2.3 Analysis Method

2.3.1 Behavioral Analyses

To ensure that participants attended to the self-paced reading task, participants answered 38 yes/no comprehension questions via a button press (described above). A participant was retained so long as they correctly answered at least 50 percent of the comprehension questions, which indicated that they were performing above chance. An independent samples t-test was used to determine if there were significant age-group differences in accuracy for the comprehension questions. This same procedure was followed for all four studies. Independent samples t-tests were conducted on the neuropsychological measures to determine if there were age-group differences in the individual language production, working memory, processing speed, and reading experience measures (see Table 3). These tests were used to characterize the sample. For subsequent analyses, composite scores were calculated for each of the four domains (language production, working memory, processing speed, and reading experience) to limit the number of predictors used in the statistical models to avoid overfitting the data.

2.3.2 Composite Score Calculation

The composite scores were calculated as the average z-scores of the individual assessments within each domain. Therefore, the language production composite score was calculated as the average z-scores of the total categorical verbal fluency score, the total phonemic verbal fluency score, picture naming accuracy, picture naming reaction time, the mean length of utterance in words for the typed elicitation task, as well as the percent of adjectives and percent of conjunctions used by participants in their typed elicitation response. The working memory composite score was calculated as the average z-scores of the forward digit span score, backward digit span score, and the partial credit load score of the reading span task. The processing speed composite score was calculated as the average z-scores of the simple speed task reaction time and the choice speed task reaction time.[1] Finally, the reading experience composite score was calculated as the average z-scores of the author recognition task total and the WAIS-III Vocabulary Subtest total.

2.3.3 Treatment of Missing Data and Outliers for Regression Analyses

Treatment of missing data and outliers was the same across all four studies and proceeded as follows. For participants missing any values used in the composite score calculation (e.g., if a participant did not complete the picture naming task) the mean of the z-scores across the other measures for that participant was used as the value for the missing metric. The composite score was then calculated following the procedure described above. This was done to preserve more datapoints for the statistical analyses, as opposed to excluding participants from any statistical analyses in which there was a missing metric that comprised a composite score.

Participants were excluded from analyses in which any scores that comprised their composite score for a given metric were outliers (e.g., if a participant's simple reaction time score was an outlier, that participant was excluded from analyses that required the processing speed composite score). Outliers were defined as values that were greater than three standard deviations from the mean (across participants) because 99 percent of normally distributed data is within 3 standard deviations of the mean.

2.3.4 Language Production Analysis Procedure

To test the hypothesis that younger adults were better at language production, an independent samples t-test was conducted on the language production composite score across age-groups.

[1] The processing speed measures were both on the same scale (milliseconds); therefore, a composite score could have also been calculated by simply taking the average reaction time. However, to be consistent across domains, the average z-scores were used here as well.

2.3.5 Self-Paced Reading Analysis Procedure

To assess the hypotheses that (1) reading times for the critical region would be faster for the predictable sentences compared to the non-predictable sentences, (2) that younger adults would have faster reading times than older adults, and (3) that better language production performance would be associated with better language prediction performance across age groups, I employed a mixed-effects regression analysis using the *lmer* function in the *lme4* package in R (version 4.1.1). Additionally, because the *lme4* package does not include p-values in model summaries, the *lmerTest* package was enabled, which uses the Satterthwaite's Method for p-value approximation (Luke, 2017). Using a multi-level approach considers individual data points while controlling for random variation across participants and items simultaneously. This is advantageous in analyzing psycholinguistic data because psycholinguistic studies typically use repeated measures designs. I followed the procedure outlined in Barr, Levy, Scheepers, and Tily (2013) and started with the full random effects structure for each model. If this model failed to converge, I systematically simplified the model according to Barr et al. (2013) until convergence was achieved. All reported results control for the other predictor variables included in the models.

Before running the models, normality of the dependent variable (i.e., reading times at the critical region) was assessed because reaction time variables are often skewed, and an assumption of regression models is that the residuals of the model are normally distributed. If the model residuals were skewed, the reading times variable was log-transformed using the natural log to normalize it. Additionally, all categorical variables (i.e., age-group and the other conditions) were converted to factor variables. Lastly, the binarized factor variables were re-coded to remove zeros to facilitate model convergence.

2.3.6 Moderation Analyses

To test the hypotheses that working memory, processing speed, and reading experience influence the relationship between language prediction and language production, three separate moderation analyses were run using linear regressions to examine the interaction between the language production composite score and either the working memory, processing speed, or reading experience composite scores. Importantly, these analyses were collapsed across age-group for two reasons. First, prior studies have found evidence suggesting that the influence working memory, processing speed, and reading experience have on the relationship between language prediction and production occurs irrespective of age. Second, collapsing across age increases statistical power which increases the likelihood of detecting individual differences effects.

Prior to conducting the analyses, the prediction score for each participant was calculated. The prediction score was operationalized as the average non-predictable reading time in the critical region minus the average predictable reading time in the critical region. Therefore, larger positive prediction score values indicate a larger difference between average non-predictable and predictable reading times for a participant, with predictable sentences being read faster than non-predictable sentences. Negative scores indicate that on average, the non-predictable sentences were read faster than the predictable sentences. The missing data and outlier procedure described above was followed for all moderation analysis models. The models listed below test for moderation effects of language production and working memory, processing speed, and reading experience, respectively.[2]

Working Memory Moderation Analysis:

[2] The moderation analysis models were consistent across all four studies. Therefore, they will only be written out here, for study one.

*Prediction Score = $\beta 0$ + $\beta 1$(language production composite score) + $\beta 2$(working memory composite score) +$\beta 3$(language production composite score*working memory composite score) + e*

Processing Speed Moderation Analysis:

*Prediction Score = $\beta 0$ + $\beta 1$(language production composite score) + $\beta 2$(processing speed composite score) +$\beta 3$(language production composite score*processing speed composite score) + e*

Reading Experience Moderation Analysis:

*Prediction Score = $\beta 0$ + $\beta 1$(language production composite score) + $\beta 2$(reading experience composite score) +$\beta 3$(language production composite score*reading experience composite score) + e*

2.4 Results

2.4.1 Reading Comprehension

During the self-paced reading task, participants were tasked with answering yes/no comprehension questions after 20% of trials, resulting in 38 total questions, to ensure they were attending to the task. An independent samples t-test indicated that the older adult group (M = 35.12, SD = 2.07) was significantly more accurate in answering these questions compared to the younger adult group (M = 33.90, SD = 2.08), $t(77.81) = -2.64$, $p = .01$, 95% CI [-2.15, -0.30].

2.4.2 Language Production Performance

To test the hypothesis that the younger adult group was better at language production than the older adult group, an independent samples t-test was performed on the language production composite scores. Four participants (one younger adult and three older adults) were excluded for having outliers on measures comprising their language production composite score. This resulted in 76 participants (38 younger adults and 38 older adults) being included in the analysis. The results indicated there was no significant difference in language production composite scores for younger (*M* = 0.008) and older (*M* = -0.016) adults, $t(72.77) = 0.30$, $p = .76$, 95% CI [-0.14, 0.19].

2.4.3 Self-Paced Reading Performance

To test the hypotheses that reading times for the critical region would be faster for the predictable sentences compared to the non-predictable sentences, that younger adults would have faster reading times than older adults, and that better language production performance would be associated with better language prediction performance across age-groups, I employed a mixed-effects regression analysis. The same participants mentioned in the section above were also excluded from this analysis. The final model included the random intercept of item and random intercept of participant to account for random variation across items and participants. Please refer to Appendix F for all statistical models used across studies.

Results indicated that the main effect of sentence type—predictable (*M* = 749.86 ms, *SD* = 299.48 ms) versus non-predictable (*M* = 753.97 ms, *SD* = 305.89 ms)—on reading times at the critical region while controlling for all other variables in the model was not significant, $\beta = -0.0015$, $t(7823) = 0.18$, $p = .86$. However, there was a significant main effect of age-group, with younger adults having significantly faster reading times than older adults at the critical region, $\beta = 0.47$, $t(73.72) = 4.99$, $p < .001$, see Figure 4 and Table 5. Additionally, the main effect of language production score on reading times at the critical region was not significant, $\beta = -0.0068$, $t(73.00) = -0.06$, $p = .95$. Lastly, the interaction between sentence type and age-group at the critical region was not significant, $\beta = -0.0086$, $t(7846) = -0.75$, $p = .45$.

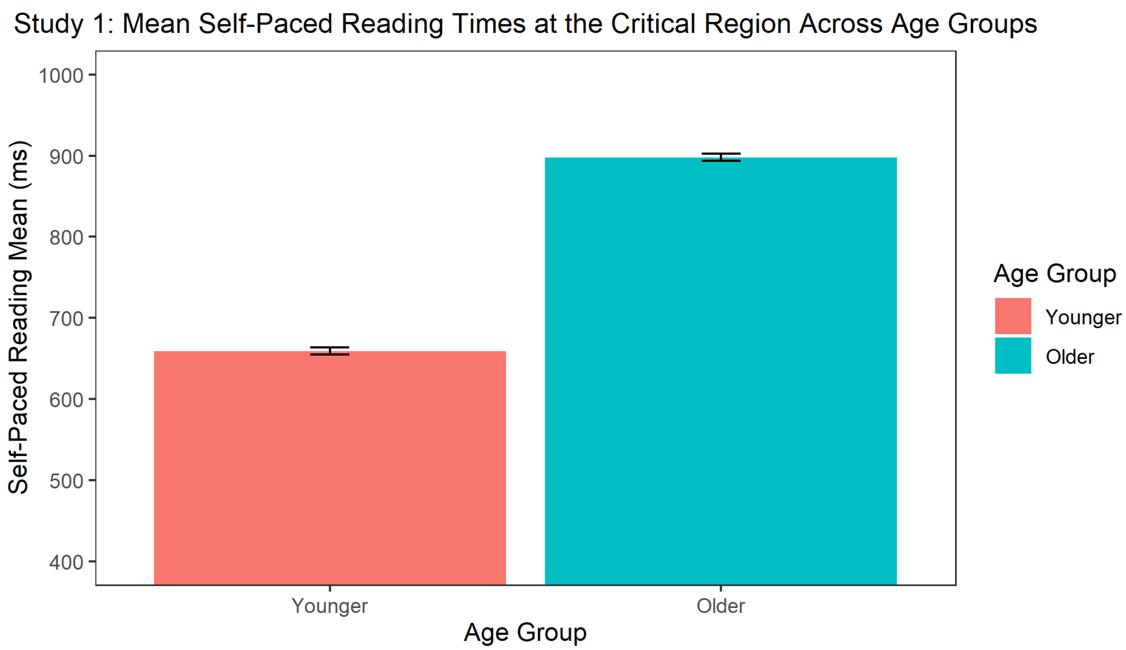

Figure 4. Mean Self-Paced Reading Times Across Age Groups in Study One. Means and standard error bars were calculated based on participant-level data and were collapsed across sentence type condition.

Table 5. Study One Mean Reading Times Across Age Groups Per Condition

	Younger	**Older**
	Mean (SD)	**Mean (SD)**
Predictable Reading Times (ms)	620.13 (265.21)	879.59 (277.28)
Non-Predictable Reading Times (ms)	622.06 (270.68)	885.88 (284.01)

Note. There were no significant differences across conditions. Means across conditions are included in the table to highlight the similarities in reading times.

2.4.4 Moderation Analyses

To test whether working memory, processing speed, and reading experience each influence the relationship between language prediction and production, three separate moderation analyses were run to examine the interaction between the language production score and working memory, processing speed, and reading experience, respectively, on participants' prediction scores.

For the working memory moderation analysis, the same four participants mentioned in the language production results were excluded as outliers. These same four participants were also excluded from the reading experience moderation analysis. Results indicated that the interaction between language production composite score and working memory composite score on prediction score was not significant, $\beta = 14.40$, $t(72) = 0.84$, $p = .41$.

For the processing speed moderation analysis, seven participants were excluded from the model—the four mentioned above as well as one younger adult and one older adult for being outliers on the simple reaction time task and one older adult for being an outlier on the choice reaction time task. Results indicated that the interaction between language production composite score and processing speed composite score on prediction score was not significant, $\beta = 3.45$, $t(69) = 0.15$, $p = .88$.

For the reading experience moderation analysis, results indicated that the interaction between language production composite score and reading experience composite score on prediction score was not significant, $\beta = -13.26$, $t(72) = -0.86$, $p = .40$.