probability for modal words compared to non-modal words. Furthermore, if incorrect predictions disrupt processing, then non-modal words should be read more slowly because of their incompatibility with the more expected modal words. L&C16's data provided no evidence for these predictions: The slope of the cloze probability function did not significantly differ between modal and non-modal words for any reading time measure of either content or function words. However, as we have discussed, the lack of any requirement for comprehension may have reduced L&C16's participants' tendency to generate specific lexical predictions.

To assess whether the present data provided evidence of differential predictability effects for expected and unexpected words, we followed L&C16's approach of testing models that included the fixed effect of modal versus non-modal word (effect-coded), task, verbal proficiency, and age group, and tested for interactions with the modality effect. Since cloze probability is definitionally higher for modal than non-modal words, like L&C16 we centered this metric on the mean for each word type so that interactions with the modal/non-modal factor tested differences in the slope of the cloze probability function (see Table 7 for LMM output).

---TABLE 7 ABOUT HERE---

Consistent with L&C16's findings, the average cloze probability effects on fixation durations did not differ between modal and non-modal words[3]. However, task significantly modulated the interaction between cloze probability and the modal/non-modal factor on gaze duration and regressions-out because the difference between modal and non-modal words was more marked in the comprehension than the proofreading task (see Figure 2).[4]

---FIGURE 2 ABOUT HERE---

Age group significantly modulated the interaction of the modal/non-modal factor with cloze probability on all fixation-duration measures because the stronger cloze probability effects for modal words were restricted to the young adult group (all $|t|$s > 1.98); older adults did not show a significant interaction on any reading time measure (all $|t|$s < 1; see Figure 3).

---FIGURE 3 ABOUT HERE---

Thus, the present data provide evidence of the stronger effects of cloze probability for modal than non-modal words that L&C16 argued "would be expected if full lexical prediction provides a boost to processing" (p. 35). No such effects were observed in their data suggesting that the lack of any requirement for comprehension reduced participants' tendency to generate specific lexical predictions. The discontinuity in the present data was more marked in the comprehension than the proofreading task, suggesting that the integrative processes required for comprehension may have increased reliance on lexical prediction. The further novel finding is that the enhanced cloze probability effect for modal words was restricted to young adults[5]. The slope of the cloze probability function for older adults was

[3] This interaction was significant on first-fixation time (t = 2.16) in the model that did not control for first/last words in each line and sentence.
[4] The direction of the cloze probability effect on regressions-out is counterintuitive in that higher cloze probability words showed higher regressions-out, particularly for modal words (see Figure 2b). The same effect was observed in L&C16's data. This pattern likely reflects the fact that the regressions-out measure is contingent on a word receiving a first-pass fixation. For words that had high skipping probabilities (such as highly predictable and/or short words), instances in which a reader fixated on the word were rare and likely reflected oculomotor error (e.g., a saccadic overshoot that was immediately corrected by a regression-out).
[5] The enhanced effects of cloze probability on modal words in younger readers were also significant in models that did not control for line and sentence position as per L&C16.

virtually identical for modal and non-modal words, and very similar to young adults' slope for non-modal words. This pattern of results is consistent with the view that younger adults are more likely to engage an additional process that selectively affects modal words (such as lexical prediction) that is rarely applied by older readers.

Are There Costs of Misprediction?

If lexical prediction is more likely to be applied by younger readers and in the comprehension task, then these factors can also be used to investigate whether costs associated with prediction failure can be observed in the eye-movement record. As reviewed earlier, the post-N400 frontal positivity observed in young adults' ERP waveforms has been attributed to processes associated with prediction failure (e.g., Kuperberg & Jaeger, 2016; Van Petten & Luka, 2012), but L&C16 found no evidence of prediction costs in their eye-movement data, and Frisson et al.'s (2017) eye-movement study of high-constraint sentence contexts also failed to find any evidence of costs when unexpected, plausible words replaced highly predictable continuations.

To evaluate whether the present data provided evidence of prediction costs, we conducted a final set of analyses that were limited to non-modal words (i.e., words that did not correspond to the most expected word in that position). L&C16 reasoned that the costs associated with misprediction should be stronger the higher the expectancy for a different word, and that reading times for unexpected, non-modal words should therefore increase with measures of the '*certainty'* of the prediction for the modal completion for that position. Certainty is inversely related to 'entropy', which is an information-complexity metric used in the sentence processing literature to quantify the diversity of cloze responses for a particular item. L&C16 found that formal measures of entropy were highly correlated (r = -.95) with the contextual constraint of the modal completion for that position. They therefore used this metric to index certainty because it was more interpretable and on the same scale as the cloze probability of the target word that actually occurred in the text (see Staub et al., 2015), while being uncorrelated with target cloze probability (*r* = .1). For non-modal target words, higher certainty means that the context strongly constrains towards a different word (the modal completion). An inhibitory effect of certainty on reading times for non-modal targets would therefore be evidence of prediction error cost. However, in L&C16's data, certainty was not significantly related to reading times for content words, and yielded facilitation rather than inhibition for function words.

To assess whether task and age modulated the impact of certainty on processing of content words in the present data, controlling for verbal proficiency, we tested models (see Table 8) that included both the certainty measure and the cloze probability for each non-modal content word. Higher certainty was associated with significant facilitative effects on skipping and all reading time measures: Skipping was higher and fixation times were shorter the stronger the context constrained towards a different word. Regressions-in also yielded a significant interaction between certainty and age because older adults showed a stronger facilitative effect of certainty on this measure than younger adults (see Figure 4). There were also significant three-way interactions between age group, task, and log cloze probability of non-modal words on regressions-in that reflected stronger cloze probability effects in older adults, particularly in the proofreading task.

---TABLE 8 & FIGURE 4 ABOUT HERE---

Thus, consistent with L&C16, the results provided no evidence of costs associated with misprediction; rather, all measures except regressions-out showed a processing *advantage* for non-modal words that replaced a more expected completion.[6] The analysis of certainty effects also complements the age group effects observed in the comparison of modal and non-modal responses by revealing that older adults were just as sensitive to the cloze probability of the most expected word as young adults on measures of fixation duration, and

[6] Models that did not control for line and sentence position did not yield the significant certainty effect on skipping rate reported by L&C16 (z = 1.53) but replicated their finding of a significant inhibitory effect of certainty on regressions-out (z = 2.01).

showed a stronger effect of certainty on regressions-in for non-modal words. However, these effects were facilitative rather than inhibitory confirming that the older group was relatively unaffected by the discrepancy between the actual target word and the most expected continuation.

General Discussion

The broad aim of this research was to extend Luke and Christianson's (2016, 2017) investigation of the impact of predictability on reading of natural texts from the Provo corpus by evaluating whether and how predictability effects are modulated by task demands and age. The general pattern of results replicated the major features of L&C16's eye-movement data. Despite the low average predictability of words in natural texts, higher cloze probability was associated with significantly higher skipping rates, shorter fixation times, and fewer regressions. However, indices of the semantic and syntactic similarity of words to the cloze predictions yielded pervasive facilitative effects across all reading measures, consistent with L&C16's conclusion that graded predictions about upcoming words play a role over and above any benefit due to specific lexical predictions. Our data also replicated L&C16's finding that cloze probability significantly facilitated processing not only for words that matched the most expected continuation, but also for unexpected words, thus providing further evidence that the benefits of predictability were not restricted to specific lexical predictions. Additional support for L&C16's conclusion that predictability effects reflect graded rather than all-or-none predictions for a specific word was provided by our failure to find any indication of processing costs for unexpected words in high-constraint positions.

The novel contributions of our research were to show that the limited impact of lexical prediction on young adults' eye movements observed by L&C16 may have reflected the absence of comprehension demands because, in the present task conditions, which required deeper semantic processing, younger adults appeared to engage anticipatory prediction processes. In contrast, although older adults showed greater overall adjustment of their eye-movement behavior in response to task demands, they appeared to rely on graded prediction in both tasks. The evidence for these conclusions and their implications for specifying the mechanisms underlying predictability effects are elaborated below.

Task and Age Effects

In contrast to most eye-movement studies, L&C16 did not explicitly assess comprehension to encourage and monitor participants' engagement of an integrative, meaning-focused reading strategy. The limited evidence of lexical prediction in their data may therefore reflect the adoption of a reading strategy that was 'good enough' (Ferreira et al., 2002) for these relatively undemanding reading conditions. To investigate this possibility, we compared a standard comprehension condition, in which comprehension questions followed approximately 50% of passages, with a proofreading task modelled on Schotter et al.'s (2014) study that required detection of TL word errors. As summarized in Table 4, young adults' first-pass reading measures for content words in the comprehension task were very similar to those reported by L&C16 (mean first-fixation and gaze durations of 216 and 259 ms, respectively, compared with 214 and 261 ms in L&C16), but our young adults showed a longer total duration and more regressions-in (355 ms and 0.21, respectively) than L&C16's sample (314 ms and 0.18), suggesting that the inclusion of comprehension questions encouraged more extensive, late integrative processing. Measures of both global and word-level reading confirmed that the proofreading task led to a more cautious, deliberate reading strategy characterized by more and longer fixations, shorter saccades, and reduced skipping. Supporting the assumption that the proofreading task encouraged greater reliance on active prediction to detect semantically incongruent words, cloze probability exerted significantly stronger effects in the proofreading than the comprehension task on skipping and regressions-in. An early effect on skipping is consistent with anticipatory generation of specific lexical predictions, but the late effect on rereading is more compatible with post-lexical integration accounts of predictability effects. The proofreading task also showed stronger effects of the LSA index of semantic similarity on gaze, total duration, and regressions-out, consistent with the involvement of graded predictions in the later stages of lexical processing and integration.

Controlling for differences in reading proficiency between older and younger age groups, the only significant age differences in average reading behavior occurred on the global measures of forward saccade length and average fixation duration, and word-level first-fixation duration: Older adults made longer forward saccades but showed longer global reading times and longer first-fixations than young adults. This trade-off is partially consistent with the meta-analytic findings of Zhang et al. (2022) and with the risky reading hypothesis of aging effects (Rayner et al., 2006), but there was no evidence of the increased skipping and regressions in older readers expected from this latter account. However, the older adults showed a stronger task effect on all reading-time measures that was principally due to a selective increase in reading time in the proofreading task; older adults' reading times were similar to those of young adults in the comprehension task (see Figure 1).

Older adults' more cautious reading in the proofreading task was mirrored in their performance: They achieved substantially higher accuracy than the young adults (d' = 3.3 vs. 1.9, respectively), particularly for correctly rejecting passages containing TL errors (85% vs. 71%, respectively). This superior performance did not appear to be due to increased reliance on anticipatory prediction: Older adults showed significantly weaker relationships between the predictability metrics and both skipping rate and reading times than younger readers. However, older readers showed stronger task modulation of the relationship between cloze probability and regressions-in than shown by younger readers, suggesting that older adults relied more on predictability during late, task-specific integration. Thus, the overall reading data reveal relatively few differences in the strategies adopted by our younger and older readers, suggesting that the observed differences in accuracy may, at least in part, occur because the older adults "tried harder" to perform the proofreading task.

The findings described above all rely on averaging the data for modal words, corresponding to words with the highest cloze probability for that position, and non-modal words that did not match the modal response and thus lower in average cloze probability. Separating these two groups of items provided insight into how task demands modulated the relative contribution of lexical and graded prediction to predictability effects and revealed qualitative differences between younger and older adults that were obscured in the averaged data.

In contrast to L&C16's young adult data, which showed equivalent effects of cloze probability on reading times for modal and non-modal content words, our young adult group showed a stronger effect of cloze probability for modal than non-modal words on all reading time measures – the pattern expected from reliance on specific lexical predictions. These effects appeared to principally reflect enhanced benefit for high cloze probability modal words, but there was some suggestion of a cost for the lower cloze probability modal words relative to non-modal words with equivalent cloze probability (see Figure 3). There was also evidence that reliance on lexical prediction was sensitive to task demands because the differential effect of cloze probability on modal and non-modal words was more marked on gaze duration and regressions-out in the comprehension than proofreading task.

Despite this evidence that the more stringent task demands of the present study led our younger sample to engage lexical prediction strategies that were not evident in L&C16's participants, our analyses of the impact of the 'certainty' of a high-constraint position on unexpected non-modal words converged with L&C16 in finding no evidence of the costs of prediction failure that have been observed in ERP studies. Indeed, our data yielded stronger evidence than L&C16 that words that replaced higher cloze probability continuations showed facilitation rather than interference. In their data, facilitative effects of certainty on reading times were observed for function words but not content words, which only showed significant certainty effects on skipping and refixations. In contrast, our content word data showed significant certainty effects on all reading time measures, demonstrating that fixation times were shorter the higher the cloze probability of a (different) word. Thus, the results converge with L&C16 and with Frisson et al.'s (2017) experimental study of high-constraint sentences in finding no evidence of disruptions to eye movements when predictions are disconfirmed.

The separation of modal and non-modal words also revealed qualitative differences between older and younger adults. There were significant age interactions on all fixation measures because the slope of the cloze probability function for the older adult group was equivalent for modal and non-modal words while our young adult sample showed a stronger effect of cloze probability on all reading time measures for modal versus non-modal words. Thus, older readers showed no evidence of the increased impact of cloze probability on modal words that L&C16 interpreted as a diagnostic of lexical prediction. Importantly, older adults' reading times were still significantly related to cloze probability, but the strength of that relationship was unaffected by whether or not the actual target was the highest cloze probability continuation for that word position. This implies that their processing was a function of the contextual fit of the word that actually occurred, regardless of other alternative continuations. However, this conclusion is qualified by the finding that older adults showed a stronger effect of certainty on regressions-in to non-modal words than young adults, implying that they *were* sensitive to the discrepancy between the presented word and the modal prediction, but that it only influenced later integration processes (see Figure 4). Consistent with this view, older adults showed stronger cloze probability effects on late processing of non-modal words in the proofreading task, suggesting that they engaged in late, post-lexical integrative processing that was sensitive to both task demands and the contextual fit of the unexpected non-modal words.

Implications for Understanding Predictability Effects

The present results broadly support L&C16's conclusion that the effects of cloze probability during online reading of natural texts are more consistent with graded prediction accounts that assume probabilistic activation of partial semantic and morphosyntactic features of upcoming words than with theories that assume specific, all-or-none lexical prediction. However, they complement and extend L&C16 by demonstrating that the procedures that they used to diagnose the involvement of lexical prediction are sensitive to the processing demands of the reading task and can reveal the application of this strategy under conditions that encourage its use. Our results show that, even in natural texts in which few words are highly predictable, younger adults do appear to engage lexical prediction strategies under more difficult processing conditions, particularly when they are required to detect semantically incongruous words. Such strategies are therefore not “circumscribed … to tightly controlled stimulus materials” (L&C16, p. 46), but are also modulated by processing demands. Younger adults also showed stronger effects of cloze probability on skipping rates, suggesting that lexical predictions influence both anticipatory pre-lexical processing and later reanalysis and integration.

The present findings strengthen L&C16’s conclusion that the eye-movement record appears to be insensitive to costs associated with prediction failures. Although our young adult data revealed use of lexical prediction strategies that were absent in L&C16’s data, we found stronger evidence that competing predictions facilitated rather than inhibited processing of unexpected words. This apparently counterintuitive facilitation implies that more constraining contexts yield benefits that are not captured by the cloze probabilities of individual words (Luke & Christianson, 2016). This view is consistent with Staub et al.’s (2015) finding that vocal cloze responses for equivalently predictable words were faster in higher constraint contexts. Such contexts narrow down the range of possible continuations and yield processing benefits that go beyond the most expected word to other, more weakly activated candidates. Staub et al. showed that the independent effects of sentence constraint and cloze probability on their vocal cloze responses were successfully simulated by an activation-based race model in which multiple candidates raced towards a response threshold

– an account that Luke and Christianson (2016) argue to be consistent with graded prediction views.

Within graded models, specific lexical predictions and graded predictions about the phonological, semantic, and/or syntactic attributes of upcoming words arise from the same processes, consistent with claims that “comprehenders make whatever linguistic predictions they can” (Pickering & Garrod, 2007, p. 341). However, anticipatory lexical prediction may reflect an independent processing strategy that potentially operates in tandem with the more automatic graded prediction mechanisms. The present evidence that both task and age modulate the differential impact of cloze probability on modal and non-modal words suggests that effects of lexical prediction are due, at least in part, to a dissociable processing strategy that is more likely to be applied in the proofreading than comprehension task, and by younger than older readers.

Implications for Understanding Age Effects on Reading

The overall reading behavior of our older adult sample was quantitatively similar to that of the younger adults, and they showed equivalently high levels of performance in the comprehension task and more accurate error detection in the proofreading task. The limited age differences in reading efficiency are presumably due, at least in part, to our recruitment of a relatively elite older sample. The majority (68%) had completed a college degree and their average vocabulary score was significantly higher than the young adult group. However, consistent with the well-established evidence that crystallized abilities are both more resistant to age-related cognitive declines and show stronger protective effects of education than do tasks that depend on memory and cognitive speed (e.g., Christensen et al., 1997), the older adults did not show a similar superiority in the standardized test of reading comprehension.

The evidence of relatively modest aging effects on reading in the comprehension task may also reflect the use of the naturalistic, multi-sentence passages of the Provo corpus as materials in the present study. Zhang et al. (2022) note that previous eye-movement research on aging has been dominated by studies of the reading of isolated sentences. Indeed, we have observed typical aging effects in an overlapping sample of older adults reading isolated sentences for comprehension (Veldre et al., 2022). This may suggest that the characteristic eye-movement signature of older readers that has been attributed to the risky reading strategy does not generalize to paragraph reading. Further research that directly compares older adults' reading of sentences versus paragraphs may shed light on this possibility.

Nevertheless, controlling for the age differences in verbal proficiency, the results provided evidence of two qualitative differences between the reading strategies of the older and younger samples. First, older adults showed stronger task effects on reading times than the younger readers, suggesting that they were more likely to adjust their strategy to meet task demands. The success of their more cautious reading strategy in the proofreading task was reflected in substantially higher accuracy than achieved by the young adults. Secondly, unlike young adults, older adults did not show the differential effects of cloze probability on reading times for modal and non-modal words that L&C16 proposed to be an index of reliance on lexical prediction. While this might be taken as evidence for the age-related reduction in sensitivity to contextual predictability that has been inferred from the reduced N400 effect observed in ERP studies of older readers (e.g., Wlotko et al., 2012), older adults showed stronger effects of the mismatch with modal expectancies on rereading of non-modal words than younger adults. This pattern of results implies that older readers made similar predictions about upcoming word to their young counterparts, but that these expectancies influenced late post-lexical integration processes rather than affecting initial online text processing. In combination, these findings suggest that older readers *can* effectively use context to generate predictions about upcoming words. However, at least for the naturalistic texts used in the present study in which few words are highly predictable, they do not appear to prioritize these top-down predictions over the extraction of bottom-up information from the text; that is, they do not make 'risky guesses' (Choi et al., 2017). Rather than committing to a specific lexical prediction, they appear to activate multiple possible continuations that are compatible with the preceding context. This reliance on probabilistic predictions may reflect a resource allocation strategy that harnesses older adults' richer crystallized linguistic knowledge to enhance lexical processing and the construction of meaning in order to compensate for declines in processing speed and working memory (Stine-Morrow et al., 2008).

This interpretation may be extended to account for evidence of enhanced predictability effects among older adults in studies that typically use isolated high-constraint sentences (e.g., Veldre et al., 2022). Although Zhang et al.'s (2022) meta-analysis did not find significant age differences in predictability effects for alphabetic scripts due to relatively few studies meeting the inclusion criteria, there were numerically larger predictability effects for older adults on gaze and total duration, but not skipping. This pattern may also reflect older adults' use of their superior linguistic knowledge to know *when* it is optimal to apply a contextually based reading strategy (i.e., in a linguistic environment with an unusually high proportion of constraining sentences). The suggestion from Zhang et al.'s meta-analysis that these age differences are restricted to late measures of reading is thus consistent with the findings of the present study that older readers typically adopt a graded prediction strategy that benefits post-lexical integration.

It is important to acknowledge that the different processing strategies adopted by the two age groups may arise from other differences between the age cohorts apart from age *per se*. Older adults have had the opportunity to acquire substantially more reading experience than young adults. In addition to enhancing their crystallized linguistic knowledge, this experience may increase their capacity to adjust their reading strategy to different task goals and reading contexts. Community samples of older readers who volunteer to participate in research for financial reimbursement may also differ motivationally from young university students who participate to meet course requirements. As well as increasing the attentional resources allocated to the task, such motivational differences may contribute to the increased adaptation to task demands observed in the present older sample.

Conclusion

The results of this research converge with those of Luke and Christianson (2016) in raising questions about whether the evidence of lexical prediction obtained in the prediction-friendly environments used in most previous research generalize to more naturalistic texts. They also add to previous evidence that, to the extent that anticipatory prediction occurs, it does not appear to yield any costs of prediction failure that are observable in the eye-movement record. Rather, more constraining sentence contexts appear to activate a range of possible continuations, consistent with graded prediction accounts. Our findings extend those of Luke and Christianson by showing that readers' use of the different levels of word predictability that are afforded by linguistic content and structure is highly nuanced, being influenced by both task demands and differences in the impact of readers' age and experience on the way that they respond to such demands. The results highlight the importance of assessing the impact of stimulus materials, task demands, and individual differences in future empirical studies of predictive processing and show that a comprehensive model of reading will need to consider such factors. Our ongoing research is directed towards developing a model of eye-movement control in reading (Reichle, 2021) that achieves that goal.

References

Altmann, G.T. & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33,* 583-609.

Andrews, S., & Veldre, A. (2021). Wrapping up sentence comprehension: The role of task demands and individual differences. *Scientific Studies of Reading*, *25*, 123-140.

Andrews, S., Veldre, A., & Clarke, I. E. (2020). Measuring lexical quality: The role of spelling ability. *Behavior Research Methods, 52*, 2257-2282.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*, 364-390.

Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions: Biological Sciences, 364,* 1235-1243.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1-48.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension, *Journal of Memory and Language*, *93*, 203-216.

Brown, J. I., Fishco, V. V., & Hanna, G. (1993). *Nelson-Denny Reading Test*. Riverside Publishing Company.

Brown, C. & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5,* 34-44.

Choi, W., Lowder, M. W., Ferreira, F., Swaab, T. Y., & Henderson, J. M. (2017). Effects of word predictability and preview lexicality on eye movements during reading: A comparison between young and older adults. *Psychology and Aging*, *32*, 232.

Christensen, H., Henderson, A. S., Griffiths, K., & Levings, C. (1997). Does ageing inevitably lead to declines in cognitive performance? A longitudinal study of elite academics. *Personality and Individual Differences, 23,* 67-78.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 1–72.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*), 181–204.

Dave, S., Brothers, T. A., Traxler, M. J., Ferreira, F., Henderson, J. M., & Swaab, T. Y. (2018). Electrophysiological evidence for preserved primacy of lexical prediction in aging. *Neuropsychologia*, *117*, 135–147.

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass, 8*, 631–645.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language, 115*, 149–161.

Federmeier, K. D., & Laszlo, S. (2009). Time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. *Psychology of Learning and Motivation, 51*, 1–44.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*, 11–15.

Frisson, S., Harvey, D.R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language, 95*, 200-214.

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*, 127–138.

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological Science, 24,* 1079-1088.

Gough, P. B., Alford, J., & Holley-Wilcox, P. (1981). Words and contexts. In O. Tzeng & H. Singer (Eds.) *Perception of print: Reading research in experimental psychology* (pp. 85–102). Hillsdale, NJ: Erlbaum.

Hale, J. (2011). What a rational parser would do. *Cognitive Science, 35*, 399–443.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626,* 118-135.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience, 31*, 19–31.

Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research, 1146,* 2-22.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16,* 262-284.

Kuperberg, G. (2013). The proactive comprehender: What event-related potentials tell us about the dynamics of reading comprehension. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling reading comprehension: Behavioral, neurobiological, and genetic components* (pp. 176–192). Baltimore, MD: Paul Brookes.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*, 32–59.

Kutas, M., DeLong, K.A. & Smith, N.J. (2011). A look around what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the Brain: Using our Past to Generate a Future* (pp. 190-207). Oxford University Press.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.

Lahar, C. J., Tun, P. A., & Wingfield, A. (2004). Sentence-final word completion norms for young, middle-aged, and older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 59*, P7–P10.

Landauer, T. L., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*, 259-284.

Laubrock, J., Kliegl, R., & Engbert, R. (2006). SWIFT explorations of age differences in eye movements during reading. *Neuroscience and Biobehavioral Reviews, 30*(6), 872-884.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126-1177.

Light, L. L., Valencia-Laver, D., & Zavis, D. (1991). Instantiation of general terms in young and older adults. *Psychology and Aging, 6*, 337.

Little, D. M., Prentice, K. J., & Wingfield, A. (2004). Adult age differences in judgments of semantic fit. *Applied Psycholinguistics, 25*, 135–143.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology, 88*, 22-60.

Luke, S. G., & Christianson, K. (2017). The Provo Corpus: A large eye-tracking corpus with predictability norms. *Behavior Research Methods, 50,* 826-833.

Madden, D. J. (1988). Adult age differences in the effects of sentence context and stimulus degradation during visual word recognition. *Psychology and Aging, 3*, 167.

McGowan, V. A., & Reichle, E. D. (2018). The “risky” reading strategy revisited: New simulations using E-Z Reader. *Quarterly Journal of Experimental Psychology, 71,* 179-189.

McKoon, G., & Ratcliff, R. (2018). Adults with poor reading skills, older adults, and college students: The meanings they understand during reading using a diffusion model analysis. *Journal of Memory and Language*, *102*, 115–129.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*, 226-254.

Paterson, K. B., McGowan, V. A., Warrington, K. L., Li, L., Li, S., Xie, F., Chang, M., Zhao, S., Pagán, A., White, S. J., & Wang, J. (2020). Effects of normative aging on eye movements during reading. *Vision, 4*, 7.

Payne, B. R., & Silcox, J. W. (2019). Aging, context processing, and comprehension. *Psychology of Learning and Motivation, 71*, 215-264.

Perfetti, C. & Helder, A. (2021). Incremental comprehension examined in event-related potentials: Word-to-text integration and structure building. *Discourse Processes, 58,* 2-21.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*, 105–110.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R.L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55-85). Hillsdale, NJ: Erlbaum.

R Core Team. (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control. *Psychological Research, 72*, 675-688.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review, 3*, 504–509.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 720-732.

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging, 21*, 448-465.

Rayner, K., Castelhano, M. S., & Yang, J. (2010). Preview benefit during eye fixations in reading for older and younger readers. *Psychology and Aging, 25*, 714-718.

Rayner, K., Yang, J., Schuett, S., & Slattery, T. J. (2013). Eye movements of older and younger readers when reading unspaced text. *Experimental Psychology, 60*, 354-361.

Reichle, E. D. (2021). *Computational models of reading: A handbook*. Oxford, UK: Oxford University Press.

Rubenstein, H., & Aborn, M. (1958). Learning, prediction, and readability. *Journal of Applied Psychology*, *42*, 28.

Schotter, E., Bicknell, K., Howard, I., Levy, R., & Rayner, K. (2014). Task effects reveal cognitive flexibility in responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition, 131*, 1-27.

Slattery, T. J. & Parker, A. J. (2019), Return sweeps in reading: Processing implications of undersweep-fixations. *Psychonomic Bulletin & Review, 26*, 1948-1957.

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass, 9*, 311-327.

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language, 82*, 1–17.

Stine-Morrow, E. A. L., Soederberg Miller, L. M., Gagne, D. D., & Herzog, C. (2008). Self-regulated reading in adulthood. *Psychology and Aging, 23*, 131-153.

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

Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*, 1266–1282.

Valencia-Laver, D. L., & Light, L. L. (2000). The occurrence of causal bridging and predictive inferences in young and older adults. *Discourse Processes, 30*, 27–56.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*, 176–190.

Veldre, A., Wong, R., & Andrews, S. (2021). Reading proficiency predicts the extent of the right, but not left, perceptual span in older readers. *Attention, Perception, & Psychophysics, 83*, 18-26.

Veldre, A., Wong, R., & Andrews, S. (2022). Predictability effects and parafoveal processing in older readers. *Psychology and Aging*, *37*, 222-238.

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading, *Cortex*, *68,* 20-32.

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging, 27*, 975-988.

Wlotko, E. W., Lee, C.-L., & Federmeier, K. D. (2010). Language of the aging brain: Event-related potential studies of comprehension in older adults. *Language and Linguistics Compass*, *4*, 623-638.

Wotschack, C., & Kliegl, R. (2013). Reading strategy modulates parafoveal-on-foveal effects in sentence reading. *Quarterly Journal of Experimental Psychology*, *66*, 548-562.

Zhang, J., Warrington, K. L., Pagán, A., Paterson, K. B., White, S. J., & McGowan, V. A. (2022). Are older adults more risky readers? Evidence from meta-analysis. *Psychology and Aging*, *37*, 239-259.

Appendix

Table A1

Mean (and standard deviations) of the reading measures for the function words (*n* = 946) for each task and age group.

Measure	**Comprehension**		**Proofreading**	
	Younger	**Older**	**Younger**	**Older**
First-fixation duration (ms)	205	206	210	220
	(26)	(24)	(26)	(29)
Gaze duration (ms)	219	216	227	234
	(30)	(28)	(31)	(35)
Total duration (ms)	271	260	284	284
	(43)	(38)	(46)	(43)
Skipping	0.62	0.66	0.60	0.61
	(0.08)	(0.08)	(0.08)	(0.09)
Regressions out	0.25	0.24	0.24	0.22
	(0.10)	(0.09)	(0.09)	(0.08)
Regressions in	0.17	0.17	0.17	0.17
	(0.05)	(0.07)	(0.05)	(0.06)

Table 1

Mean (and standard deviation) of scores on the Vocabulary and Reading Comprehension sub-tests of the Nelson-Denny Reading Test and performance in the comprehension and proofreading tasks, for the groups of younger and older adults.

Measure	*Younger* *M (SD)*	*Older* *M (SD)*	*t*	*p*
Nelson-Denny Vocabulary	0.72 (0.14)	0.86 (0.12)	**5.27**	**<.001**
Nelson-Denny Reading Rate (wpm)	264 (78)	297 (86)	1.92	.058
Nelson-Denny Reading Comprehension	0.41 (0.16)	0.37 (0.16)	-1.30	.198
Comprehension accuracy (filler passages)	0.91 (0.08)	0.90 (0.08)	-0.26	.797
Proofreading accuracy (filler passages)	0.71 (0.14)	0.85 (0.10)	**5.37**	**<.001**
Proofreading accuracy (Provo passages)	0.88 (0.09)	0.94 (0.06)	**3.79**	**<.001**
Proofreading *d'*	1.92 (0.74)	3.30 (1.54)	**5.88**	**<.001**

Note. Nelson-Denny Vocabulary and Reading Comprehension sub-tests were administered for half the standard time limit; scores are proportion correct. Significant group differences are indicated in bold.

Table 2

Examples of stimulus materials showing a Provo passage and two filler passages, one from the comprehension block and one from the proofreading block.

CRITICAL PROVO PASSAGE

When it comes to having a lasting and fulfilling relationship, common wisdom says that feeling close to your romantic partner is paramount. But a new study finds that it's not how close you feel that matters most, it's whether you are as close as you want to be, even if that's really not close at all.

FILLER PASSAGE: COMPREHENSION BLOCK

Before the exam, Janice felt very relaxed and **calm** because she had studied hard. But she had the vague sense of a headache brewing so she decided to have an early night's sleep straight after dinner. Her dad had also kindly offered to drive her to school first thing in the morning.
COMPREHENSION QUESTION: *Her father was trying to be… (1) Supportive (2) Difficult.*

FILLER PASSAGE: PROOFREADING BLOCK

Before the exam, Janice felt very relaxed and **clam** because she had studied hard. But she had the vague sense of a headache brewing so she decided to have an early night's sleep straight after dinner. Her dad had also kindly offered to drive her to school first thing in the morning.
QUESTION: *Was there an error in this passage? (1) Yes (2) No.*

Note. The target word and its corresponding TL word are indicated here in bold.

Table 3

Mean (and standard deviation) of the global reading measures for younger and older readers in the comprehension and proofreading tasks.

	Comprehension		Proofreading	
Measure	*Younger*	*Older*	*Younger*	*Older*
Total passage reading duration (ms)	13,099	12,587	14,472	14,526
	(1,896)	(2,116)	(1,894)	(2,367)
Fixation count	58.22	54.13	62.52	60.77
	(9.29)	(12.32)	(8.58)	(13.65)
Average fixation duration (ms)	204	207	216	225
	(25)	(25)	(24)	(29)
Forward saccade length (letters)	8.58	9.62	7.93	8.61
	(1.50)	(1.68)	(1.32)	(1.48)
Regression count	13.73	13.34	14.71	14.87
	(4.45)	(5.47)	(4.13)	(5.79)

Table 4

LMM summaries for analyses of global reading measures.

Measure	Fixed effect	*b*	*SE*	*t/z*
Total passage duration	**Intercept**	**14979.50**	**318.40**	**47.04**
	Age group	353.10	531.70	0.66
	Task	**2095.90**	**246.50**	**8.50**
	Verbal proficiency	**-2029.90**	**310.20**	**-6.54**
	Age × Task	822.90	451.10	1.82
Forward saccade length	**Intercept**	**8.68**	**0.15**	**59.73**
	Age group	**0.64**	**0.28**	**2.26**
	Task	**-0.80**	**0.08**	**-9.79**
	Verbal proficiency	**0.67**	**0.16**	**4.22**
	Age × Task	**-0.38**	**0.15**	**-2.53**
Fixation count	**Intercept**	**4.06**	**0.02**	**207.39**
	Age group	-0.03	0.03	-1.01
	Task	**0.10**	**0.01**	**6.55**
	Verbal proficiency	**-0.09**	**0.02**	**-4.86**
	Age × Task	0.04	0.03	1.43
Average fixation duration	**Intercept**	**213.02**	**2.49**	**85.68**
	Age group	**9.67**	**4.94**	**1.96**
	Task	**15.47**	**1.10**	**14.01**
	Verbal proficiency	**-11.04**	**2.90**	**-3.81**
	Age × Task	**5.57**	**2.14**	**2.61**
Regression count	**Intercept**	**2.59**	**0.04**	**70.24**
	Age group	-0.02	0.07	-0.23
	Task	**0.11**	**0.02**	**4.86**
	Verbal proficiency	-0.05	0.04	-1.38
	Age × Task	0.03	0.04	0.70

***Note*.** Significant effects indicated in bold.

Table 5

Mean (and standard deviations) of the reading measures for the content word (*n* = 1,444) for each task and age group.

Measure	**Comprehension**		**Proofreading**	
	Younger	**Older**	**Younger**	**Older**
First-fixation duration (ms)	216	220	226	239
	(24)	(27)	(23)	(31)
Gaze duration (ms)	259	249	283	289
	(32)	(37)	(30)	(49)
Total duration (ms)	355	339	400	397
	(59)	(67)	(54)	(78)
Skipping	0.30	0.31	0.27	0.26
	(0.07)	(0.10)	(0.07)	(0.08)
Regressions out	0.27	0.26	0.26	0.27
	(0.08)	(0.07)	(0.07)	(0.08)
Regressions in	0.21	0.21	0.22	0.21
	(0.06)	(0.07)	(0.06)	(0.07)

Table 6

LMM summaries of analyses of predictability effects on local eye-movement measures of reading of content words.

Measure	Fixed effect	*b*	*SE*	*t/z*
(log) First fixation duration	**Intercept**	**5.35**	**0.01**	**483.62**
	Verbal proficiency	**-0.05**	**0.01**	**-3.78**
	Line initial	**-0.09**	**0.01**	**-7.12**
	Line final	**-0.07**	**0.01**	**-5.77**
	Sentence final	0.02	0.01	1.63
	Passage final	0.01	0.02	0.41
	Task	**0.06**	**0.00**	**15.09**
	Age group	**0.06**	**0.02**	**2.70**
	(log) Cloze probability	**-0.01**	**0.00**	**-2.67**
	LSA similarity	**-0.03**	**0.00**	**-6.31**
	(log) PoS match	-0.01	0.00	-1.71
	Age × Task	**0.03**	**0.01**	**4.05**
	Task × (log) Cloze probability	0.00	0.00	-1.20
	Task × LSA similarity	0.00	0.00	-0.13
	Task × (log) PoS match	0.00	0.00	-0.20
	Age × (log) Cloze probability	0.01	0.00	1.25
	Age × LSA similarity	0.00	0.00	0.85
	Age × (log) PoS match	0.00	0.00	-0.89
	Task × Age × (log) Cloze probability	0.00	0.01	0.17
	Task × Age × LSA similarity	0.00	0.01	-0.46
	Task × Age × (log) PoS match	0.01	0.00	1.31
(log) Gaze duration	**Intercept**	**5.46**	**0.01**	**448.46**
	Verbal proficiency	**-0.07**	**0.01**	**-5.04**
	Line initial	**0.29**	**0.02**	**16.55**
	Line final	**-0.06**	**0.02**	**-3.51**
	Sentence final	**0.05**	**0.02**	**2.63**
	Passage final	**0.06**	**0.03**	**2.03**
	Task	**0.10**	**0.01**	**15.31**
	Age group	0.03	0.02	1.35
	(log) Cloze probability	**-0.02**	**0.01**	**-3.32**
	LSA similarity	**-0.06**	**0.01**	**-10.99**
	(log) PoS match	**-0.01**	**0.00**	**-2.59**
	Age × Task	**0.05**	**0.01**	**3.49**
	Task × (log) Cloze probability	0.00	0.00	-0.19
	Task × LSA similarity	**-0.02**	**0.00**	**-4.75**
	Task × (log) PoS match	0.00	0.00	-0.17
	Age × (log) Cloze probability	0.01	0.01	1.54
	Age × LSA similarity	**0.01**	**0.00**	**3.25**
	Age × (log) PoS match	0.00	0.00	0.26
	Task × Age × (log) Cloze probability	0.00	0.01	0.18
	Task × Age × LSA similarity	0.00	0.01	-0.50
	Task × Age × (log) PoS match	0.01	0.01	1.73
(log) Total duration	**Intercept**	**5.77**	**0.02**	**341.68**
	Verbal proficiency	**-0.10**	**0.02**	**-5.64**
	Line initial	0.03	0.02	1.26
	Line final	**-0.14**	**0.02**	**-6.09**
	Sentence final	0.03	0.03	1.06

	Passage final	**0.22**	**0.04**	**5.13**
	Task	**0.15**	**0.01**	**22.58**
	Age group	0.03	0.03	0.95
	(log) Cloze probability	**-0.04**	**0.01**	**-4.94**
	LSA similarity	**-0.11**	**0.01**	**-12.45**
	(log) PoS match	-0.01	0.01	-1.95
	Age × Task	**0.05**	**0.01**	**5.46**
	Task × (log) Cloze probability	0.02	0.01	1.94
	Task × LSA similarity	**-0.04**	**0.01**	**-6.37**
	Task × (log) PoS match	0.00	0.00	-0.91
	Age × (log) Cloze probability	**0.01**	**0.00**	**2.56**
	Age × LSA similarity	0.00	0.01	0.25
	Age × (log) PoS match	0.01	0.00	1.49
	Task × Age × (log) Cloze probability	0.02	0.01	1.21
	Task × Age × LSA similarity	0.00	0.01	-0.05
	Task × Age × (log) PoS match	-0.01	0.01	-0.87
Skipping	**Intercept**	**-1.25**	**0.06**	**-21.76**
	Verbal proficiency	**0.16**	**0.06**	**2.70**
	Line initial	**1.87**	**0.11**	**16.74**
	Line final	**-0.28**	**0.12**	**-2.30**
	Sentence final	**-0.47**	**0.14**	**-3.45**
	Passage final	-0.18	0.22	-0.80
	Task	**-0.28**	**0.03**	**-9.64**
	Age group	-0.11	0.10	-1.10
	(log) Cloze probability	-0.08	0.04	-1.87
	LSA similarity	**0.57**	**0.04**	**13.71**
	(log) PoS match	**0.07**	**0.03**	**2.53**
	Age × Task	**-0.13**	**0.06**	**-2.29**
	Task × (log) Cloze probability	**0.04**	**0.02**	**2.13**
	Task × LSA similarity	0.03	0.02	1.47
	Task × (log) PoS match	-0.01	0.02	-0.41
	Age × (log) Cloze probability	**-0.12**	**0.02**	**-5.97**
	Age × LSA similarity	**0.07**	**0.03**	**2.35**
	Age × (log) PoS match	0.02	0.02	1.03
	Task × Age × (log) Cloze probability	0.07	0.04	1.72
	Task × Age × LSA similarity	-0.05	0.04	-1.12
	Task × Age × (log) PoS match	0.01	0.03	0.32
Regressions out	**Intercept**	**-1.32**	**0.05**	**-25.15**
	Verbal proficiency	0.00	0.06	0.08
	Line initial	**-1.53**	**0.13**	**-12.03**
	Line final	-0.19	0.10	-1.86
	Sentence final	**0.63**	**0.12**	**5.42**
	Passage final	**3.36**	**0.19**	**17.52**
	Task	**0.04**	**0.02**	**2.43**
	Age group	-0.02	0.09	-0.25
	(log) Cloze probability	0.03	0.04	0.80
	LSA similarity	**-0.09**	**0.04**	**-2.54**
	(log) PoS match	0.02	0.03	0.74
	Age × Task	**0.11**	**0.03**	**3.13**
	Task × (log) Cloze probability	0.03	0.03	1.32
	Task × LSA similarity	**-0.05**	**0.03**	**-2.06**
	Task × (log) PoS match	0.02	0.02	0.86
	Age × (log) Cloze probability	0.00	0.03	0.13

	Age × LSA similarity	-0.02	0.03	-0.63
	Age × (log) PoS match	-0.01	0.02	-0.54
	Task × Age × (log) Cloze probability	0.03	0.05	0.62
	Task × Age × LSA similarity	-0.09	0.05	-1.60
	Task × Age × (log) PoS match	-0.03	0.04	-0.81
Regressions in	**Intercept**	**-1.25**	**0.05**	**-24.23**
	Verbal proficiency	0.07	0.06	1.17
	Line initial	**2.10**	**0.09**	**22.76**
	Line final	**-2.32**	**0.13**	**-18.51**
	Sentence final	**-0.86**	**0.12**	**-7.45**
	Passage final	-15.68	54.42	-0.29
	Task	0.02	0.02	1.24
	Age group	-0.08	0.09	-0.84
	(log) Cloze probability	**-0.23**	**0.03**	**-6.64**
	LSA similarity	**0.13**	**0.03**	**3.66**
	(log) PoS match	-0.01	0.02	-0.33
	Age × Task	-0.02	0.03	-0.64
	Task × (log) Cloze probability	**0.12**	**0.02**	**4.83**
	Task × LSA similarity	-0.03	0.02	-1.26
	Task × (log) PoS match	**-0.08**	**0.02**	**-4.81**
	Age × (log) Cloze probability	-0.01	0.02	-0.51
	Age × LSA similarity	-0.05	0.03	-1.75
	Age × (log) PoS match	0.03	0.02	1.84
	Task × Age × (log) Cloze probability	**0.15**	**0.05**	**2.95**
	Task × Age × LSA similarity	-0.06	0.05	-1.32
	Task × Age × (log) PoS match	**-0.07**	**0.03**	**-2.21**

Note. Significant effects indicated in bold.

Table 7

LMM summaries of analyses of predictability effects on local eye-movement measures of reading of expected versus unexpected content words

Measure	Fixed effect	*b*	*SE*	*t/z*
(log) First fixation duration	**Intercept**	**5.33**	**0.01**	**473.43**
	Verbal proficiency	**-0.05**	**0.01**	**-3.78**
	Line initial	**-0.09**	**0.01**	**-6.76**
	Line final	**-0.07**	**0.01**	**-5.68**
	Sentence final	**0.03**	**0.01**	**1.96**
	Passage final	0.01	0.02	0.26
	Age group	**0.06**	**0.02**	**2.91**
	Task	**0.06**	**0.00**	**21.73**
	Modal response	**-0.06**	**0.01**	**-9.48**
	(log) Cloze probability (centered)	**-0.02**	**0.01**	**-4.59**
	Age × Task	**0.03**	**0.01**	**6.31**
	Age × Modal response	**0.01**	**0.01**	**2.73**
	Task × Modal response	-0.01	0.01	-1.49
	Age × (log) Cloze probability	**0.01**	**0.00**	**2.67**
	Task × (log) Cloze probability	0.01	0.00	1.31
	Modal response × (log) Cloze probability	-0.02	0.01	-1.70
	Age × Task × Modal response	-0.01	0.01	-0.58
	Age × Task × (log) Cloze probability	0.01	0.01	1.23
	Age × Modal response × (log) Cloze	**0.02**	**0.01**	**1.98**
	Task × Modal response × (log) Cloze	0.01	0.01	1.82
	Age × Task × Modal response × (log) Cloze	0.01	0.02	0.69
(log) Gaze duration	**Intercept**	**5.41**	**0.01**	**430.74**
	Verbal proficiency	**-0.07**	**0.01**	**-5.04**
	Line initial	**0.29**	**0.02**	**16.42**
	Line final	**-0.06**	**0.02**	**-3.39**
	Sentence final	**0.06**	**0.02**	**3.18**
	Passage final	0.05	0.03	1.55
	Age group	0.04	0.02	1.85
	Task	**0.10**	**0.00**	**29.18**
	Modal response	**-0.13**	**0.01**	**-12.83**
	(log) Cloze probability (centered)	**-0.05**	**0.01**	**-6.34**
	Age × Task	**0.05**	**0.01**	**7.49**
	Age × Modal response	**0.04**	**0.01**	**6.03**
	Task × Modal response	**-0.02**	**0.01**	**-2.06**
	Age × (log) Cloze probability	**0.03**	**0.00**	**5.80**
	Task × (log) Cloze probability	0.00	0.00	0.06
	Modal response × (log) Cloze probability	-0.02	0.01	-1.26
	Age × Task × Modal response	-0.01	0.02	-0.46
	Age × Task × (log) Cloze probability	**0.02**	**0.01**	**2.03**
	Age × Modal response × (log) Cloze	**0.04**	**0.01**	**4.57**
	Task × Modal response × (log) Cloze	**0.02**	**0.01**	**2.20**
	Age × Task × Modal response × (log) Cloze	0.02	0.02	1.25
(log) Total duration	**Intercept**	**5.64**	**0.02**	**363.07**
	Verbal proficiency	**-0.09**	**0.02**	**-5.50**
	Line initial	0.04	0.02	1.72
	Line final	**-0.14**	**0.02**	**-5.84**
	Sentence final	0.05	0.03	1.85

	Passage final	**0.19**	**0.04**	**4.32**
	Age group	0.04	0.03	1.27
	Task	**0.12**	**0.00**	**31.60**
	Modal response	**-0.21**	**0.01**	**-15.60**
	(log) Cloze probability (centered)	**-0.07**	**0.01**	**-7.40**
	Age × Task	**0.05**	**0.01**	**7.00**
	Age × Modal response	**0.03**	**0.01**	**4.34**
	Task × Modal response	-0.03	0.02	-1.74
	Age × (log) Cloze probability	**0.03**	**0.01**	**4.72**
	Task × (log) Cloze probability	-0.01	0.01	-1.20
	Modal response × (log) Cloze probability	-0.02	0.02	-1.20
	Age × Task × Modal response	0.01	0.03	0.30
	Age × Task × (log) Cloze probability	0.01	0.01	1.24
	Age × Modal response × (log) Cloze	**0.04**	**0.01**	**3.82**
	Task × Modal response × (log) Cloze	0.00	0.01	-0.38
	Age × Task × Modal response × (log) Cloze	0.00	0.02	0.06
Skipping	**Intercept**	**-1.00**	**0.06**	**-16.56**
	Verbal proficiency	**0.16**	**0.06**	**2.79**
	Line initial	**1.79**	**0.12**	**15.05**
	Line final	**-0.27**	**0.13**	**-2.06**
	Sentence final	**-0.60**	**0.14**	**-4.14**
	Passage final	-0.07	0.23	-0.31
	Age group	-0.14	0.10	-1.44
	Task	**-0.24**	**0.02**	**-15.73**
	Modal response	**0.72**	**0.07**	**10.44**
	(log) Cloze probability (centered)	**0.24**	**0.05**	**4.60**
	Age × Task	**-0.11**	**0.03**	**-3.46**
	Age × Modal response	**-0.08**	**0.03**	**-2.64**
	Task × Modal response	**0.09**	**0.04**	**2.23**
	Age × (log) Cloze probability	-0.05	0.03	-1.86
	Task × (log) Cloze probability	0.03	0.02	1.43
	Modal response × (log) Cloze probability	0.17	0.10	1.64
	Age × Task × Modal response	0.08	0.09	0.88
	Age × Task × (log) Cloze probability	0.03	0.05	0.59
	Age × Modal response × (log) Cloze	0.00	0.05	0.06
	Task × Modal response × (log) Cloze	-0.01	0.05	-0.29
	Age × Task × Modal response × (log) Cloze	0.01	0.10	0.12
Regressions out	**Intercept**	**-1.35**	**0.06**	**-24.28**
	Verbal proficiency	0.00	0.06	0.08
	Line initial	**-1.52**	**0.13**	**-11.92**
	Line final	-0.20	0.10	-1.94
	Sentence final	**0.66**	**0.12**	**5.73**
	Passage final	**3.33**	**0.19**	**17.31**
	Age group	-0.02	0.10	-0.24
	Task	0.04	0.02	1.68
	Modal response	-0.09	0.06	-1.63
	(log) Cloze probability (centered)	0.03	0.04	0.77
	Age × Task	**0.10**	**0.04**	**2.33**
	Age × Modal response	0.00	0.04	-0.05
	Task × Modal response	-0.03	0.04	-0.62
	Age × (log) Cloze probability	0.03	0.03	0.88
	Task × (log) Cloze probability	**-0.07**	**0.03**	**-2.29**

	Modal response × (log) Cloze probability	0.06	0.09	0.70
	Age × Task × Modal response	0.01	0.08	0.09
	Age × Task × (log) Cloze probability	-0.09	0.06	-1.37
	Age × Modal response × (log) Cloze	0.08	0.06	1.24
	Task × Modal response × (log) Cloze	**-0.16**	**0.06**	**-2.58**
	Age × Task × Modal response × (log) Cloze	-0.08	0.13	-0.62
Regressions in	**Intercept**	**-1.33**	**0.05**	**-24.55**
	Verbal proficiency	0.07	0.06	1.20
	Line initial	**2.08**	**0.09**	**22.49**
	Line final	**-2.31**	**0.13**	**-18.33**
	Sentence final	**-0.90**	**0.12**	**-7.80**
	Passage final	-15.61	54.94	-0.28
	Age group	-0.09	0.10	-0.89
	Task	**0.04**	**0.02**	**2.14**
	Modal response	**-0.20**	**0.06**	**-3.57**
	(log) Cloze probability (centered)	**-0.13**	**0.04**	**-3.06**
	Age × Task	0.02	0.04	0.41
	Age × Modal response	-0.01	0.04	-0.29
	Task × Modal response	0.04	0.04	0.98
	Age × (log) Cloze probability	0.02	0.03	0.73
	Task × (log) Cloze probability	0.01	0.03	0.25
	Modal response × (log) Cloze probability	-0.06	0.08	-0.70
	Age × Task × Modal response	0.08	0.08	0.97
	Age × Task × (log) Cloze probability	0.03	0.06	0.50
	Age × Modal response × (log) Cloze	0.11	0.06	1.79
	Task × Modal response × (log) Cloze	-0.11	0.06	-1.79
	Age × Task × Modal response × (log) Cloze	-0.06	0.12	-0.45

Note. Significant effects indicated in bold.

Table 8

LMM summaries of analyses of predictability and certainty effects on local eye-movement measures of reading of unexpected content words

Measure	Fixed effect	*b*	*SE*	*t/z*
(log) First fixation duration	**Intercept**	**5.35**	**0.01**	**468.09**
	Verbal proficiency	**-0.05**	**0.01**	**-3.85**
	Line initial	**-0.10**	**0.01**	**-6.58**
	Line final	**-0.06**	**0.01**	**-3.83**
	Sentence final	**0.05**	**0.02**	**2.63**
	Passage final	-0.01	0.03	-0.25
	Age group	**0.06**	**0.02**	**2.66**
	Task	**0.06**	**0.00**	**13.43**
	(log) Cloze probability	**-0.03**	**0.00**	**-5.30**
	(log) Certainty	**-0.01**	**0.00**	**-2.30**
	Age × Task	**0.04**	**0.01**	**4.17**
	Age × (log) Cloze probability	0.01	0.00	1.13
	Age × (log) Certainty	0.00	0.00	0.03
	Task × (log) Cloze probability	0.00	0.00	-1.33
	Task × (log) Certainty	0.00	0.00	1.58
	Age × Task × (log) Cloze probability	0.01	0.01	1.03
	Age × Task × (log) Certainty	0.00	0.01	0.54
(log) Gaze duration	**Intercept**	**5.45**	**0.01**	**413.98**
	Verbal proficiency	**-0.07**	**0.01**	**-5.09**
	Line initial	**0.27**	**0.02**	**12.77**
	Line final	-0.04	0.02	-1.71
	Sentence final	**0.09**	**0.03**	**3.64**
	Passage final	0.03	0.04	0.76
	Age group	0.03	0.02	1.14
	Task	**0.10**	**0.00**	**27.67**
	(log) Cloze probability	**-0.07**	**0.01**	**-9.35**
	(log) Certainty	**-0.02**	**0.01**	**-3.96**
	Age × Task	**0.06**	**0.01**	**7.79**
	Age × (log) Cloze probability	**0.01**	**0.00**	**2.26**
	Age × (log) Certainty	0.00	0.00	0.74
	Task × (log) Cloze probability	**-0.02**	**0.01**	**-2.46**
	Task × (log) Certainty	0.00	0.00	-0.44
	Age × Task × (log) Cloze probability	0.02	0.01	1.11
	Age × Task × (log) Certainty	-0.01	0.01	-0.76
(log) Total duration	**Intercept**	**5.71**	**0.02**	**344.34**
	Verbal proficiency	**-0.10**	**0.02**	**-5.55**
	Line initial	0.00	0.03	-0.12
	Line final	**-0.12**	**0.03**	**-3.91**
	Sentence final	**0.08**	**0.04**	**2.22**
	Passage final	**0.19**	**0.06**	**2.99**
	Age group	0.02	0.03	0.76
	Task	**0.13**	**0.00**	**30.97**
	(log) Cloze probability	**-0.11**	**0.01**	**-11.34**
	(log) Certainty	**-0.03**	**0.01**	**-4.24**
	Age × Task	**0.06**	**0.01**	**6.49**
	Age × (log) Cloze probability	0.01	0.00	1.75
	Age × (log) Certainty	0.00	0.00	-0.50

	Task × (log) Cloze probability	-0.01	0.01	-0.84
	Task × (log) Certainty	0.00	0.00	0.81
	Age × Task × (log) Cloze probability	0.03	0.02	1.34
	Age × Task × (log) Certainty	-0.01	0.01	-0.86
Skipping	**Intercept**	**-1.28**	**0.06**	**-19.78**
	Verbal proficiency	**0.17**	**0.06**	**2.75**
	Line initial	**2.06**	**0.13**	**15.69**
	Line final	**-0.49**	**0.15**	**-3.24**
	Sentence final	**-0.65**	**0.18**	**-3.72**
	Passage final	0.05	0.31	0.16
	Age group	-0.13	0.10	-1.26
	Task	**-0.27**	**0.03**	**-7.90**
	(log) Cloze probability	**0.28**	**0.05**	**5.67**
	(log) Certainty	**0.10**	**0.04**	**2.85**
	Age × Task	**-0.14**	**0.07**	**-2.01**
	Age × (log) Cloze probability	**-0.09**	**0.04**	**-2.69**
	Age × (log) Certainty	-0.02	0.02	-0.91
	Task × (log) Cloze probability	**0.07**	**0.02**	**3.14**
	Task × (log) Certainty	0.01	0.02	0.70
	Age × Task × (log) Cloze probability	0.05	0.05	0.94
	Age × Task × (log) Certainty	-0.03	0.04	-0.72
Regressions out	**Intercept**	**-1.30**	**0.06**	**-22.68**
	Verbal proficiency	-0.01	0.06	-0.18
	Line initial	**-1.80**	**0.15**	**-12.04**
	Line final	-0.14	0.12	-1.18
	Sentence final	**0.78**	**0.14**	**5.50**
	Passage final	**3.25**	**0.26**	**12.64**
	Age group	-0.02	0.10	-0.21
	Task	0.06	0.03	1.77
	(log) Cloze probability	0.00	0.04	0.01
	(log) Certainty	0.01	0.03	0.30
	Age × Task	0.07	0.06	1.18
	Age × (log) Cloze probability	-0.02	0.03	-0.83
	Age × (log) Certainty	0.02	0.02	0.94
	Task × (log) Cloze probability	0.02	0.03	0.61
	Task × (log) Certainty	0.00	0.02	-0.23
	Age × Task × (log) Cloze probability	-0.09	0.05	-1.57
	Age × Task × (log) Certainty	0.00	0.04	-0.02
Regressions in	**Intercept**	**-1.28**	**0.06**	**-22.83**
	Verbal proficiency	0.07	0.06	1.27
	Line initial	**2.05**	**0.10**	**19.77**
	Line final	**-2.20**	**0.14**	**-15.50**
	Sentence final	**-0.98**	**0.14**	**-6.80**
	Passage final	-15.39	77.54	-0.20
	Age group	-0.11	0.10	-1.11
	Task	0.07	0.03	1.91
	(log) Cloze probability	**-0.17**	**0.04**	**-4.56**
	(log) Certainty	-0.02	0.03	-0.73
	Age × Task	-0.01	0.07	-0.11
	Age × (log) Cloze probability	-0.05	0.03	-1.60
	Age × (log) Certainty	**-0.07**	**0.02**	**-3.80**
	Task × (log) Cloze probability	**0.11**	**0.03**	**4.28**

	Task × (log) Certainty	0.03	0.02	1.67
	Age × Task × (log) Cloze probability	**0.11**	**0.05**	**2.11**
	Age × Task × (log) Certainty	-0.04	0.04	-1.04

***Note*.** Significant effects indicated in bold.

Figure 1

Mean first-pass fixation durations on content words as a function of age group and task.

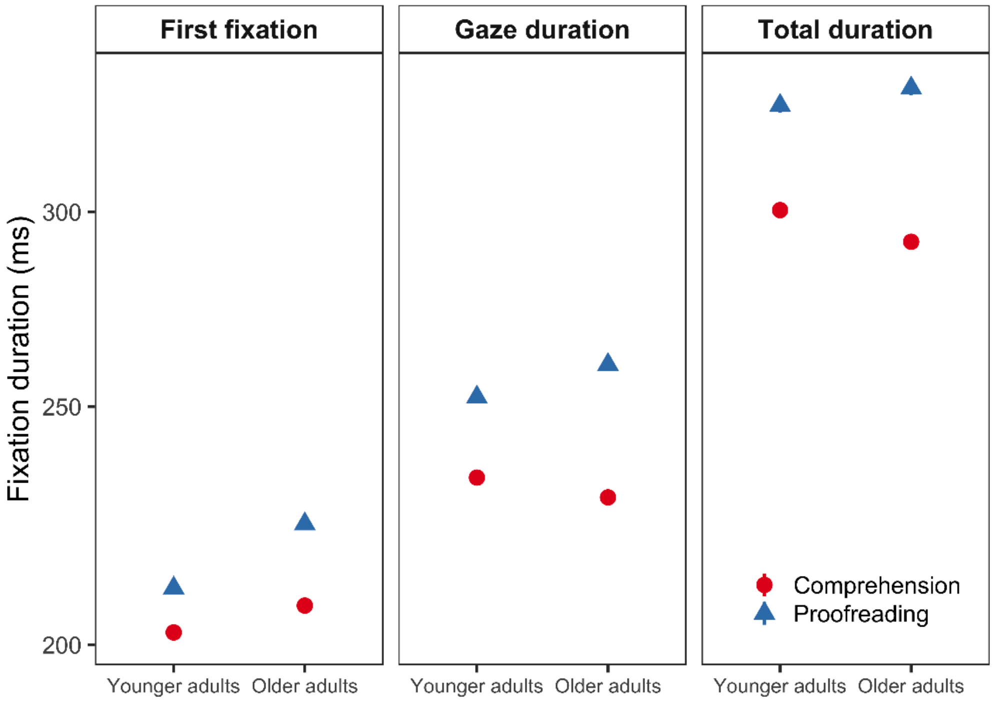

Figure 2a

Relationship between log cloze probability and log gaze duration as a function of task and whether the word was a modal or non-modal response in the cloze task.

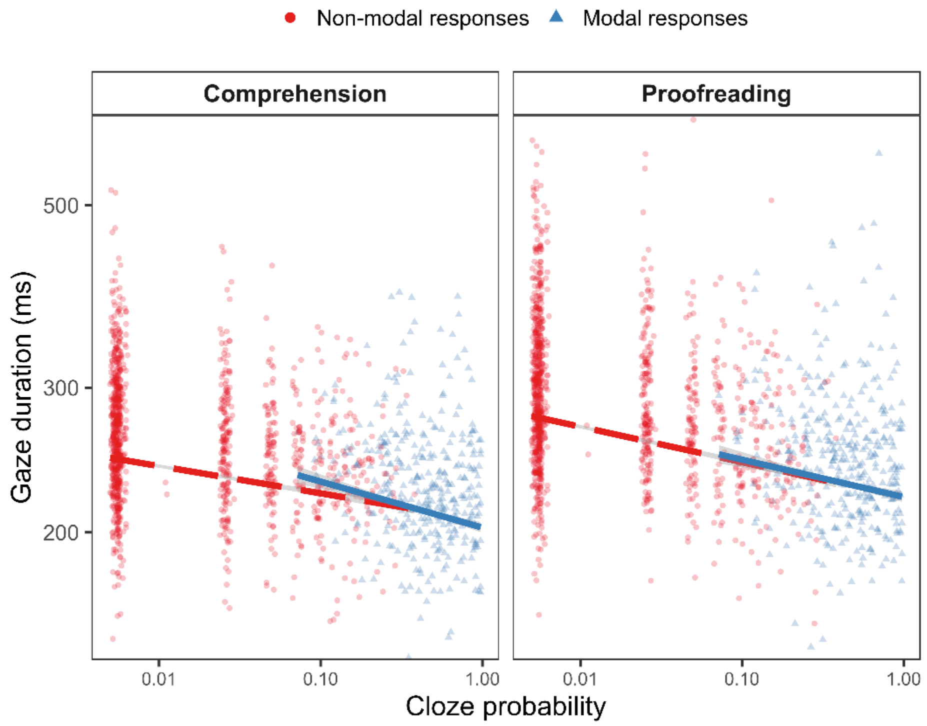

Figure 2b

Relationship between log cloze probability and regressions-out as a function of task and whether the word was a modal or non-modal response in the cloze task.

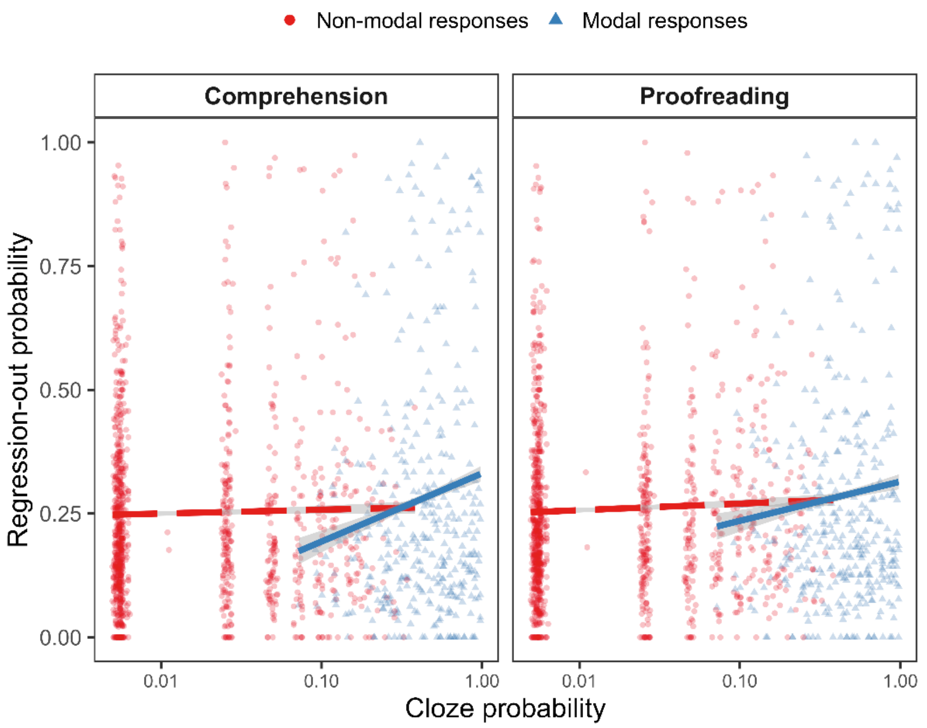

Figure 3a

Relationship between log cloze probability and log gaze duration as a function of age group and whether the word was a modal or non-modal response in the cloze task.

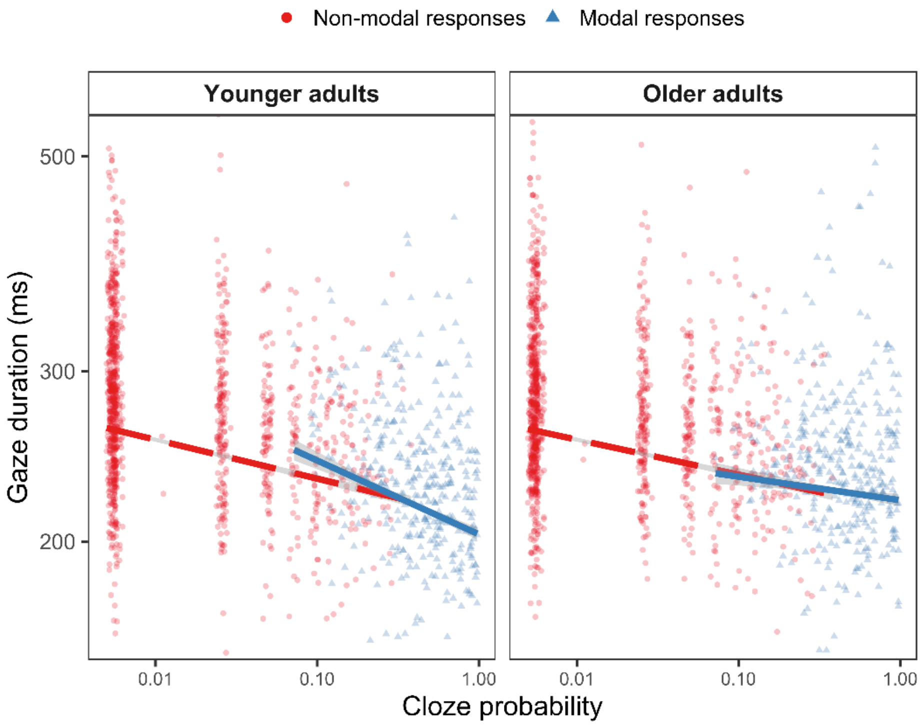

Figure 3b

Relationship between log cloze probability and log total duration as a function of age group and whether the word was a modal or non-modal response in the cloze task.

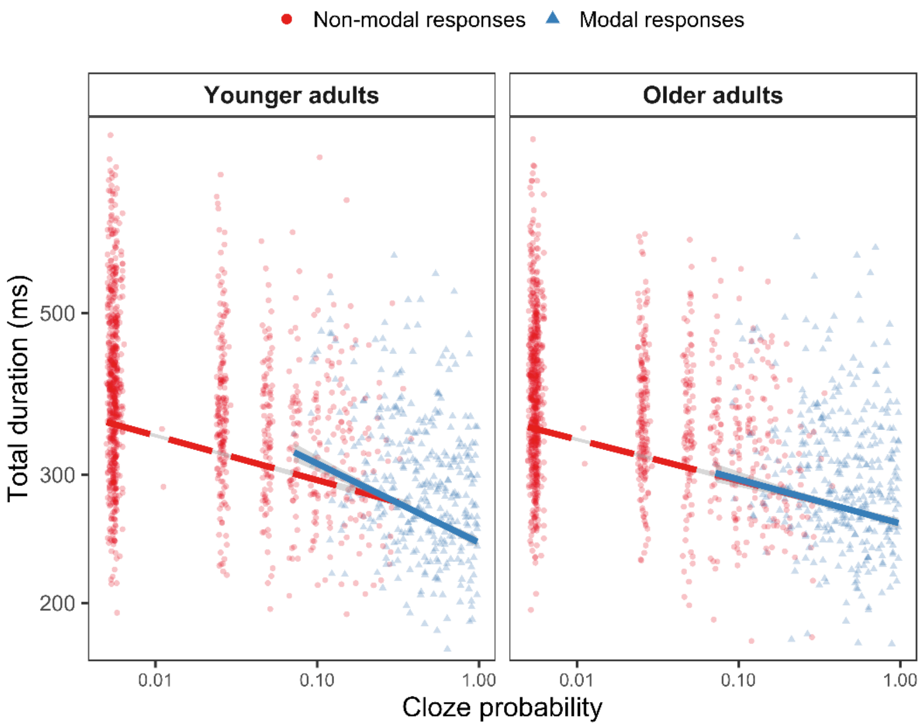

Figure 4

Relationship between log certainty and regressions-in as a function of age group for words that were not the modal response in the cloze task.