Supplementary Material

ORTHO-SEMANTIC LEARNING OF NOVEL WORDS: AN EVENT-RELATED POTENTIAL STUDY OF GRADE 3 CHILDREN

Alena Galilee, Lisa J. Beck, Clara J. Lownie, Jennika Veinot, Catherine Mimeau, Tammy Dempster, Laura Elliott, S. Hélène Deacon, Aaron J. Newman

1 SUPPLEMENTARY METHODS

1.1 Reading Skill Assessments

Three standardized tasks were used to assess reading skills. The Sight Word Efficiency and Phonemic Decoding sub-tests from the Test of Word Reading Efficiency (TOWRE-2; Torgesen et al., 1999) were administered to assess reading fluency (speed and accuracy) and phonemic decoding skills. In these tasks, participants read aloud as many words or pseudowords as they could during a 45-second time interval from a list that gradually increased in length and difficulty.

The Word Identification subtest from the Woodcock Reading Mastery Test-Revised (WRMT-R; Woodcock, 1998) was used to assess word reading accuracy. This subtest required children to read aloud words of increasing difficulty.

The Passage Comprehension subtest, also from the WRMT-R, was used as a measure of comprehension of written text. During this task, participants had to silently read sentences or short passages of increasing length and complexity, all of which contained a missing word. Participants were asked to provide an appropriate word to fill the missing blank.

1.1.1 Phonological Skills

The Elision subset of the Comprehensive Test of Phonological Processing (C-TOPP-2; Wagner et al., 2013), a measure of phonological awareness, assessed participants' ability to remove phonological segments from spoken words to form different words. The participants were verbally instructed to repeat the original word, and then to remove phonemes from that word and generate the new word aloud.

1.2 Orthographic and Semantic Knowledge

These non-standardized tasks were the same as those used by Mimeau et al. 2018 and adapted from Olson et al. (Olson et al., 1985). Homophone judgment tasks were used to characterize the participants' existing orthographic and semantic knowledge. First, participants chose the correct of two alternative spellings for a known word (e.g., *The boy liked to read stories with mysterey*/textitmystery). Participants then selected the more appropriate of two real words given the context of the sentence (e.g., *The father read a story to his youngest son*/sun.). The participants also completed a modified version of the Peabody Picture Vocabulary Test (M-PPVT-3 Dunn and Dunn, 2007; Wang et al., 2009; Pasquarella et al., 2011). Participants were shown four black and white images at a time and were asked to point to the image that represented a spoken word provided verbally by the experimenter.

1.2.1 Nonverbal Intelligence

The Matrix Reasoning subtest of the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999, WASI;) was used as a control measure of children's non-verbal intelligence. In this task, participants were shown pictures containing incomplete patterns of increasing difficulty. Participants were asked to select the correct piece out of five choices that finished each pattern.

1.2.2 Working Memory

The Digit Span subtest of the Wechsler , (WISC-4; Digit Span subtest) was used to assess working memory . This subtest required participants to repeat lists of digits, increasing in length, in either the same order as dictated, or in the reverse order.

1.3 Ortho-Semantic Learning Task

Each novel word taught in the study was monosyllabic with four letters and contained consonant sounds in the word-initial and word-final positions. Each novel word indicated a target phoneme that was associated with more than one appropriate spelling in English (e.g., target = */i/*, spelling pattern = *ea* or *ee*, see Mimeau et al., 2018, for the full list of stimuli). The novel words were searched in the Children's Printed Word Database (Masterson et al., 2010) to ensure that they did not previously exist in the English language and that children would not have previously had exposure to them. It was ensured that the spelling of each novel word followed a regular grapheme-phoneme correspondence based on rules provided in Rastle and Coltheart (Rastle and Coltheart, 1999) and as discussed by (Ricketts et al., 2011). The words were embedded in short stories as described in the main article. An example story is shown in Figure S1.

Children were given the following instructions verbally by an experimenter (square brackets indicated actions performed by experimenter): *Now, I am going to talk to you about someone called Professor Parsnip.* [show the picture of Professor Parsnip] *Professor Parsnip likes to invent new things, and he has an invention factory. He named each and every one of his inventions with words you have never heard or seen before. I want you to read about 12 of his latest inventions. I will use this recorder* [show the recorder] *because Professor Parsnip likes to listen to these stories. So please read the stories out loud, and as clearly and carefully as you can. After we learn about Professor Parsnip's neat inventions, you will do some activities that involve the inventions and their names. I will ask you to spell the names of the inventions and to tell me what they are used for. So when you read the stories, try and remember how the names of the inventions are spelled and what they are used for.*

In the *orthographic choice task*, children were shown four written non-words on a page (see Figure S1 for an example) and given the following instructions: *Now, I am going to show you different words. I want you to look carefully at each word, one at a time, and to show me the spelling of the inventions you read about. You need to show me the spellings that are the same as what you saw in the stories. I will read out loud the name of each invention. There is only one right answer. Show me the spelling of* _____.; the blank at the end was replaced with the target word. The three distractors consisted of the homophonic variant of the target (e.g., *klet*, if the target was *clet*), and two non-words that were identical to the target except for either the first or last consonant (e.g., *kleb* and *cleb*).

In the *semantic choice* task, for each novel word children were shown four pictures on a page (see Figure S1 for an example) and given the following instructions: *Now, I am going to show you different pictures. I want you to look carefully at each picture, one at a time, and to show me the picture of the inventions you read about. You need to show me the pictures that represent the inventions that were in the stories. I will read out loud the name of each invention. There is only one right answer. Show me the picture* *of a* _____.; the blank at the end was replaced with the target word. One of the four pictures depicted the function of the novel word as described in the preceding story; one distractor shared some semantic information with the target, depicting an invention involving the same object (e.g., when the target was a juice remover from oranges, the related distractor was an orange peeler); and the additional two distractors were foils that corresponded to the same unrelated object (e.g., a tummy-ache fixer and a tummy remover for shirts). The positions on the page of the four items for both tasks was randomized.

1.4 Lexical Decision Task

1.4.1 Stimuli

1.4.1.1 False Fonts

Fifty false font strings were created based on a protocol developed and implemented by (Grossi and Coch, 2005). Using Adobe Photoshop Software, the letters from real word stimuli were individually rotated by either 90 or 180 degrees. Additionally, letters that had either vertical or horizontal symmetry (e.g., *u*), were vertically altered (i.e., expanded or diminished) on one side. Overall spatial characteristics of the original letters (height, width, and inter-letter distance) were maintained for all false font stimuli.

1.4.1.2 Consonant Strings

Fifty consonant strings were generated using a random consonant generator (http://www.dave-reed.com/Nifty/randSeq.html) with strings ranging from 3 to 5 (mean = 3.99) — consistent with the qualities of the real word stimuli.

1.4.1.3 Non-words

The non-words were generated using the same list of 50 real words from the Children's Printed Word Database by replacing either one or two phonemes in the word, such that the modified stimuli were not real words, but were still pronounceable. These non-word stimuli were more reflective of typical pseudo-words, as they followed typical orthographic rules.

1.4.1.4 Novel Words

The 24 novel words used in the ortho-semantic learning task were also included in the lexical decision task. To create a comparable number of trials to the other conditions, each novel word was repeated twice in the lexical decision task, both times in the block of trials immediately following when that word was presented in the ortho-semantic learning task.

1.4.1.5 Real Words

A list of 100 single-syllable words was generated using the Children's Printed Word Database (Masterson et al., 2010). All real words were nouns which were controlled for word frequency (frequency per million; range = 103 to 1,880, average = 323.97), word length (range = 3 to 5, average = 3.99), number of phonological neighbors (range = 0-29, average = 12.69) and orthographic neighbors (range = 0 to 17, average = 6.49).

1.5 Procedure

The instructions given to children for the lexical decision task were as follows: *Welcome to our EEG study! In this study, you will see words on the screen. Some are real words, and some are made-up ones. Press YES if you think the word is a real word (This includes the new words you just learned!) Press NO if* *you think it's a made-up word, or if it doesn't look like a word at all. Please ask if you have any questions. When you're ready to start and the helpers say it's OK, press the SPACEBAR to see the first word!*

1.6 ERP Preprocessing

First, a 1–30 Hz bandpass zero-phase finite impulse response filter with a hamming window was applied to the continuous EEG data, with the filter length and transition bandwidths selected automatically. The data were then segmented into epochs time-locked to the onset of each lexical decision task stimulus, from 250 ms prior to stimulus onset to 1250 ms post-stimulus onset. The epochs were then submitted to the AutoReject algorithm (v0.4.0; Jas et al., 2017) to identify and mark any bad channels or segments of data for exclusion from the subsequent preprocessing step, independent components analysis (ICA). ICA was then applied to the epochs, using the *fastica* algorithm (Hyvarinen, 1999) with the number of components set to explain 95% of the variance. The EOG channels were re-referenced as bipolar channels (horizontal: E128-E125; vertical left: E21-E127; vertical right: E14-E126), and the time course of each independent component was correlated with each bipolar EOG channel to identify components reflecting blinks or eye movements. The correlation threshold was initially set at $z > 3.0$ and any independent component that exceeded this threshold was marked for exclusion; if less than three independent components met this criterion, then the z threshold was lowered by 0.1 and the correlation performed again. This stepwise threshold-lowering procedure was repeated until at least three components were marked for exclusion. Due to the inherent noisiness of EEG data collected from children, as well as the non-stationarity of the EEG recordings due to drying of the electrode sponges as well as other factors, each ICA decomposition was visually inspected by the last author and additional components were excluded if they appeared to reflect variance at only a single electrode, or on a very small number of trials. The complete ICA decomposition for each participant, as well as the components manually marked for exclusion, are included in the online open repository for this study.

Once ICA components were marked for exclusion, the raw, continuous EEG data were filtered using a 0.1–30 Hz bandpass and segmented into epochs from -250–1250 ms relative to the onset of each target stimulus. The ICA decomposition obtained by the prior steps was applied to the data to remove variance related to ocular and other artifacts. Then, the AutoReject algorithm was applied, this time to correct bad channels (via interpolation) or segments where possible, and excluding any segments that exceeded the criteria for correction (Jas et al., 2017). On average, 9.9% of trials were rejected at this stage (range: 0–31%). The epochs were then re-referenced to the average of all electrodes for visualization and analysis of the N170 component. For visualization and analysis of the N400, the data were re-referenced to the average of the clusters of electrodes over the mastoid processes bilaterally (electrodes 56, 57, 50, 63, 101, 100, 107, and 99). This was done because mastoid referencing is the most common way in which this component is analyzed in the literature (Šoškić et al., 2022).

Finally, the timing of the event code (i.e., 0 ms in each epoch) was shifted forward in time by 40 ms. This was done to account for the average delay between receipt of the event code by the EEG recording computer, and actual appearance of the stimulus item on the screen, as measured using a photocell connected to the EEG amplifier (mean delay = 39.94 ms, SD = 4.41 ms). The preprocessed epochs were then saved for later analysis.

2 SUPPLEMENTARY RESULTS

2.1 Standardized Tests

Table S1 shows demographic information and scores on all reading skills tests. Figure S2 shows the distribution of scores on these tests for each individual child who participated.

Table S1. Summary of demographic information and standardized test scores

Measure	Count	Mean	Std. Dev.	Min.	Max.
Age	34	8.7	0.5	7.5	9.4
SWE_scaled	33	104.7	12.4	79.0	128.0
SWE_percentile	33	60.5	26.2	8.0	97.0
SWE_age_equiv	33	9.8	1.8	7.5	14.0
PDE_scaled	33	104.4	13.5	74.0	126.0
PDE_percentile	33	60.0	27.5	4.0	96.0
PDE_age_equiv	33	10.0	3.2	1.5	17.5
TOWRE_scaled_sum	33	209.1	24.7	156.0	252.0
TOWRE_idx	33	104.9	13.0	77.0	127.0
WordID_percentile	33	74.1	26.5	4.0	99.9
WordID_age_equiv	33	10.2	1.7	7.8	16.6
WordID_std	33	112.6	14.6	74.0	149.0
PsgComp_percentile	34	71.9	19.5	33.0	98.0
PsgComp_age_equiv	34	10.1	1.6	7.6	13.8
PsgComp_std	34	110.7	9.9	93.0	131.0
PPVT_raw	33	34.0	4.8	21.0	42.0
CTOPP_scaled	33	10.0	4.1	1.0	18.0
CTOPP_percentile	33	50.8	35.7	1.0	100.0
CTOPP_age_equiv	33	10.1	3.4	5.5	15.0
WASI_scaled	32	11.8	3.1	6.0	19.0
WASI_age_equiv	28	10.1	2.5	5.8	16.5
WISC_sum	33	14.7	3.0	11.0	26.0
WISC_scaled	33	10.8	2.7	7.0	19.0

2.2 Ortho-Semantic Tasks

Descriptive statistics for each subtest are shown in Table S2.

Table S2. Descriptive statistics for the ortho-semantic tasks, expressed as percentage of correct trials.

Test	Mean (%)	Std. Dev.	Min.	Max.
Orthographic Choice	65.5	10.0	47.8	83.3
Orthographic Knowledge	82.1	13.4	48.0	100
Semantic Choice	76.8	14.6	33.3	100
Semantic Knowledge	90.0	11.8	56.0	100

Table S3. Descriptive statistics for accuracy and reaction time (RT, in seconds) on the lexical decision task.

Condition	Accuracy Mean (%)	Std. Dev.	Min.	Max.	RT Mean (s)	Std Dev	Min	Max
ConsonantString	93.9	10.2	52.2	100	1.20	0.44	0.56	2.23
FalseFont	97.2	6.6	72.7	100	1.09	0.38	0.59	2.07
NovelWord	63.8	33.6	0.0	100	1.38	0.45	0.59	2.43
PseudoWord	65.8	25.9	17.4	100	1.42	0.51	0.56	2.54
RealWord	93.3	9.4	51.3	98.8	1.19	0.38	0.64	2.23

2.3 Lexical Decision Task

2.3.1 Accuracy

Table S3 shows descriptive statistics for accuracy and reaction time (RT) in the lexical decision task (LDT) performed during EEG data collection. Table S4 shows the statistical results of contrasting accuracy between each pair of conditions.

Table S4. Between-condition comparisons of accuracy on the lexical decision task performed during EEG data collection, from the linear mixed effects analysis. All p values are corrected for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR) method

Contrast	Est. Diff.	SE	z	p (corr.)
FalseFont - ConsonantString	1.02	0.755	1.350	$.736$
FalseFont - PseudoWord	4.43	0.677	6.540	$< .001$
FalseFont - NovelWord	4.58	0.851	5.390	$< .001$
FalseFont - RealWord	1.89	0.606	3.120	$.022$
ConsonantString - PseudoWord	3.41	0.421	8.100	$< .001$
ConsonantString - NovelWord	3.56	0.760	4.690	$< .001$
ConsonantString - RealWord	0.88	0.520	1.690	$.551$
PseudoWord - NovelWord	0.16	0.689	0.220	$.999$
PseudoWord - RealWord	-2.53	0.416	-6.090	$< .001$
NovelWord - RealWord	-2.69	0.479	-5.610	$< .001$

2.3.2 Sensitivity and Response Bias

Figure S3 plots the signal detection analysis results for sensitivity (A'), and response bias (B''), for each child in the study.

2.3.3 Reaction Time

Table S5 shows the statistical results of contrasting reaction time (RT) between each pair of conditions, using all trials (both correct and incorrect responses).

Table S5. Results of contrasts from linear mixed effects analysis of reaction time, using all trials (correct and incorrect). All p values are corrected for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR) method

Contrast	Est. Diff.	SE	z	p (corr.)
FalseFont - ConsonantString	0.196	0.043	4.520	$< .001$
FalseFont - PseudoWord	0.459	0.062	7.410	$< .001$
FalseFont - NovelWord	0.472	0.069	6.830	$< .001$
FalseFont - RealWord	0.255	0.054	4.690	$< .001$
ConsonantString - PseudoWord	0.264	0.053	4.970	$< .001$
ConsonantString - NovelWord	0.276	0.061	4.540	$< .001$
ConsonantString - RealWord	0.059	0.047	1.270	.786
PseudoWord - NovelWord	0.013	0.032	0.410	.994
PseudoWord - RealWord	-0.204	0.045	-4.520	$< .001$
NovelWord - RealWord	-0.217	0.044	-4.970	$< .001$

Table S6 shows the statistical results of contrasting reaction time (RT) between each pair of conditions, using only correct trials.

Table S6. Results of contrasts from linear mixed effects analysis of reaction time, using only correct trials. All p values are corrected for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR) method

Contrast	Est. Diff.	SE	z	p (corr.)
FalseFont - ConsonantString	0.169	0.051	3.330	.011
FalseFont - PseudoWord	0.499	0.069	7.260	$< .001$
FalseFont - NovelWord	0.413	0.082	5.040	$< .001$
FalseFont - RealWord	0.268	0.063	4.260	$< .001$
ConsonantString - PseudoWord	0.330	0.054	6.070	$< .001$
ConsonantString - NovelWord	0.244	0.074	3.320	.011
ConsonantString - RealWord	0.099	0.067	1.480	.573
PseudoWord - NovelWord	-0.086	0.051	-1.690	.492
PseudoWord - RealWord	-0.231	0.068	-3.410	.011
NovelWord - RealWord	-0.145	0.071	-2.050	.304

2.4 Event-Related Potentials

Grand-averaged ERP waveforms time-locked to the onset of printed words in each experimental condition are shown in Figure S4, along with scalp topographic maps are time points corresponding to periods of peak variance in the signal.

2.4.1 N400

2.4.1.1 Late N400 Time Window (400–500 ms)

In the later N400 window, false fonts elicited a larger negativity than any other condition. However, from the waveform plot in Figure **??** the differences between other conditions are somewhat difficult to assess visually, as their relative amplitudes varied over the time window. As in the preceding time window, the best fitting LME model included fixed effects of condition and baseline (but no interaction between them), random intercepts for each participant, and random channel-by-participant and condition-by-participant slopes. The model-estimated means for each condition are shown in the main article, and the results of planned between-condition contrasts are shown in Table S7.

Table S7. Between-conditions contrasts for each condition from the linear mixed effects analysis of the N400 component from 400-500 ms. All p values are corrected for multiple comparisons using Tukey's method.

Contrast	Estimate (μV)	SE	t	p
Print Tuning	-1.66	0.42	-3.97	$< .001$
Lexical Tuning	0.47	0.41	1.13	.257
Real vs. Pseudoword	0.40	0.41	0.97	.331
Real vs. Novel	-0.74	0.41	-1.79	.074
Pseudoword vs. Novel	1.14	0.42	2.74	.006

False fonts elicited a significantly larger N400 than consonant strings. Additionally, as in the preceding time window, the N400 was significantly larger for pseudowords than novel words. Looking closely at the waveforms (see main article), this difference appears to have occurred primarily in the early part of this 400–500 ms time window. None of the planned regressions relating behavioral measures to the N400 difference between pseudowords and novel words were significant in this later time window.

2.4.1.2 Full a priori N400 Time Window (300–500 ms)

In the main article, we analyzed the N400 in two separate time windows, based on the observation of different between-condition differences from 300–400 versus 400–500 ms. Below are the results of the linear mixed effects (LME) analysis of mean amplitude across the a priori planned 300–500 ms time window. Figure S5 shows the model-estimated mean amplitudes for each condition, and Table S8 shows the statistical results for each planned contrast. Figure S6 plots the significant interaction between N400 amplitude and the Passage Comprehgension subtest of the WRMT-R.

Table S8. Between-conditions contrasts for each condition from the linear mixed effects analysis of the N400 component from 400-500 ms. All p values are corrected for multiple comparisons using Tukey's method.

Contrast	Estimate (μV)	SE	t	p
Print Tuning	-1	.38	-2.61	.009
Lexical Tuning	.18	.38	.48	.634
Real vs. Pseudo	.7	.38	1.85	.065
Real vs. Novel	$-.44$	.38	-1.16	.247
Novel vs. Pseudo	1.14	.38	2.99	.003

REFERENCES

Dunn, D. M. and Dunn, L. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition, manual*. NCS Pearson, Inc. (NCS Pearson, Inc.)

Grossi, G. and Coch, D. (2005). Automatic word form processing in masked priming: An ERP study. *Psychophysiology* 42, 343–355. doi:10.1111/j.1469-8986.2005.00286.x

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks* 10, 626–634. doi:10.1109/72.761722

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., and Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage* 159, 417–429. doi:10.1016/j.neuroimage.2017.06.030

Masterson, J., Stuart, M., Dixon, M., and Lovejoy, S. (2010). Children's printed word database: Continuities and changes over time in children's early reading vocabulary. *British Journal of Psychology* 101, 221–242. doi:10.1348/000712608x371744

Mimeau, C., Ricketts, J., and Deacon, S. H. (2018). EnglishThe Role of Orthographic and Semantic Learning in Word Reading and Reading Comprehension. *Scientific Studies of Reading* 22, 1–17. doi:10.1080/10888438.2018.1464575

Olson, R. K., Kliegl, R., Davidson, B. J., and Foltz, G. (1985). Individual and developmental differences in reading disability. In *Reading research: Advances in theory and practice*, eds. G. E. MacKinnon and T. G. Waller (Academic Press), vol. 4. 1–64

Pasquarella, A., Chen, X., Lam, K., Luo, Y. C., and Ramirez, G. (2011). Cross-language transfer of morphological awareness in Chinese–English bilinguals. *Journal of Research in Reading* 34, 23–42. doi:10.1111/j.1467-9817.2010.01484.x

Rastle, K. and Coltheart, M. (1999). Serial and Strategic Effects in Reading Aloud. *Journal of Experimental Psychology: Human Perception and Performance* 25, 482–503. doi:10.1037/0096-1523.25.2.482

Ricketts, J., Bishop, D. V. M., Pimperton, H., and Nation, K. (2011). The Role of Self-Teaching in Learning Orthographic and Semantic Aspects of New Words. *Scientific Studies of Reading* 15, 47–70. doi:10.1080/10888438.2011.536129

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., et al. (1999). Preventing Reading Failure in Young Children With Phonological Processing Disabilities: Group and Individual Responses to Instruction. *Journal of Educational Psychology* 91, 579–593. doi:10.1037/0022-0663.91.4.579

Wagner, R., Torgesen, J., Rashotte, C., and Pearson, N. (2013). *Comprehensive Test of Phonological Processing–2nd ed. (CTOPP-2)* (Austin, TX: Pro-Ed)

Wang, M., Yang, C., and Cheng, C. (2009). The contributions of phonology, orthography, and morphology in Chinese–English biliteracy acquisition. *Applied Psycholinguistics* 30, 291–314. doi:10.1017/s0142716409090122

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence* (Pro-Ed)

Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests – Revised Normative Update forms G and H examiner's manual* (Pearson Assessment)

Šoškić, A., Jovanović, V., Styles, S. J., Kappenman, E. S., and Ković, V. (2022). How to do Better N400 Studies: Reproducibility, Consistency and Adherence to Research Standards in the Existing Literature. *Neuropsychology Review* 32, 577–600. doi:10.1007/s11065-021-09513-4

2.5 Figures

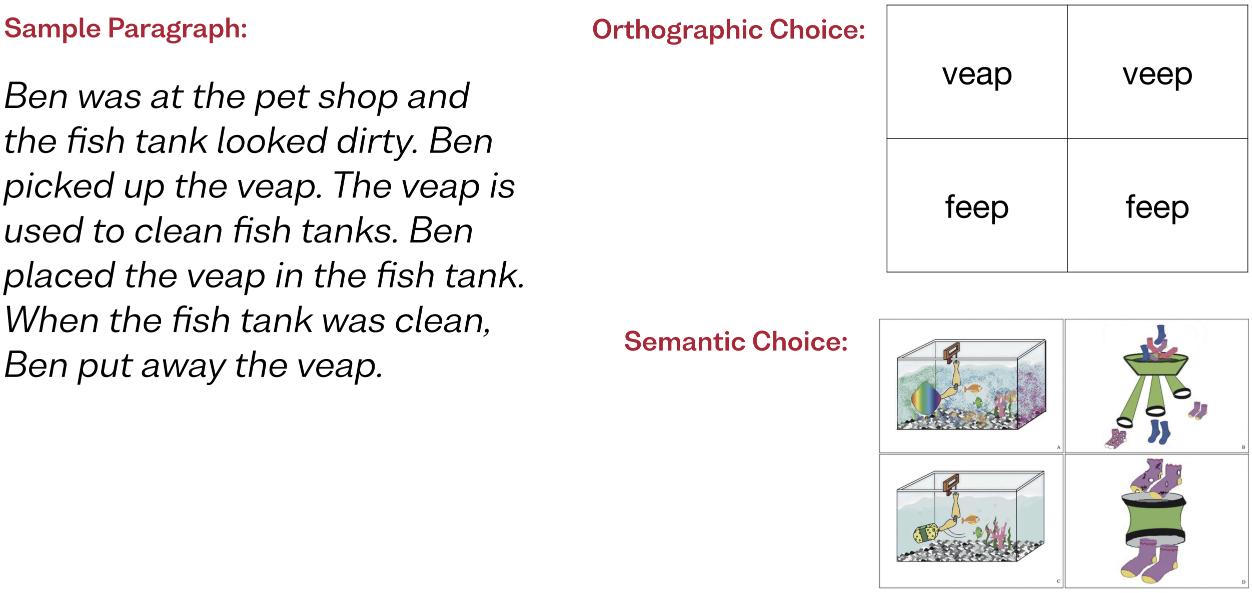

Figure S1. An example paragraph used in the ortho-semantic learning task (left), along with the two tasks used to assess learning of the novel word introduced in that paragraph (right).

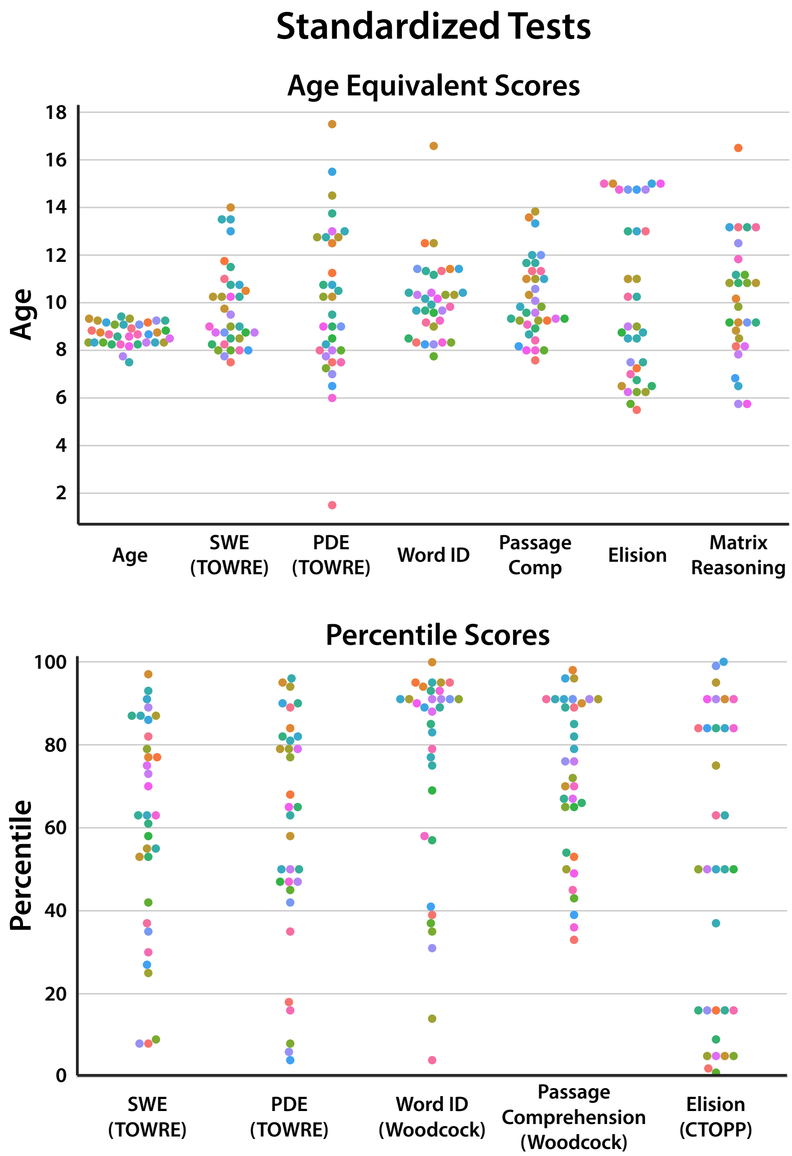

Figure S2. Swarm plots showing the distribution of age-equivalent and percentile scores on the standardized tests administered. Different hues are used for individual participants

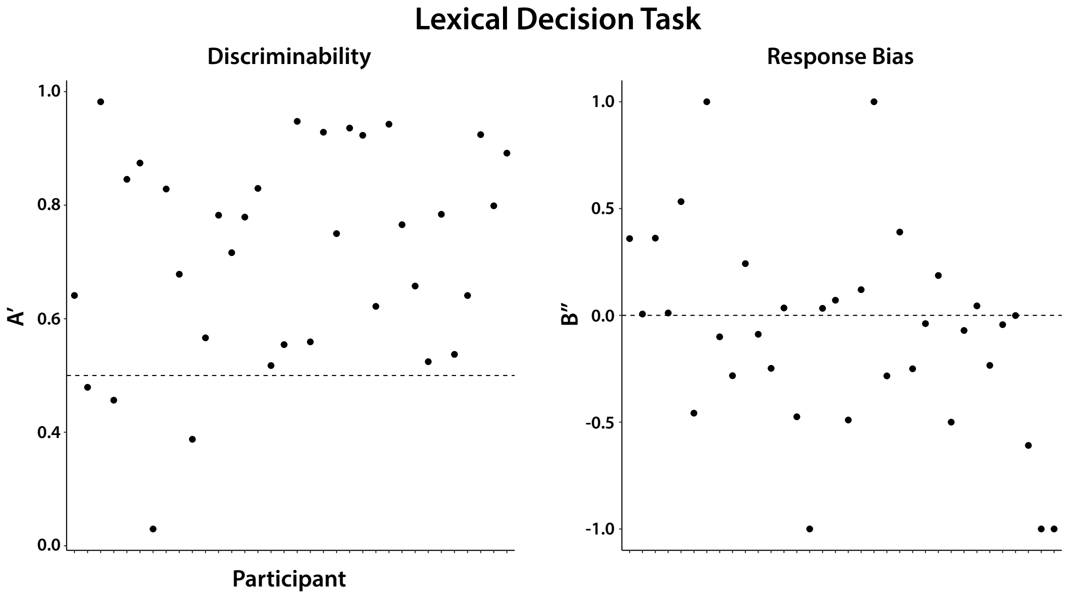

Figure S3. Results of signal detection theory analysis of accuracy in the lexical decision task. A' is a nonparametric measure of discriminability, where 1 represents good discriminability, and a value near 0.5 (indicated with a dashed line) indicates chance performance. B'' is a nonparametric measure of response bias, where positive values reflect a conservative bias (a tendency to say ”no”, i.e., classify items as non-words), and negative values reflect a liberal bias (a tendency to say ”yes”, i.e. classify items as words); zero (indicated by a dashed line) reflects unbiased performance. Participants are ordered by ID code along the x axis

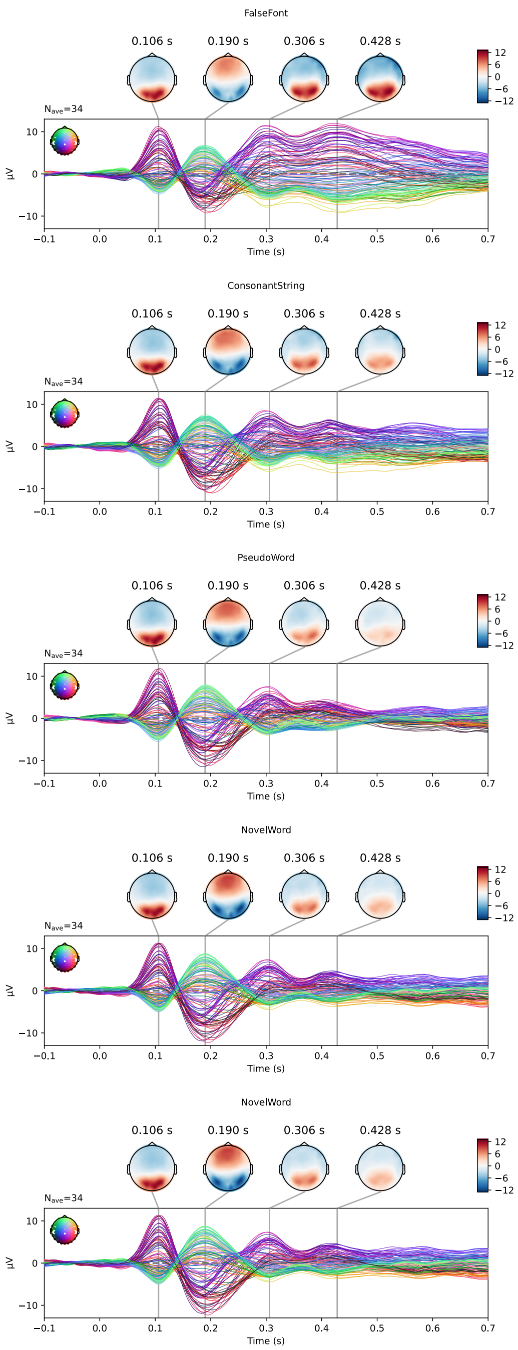

Figure S4. Grand averaged ERP waveforms and scalp topographic maps for each condition. Each channel is represented with a separate line, with line color coding the channel location. Data are referenced to the average of all electrodes.

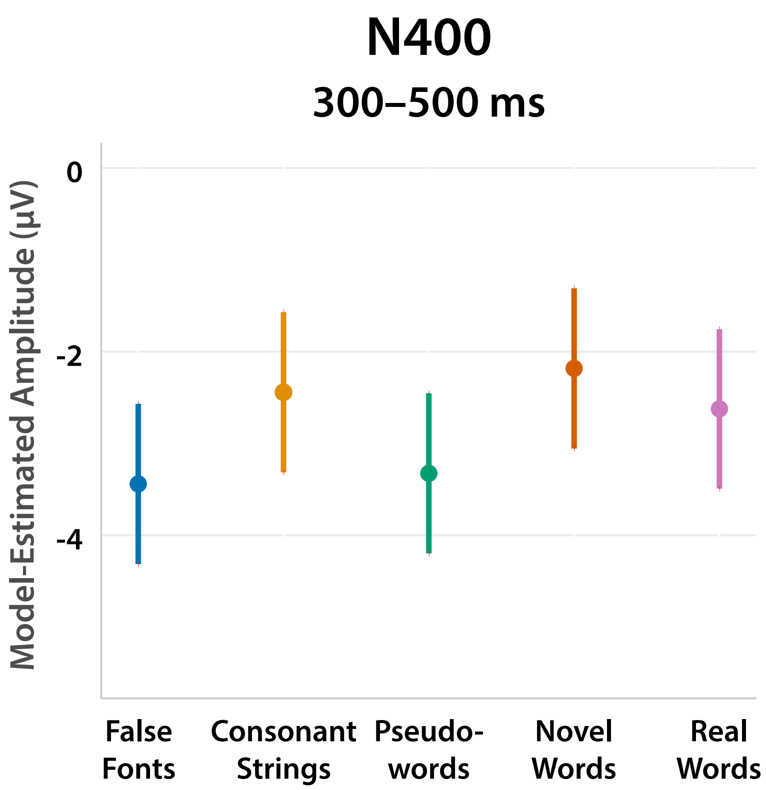

Figure S5. Model-derived plots from the linear mixed effects analysis of the N400 over the 300-500 ms time window. Points represent estimated means and error bars show 95% confidence intervals for each condition

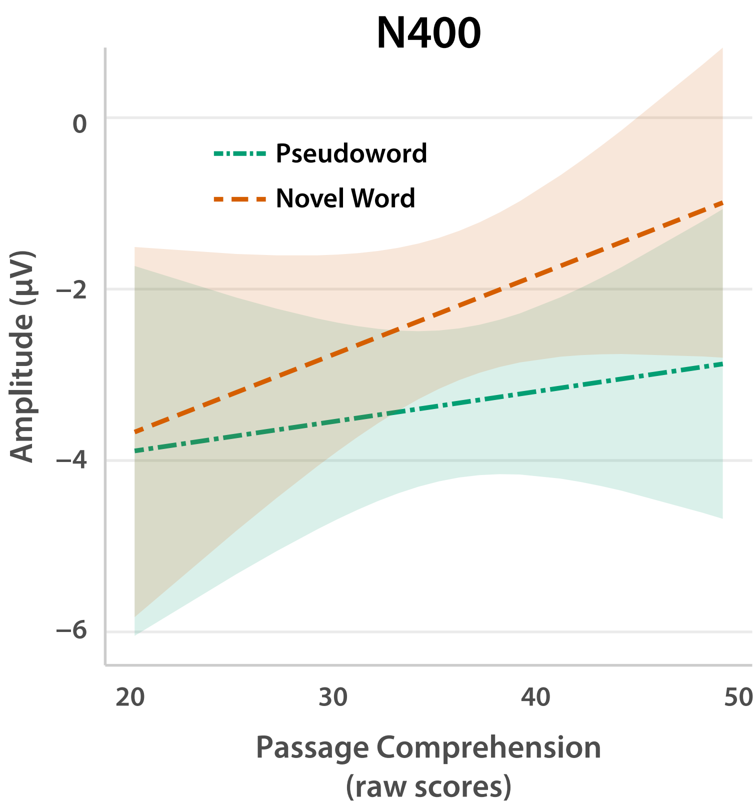

Figure S6. Model-derived plot showing the relationship between passage comprehension (from the WRMT-R), and N400 amplitude (300-500 ms) for the contrast between novel words and pseudowords. Shaded areas represent 95% CIs