Published in final edited form as:

Psychophysiology. 2016 February ; 53(2): 115–128. doi:10.1111/psyp.12567.

N1 and P2 to Words and Wordlike Stimuli in Late Elementary School Children and Adults

Donna Coch and Gabriela Meade

Dartmouth College

Abstract

In an investigation of the development of fine-tuning for word processing across the late elementary school years as indexed by the posterior N1 and P2 components of the event-related potential waveform, third, fourth, and fifth graders and a comparison group of adults viewed words, pseudowords, nonpronounceable letter strings, and false font strings in a semantic categorization task. In adults, N1 was larger to and P2 was later to words as compared to pseudowords, a finely tuned effect of lexicality reflecting specialization for word processing. In contrast, in each group of children, N1 was larger to letter strings than false font strings and P2 was larger to false font strings than letter strings, reflecting coarse encoding for orthography. In regression analyses, scores on standardized behavioral test measures of orthographic knowledge, decoding skill, and fluency predicted N1 amplitude; these effects were not significant with age included as a separate predictor. None of the behavioral scores, in models including or not including age, predicted P2 amplitude. In direct comparisons between groups, there were multiple differences between the child and adult groups for both N1 and P2 amplitude effects, and only a single significant difference between two child groups. Overall, the findings suggest a lengthy developmental time course for the fine-tuning of early word processing as indexed by N1 and P2.

In the adult event-related potential (ERP) literature, the early posterior component N1 has been interpreted as an index of specialization for word processing (e.g., McCandliss, Cohen, & Dehaene, 2003), and the posterior P2 has been associated with word processing, as well (e.g., Savill & Thierry, 2012). A number of ERP investigations have considered N1 development in beginning readers (e.g., Brem et al., 2013; Maurer et al., 2006) and adolescents (e.g., Brem et al., 2006; Kronschnabel, Schmid, Maurer, & Brandies, 2013), but curiously few have addressed the time in between. Here, we investigated the development of word processing as indexed by the N1 and P2 between beginning and fully fluent reading, in late elementary students, by recording ERPs to words, pseudowords, nonpronounceable letter strings, and false font strings.

The N1 Component

ERP studies with adults have associated a family of early negative components, alternately termed N1, N170, or N200, peaking between about 140 and 200 ms and observed primarily

Corresponding author: Donna Coch, Dartmouth College, Department of Education, Reading Brains Lab, 5 Maynard Street, Raven House, HB 6103, Hanover, NH, 03755. Telephone: 603.646.3282. FAX: 603.646.3968. donna.coch@dartmouth.edu.

The authors declare no conflicts of interest.

at occipitotemporal sites, with automatic orthographic processing at both a coarse- and fine-grained level (e.g., Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999; Coch & Mitra, 2010; Nobre, Allison, & McCarthy, 1994). At a coarse-grained level, stimulus strings composed of real letters elicit a larger amplitude (more negative) N1 than other symbol strings in adults, with a similar latency (e.g., Appelbaum, Liotti, Perez III, Fox, & Woldorff, 2009; Bentin et al., 1999; Brem et al., 2006; Mahé, Bonnefond, Gavens, Dufour, & Doignon-Camus, 2012; Maurer, Brandeis, & McCandliss, 2005). At a fine-grained level, N1 amplitude is sensitive to how letters pattern together within strings, such that real words and pseudowords elicit more negative N1s than unpronounceable letter strings in adults (e.g., Coch & Mitra, 2010; Compton, Grossenbacher, Posner, & Tucker, 1991; C. D. Martin, Nazir, Thierry, Paulignan, & Démonet, 2006; McCandliss, Posner, & Givón, 1997). At an even finer grain of lexicality, words elicit a more negative N1 than pseudowords in some studies with adults (e.g., Mahé et al., 2012; Maurer, Brandeis, et al., 2005), although not others (e.g., Bentin et al., 1999; Grossi & Coch, 2005; Maurer, Brem, Bucher, & Brandeis, 2005).

In studies of beginning reading, N1 amplitude does not distinguish between words and false font strings in pre-readers performing a repetition detection task (e.g., Brem et al., 2010; Maurer et al., 2006). However, after a few hours of grapheme-phoneme correspondence training, words do elicit a greater amplitude (more negative) N1 than false font strings in these children (Brem et al., 2010), similar to the pattern observed in adults (e.g., Maurer, Brem, et al., 2005). This rapid emergence of N1 differentiation between words and false font strings appears to occur when orthographic and phonological subsystems are connected (e.g., Maurer & McCandliss, 2007). A similar pattern is apparent in first graders: After a year of formal reading instruction, N1 peaks earlier to words and N1 amplitude differentiates words and pseudowords as compared to false font strings; further, the N1 amplitude difference between words and false font strings is correlated with word reading fluency and vocabulary measures (Eberhard-Moscicka, Jost, Raith, & Maurer, 2015; see also Parviainen, Helenius, Poskiparta, Niemi, & Salmelin, 2006). In second grade, words continue to elicit a larger N1 than symbol strings in typically developing readers (e.g., Brem et al., 2013; Brem et al., 2009; Hasko, Groth, Bruder, Bartling, & Schulte-Körne, 2013), with the difference in N1 amplitude correlated with reading speed (e.g., Maurer et al., 2006). Overall, coarse-grained N1 processing appears to emerge early, with initial reading experiences in kindergarten and the primary grades.

Fewer studies have investigated N1 as an index of fine-grained processing in beginning readers. Zhao and colleagues (2014) recently reported that, in 7-year-old German children performing a repetition detection task, words, pseudowords, and consonant strings all elicited longer-latency and more negative N1s than symbol strings. N1 amplitude also differentiated among the types of letter-based stimuli (except words and pseudowords), such that both words and (marginally) pseudowords elicited larger N1s than consonant strings, but only in better readers. Further, children with faster behavioral pseudoword reading speeds showed larger N1 differences between pseudowords and consonant strings at left hemisphere sites (Zhao et al., 2014), again suggesting an association between N1 processing and fluency in beginning readers. In contrast, Posner and McCandliss (1999) had previously reported no differences in mean N1 amplitude to consonant strings and high frequency

words in 7-year-old children during a passive reading task. Given that they did not report on the strength of their readers, it could be the case that these findings are not conflicting, and that differentiation is only seen in better readers at this age.

Beyond the primary years, there has been little investigation of the typical course of N1 development in older elementary school children (e.g., Brem et al., 2009; Coch, Mitra, & George, 2012; Posner & McCandliss, 1999; Spironelli & Angrilli, 2009), along with a few studies focused on older elementary readers with dyslexia (e.g., Araújo, Bramão, Faísca, Petersson, & Reis, 2012; Hasko et al., 2013; Kast, Elmer, Jancke, & Meyer, 2010); no studies, to our knowledge, have investigated development across the upper elementary school years. In a word-pair paradigm, Spironelli and Angrilli (2009) reported a more bilateral posterior N1 in 10-year-olds than in young or middle-aged adults, regardless of task (orthographic, phonological, or semantic). In an implicit reading (matching) task, Araújo et al. (2012) reported similar N1 amplitudes to words and pseudowords, but smaller (less negative) N1s to symbol strings, in a group of 9- to 13-year-old typical readers (see also Kast et al., 2010). Indeed, Brem and colleagues (2009) found that this coarse differentiation between words and symbol strings in terms of N1 amplitude was greater in 10-year-old children than in adolescents or adults (Brem et al., 2006); however, controlling for age, N1 amplitude to words did not correlate with reading skills (Brem et al., 2009). At a finer grain, Posner and McCandliss (1999) reported a smaller (less negative) N1 for known words as compared to unfamiliar stimuli (unknown words and consonant strings) in 10-year-olds in a task requiring detection of a thickened letter segment, and no fine-tuning differentiating consonant strings from words overall, as observed in adults (e.g., McCandliss et al., 1997).

These findings, although inconsistent, suggest development of N1 tuning effects in typical readers during the late elementary school years, but few studies have been designed to investigate this directly. Here, we focused on the N1 as an index of fine-grained word processing across the upper elementary grades (in groups of third, fourth, and fifth graders) by presenting words and various types of word-like stimuli in a semantic categorization task. Further, given recent evidence that N1 fine-tuning in beginning readers may be related to reading skill (Zhao et al., 2014), we investigated whether various behavioral measures of reading skill would predict N1 amplitude to these different stimuli.

The P2 Component

A few studies in the adult ERP literature have considered the posterior P2 as an additional index of word processing, although the P2 is not specific to words or word-like stimuli (e.g., F. H. Martin, Kaine, & Kirby, 2006; McCandliss et al., 1997; Savill & Thierry, 2012; Sereno, Rayner, & Posner, 1998). The posterior visual P2 has been related to encoding (e.g., B. R. Dunn, Dunn, Languis, & Andrews, 1998), but little is known about its functional significance and development in word processing. Across languages and paradigms in studies with adults, the P2 tends to be larger to stimuli that are impossible or difficult (e.g., irregular words, consonant strings, false font or symbol strings) to decode than to stimuli for which orthography and phonology are more easily integrated (e.g., regular words and pseudowords; Appelbaum et al., 2009; F. H. Martin et al., 2006; McCandliss et al., 1997; Sereno et al., 1998). For example, in a training study with adults, consonant strings elicited a

larger posterior P2 than words, more so in a semantic judgment task than a passive task (McCandliss et al., 1997), whereas pseudowords and low frequency words elicit similar-amplitude posterior P2s (Carreiras, Vergara, & Barber, 2005), as do words and pseudohomophones (Savill & Thierry, 2012). Moreover, the posterior P2 elicited by pseudowords is larger in adults who are poorer decoders (i.e., make more errors on a pseudoword reading task; F. H. Martin et al., 2006).

To our knowledge, just a handful of ERP reading studies have reported on the posterior P2 in typically developing children (e.g., Holcomb, Coffey, & Neville, 1992; Meng, Jian, Shu, Tian, & Zhou, 2008). For example, in a semantic judgment task in Chinese, Meng and colleagues (2008) found a stronger graded effect of orthographic mismatch on posterior P2 amplitude to sentence-terminal characters in fourth and fifth graders as compared to adults. However, in English, Holcomb and colleagues (1992) reported a smaller and earlier occipital P2 to anomalous, as compared to best completion, sentence-terminal words in both children and adults. This limited evidence is equivocal regarding developmental differences in P2 word processing in upper elementary students and adults; here, we explored this more directly.

The Present Study

We investigated the development of fine-tuning for word processing as indexed by the posterior N1 and P2 components across the late elementary school years by comparing third, fourth, and fifth grade students and a control group of adults. We focused on the sensitivity of the N1 and P2 to semantics, by comparing ERPs to words and pseudowords; to phonology, by comparing ERPs to pseudowords and letter strings; and to orthography, by comparing ERPs to letter strings and false font strings. Further, we explored associations between these ERP measures and standardized behavioral measures of orthographic, phonological, and semantic skill. In adults, we predicted that N1 amplitude would differentiate between letter strings and false font strings (e.g., Appelbaum et al., 2009), reflecting coarse tuning, as well as words and pseudowords (e.g., Mahé et al., 2012; Maurer, Brandeis, et al., 2005) and pseudowords and letter strings, reflecting fine-tuning for word processing. In children, we predicted similar evidence of coarse tuning, given previous findings of early development in younger readers, perhaps related to behavioral reading fluency (e.g., Brem et al., 2010; Eberhard-Moscicka et al., 2015; Maurer et al., 2006). Given limited evidence for N1 fine-tuning in children (Zhao et al., 2014; cf. Posner & McCandliss, 1999), we predicted differentiation between words and pseudowords, as well as pseudowords and letter strings, in older and better readers. With respect to the P2, given previous findings in adults, we predicted differentiation of pseudowords and letter strings, but not words and pseudowords (e.g., Savill & Thierry, 2012). Posterior P2 analyses in the child groups were more exploratory, given the lack of previous research on developmental fine-tuning for word processing for the P2.

Method

Participants

Participants included 24 undergraduate students (12 female, average age 20;3 years, *SD* 21 months), 24 fifth grade students (12 female, average 11;1, *SD* 4 months), 24 fourth grade students (12 female, average 10;0, *SD* 6 months), and 24 third grade students (13 female, average 8;9, *SD* 5 months) from schools in the northeastern United States. All participants were right-handed (Oldfield, 1971), monolingual English speakers with no history of neurological dysfunction or diagnosed language or reading disorders by self or parent report. None were currently taking medication that would affect brain function. All had normal or corrected-to-normal binocular vision (20/40 or better) as tested with a standard Snellen chart. All participants were volunteers paid \$20 for their time; children also received small prizes. Data from these participants related to the N400 component have been reported previously (Coch, 2015; Coch & Benoit, 2015).

Materials and Procedure

Standardized behavioral tests—Participants were given a battery of standardized tests measuring various aspects of reading-related skills. The Spelling Accuracy composite of the Test of Orthographic Competence (TOC, Mather, Roberts, Hammill, & Allen, 2008) provided a measure of orthographic and orthographic-to-phonological mapping knowledge. This test is normed through age 17;11; the norms for this age group were used for adult participants here. The Phonological Awareness composite of the Comprehensive Test of Phonological Processing (CTOPP, Wagner, Torgesen, & Rashotte, 1999) and the Phonemic Decoding Efficiency subtest of the Test of Word Reading Efficiency (TOWRE, Torgesen, Wagner, & Rashotte, 1999) measured phonological and decoding knowledge. The Peabody Picture Vocabulary Test IIIA (PPVT, L. M. Dunn & Dunn, 1997) provided a measure of receptive vocabulary, while comprehension was measured by the Passage Comprehension subtest of the Woodcock Reading Mastery Test – Revised (WRMT, Woodcock, 1987). The Letters and Numbers subtests from the Rapid Automatized Naming and Rapid Alternating Stimulus Tests (RAN/RAS, Wolf & Denckla, 2005) served as measures of fluency. Finally, the Memory for Digits subtest of the CTOPP (Wagner et al., 1999) was used as a measure of short-term memory. Administration of these tests took about one hour. Subjects completed behavioral and electrophysiological testing in one or two sessions. Due to experimenter error, scores were not available for one child for the TOC composite or another for the PPVT.

ERP stimuli—The stimulus list consisted of 60 instances of each of five types of stimuli: real words (e.g., *bed*, *bring*), pseudowords (e.g., *bem*, *fring*), nonpronounceable letter strings (e.g., *mbe*, *nrfgi*), false font strings (e.g., *Dech, fuhf*), and target animal names (e.g., *cat*, *duck*) for the semantic categorization task. Real words were the 20 most frequent 3-, 4-, and 5-letter open-class words in grades 1 through 5 (Zeno, Ivens, Millard, & Duvvuri, 1995) that were regular and decodable, single syllable, two to five phonemes, and not animal names. Pseudowords were matched to real words on number of letters and phonemes, orthographic neighborhood size, and constrained and unconstrained bigram and trigram counts (Medler & Binder, 2005). They were formed by changing one letter of each real word to create a

decodable string with no homophones and a single preferred pronunciation. Letter strings were formed by randomly rearranging letters from the pseudowords into orthographically and phonologically illegal strings that were not pronounceable. Letter strings were significantly different from pseudowords on all orthographic neighborhood, bigram, and trigram measures (all $ps < .001$), but had the same number of phonemes as pseudowords. False font stimuli were composed of characters based on each letter in a pseudoword, systematically altered to retain the features of each letter, rearranged (see Grossi & Coch, 2005).

Stimuli were presented using Presentation software (Neurobehavioral Systems) singly, in list form, in white Times New Roman font on a black background. Stimuli were presented in pseudorandom order such that related stimuli (e.g., *fring* and *nrfgi*) were separated by at least four intervening stimuli.

Procedure

Participants were given a brief overview and questions were addressed before undergraduates signed a consent form or parents signed a permission form and children signed an assent form. The standardized behavioral tests were administered, and participants were fitted with an electrode cap (Electro-Cap International, Eaton, Ohio) for EEG recording. Active electrodes included Fz, Cz, Pz, FP1/2, F7/8, FT7/8, F3/4, FC5/6, C3/4, C5/6, T3/4, CT5/6, P3/4, T5/6, TO1/2, and O1/2; only recordings from posterior sites P3/4, T5/6, TO1/2, and O1/2 are analyzed here, but data from all sites are included in the voltage maps. During on-line recording, the right mastoid served as reference; during analysis, recordings were re-referenced to averaged mastoids. Electrodes located beside and below the eyes were used to monitor horizontal eye movements and blinks. Impedances for mastoid and scalp electrodes were maintained below 5 K Ω , and eye electrode impedances below 10 K Ω .

Participants were seated in a comfortable chair in a sound attenuating and electrically shielded booth for the ERP semantic categorization task. As described previously (Coch, 2015), stimuli appeared singly on an LCD monitor and subtended about 0.64° of vertical visual angle and 1.5° of horizontal visual angle, minimizing the need for scanning eye movements. The sequence of events began with the presentation of an outline of a white rectangle, within which a stimulus appeared 500 ms later. Duration of the stimulus was 350 ms for college students and 500 ms for children (based on piloting and consistent with evidence that children take longer to read letter strings than adults, e.g., Greene & Royer, 1994). The white rectangle outline remained on the screen for 700 ms after the stimulus disappeared, followed by an asterisk for 2000 ms, followed by a blank screen for 500 ms, and the beginning of the next trial. Participants were instructed that they would see different kinds of words appear on the screen and to press a button on a game controller whenever they saw an animal name; response hand was counterbalanced across participants. They were further instructed to sit as still and relaxed as possible, to keep their eyes at the center of the screen, to try not to blink or move while the white rectangle was on the screen, and to save their blinks and wiggles for when the star (asterisk) was on the screen. A total of 300 stimuli were presented, with recording time of approximately 20 minutes. Breaks were given

at each quarter, and more often as needed. A practice list with 10 items, not including any stimulus used in the actual experiment, was run prior to presentation of the experimental list.

Data Analysis

As previously reported (Coch, 2015), EEG was amplified with SA Instrumentation bioamplifiers (bandpass 0.01 to 100 Hz) and digitized on-line (sampling rate 4 ms). ERPs were time-locked to the onset of each stimulus. Off-line, separate ERPs to words, pseudowords, letter strings, and false font strings were averaged for each subject at each electrode site over a 1000 ms epoch, using a 200 ms pre-stimulus-onset baseline. ERPs to animal name trials, which required a button press response in the semantic categorization task, are not reported on here; accuracy across groups ranged from 90% to 98%, confirming attentiveness and reading ability. Trials contaminated by eye movements, muscular activity, or electrical noise were not included in analyses. Standard artifact rejection parameters were initially employed, and data were subsequently analyzed on an individual basis for artifact rejection. Automated programs identified artifacts through a peak-to-peak amplitude function: Trials were rejected if the amplitude value between the maximum and minimum data points in the specified time window was larger or smaller than an established threshold. The average number of trials included in individual ERP averages for each of the four conditions by group ranged from 43.0 (*SD* 9.6) to 52.8 (*SD* 6.8); pair-wise comparisons across the four groups for the four conditions of interest (Bonferroni-corrected $p = .002$, .05/24), yielded only that the fourth grade group had fewer word trials than the adult group ($p = .001$).

The local peak amplitude and latency of the N1 and P2 were measured at posterior sites P3/4, T5/6, TO1/2, and O1/2, where these components were most clearly identifiable. The N1 was measured in the 150 to 250 ms epoch in all groups; the local peak was defined as the most negative data point within this time window, such that the three preceding and following points were less negative. The P2 was measured in the 250 to 350 ms epoch in the child groups, and in the 200 to 250 ms epoch in adults; local peak was defined as the most positive data point within the time window, such that the three preceding and following points were less positive. Time windows of measurement were consistent with visual inspection of both the individual and grand average waveforms.

Within-group analyses of the ERP data were conducted first in order to quantify the effects of semantics, phonology, and orthography on N1 and P2 local peak amplitude and latency within each group. For each group (adults, fifth graders, fourth graders, or third graders), component (N1 or P2), and measure (amplitude or latency), a series of three ANOVAs was run with factors condition, anterior/posterior [2 levels: temporoparietal (T5/6, P3/4), occipital (TO1/2, O1/2)], lateral/medial, and hemisphere (left, right). Across the three analyses, levels of the condition factor included words and pseudowords (fine-tuning for semantics), pseudowords and letter strings (fine-tuning for phonology), or letter strings and false font strings (coarse tuning for orthography). For these within-group analyses, results were considered significant at the Bonferroni-corrected p of .017.

In addition, between-group analyses were conducted with the N1 peak latency data from the averaged waveforms (because the P2 peak was measured in different time windows for

children and adults, between-group P2 latency analyses were not conducted). An ANOVA with between-groups factor group (adults, fifth graders, fourth graders, third graders) and within-group factors anterior/posterior [2 levels: temporoparietal (T5/6, P3/4), occipital (TO1/2, O1/2)], lateral/medial, and hemisphere was conducted with the N1 latency data. Follow-up pair-wise comparisons by group were conducted with a Bonferroni-corrected p of .008.

Further, difference waves were created by subtracting the ERPs to pseudowords from the ERPs to words to index an effect of semantics; the ERPs to letter strings from the ERPs to pseudowords to index an effect of phonology; and the ERPs to false font strings from the ERPs to letter strings to index an effect of orthography (for a similar approach, see, e.g., Petersen, Fox, Snyder, & Raichle, 1990). Topographical voltage maps based on local peak amplitude measures of the difference waves in each time window of interest were created using a spherical spline interpolation (Perrin, Pernier, Bertrand, & Echallier, 1989) in order to visualize the distributions of these three effects by group. Creation of difference waves served as a kind of normalization that allowed for direct comparison of the effects of stimulus semantics, phonology, and orthography on the N1 and P2 across groups. Local peak amplitude of the N1 and P2 effects was measured in the difference waves using the same approach described above for the individual average waves.

A second set of analyses was conducted with the difference wave data in order to compare effects between groups. For each effect (semantic, phonological, or orthographic) and component (N1 or P2), an ANOVA with between-groups factor group (adults, fifth graders, fourth graders, third graders), and within-group factors effect, anterior/posterior [2 levels: temporoparietal (T5/6, P3/4), occipital (TO1/2, O1/2)], lateral/medial, and hemisphere was conducted. For significant results involving group, follow-up pair-wise comparisons by group were conducted with a Bonferroni-corrected p of .008. For both within- and between-group analyses, the Greenhouse-Geisser correction was applied to all within-subjects measures with more than one degree of freedom; epsilon (ϵ) values and corrected p -values are reported below. Partial eta squared (η^2_p) values are reported as estimates of effect size.

Finally, multiple regressions were undertaken with the ERP and behavioral test data in order to explore potential relations among these variables. In the first set, standard or composite scores on the reading measures, which control for age, were simultaneously entered as predictors of the average local peak amplitude (measured across the eight posterior sites) of the N1 and P2 to words, pseudowords, letter strings, and false font strings. In the second set, age was included as an additional predictor in each model. Results were considered significant at the .05 level.

Results

Standardized Behavioral Tests

Standard scores for each of the reading-related standardized behavioral tests for each group, as well as overall, are summarized in Table 1.

Electrophysiological Data

Within-group analyses—Grand average waveforms at posterior sites for the four conditions for each group are presented in Figure 1. A summary of the results of the within-group analyses can be found in Table 2.

Adults

N1: The amplitude of the N1 was more negative for words than pseudowords, particularly at medial, anterior sites (P3 and P4), Condition $\times$ Anterior/Posterior $\times$ Lateral/Medial, $F(1, 23) = 7.80$, $p = .01$, $\eta^2_p = .25$. N1 amplitude in adults did not significantly differ for pseudowords and letter strings or letter strings and false font strings. N1 latency in adults did not differ between words and pseudowords, pseudowords and letter strings, or letter strings and false font strings.

P2: The amplitude of the P2 in adults did not differ between words and pseudowords, pseudowords and letter strings, or letter strings and false font strings. The P2 peaked later for words than pseudowords at left hemisphere sites, Condition $\times$ Hemisphere, $F(1, 23) = 7.67$, $p = .011$, $\eta^2_p = .25$, but did not differ for pseudowords and letter strings or letter strings and false font strings.

Fifth graders

N1: The amplitude of the N1 was more negative for letter strings than for false font strings in fifth graders, particularly at lateral sites, Condition $\times$ Lateral/Medial, $F(1, 23) = 21.40$, $p = .001$, $\eta^2_p = .48$. N1 amplitude did not differ for words and pseudowords or pseudowords and letter strings. N1 latency in fifth graders did not differ between words and pseudowords, pseudowords and letter strings, or letter strings and false font strings.

P2: The amplitude of the P2 in fifth graders was greater for false font strings than letter strings, particularly at lateral sites, Condition $\times$ Lateral/Medial, $F(1, 23) = 39.66$, $p = .001$, $\eta^2_p = .63$. P2 amplitude did not differ for words and pseudowords or pseudowords and letter strings. The P2 peaked later for letter strings than pseudowords, condition, $F(1, 23) = 8.32$, $p = .008$, $\eta^2_p = .27$, but P2 latency did not differ for words and pseudowords or letter strings and false font strings.

Fourth graders

N1: The amplitude of the N1 in fourth graders was more negative for letter strings than for false font strings, particularly at lateral, posterior sites (TO1 and TO2), Condition $\times$ Anterior/Posterior, $F(1, 23) = 6.81$, $p = .016$, $\eta^2_p = .23$, Condition $\times$ Lateral/Medial, $F(1, 23) = 26.24$, $p = .001$, $\eta^2_p = .53$, Condition $\times$ Anterior/Posterior $\times$ Lateral/Medial, $F(1, 23) = 10.34$, $p = .004$, $\eta^2_p = .31$. The amplitude of the N1 elicited by words and pseudowords, as well as pseudowords and letter strings, was not significantly different. The N1 peaked earlier to words than pseudowords, condition, $F(1, 23) = 10.96$, $p = .003$, $\eta^2_p = .32$, but N1 latency did not differentiate pseudowords and letter strings or letter strings and false font strings.

P2: The amplitude of the P2 in fourth graders was larger to words than pseudowords, condition, $F(1, 23) = 7.84$, $p = .01$, $\eta^2_p = .25$; larger to letter strings than pseudowords,

particularly over right hemisphere sites, condition, $F(1, 23) = 8.10, p = .009, \eta^2_p = .26$, Condition $\times$ Hemisphere, $F(1, 23) = 6.81, p = .016, \eta^2_p = .23$; and larger to false font strings than letter strings, particularly at the most posterior and lateral sites, condition, $F(1, 23) = 11.44, p = .003, \eta^2_p = .33$, Condition $\times$ Anterior/Posterior, $F(1, 23) = 19.57, p = .001, \eta^2_p = .46$, Condition $\times$ Lateral/Medial, $F(1, 23) = 26.08, p = .001, \eta^2_p = .53$. The peak latency of the P2 was not significantly different for words and pseudowords, pseudowords and letter strings, or letter strings and false font strings in fourth graders.

Third graders

N1: The amplitude of the N1 was more negative for letter strings than false font strings in third graders, particularly at the most posterior and lateral sites (TO1 and TO2), Condition $\times$ Anterior/Posterior, $F(1, 23) = 7.36, p = .012, \eta^2_p = .24$, Condition $\times$ Lateral/Medial, $F(1, 23) = 9.07, p = .006, \eta^2_p = .28$. The amplitude of the N1 elicited by words and pseudowords, as well as pseudowords and letter strings, was not significantly different. Latency of the N1 was not significantly different for words and pseudowords, pseudowords and letter strings, or letter strings and false font strings.

P2: The amplitude of the P2 in third graders was greater for letter strings than pseudowords, particularly at lateral sites, Condition $\times$ Lateral/Medial, $F(1, 23) = 6.57, p = .017, \eta^2_p = .22$, and was greater for false font strings than letter strings, particularly at left hemisphere, posterior and lateral sites, Condition $\times$ Lateral/Medial, $F(1, 23) = 9.04, p = .006, \eta^2_p = .28$, Condition $\times$ Anterior/Posterior, $F(1, 23) = 6.68, p = .017, \eta^2_p = .23$, Condition $\times$ Anterior/Posterior $\times$ Hemisphere, $F(1, 23) = 6.79, p = .016, \eta^2_p = .28$. P2 amplitude did not differentiate between words and pseudowords in third graders, and P2 latency was not significantly different for words and pseudowords, pseudowords and letter strings, or letter strings and false font strings.

Between-group analyses—Between-group analyses focused on N1 latency in the averaged waveforms and amplitude effects in the difference waves.

N1 latency

Words: Overall, the peak latency of the N1 elicited by words differed by group, $F(3, 92) = 3.51, p = .018, \eta^2_p = .10$, with an average of 203 ms for third graders, 198 ms for fourth graders, 205 ms for fifth graders, and 190 ms for adults. Pair-wise comparisons indicated only a marginal difference between fifth graders and adults, $F(1, 46) = 7.73, p = .008, \eta^2_p = .14$.

Pseudowords: The peak latency of the N1 elicited by pseudowords also differed by group, $F(3, 92) = 6.84, p = .001, \eta^2_p = .18$, Group $\times$ Lateral/Medial, $F(3, 92) = 4.68, p = .004, \eta^2_p = .13$, Group $\times$ Lateral/Medial $\times$ Anterior/Posterior, $F(3, 92) = 4.96, p = .003, \eta^2_p = .14$, with average 204 ms for third graders, 203 ms for fourth graders, 208 ms for fifth graders, and 187 ms for adults. Paired comparisons indicated no differences among the child groups, but the N1 to pseudowords peaked later in third graders, $F(1, 46) = 9.61, p = .003, \eta^2_p = .17$, especially at posterior, lateral sites, Group $\times$ Lateral/Medial, $F(1, 46) = 10.34, p = .002, \eta^2_p = .18$, Group $\times$ Lateral/Medial $\times$ Anterior/Posterior, $F(1, 46) = 14.12, p = .001, \eta^2_p = .24$;

fourth graders, $F(1, 46) = 9.12, p = .004, \eta^2_p = .17$; and fifth graders, $F(1, 46) = 15.13, p = .001, \eta^2_p = .25$, than in adults.

Letter strings: Similarly, N1 peak latency to letter strings differed by group, $F(3, 92) = 4.57, p = .005, \eta^2_p = .13$, with average 199 ms in third graders, 202 ms in fourth graders, 203 ms in fifth graders, and 186 ms in adults. Paired comparisons yielded no differences among the child groups or third graders and adults, but fourth, $F(1, 46) = 9.03, p = .004, \eta^2_p = .16$, and fifth, $F(1, 46) = 10.16, p = .003, \eta^2_p = .18$, graders showed longer-latency N1s to letter strings than adults.

False font strings: Finally, N1 peak latency to false font strings also differed by group, $F(3, 92) = 8.79, p = .001, \eta^2_p = .22$, with average peak at 204 ms in third graders, 197 ms in fourth graders, 199 ms in fifth graders, and 181 ms in adults. Paired comparisons indicated no differences among the child groups, but third graders, $F(1, 46) = 22.13, p = .001, \eta^2_p = .33$, fourth graders, $F(1, 46) = 11.80, p = .001, \eta^2_p = .20$, and fifth graders, $F(1, 46) = 11.34, p = .002, \eta^2_p = .20$, particularly at anterior lateral and posterior medial sites, Group $\times$ Lateral/Medial $\times$ Anterior/Posterior, $F(1, 46) = 9.46, p = .004, \eta^2_p = .17$, all showed later N1s to false font strings than adults.

Difference waves: Between-group analyses of amplitude effects were conducted with the difference waves. Topographical voltage maps based on the difference waves and representing the effect of stimulus semantics (words-pseudowords), phonology (pseudowords-letter strings), and orthography (letter strings-false font strings) in each group are presented for the N1 in Figure 2 and for the P2 in Figure 3. A summary of the significant between-group comparisons is provided in Table 3.

Semantic effect

N1: Overall, the distribution of the semantic effect on N1 amplitude differed by group, Group $\times$ Lateral/Medial, $F(3, 92) = 9.45, p = .030, \eta^2_p = .09$, Group $\times$ Lateral/Medial $\times$ Hemisphere, $F(3, 92) = 3.38, p = .022, \eta^2_p = .10$. However, follow-up pair-wise comparisons with a Bonferroni-corrected p revealed no significant differences between any groups.

P2: Overall, the amplitude of the P2 in the semantic effect difference waves differed by group, $F(3, 92) = 9.62, p = .001, \eta^2_p = .24$, such that it was largest in fourth graders and smallest in adults. Whereas follow-up paired comparisons indicated no significant differences among the child groups, the P2 semantic effect was smaller in adults than in third graders, $F(1, 46) = 12.59, p = .001, \eta^2_p = .22$; fourth graders, $F(1, 46) = 23.45, p = .001, \eta^2_p = .34$, Group $\times$ Lateral/Medial, $F(1, 46) = 10.26, p = .002, \eta^2_p = .18$; and fifth graders, $F(1, 36) = 15.83, p = .001, \eta^2_p = .26$.

Phonological effect

N1: Overall, the main effect of group was significant, such that the amplitude of the N1 in the phonological effect difference waves was largest for fifth graders, then fourth and third graders, then adults, $F(1, 92) = 2.71, p = .050, \eta^2_p = .08$. Follow-up paired comparisons

yielded only a larger N1 effect in fifth graders than adults, $F(1, 46) = 13.72, p = .001, \eta^2_p = .23$.

P2: The amplitude of the P2 in the phonological effect difference waves differed by group, Group $\times$ Hemisphere, $F(3, 92) = 4.09, p = .009, \eta^2_p = .12$. Follow-up comparisons indicated that this was driven primarily by the third grade group. Third graders differed from fourth graders such that the P2 was larger over the right hemisphere than the left, opposite the effect in the fourth graders, Group $\times$ Hemisphere, $F(1, 46) = 9.42, p = .004, \eta^2_p = .17$. Third graders also differed from the adult group such that the P2 was larger over medial sites in third graders, but more similarly distributed across lateral and medial sites in adults, Group $\times$ Lateral/Medial, $F(1, 46) = 7.74, p = .008, \eta^2_p = .14$.

Orthographic effect

N1: Overall, the amplitude of the N1 in the orthographic effect difference waves varied by group across lateral and medial sites, Group $\times$ Lateral/Medial, $F(3, 92) = 4.91, p = .003, \eta^2_p = .14$. In follow-up comparisons, both third, Group $\times$ Lateral/Medial, $F(1, 46) = 10.10, p = .003, \eta^2_p = .18$, and fourth, Group $\times$ Lateral/Medial, $F(1, 46) = 23.02, p = .001, \eta^2_p = .33$, graders showed a greater (more negative) effect at lateral sites than adults.

P2: The amplitude of the P2 in the orthographic effect difference waves differed by group, Group $\times$ Lateral/Medial, $F(3, 92) = 11.96, p = .001, \eta^2_p = .28$. Follow-up paired comparisons indicated that this was driven by the adult group, as there were no differences between child groups, but all child groups showed a larger P2 at medial than lateral sites, whereas adults showed a more evenly distributed P2 for this effect: third graders and adults, Group $\times$ Lateral/Medial, $F(1, 46) = 21.14, p = .001, \eta^2_p = .32$; fourth graders and adults, Group $\times$ Lateral/Medial, $F(1, 46) = 48.68, p = .001, \eta^2_p = .51$; and fifth graders and adults, Group $\times$ Lateral/Medial, $F(1, 46) = 44.32, p = .001, \eta^2_p = .49$.

Regressions: Behavioral and Electrophysiological Data

In the first series of multiple regressions, standard or composite scores from each of the reading-related standardized behavioral tests (see Table 1) were simultaneously entered as predictors of average N1 and P2 peak amplitude to words, pseudowords, letter strings, and false font strings.

Average N1 local peak amplitude—The test scores explained a significant amount of the variance in N1 amplitude to words, $F(7, 86) = 2.49, p = .022, R^2 = .17, R^2_{adj} = .10$, with both TOWRE Phonemic Decoding scores, $\beta = -.375, t(95) = -2.82, p = .006$, and RAN Letters scores, $\beta = .347, t(95) = 2.22, p = .029$, as significant predictors. The test scores also predicted N1 amplitude to pseudowords, $F(7, 86) = 2.54, p = .020, R^2 = .17, R^2_{adj} = .10$, with both TOC Spelling Accuracy, $\beta = .265, t(95) = 2.29, p = .025$, and TOWRE Phonemic Decoding, $\beta = -.379, t(95) = -2.86, p = .005$, scores as significant predictors. This same pattern held for N1 amplitude to letter strings, $F(7, 86) = 3.01, p = .007, R^2 = .20, R^2_{adj} = .13$, with both TOC Spelling Accuracy, $\beta = .239, t(95) = 2.10, p = .039$, and TOWRE Phonemic Decoding, $\beta = -.407, t(95) = -3.12, p = .002$, scores as predictors. N1 amplitude to false font strings was also predicted by the test scores, $F(7, 86) = 3.35, p = .003, R^2 = .21$,

$R^2_{\text{adj}} = .15$, but only by TOWRE Phonemic Decoding scores, $\beta = -.438$, $t(95) = -3.39$, $p = .001$.

Average P2 local peak amplitude—The behavioral test scores were not significant predictors of P2 amplitude; none of the models for P2 amplitude – to words ($p = .149$), pseudowords ($p = .256$), letter strings ($p = .465$), or false font strings ($p = .653$) – reached significance.

Including age as a predictor—In a second series of multiple regressions, age was included as a separate predictor in addition to the standard or composite behavioral test scores. In this case, the models accounted for a significant amount of the variance in N1 amplitude to words, $F(8, 85) = 2.75$, $p = .01$, $R^2 = .21$, $R^2_{\text{adj}} = .13$, pseudowords, $F(8, 85) = 3.01$, $p = .005$, $R^2 = .22$, $R^2_{\text{adj}} = .15$, letter strings, $F(8, 85) = 3.12$, $p = .004$, $R^2 = .23$, $R^2_{\text{adj}} = .15$, and false font strings, $F(8, 85) = 3.53$, $p = .001$, $R^2 = .25$, $R^2_{\text{adj}} = .18$, but only age was a significant predictor for N1 amplitude to words, $\beta = .288$, $t(95) = 1.99$, $p = .05$, pseudowords, $\beta = .333$, $t(95) = 2.32$, $p = .023$, and false font strings, $\beta = .282$, $t(95) = 2.01$, $p = .048$. Including age as a predictor did not improve the models for P2 amplitude to words ($p = .120$), pseudowords ($p = .245$), letter strings ($p = .163$), or false font strings ($p = .159$).

Discussion

In an investigation of the development of fine-tuning for word processing in typical readers across the late elementary school years as indexed by the posterior N1 and P2, there was ample evidence of developmental differences between children and adults, as well as some evidence for development across the upper elementary grades. In adults, N1 was finely tuned to real words. In contrast, in third, fourth, and fifth graders, N1 showed evidence of coarse tuning for orthography. With respect to P2, latency distinguished words from pseudowords in adults, whereas amplitude distinguished letter strings from false font strings in all three groups of children, letter strings from pseudowords in third and fourth graders, and words from pseudowords only in fourth graders. Taken together, these findings suggest a long time course, extending beyond elementary school, for the development of fine-tuning in early word processing.

N1 Peak Amplitude and Latency

Whereas N1 amplitude did not differentiate between letter strings and pseudowords or false font strings in adults, as we had predicted, words did elicit a larger N1 than pseudowords at parietal sites. This exquisite fine-tuning for words, reflecting N1 sensitivity to lexicality, has been reported in some previous studies with adults (e.g., Mahé et al., 2012; Maurer, Brandeis, et al., 2005). One of these involved a lexical decision task in French with low frequency words (Mahé et al., 2012), which may have drawn on semantic processing similarly to the semantic categorization task used here; however, the other used a one-back task in English with high frequency words (Maurer, Brandeis, et al., 2005), which would not require semantic processing. Still another study has reported lexicality effects in adults in this epoch, but with pseudowords eliciting a larger negativity than words (Hauk et al., 2006). Others have reported no such effects (e.g., Bentin et al., 1999; Grossi & Coch, 2005;

Maurer, Brem, et al., 2005). Overall, the stimulus and paradigm characteristics that support N1 lexicality effects in adults, and their direction, are unclear. Here, with simple, high frequency words; pseudowords created by a single letter swap; and a semantic categorization task, we observed a more negative N1 peak amplitude tuned to words in the adult group.

In contrast, there was no lexicality effect for N1 amplitude in any of the child groups, consistent with previous findings with kindergartners, 8-year-olds, and 9- to 13-year-olds (e.g., Araújo et al., 2012; Hasko et al., 2013; Kast et al., 2010; Maurer, Brem, et al., 2005). However, as predicted based on findings with younger beginning readers (e.g., Brem et al., 2010; Eberhard-Moscicka et al., 2015; Maurer et al., 2006), N1 amplitude was sensitive to coarse orthography: Letter strings elicited a larger (more negative) N1 than false font strings in all three groups of children, particularly at lateral sites. This comparison is a more stringent test of orthographic sensitivity than the typical comparison between words and false font or symbol strings in previous studies (e.g., Brem et al., 2010; Eberhard-Moscicka et al., 2015; Maurer et al., 2006), as words and false font strings differ not only in terms of orthography, but also in terms of phonology and semantics. In comparison, the letter strings used here had no meaning and were conventionally unpronounceable as strings (although single grapheme-to-phoneme correspondences could be applied); the primary difference from false font strings was in the orthographic construction of real letters (indeed, the same letters as in the pseudowords, in rearranged order) forming the string. Despite the more stringent comparison, our findings are consistent with a previous finding of greater coarse differentiation (between words and symbol strings) in terms of N1 amplitude in 10-year-olds than adults (Brem et al., 2006).

In their study with 7-year-olds, Zhao and colleagues (2014) argued that a comparison of words, pseudowords, and consonant strings was more appropriate than a comparison to false font or symbol strings for determining fine-tuning for word processing. As noted, we found similar N1 amplitude for words and pseudowords in children, but we also found similar N1 processing for pseudowords and letter strings in third, fourth, and fifth graders, suggesting little fine-tuning. Similarly, Zhao et al. (2014) reported no significant N1 amplitude difference between words and pseudowords in 7-year-old good readers, and only a marginal difference between pseudowords and consonant strings. As the presence of vowels slows lexical decisions in children (e.g., Henderson & Chard, 1980), it is not surprising that consonant strings may be differentiated from pseudowords more easily (and earlier in developmental time) than letter strings including vowels, as used here. Overall, these findings suggest a relatively short developmental time course, coincident with beginning reading, for coarse orthographic tuning of the N1 (i.e., strings composed of real letters as compared to not), but a relatively long developmental time course, extending beyond elementary school, for specialization of the N1 for other elements of word processing, including phonology and lexicality.

That third, fourth, and fifth graders showed longer-latency N1s to pseudowords, letter strings (except third graders), and false font strings than adults supports the peak amplitude evidence suggesting a long developmental time course for the processing (including phonological and lexical) indexed by this component. However, N1 peak latency to words

was essentially adult-like in the child groups (with a marginal effect for fifth graders). This might suggest some early tuning for words, but such latency effects in the context of a lack of amplitude effects are difficult to interpret; perhaps the timing is on, but the tuning is off.

Consistent with early coarse orthographic processing and eventual specialization of the N1 for word processing, without including age as a separate predictor, scores on the TOC Spelling Accuracy Composite predicted N1 amplitude to pseudowords and letter strings across participants, with higher scores associated with a smaller (less negative) N1. For children, this composite reflects scores on a sight spelling subtest (including sight words like *should*) and a homophone choice subtest (e.g., does *wail* or *whale* match a given picture), both of which emphasize orthographic skill. Better orthographic awareness (i.e., knowledge about the way words look) appears to be related to smaller N1 responses to pseudowords and letter strings (i.e., not-words), consistent with eventual specialization toward the adult pattern of N1 fine-tuning for words in this paradigm.

However, N1 amplitude was not exclusively associated with orthographic awareness in regression analyses not including age as a predictor: Decoding skills as measured by the TOWRE Phonemic Decoding subtest (Torgesen et al., 1999) were also a significant predictor of N1 amplitude, to words, pseudowords, letter strings, and false font strings; better decoding was associated with a larger (more negative) N1. This seems consistent with the reported emergence of a larger N1 to words than false font strings with growing grapheme-to-phoneme correspondence knowledge in beginning readers (e.g., Brem et al., 2010; Eberhard-Moscicka et al., 2015; Zhao et al., 2014), and indicates that N1 amplitude indexes the interaction between orthographic and phonological knowledge into the late elementary school years and beyond. Finally, as a measure of fluency, better RAN Letters (Wolf & Denckla, 2005) scores were associated with larger (more negative) N1 amplitudes to words only. Fluency and N1 amplitude have been associated in previous studies with younger readers (e.g., Eberhard-Moscicka et al., 2015; Maurer et al., 2006; Zhao et al., 2014); this finding extends this developmental relationship through the late elementary school years and into adulthood, and suggests that it is specific to *word* processing.

In contrast, in regression models including age as a separate predictor, only age was a significant predictor of N1 amplitude to words, pseudowords, and false font strings. This is similar to a previous finding of a lack of correlation between reading skill and word-specific N1 amplitude in 10-year-olds, adolescents, and adults when controlling for age (Brem et al., 2009). Using standard or composite scores controls for age on the behavioral test measures, whereas including age as a separate predictor controls for effects of age across both the behavioral and electrophysiological measures. Thus, including age as a predictor potentially addresses maturational issues such as typically larger ERP waveforms in children as compared to adults, but it may also compromise the very effects of interest: how ERPs change over developmental time. That age was not a significant predictor of N1 amplitude to letter strings suggests that “age” as a variable here potentially encompasses both maturational and condition-related effects. Similarly, that including age in the P2 models did not improve their predictability would also seem to argue against age as a purely maturational variable. In addition, that the behavioral scores that were predictive without age included as a separate predictor were consistent with the literature on N1 and grapheme-

to-phoneme development seems important. Therefore, we included both sets of regression analyses here in order to present a full view.

Finally for the N1, in direct comparison of difference waves across groups, the phonological effect on N1 amplitude was larger for fifth graders than for adults and the orthographic effect was larger in third and fourth graders than adults at lateral sites. As to the latter, a larger difference in N1 amplitude between words and symbol strings in 10-year-olds than adults has previously been interpreted as a reflection of a process that becomes refined with reading practice and automatized over time (e.g., Brem et al., 2009; Maurer et al., 2006); specifically, a process related to coarse orthographic encoding (e.g., Brem et al., 2010; Eberhard-Moscicka et al., 2015; Maurer et al., 2006). If the comparison of letter and false font strings here indexes at least in part this same sort of orthographic processing, our orthographic N1 effect findings confirm that this processing is not adult-like in third or fourth graders; however, the similar degree of N1 differentiation between letter and false font strings in fifth graders and adults suggests that such processing may become adult-like during fifth grade. Indeed, the overall pattern of findings in the N1 difference wave analyses suggests that the sensitivity of the N1 may change over developmental time across the late elementary school years, reflecting a greater influence of orthography in third and fourth grade and a greater influence of phonology in fifth grade.

This may be indicative of a shift over time from coarse orthographic sensitivity to additional fine-grained sensitivity for the N1. This is an interesting developmental sequence, as behavioral studies have indicated that younger readers tend to rely more on phonology, gradually shifting to a greater reliance on orthography across the elementary school years as fluency in word reading develops (e.g., Doctor & Coltheart, 1980). However, task differences may contribute to these differing patterns; for example, Doctor and Coltheart reported that visual encoding becomes progressively more important with age from 6 to 10 based on children's responses to homophones in sentence contexts (younger children found sentences such as *He ran threw the street* acceptable more often than older children). In context-less reading of single words, phonological encoding as indexed by N1 amplitude in fifth graders may be part of the process of development into an adult N1 response.

P2 Peak Amplitude and Latency

In adults, P2 amplitude was not sensitive to semantics, as predicted (e.g., Savill & Thierry, 2012), but, contrary to expectations, it was also not sensitive to phonology or orthography; nor was P2 latency sensitive to phonology or orthography. However, the P2 peaked later to words than to pseudowords at left hemisphere sites, again revealing a finely tuned effect of lexicality in adults. This contrasts with the findings from a visual oddball study with controls and adults with dyslexia in which P2 latency was similar to words and pseudowords (Savill & Thierry, 2012). In that study, pseudowords included pseudohomophones of animal targets (e.g., *bair*), a phonological manipulation not used in the current design that may have contributed to the difference in results. Here, both N1 amplitude and P2 latency appeared to be finely tuned to *word* processing for simple, monosyllabic, high frequency words in adults, whereas all of the other word-like but not-word stimuli, in select planned comparisons, were indistinguishable in terms of N1 and P2 processing.

Similar to the N1 in children, the peak amplitude of the P2 was sensitive to whether strings were composed of Roman letters or not: In all three groups of children, false font strings elicited larger P2s than letter strings at lateral sites, continuing the greater positivity elicited by false font strings observed in the N1 epoch into the P2 epoch. Meng and colleagues (2008) also reported orthographic effects on P2 amplitude in fourth and fifth graders reading Chinese sentence-terminal words, although their made-up characters elicited a smaller P2 than real characters. Here, in third and fourth graders, letter strings also elicited larger P2s than pseudowords; a similar pattern was seen in fifth graders, but in terms of P2 latency (a longer latency to letter strings than pseudowords, opposite the pattern in adults showing a longer latency P2 to more word-like than less word-like stimuli). Overall, this pattern suggests that P2 (in terms of latency or amplitude) is not only sensitive to letter-level orthography in children, but also to sequences and combinations of letters, in a way that it is not in adults in this paradigm. Thus, oddly, the pattern of P2 findings that we observed in children is consistent with the general pattern of P2 findings in the adult literature: The component was larger to stimuli that were more difficult to decode or integrate (e.g., Appelbaum et al., 2009; F. H. Martin et al., 2006; McCandliss et al., 1997). An exception to this pattern in the child data was the finding of a larger P2 to words than to pseudowords, exclusive to fourth graders, which is difficult to interpret.¹ That interpretation is not aided by the behavioral test data, as none of the scores were predictive of P2 amplitude, with or without age included as a predictor.

In difference wave analyses, whereas the semantic effect on P2 amplitude was similar in the child groups, it was larger in children than adults; although words elicited a significantly larger P2 than pseudowords only in fourth graders, all three groups of children showed greater P2 differentiation between these stimulus types than the adult group. P2 amplitude has been associated with semantic processing in children previously, such that words that were more difficult to integrate with a sentential context elicited similarly smaller P2s in children and adults (Holcomb et al., 1992). Whereas such an effect may in part be semantic, it may also reflect enhanced attention to predicted (i.e., by the semantic context) items, as suggested by divided visual field studies with adults (e.g., Federmeier & Kutas, 2002; Wlotko & Federmeier, 2007). Given the lack of semantic constraint and predictability here, we speculate that the current findings suggest that the posterior P2 may be sensitive to developmental changes in semantic processing as indexed by lexicality, potentially encompassing greater attention to the expected (words), even in isolated single-word processing.

In addition to semantics, P2 amplitude was sensitive to phonology and orthography in the difference wave analyses. The third grade group appeared to drive developmental differences in the phonological effect for P2 amplitude (letter strings elicited a larger P2 than pseudowords), which was more right-lateralized in third graders than in fourth graders

¹Whereas the fourth grade group showed the largest average difference between P2 peak amplitude to words and pseudowords, and the largest standard deviation for this contrast, there was, as might be expected, substantial variability within each group (standard deviation for third graders: 2.7 μ V, fourth graders: 3.7 μ V, fifth graders: 2.2 μ V, and adults: 2.2 μ V). Every fourth grader except one showed a positive effect, as did every fifth grader except one, and every third grader except two. It is possible that fewer word trials in the fourth grade group as compared to the adult group could have contributed to the difference between these groups, but not to the differences between the child groups.

and more medial in third graders than adults. Such right-lateralization has previously been reported in children for the N1 (Maurer, Brem, et al., 2005; Spironelli & Angrilli, 2009). However, the lack of difference between adults and fifth graders for P2 amplitude as related to the phonological effect contrasts with the findings for N1, suggesting different developmental timelines for the influence of phonology on these two early components. Finally, the orthographic effect on P2 amplitude in the difference wave analyses was similar in the child groups, but more medially distributed in children than adults. These distributional differences for the phonological and orthographic effects indicate some development across the late elementary school years and into adulthood, perhaps related to differing orientations of generators with maturation.

Limitations

A number of methodological limitations may have influenced the results of this study. First, children in the study tended to be strong readers; while we purposely screened for a history of language and reading disorders, our “typical” readers were, for the most part, more above grade level than at grade level. Second, stimulus duration was longer for children than for adults. Piloting indicated that the adult duration was too short for children to accurately read the stimuli without stress and the child duration was too long for adults not to blink and lose attention. This is likely related to differences in the automaticity of word processing between groups, and is therefore relevant to the results; however, practical considerations necessarily motivated this design. Third, mindful of the need to statistically correct for multiple comparisons, we limited our paired contrasts among the four conditions in a principled way: to focus on effects of orthography, phonology, and semantics. In some cases, this disallowed for direct comparison to previous work; for example, in terms of the ubiquitous word versus symbol string contrast used as a measure of coarse orthographic processing. Instead, we have argued that our letter string versus false font string contrast is a more conservative measure of such coarse orthographic encoding (e.g., Zhao et al., 2014). Fourth and finally, consistent with some (e.g., Appelbaum et al., 2009; Grossi & Coch, 2005; Holcomb et al., 1992; Meng et al., 2008; Savill & Thierry, 2012) but not other (e.g., Brem et al., 2010; Brem et al., 2013; Maurer et al., 2006; Zhao et al., 2014) previous studies of N1 and P2, we have used an averaged mastoid reference, as with the N400 data previously reported for these same participants (Coch, 2015; Coch & Benoit, 2015). Overall, our findings do not seem incompatible with previous findings from studies employing a different reference.

Summary and Conclusion

To our knowledge, this is the first developmental investigation of fine-tuning for word processing as indexed by the posterior N1 and P2 components across consecutive late elementary school years. We found remarkable evidence for fine-tuning in adults in this paradigm: Words elicited a larger (more negative) N1 and a later P2 than pseudowords, with processing of other stimulus types indistinguishable in planned comparisons. In contrast, in children, we found evidence for coarse orthographic processing in terms of N1, and both coarse and finer processing in terms of P2. In direct comparisons between groups, there were few significant differences among the child groups but marked differences between the child groups and the adult group. Overall, our results are consistent with development beyond the fifth grade in fine-tuning for word processing as indexed by N1 and P2. Thus,

our findings contribute to a growing ERP literature suggesting a lengthy developmental time course for the automatic processes underlying fluent reading (e.g., Brem et al., 2009; Coch et al., 2012; Eddy, Grainger, Holcomb, Mitra, & Gabrieli, 2014).

Acknowledgments

We thank the children, parents, and college students who chose to participate, and the schools, libraries, businesses, and programs that allowed us to share information about the study. We appreciate the help of research assistants Jennifer Bares, Clarisse Benoit, Natalie Berger, Sarah Brim, Ayesha Dholakia, Elyse George, Emily Jasinski, Priya Mitra, Erin Rokey, and Anna Roth, and programmers Ray Vukcevic and Mark Dow.

This research was supported by Grant R03HD058613 from the National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development.

References

- Appelbaum LG, Liotti M, Perez R III, Fox SP, Woldorff MG. The temporal dynamics of implicit processing of non-letter, letter, and word-forms in the human visual cortex. *Frontiers in Human Neuroscience*. 2009; 3(Article 56):1–11.10.3389/neuro.09.056.2009 [PubMed: 19255629]
- Araújo S, Bramão I, Faísca L, Petersson KM, Reis A. Electrophysiological correlates of impaired reading in dyslexic pre-adolescent children. *Brain and Cognition*. 2012; 79:79–88.10.1016/j.bandc.2012.02.010 [PubMed: 22466501]
- Bentin S, Mouchetant-Rostaing Y, Giard MH, Echallier JF, Pernier J. ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *Journal of Cognitive Neuroscience*. 1999; 11(3):235–260.10.1162/089892999563373 [PubMed: 10402254]
- Brem S, Bach S, Kucian K, Guttorm T, Martin E, Lyytinen H, Richardson U. Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proceedings of the National Academy of Sciences*. 2010; 107(17):7939–7944.10.1073/pnas.0904402107
- Brem S, Bach S, Kujala JV, Maurer U, Lyytinen H, Richardson U, Brandeis D. An electrophysiological study of print processing in kindergarten: the contribution of the visual N1 as a predictor of reading outcome. *Developmental Neuropsychology*. 2013; 38(9):567–594.10.1080/87565641.2013.828729 [PubMed: 24219696]
- Brem S, Bucher K, Halder P, Summers P, Dietrich T, Martin E, Brandeis D. Evidence for developmental changes in the visual word processing network beyond adolescence. *Neuroimage*. 2006; 29:822–837.10.1016/j.neuroimage.2005.09.023 [PubMed: 16257546]
- Brem S, Halder P, Bucher K, Summers P, Martin E, Brandeis D. Tuning of the visual word processing system: distinct developmental ERP and fMRI effects. *Human Brain Mapping*. 2009; 30:1833–1844.10.1002/hbm.20751 [PubMed: 19288464]
- Carreiras M, Vergara M, Barber H. Early event-related potential effects of syllabic processing during visual word recognition. *Journal of Cognitive Neuroscience*. 2005; 17(11):1803–1817.10.1162/089892905774589217 [PubMed: 16269115]
- Coch D. The N400 and the fourth grade shift. *Developmental Science*. 2015; 18(2):254–269.10.1111/desc.12212 [PubMed: 25041502]
- Coch D, Benoit C. N400 event-related potential and standardized measures of reading in late elementary school children: correlated or independent? *Mind, Brain, and Education*. 2015; 9(3):145–153.10.1111/mbe.12083
- Coch D, Mitra P. Word and pseudoword superiority effects reflected in the ERP waveform. *Brain Research*. 2010; 1329:159–174.10.1016/j.brainres.2010.02.084 [PubMed: 20211607]
- Coch D, Mitra P, George E. Behavioral and ERP evidence of word and pseudoword superiority effects in 7- and 11-year-olds. *Brain Research*. 2012; 1486:68–81.10.1016/j.brainres.2012.09.041 [PubMed: 23036274]

- Compton PE, Grossenbacher P, Posner MI, Tucker DM. A cognitive-anatomical approach to attention and lexical access. *Journal of Cognitive Neuroscience*. 1991; 3:304–312.10.1162/jocn.1991.3.4.304 [PubMed: 23967809]
- Doctor EA, Coltheart M. Children's use of phonological encoding when reading for meaning. *Memory and Cognition*. 1980; 8(3):195–209.10.3758/BF03197607 [PubMed: 7392946]
- Dunn BR, Dunn DA, Languis M, Andrews D. The relation of ERP components to complex memory processing. *Brain and Cognition*. 1998; 36(3):355–376.10.1006/brcg.1998.0998 [PubMed: 9647684]
- Dunn, LM.; Dunn, LM. Peabody picture vocabulary test. 3. Circle Pines, MN: American Guidance Service; 1997.
- Eberhard-Moscicka AK, Jost LB, Raith M, Maurer U. Neurocognitive mechanisms of learning to read: print tuning in beginning readers related to word-reading fluency and semantics but not phonology. *Developmental Science*. 2015; 18(1):106–118.10.1111/desc.12189 [PubMed: 24863157]
- Eddy MD, Grainger J, Holcomb PJ, Mitra P, Gabrieli JDE. Masked priming and ERPs dissociate maturation of orthographic and semantic components of visual word recognition in children. *Psychophysiology*. 2014; 51(2):136–141.10.1111/psyp.12164 [PubMed: 24313638]
- Federmeier KD, Kutas M. Picture the difference: electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia*. 2002; 40:730–747.10.1016/S0028-3932(01)00193-2 [PubMed: 11900725]
- Greene BA, Royer JM. A developmental review of response time data that support a cognitive components model of reading. *Educational Psychology Review*. 1994; 6(2):141–172.10.1007/BF02208971
- Grossi G, Coch D. Automatic word form processing in masked priming: an ERP study. *Psychophysiology*. 2005; 42:343–355.10.1111/j.1469-8986.2005.00286.x [PubMed: 15943688]
- Hasko S, Groth K, Bruder J, Bartling J, Schulte-Körne G. The time course of reading processes in children with and without dyslexia: an ERP study. *Frontiers in Human Neuroscience*. 2013; 7 Article 570. 10.3389/fnhum.2013.00570
- Hauk O, Patterson K, Woollams A, Watling L, Pulvermüller F, Rogers TT. [Q:] When would you prefer a SOSSAGE to a SAUSAGE? [A:] At about 100 msec. ERP correlates of orthographic typicality and lexicality in written word recognition. *Journal of Cognitive Neuroscience*. 2006; 18(5):818–832.10.1162/jocn.2006.18.5.818 [PubMed: 16768380]
- Henderson, L.; Chard, J. The reader's implicit knowledge of orthographic structure. In: Frith, U., editor. *Cognitive processes in spelling*. NY: Academic Press; 1980. p. 85–116.
- Holcomb PJ, Coffey SA, Neville HJ. Visual and auditory sentence processing: a developmental analysis using event-related brain potentials. *Developmental Neuropsychology*. 1992; 8:203–241.10.1080/87565649209540525
- Kast M, Elmer S, Jancke L, Meyer M. ERP differences of pre-lexical processing between dyslexic and non-dyslexic children. *International Journal of Psychophysiology*. 2010; 77:56–69.10.1016/j.ijpsycho.2010.04.003
- Kronsnabel J, Schmid R, Maurer U, Brandies D. Visual print tuning deficits in dyslexic adolescents under minimized phonological demands. *Neuroimage*. 2013; 74:58–69.10.1016/j.neuroimage.2013.02.014 [PubMed: 23428569]
- Mahé G, Bonnefond A, Gavens N, Dufour A, Doignon-Camus N. Impaired visual expertise for print in French adults with dyslexia as shown by N170 tuning. *Neuropsychologia*. 2012; 50:3200–3206.10.1016/j.neuropsychologia.2012.10.013 [PubMed: 23088819]
- Martin CD, Nazir T, Thierry G, Paulignan Y, Démonet JF. Perceptual and lexical effects in letter identification: an event-related potential study of the word superiority effect. *Brain Research*. 2006; 1098:153–160.10.1016/j.brainres.2006.04.097 [PubMed: 16774747]
- Martin FH, Kaine A, Kirby M. Event-related potentials elicited during word recognition by adult good and poor phonological decoders. *Brain and Language*. 2006; 96(1):1–13.10.1016/j.bandl.2005.04.009 [PubMed: 15963561]
- Mather, N.; Roberts, R.; Hammill, DD.; Allen, EA. *Test of orthographic competence*. Austin, TX: Pro-Ed; 2008.

- Maurer U, Brandeis D, McCandliss BD. Fast, visual specialization for reading in English revealed by the topography of the N170 ERP response. *Behavioral and Brain Functions*. 2005; 1(13):10.1186/1744-9081-1-13
- Maurer U, Brem S, Bucher K, Brandeis D. Emerging neurophysiological specialization for letter strings. *Journal of Cognitive Neuroscience*. 2005; 17(10):1532–1552.10.1162/089892905774597218 [PubMed: 16269095]
- Maurer U, Brem S, Kranz F, Bucher K, Benz R, Halder P, Brandeis D. Coarse neural tuning for print peaks when children learn to read. *Neuroimage*. 2006; 33:749–758.10.1016/j.neuroimage.2006.06.025 [PubMed: 16920367]
- Maurer, U.; McCandliss, BD. The development of visual expertise for words: the contribution of electrophysiology. In: Grigorenko, EL.; Naples, AJ., editors. *Single-word reading: behavioral and biological perspectives*. New York: Lawrence Erlbaum Associates; 2007. p. 43-64.
- McCandliss BD, Cohen L, Dehaene S. The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*. 2003; 7(7):293–299.10.1016/S1364-6613(03)00134-7 [PubMed: 12860187]
- McCandliss BD, Posner MI, Givón T. Brain plasticity and learning visual words. *Cognitive Psychology*. 1997; 33(1):88–110.10.1006/cogp.1997.0661
- Medler, DA.; Binder, JR. MCWord: an on-line orthographic database of the English language. 2005. <http://www.neuro.mcu.edu/mcword/>
- Meng X, Jian J, Shu H, Tian X, Zhou X. ERP correlates of the development of orthographical and phonological processing during Chinese sentence reading. *Brain Research*. 2008; 1219:91–102.10.1016/j.brainres.2008.04.052 [PubMed: 18539265]
- Nobre AC, Allison T, McCarthy G. Word recognition in the human inferior temporal lobe. *Nature*. 1994; 372:260–263.10.1038/372260a0 [PubMed: 7969469]
- Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*. 1971; 9:97–113.10.1016/0028-3932(71)90067-4 [PubMed: 5146491]
- Parviainen T, Helenius P, Poskiparta E, Niemi P, Salmelin R. Cortical sequence of word perception in beginning readers. *Journal of Neuroscience*. 2006; 26(22):6052–6061.10.1523/JNEUROSCI.0673-06.2006 [PubMed: 16738248]
- Perrin F, Pernier J, Bertrand O, Echallier JF. Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*. 1989; 72:184–187.10.1016/0013-4694(89)90180-6 [PubMed: 2464490]
- Petersen SE, Fox PT, Snyder A, Raichle ME. Activation of prestriate and frontal cortical activity by words and word-like stimuli. *Science*. 1990; 249:1041–1044.10.1126/science.2396097 [PubMed: 2396097]
- Posner, MI.; McCandliss, B. Brain circuitry during reading. In: Klein, RM.; McMullen, P., editors. *Converging methods for understanding reading and dyslexia*. Cambridge, MA: MIT Press; 1999. p. 305-337.
- Savill NJ, Thierry G. Decoding ability makes waves in reading: deficient interactions between attention and phonological analysis in developmental dyslexia. *Neuropsychologia*. 2012; 50:1553–1564.10.1016/j.neuropsychologia.2012.03.008 [PubMed: 22426204]
- Sereno SC, Rayner K, Posner MI. Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport*. 1998; 9(10):2195–2200.10.1097/00001756-199807130-00009 [PubMed: 9694199]
- Spironelli C, Angrilli A. Developmental aspects of automatic word processing: language lateralization of early ERP components in children, young adults and middle-aged subjects. *Biological Psychology*. 2009; 80:35–45.10.1016/j.biopsycho.2008.01.012 [PubMed: 18343558]
- Torgesen, JK.; Wagner, R.; Rashotte, C. *Test of word reading efficiency*. Austin, TX: Pro-Ed; 1999.
- Wagner, RK.; Torgesen, JK.; Rashotte, CA. *The comprehensive test of phonological processing*. Austin, TX: Pro-Ed; 1999.
- Wlotko EW, Federmeier KD. Finding the right word: hemispheric asymmetries in the use of sentence context information. *Neuropsychologia*. 2007; 45:3001–3014.10.1016/j.neuropsychologia.2007.05.013 [PubMed: 17659309]

- Wolf, M.; Denckla, MB. Rapid automatized naming and rapid alternating stimulus tests (RAN/RAS). Austin, TX: Pro-Ed; 2005.
- Woodcock, RW. The Woodcock reading mastery tests - revised. Circle Pines, MN: American Guidance Service; 1987.
- Zeno, SM.; Ivens, SH.; Millard, RT.; Duvvuri, R. The educator's word frequency guide. Brewster, NY: Touchstone Applied Science Associates, Inc; 1995.
- Zhao J, Kipp K, Gaspar C, Maurer U, Weng X, Mecklinger A, Li S. Fine neural tuning for orthographic properties of words emerges early in children reading alphabetic script. *Journal of Cognitive Neuroscience*. 2014; 26(11):2431–2442.10.1162/jocn_a_00660 [PubMed: 24800627]

Figure 1.

Grand average ERP waveforms elicited by words (solid black line), pseudowords (solid gray line), letter strings (dashed black line), and false font strings (dashed gray line) at posterior sites for each of the four groups. The N1 and P2 peaks are identified at site O2 in each of the groups. Left hemisphere sites are on the left, lateral sites are toward the outer edges of the figure, each vertical tick marks 100 ms, the calibration bar marks 2 μ V and stimulus onset, and negative is plotted up.

Figure 2.

Topographical voltage maps illustrating the effect of semantics (words-pseudowords), phonology (pseudowords-letter strings), and orthography (letter strings-false font strings) on N1 amplitude for the four groups. A spherical spline interpolation (Perrin et al., 1989) was used to interpolate the potential on the surface of an idealized, spherical head based on the local peak amplitude voltages measured from the difference waves at each electrode location within the N1 epoch (150–250 ms across groups). Posterior electrode sites included in analyses are identified by gray circles in the adult panel, semantic effect.

Figure 3.

Topographical voltage maps illustrating the effect of semantics (words-pseudowords), phonology (pseudowords-letter strings), and orthography (letter strings-false font strings) on P2 amplitude for the four groups. A spherical spline interpolation (Perrin et al., 1989) was used to interpolate the potential on the surface of an idealized, spherical head based on the local peak amplitude voltages measured from the difference waves at each electrode location within the P2 epoch (250–350 ms in the child groups, 200–250 ms in the adult group).

Posterior electrode sites included in analyses are identified by gray circles in the adult panel, semantic effect.

Table 1

Standard/composite scores (SD) on the reading-related standardized behavioral tests by group and overall

Group	TOC SPACC	CTOPP PA	PPVT	WRMT PC	TOWRE PD	RAN NUM	RAN LETT
Adults	120.8 (8.4)	106.4 (7.2)	126.7 (10.1)	118.8 (9.0)	101.2 (8.2)	114.3 (3.9)	111.8 (4.2)
Fifth Graders	116.7 (19.9)	109.8 (9.0)	123.8 (16.4)	115.3 (12.2)	109.4 (15.1)	104.8 (11.3)	104.6 (11.9)
Fourth Graders	114.0 (26.7)	113.6 (9.4)	123.3 (11.6)	119.2 (10.6)	117.1 (14.0)	108.0 (9.4)	105.2 (8.2)
Third Graders	112.8 (13.8)	114.0 (10.9)	125.5 (10.7)	119.5 (9.4)	114.3 (10.3)	105.8 (10.3)	103.0 (8.5)
Overall	116.0 (18.5)	111.0 (9.6)	124.9 (12.3)	118.3 (10.3)	110.6 (13.5)	108.2 (9.8)	106.0 (9.2)

Note. TOC SPACC = Test of Orthographic Competence Spelling Accuracy Composite (Mather et al., 2008); CTOPP PA = Comprehensive Test of Phonological Processing Phonological Awareness Composite (Wagner et al., 1999); PPVT = Peabody Picture Vocabulary Test IIIA (L. M. Dunn & Dunn, 1997); WRMT PC = Woodcock Reading Mastery Test – Revised Passage Comprehension (Woodcock, 1987); TOWRE PD = Test of Word Reading Efficiency Phonemic Decoding Efficiency (Torgesen et al., 1999); RAN NUM = Rapid Automatized Naming and Rapid Alternating Stimulus Tests Numbers (Wolf & Denckla, 2005); RAN LETT = Rapid Automatized Naming and Rapid Alternating Stimulus Tests Letters (Wolf & Denckla, 2005).

Table 2
Summary of within-group analysis results for word (W), pseudoword (PS), letter string (LS), and false font string (FF) comparisons

	N1 Amplitude				N1 Latency				P2 Amplitude				P2 Latency			
	W/PS	PS/LS	LS/FF	W/PS	PS/LS	LS/FF	W/PS	PS/LS	W/PS	PS/LS	LS/FF	W/PS	PS/LS	LS/FF	W/PS	PS/LS
Adults	w>ps**	---	---	---	---	---	---	---	---	---	---	w>ps*	---	---	---	---
Fifth Graders	---	---	ls>ff***	---	---	---	---	---	---	---	ff>ls***	---	ls>ps**	---	---	---
Fourth Graders	---	---	ls>ff**	ps>w**	---	---	w>ps**	ls>ps*	---	---	ff>ls***	---	---	---	---	---
Third Graders	---	---	ls>ff**	---	---	---	---	ls>ps*	---	---	ff>ls*	---	---	---	---	---

Note. The greater than symbol (>) indicates a larger peak for amplitude (for N1, a more negative peak; for P2, a more positive peak) and a later peak for latency.

* $p < .05$,

** $p < .01$,

*** $p < .001$ (highest level interaction)

Table 3

Summary of significant between-group analysis results (all possible pair-wise comparisons) for the semantic (words-pseudowords), phonological (pseudowords-letter strings), and orthographic (letter strings-false font strings) effects measured from the difference waves

	N1 Amplitude	P2 Amplitude
Semantic Effect	---	third graders/adults **
		fourth graders/adults **
		fifth graders/adults **
Phonological Effect	fifth graders/adults **	third graders/fourth graders *
		third graders/adults *
Orthographic Effect	third graders/adults *	third graders/adults **
	fourth graders/adults **	fourth graders/adults **
		fifth graders/adults **

* p .008 (Bonferroni-corrected),

** p .001