



## Brain-based individual difference measures of reading skill in deaf and hearing adults



Alison S. Mehravari<sup>a,\*</sup>, Karen Emmorey<sup>b</sup>, Chantel S. Prat<sup>a,c,d</sup>, Lindsay Klarman<sup>d</sup>, Lee Osterhout<sup>a,c</sup>

<sup>a</sup> Program in Neuroscience, University of Washington, Seattle, WA 98195, United States

<sup>b</sup> School of Speech, Language and Hearing Sciences, San Diego State University, San Diego, CA 98182, United States

<sup>c</sup> Department of Psychology, University of Washington, Seattle, WA 98195, United States

<sup>d</sup> Institute for Learning and Brain Sciences, University of Washington, Seattle, WA 98195, United States

### ARTICLE INFO

#### Keywords:

Event-related potential

Reading

N400

P600

Deaf

Individual differences

### ABSTRACT

Most deaf children and adults struggle to read, but some deaf individuals do become highly proficient readers. There is disagreement about the specific causes of reading difficulty in the deaf population, and consequently, disagreement about the effectiveness of different strategies for teaching reading to deaf children. Much of the disagreement surrounds the question of whether deaf children read in similar or different ways as hearing children. In this study, we begin to answer this question by using real-time measures of neural language processing to assess if deaf and hearing adults read proficiently in similar or different ways. Hearing and deaf adults read English sentences with semantic, grammatical, and simultaneous semantic/grammatical errors while event-related potentials (ERPs) were recorded. The magnitude of individuals' ERP responses was compared to their standardized reading comprehension test scores, and potentially confounding variables like years of education, speechreading skill, and language background of deaf participants were controlled for. The best deaf readers had the largest N400 responses to semantic errors in sentences, while the best hearing readers had the largest P600 responses to grammatical errors in sentences. These results indicate that equally proficient hearing and deaf adults process written language in different ways, suggesting there is little reason to assume that literacy education should necessarily be the same for hearing and deaf children. The results also show that the most successful deaf readers focus on semantic information while reading, which suggests aspects of education that may promote improved literacy in the deaf population.

### 1. Introduction

Reading can be difficult for many people who are deaf. Reading outcomes are generally poor for deaf individuals, but some deaf people do nonetheless achieve high levels of reading proficiency (Allen, 1986; Goldin-Meadow and Mayberry, 2001; Qi and Mitchell, 2012; Traxler, 2000). To improve the potential for all deaf individuals to read well, we must determine what allows some to become proficient readers, while many others struggle (Mayberry et al., 2011). Though there have been decades of research on the causes of reading difficulty in deaf individuals, conflicting results prevent a clear consensus (Allen et al., 2009; Mayberry et al., 2011; Mayer and Trezek, 2014; Paul et al., 2009).

The overarching question in this debate has been, do deaf children read in the same ways as hearing children, albeit with reduced access to sound, or do they read in qualitatively different ways (Hanson, 1989; Mayer and Trezek, 2014; Perfetti and Sandak, 2000; Wang et al., 2008;

Wang and Williams, 2014)? The answer has profound implications for education; if deaf children read proficiently in different ways from hearing children, they may learn best in different ways as well. One potential answer to this question is that proficient reading in deaf and hearing individuals is dependent on the same types of (neuro)cognitive capacities and skills. For example, proficient reading is often claimed to be fundamentally grounded in an individual's ability to compute, in real time, syntactic representations of sentence structure (Russell et al., 1976; Trybus and Buchanan, 1973). However, it is known that deaf children have considerable difficulty understanding syntactically non-canonical or complex structures, such as passive constructions (Power and Quigley, 1973) and relative clauses (Quigley et al., 1974). Faced with these realities, one pedagogical strategy has been to withhold syntactically complex sentences from pedagogical materials for the deaf (Shulman and Decker, 1980) and to gradually introduce a theory-motivated progression of grammatical structures. Even when this pedagogical approach is employed, however, literacy levels in deaf

\* Correspondence to: University of Washington, Box 351800, Seattle, WA 98195, United States.  
E-mail address: [amehrava@uw.edu](mailto:amehrava@uw.edu) (A.S. Mehravari).

students remain low.

An alternative approach is predicated on the possibility that deaf individuals might be able to achieve significant gains in literacy through different means. Specifically, this approach is motivated by evidence that deaf readers can make significant gains in literacy by focusing on semantic cues, even while remaining insensitive to key grammatical aspects of sentences (Cohen, 1967; Sarachan-Deily, 1980; Yurkowski and Ewoldt, 1986). The notion that deaf individuals rely more on meaning and less on syntax has previously been proposed (e.g. Ewoldt, 1981). The general claim is that relatively proficient deaf readers employ a semantically-driven predictive comprehension strategy that works most effectively when they are familiar with the semantic domain of the text (cf. Boudewyn et al., 2015; Pickering and Garrod, 2007). A deaf individual's reading proficiency would then be a function of her or his ability to extract the intended meanings from sentences and larger units of text, rather than the ability to construct precise syntactic representations of sentences that the individual reads.

Prior research on deaf literacy has primarily used behavioral tasks, such as reaction time measures and standardized reading tests. While much has been learned from this work, the field lacks detailed information on how the brains of deaf readers process written language in real time. Such data would help identify the neurocognitive mechanisms by which deaf people read successfully, by providing critical information about which aspects of language a deaf reader is sensitive to when processing written text. Event-related potentials (ERPs), recorded while a subject reads, provide a unique way to better understand how deaf readers read. ERPs are especially well-suited for studying reading for two reasons. First, ERPs respond to specific aspects of language. Grammatical errors in sentences typically elicit a positive-going component starting around 500–600 ms in an ERP response (the P600 effect), while semantic errors in sentences elicit a negative-going component peaking around 400 ms (the N400 effect) (Kaan et al., 2000; Kutas and Federmeier, 2000; Kutas and Hillyard, 1984, 1980; Osterhout et al., 1994; Osterhout and Holcomb, 1992). When a word in a sentence is anomalous in both grammar and semantics, both effects are elicited in a nearly additive fashion (Osterhout and Nicol, 1999). Second, mounting evidence links ERP response variability to individual differences in linguistic abilities, and the size of an individual's ERP response can be viewed as an index of their sensitivity to a particular type of information (McLaughlin et al., 2010; Newman et al., 2012; Ojima et al., 2011; Pakulak and Neville, 2010; Rossi et al., 2006; Tanner et al., 2014, 2013; Weber-Fox et al., 2003; Weber-Fox and Neville, 1996). Prior ERP research in deaf readers (Skotara et al., 2012, 2011) has not explored individual differences in participants' responses, nor how ERP responses relate to subjects' reading skill. The answers to these questions have the potential to shed light on how some deaf individuals read more proficiently than others, and whether proficient deaf and hearing individuals read in similar or different ways.

In the present study, we used the systematic differences in individuals' ERP responses to better understand similarities and differences in how deaf and hearing adults read. Participants read sentences with semantic, grammatical, and simultaneous semantic-grammatical errors while ERPs were recorded. We compared the magnitude of participants' N400 and P600 responses to their performance on a standardized reading comprehension test. Because many factors contribute to how well someone reads, we used multiple regression models to control for potentially confounding variables. If deaf and hearing participants read proficiently using similar strategies or mechanisms, we would expect to see similar relationships between reading skill and sensitivity to semantic and grammatical information, as reflected by N400 and P600 size, in both groups. However, if deaf and hearing participants showed different relationships between reading skill and ERP response sizes, it would indicate that the two groups were reading proficiently using different strategies or mechanisms.

## 2. Materials and methods

### 2.1. Participants

Participants were 42 deaf (27 female) and 42 hearing (27 female) adults. The number of participants needed was determined via power analysis (see Section 2.6). All deaf participants were severely or profoundly deaf (hearing loss of 71 dB or greater, self-reported), except for one participant with profound (95 dB) hearing loss in the left ear and moderate (65 dB) hearing loss in the right ear. All deaf participants lost their hearing by the age of two. Thirty-three of the 42 deaf participants reported being deaf from birth. Three of the remaining deaf participants reported that it was likely they were deaf from birth but had not been diagnosed until later (still by age two). The final six deaf participants reported clear causes of deafness that occurred after birth but before age two. All deaf participants reported having worn hearing aids in one or both ears at some point in life; 22 participants still wore hearing aids, 5 participants only wore them occasionally or in specific circumstances, and 15 participants no longer wore them. One participant, age 28.5 years, had a unilateral cochlear implant, but it was implanted late in life (at age 25.8 years) and the participant reported rarely using it. Other than that, individuals with cochlear implants did not take part in this study. The average age of deaf participants was 38.6 years (range: 19–62 years) and the average age of hearing participants was also 38.6 years (range: 19–63 years); there was no significant difference in the ages of the two groups ( $t = -0.011$ ,  $p = 0.991$ ). All participants had normal or corrected-to-normal vision, except for one deaf participant with reduced peripheral vision due to Usher syndrome. The deaf participant with Usher syndrome did not have any difficulty in completing any of the study procedures. No participants had any history of significant head injury or epilepsy. While most participants were right-handed, two of the deaf participants and seven of the hearing participants were left-handed, as assessed by an abridged version of the Edinburgh Handedness Inventory (Oldfield, 1971). One deaf participant and one hearing participant reported being ambidextrous.

All participants filled out a detailed life history questionnaire that asked about their language background and education history. Hearing participants had completed an average of 17.3 years of education (standard deviation 2.5 years) and deaf participants an average of 16.5 years (standard deviation 2.1 years); there was not a significant difference in years of education between the two groups ( $t = -1.611$ ,  $p = 0.111$ ). The first language of all hearing participants was English, and English was the only language that had been used in their homes while they were growing up. Deaf participants came from a wide variety of language backgrounds, and were asked in detail about their spoken and manual/signed language exposure and use throughout their life. On a 1–7 scale, where 1 = all oral communication, 7 = all manual/signed communication, and 4 = equal use of both, deaf participants were asked about their method(s) of communication at the following points in their life: a) overall while they were growing up (incorporating language use both in school and in the home), and b) at the current point in time. Importantly, a '7' on this scale did not distinguish between the use of American Sign Language (ASL) and manually coded forms of English (i.e., Signed Exact English, SEE, or Pidgin Sign English, PSE). Participants also wrote descriptions of their language use at each of these points in time, which served two purposes. First, it allowed us to confirm that the participants' ratings on the 1–7 scales generally corresponded to what they described – and if the ratings did not seem to correspond, the participant was asked for clarification. Second, these descriptions allowed us to distinguish between participants who grew up using and being exposed to ASL versus those who grew up using and being exposed to forms of Manually Coded English. The language backgrounds of the deaf participants were extremely diverse. On both

scales, responses ranged from 1 to 7. The average response for language use growing up was 4.0 (standard deviation 2.1) indicating a nearly equal mixture of participants who grew up in a more spoken language environment versus a more manual/signed language environment. Though many participants reported using ASL, SEE, and/or PSE at some point while growing up, only four participants had deaf parents or other family members who communicated with them in fluent ASL from birth or a young age. The average response on the scale for current language use was 5.4 (standard deviation 1.9), indicating that at the current point in time, there was greater use of manual/signed communication than spoken communication by the deaf participants. All experimental procedures were approved by the University of Washington Institutional Review Board, and informed consent was obtained from all participants.

## 2.2. Standardized reading comprehension

Standardized reading comprehension was measured using the Word and Passage Comprehension sections of the Woodcock Reading Mastery Tests, Third Edition (WRMT-III), Form A (Woodcock, 2011). Though the test typically requires verbal answers, deaf participants were instructed to respond in their preferred method of communication; this test is often administered in this way to deaf individuals (Easterbrooks and Huston, 2008; Kroese et al., 1986; Spencer et al., 2003). When the English word being signed was unclear, deaf participants were asked to fingerspell the English word they were intending. The maximum possible total raw score was 124. The average score of deaf participants (mean = 82.8, standard deviation = 18.8) was significantly lower than the average score of hearing participants (mean = 101.7, standard deviation = 12.4;  $t = -5.449$ ,  $p < 0.001$ ). However, the highest scoring deaf participants scored as high as the highest scoring hearing participants. Deaf participants ranged in score from 40 to 115. Hearing participants effectively ranged in score from 80 to 116; a single hearing participant had a score of 46 (Fig. 1). Twenty-four of the 42 deaf participants had scores within the effective range of hearing participants. Thus, while the deaf participants had a much wider range of scores, the highest scoring participants in both groups were at the same reading level. There was no significant correlation between deaf participants' language background while growing up and their standardized reading comprehension scores ( $r = -0.168$ ,  $p = 0.288$ ).

In addition to raw scores, WRMT-III scores can be standardized based on each participant's age, and a percentile rank determined. When raw scores were standardized by participant age, the hearing participants effectively ranged in percentile for their ages from the 16th through the 98th percentile (the single hearing participant with the

notably lower reading test score was in the 0.4th percentile). The average score of the hearing participants corresponded to a score in the 75th percentile. Norming scores for deaf participants is problematic because they are being compared to norms for hearing readers, but based on these norms, the deaf participants ranged from the 0.2nd through the 96th percentile, with the group's average score corresponding to a score in the 30th percentile of normed hearing readers.

## 2.3. Speechreading skill

Standardized speechreading skill was measured in all participants, because speechreading skill is often highly correlated with reading proficiency in deaf individuals (Kyle and Harris, 2010; Mohammed et al., 2006). Speechreading skill was measured with the National Technical Institute for the Deaf's "Speechreading: CID Everyday Sentences Test" (Sims, 2009). In this test, participants watch a video, without sound, of a speaker saying 10 sentences. After each sentence, participants write what they believe the speaker said. The test is scored based on the number of correct words identified. The test comes with 10 lists of 10 sentences each; participants were evaluated on List 1 and List 6, and the average score from the two lists was taken. For all possible raw scores in each list, there was also a corrected score, used to normalize slight difficulty differences between lists. Lists 1 and 6 were used because the relationship between their raw and corrected scores was similar. Corrected scores are reported here. The maximum average speechreading test score was 100. Overall, the group of deaf participants performed significantly better on the speechreading test than the hearing participants (deaf: mean = 59.3, standard deviation = 21.4; hearing: mean = 43.7, standard deviation = 13.6;  $t = 3.984$ ,  $p < 0.001$ ). However, there was large variability within each group. Deaf participants' scores ranged from 13.5 to 89.5; hearing participants' scores ranged from 15 to 73.5. As expected, in deaf participants, higher speechreading test score was strongly correlated with higher standardized reading comprehension score ( $r = 0.702$ ,  $p < 0.001$ ). For hearing participants, there was no significant relationship between speechreading test score and standardized reading comprehension score ( $r = 0.173$ ,  $p = 0.273$ ).

## 2.4. ERP stimuli

Sentence stimuli were 120 sentence quadruplets in a fully crossed 2 (semantic correctness) by 2 (grammaticality) design. Sentences were either grammatically correct or contained a subject-verb agreement violation, and were also either semantically well-formed or contained a semantic anomaly. This resulted in four types/conditions of sentences: 1) well-formed sentences, 2) sentences with a grammatical violation alone, 3) sentences with a semantic violation alone, and 4) sentences with a double violation – a simultaneous error of grammar and semantics. All violations, semantic and/or grammatical, occurred on the critical word in the sentence. Critical words were either verbs in their base/uninflected form (e.g., belong) or in their third person singular present tense form (e.g., belongs). Each sentence condition had two versions – one with the critical word in the base form and one in the –s form (see Table 1 for an example). The stimuli were designed in this way so that the singular/plural status of the subject of the sentence (the noun preceding the critical verb) could not be used to predict whether or not the sentence would contain an error. Thus, while there were four conditions of sentences (well-formed, grammatical violation, semantic violation, double violation), there were eight versions of each sentence (Table 1). Sentences that were well-formed or contained grammatical violations used a set of 120 unique verbs (in both their base and –s forms), and sentences that contained a semantic or double violation used a different set of 120 unique verbs. The two sets of verbs were chosen so that the average written word-form log frequency (from the CELEX2 database (Baayen et al., 1995)) of the verbs in the two sets was not significantly different (well-formed/

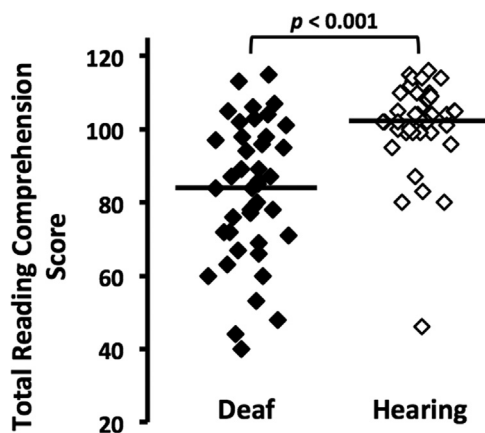


Fig. 1. Distribution of Woodcock Reading Mastery Test III Reading Comprehension scores (sum of Word and Passage Comprehension scores) for deaf and hearing participants. Horizontal lines indicate the average score for each group. Hearing participants scored significantly higher than deaf participants ( $t = -5.449$ ,  $p < 0.001$ ).

**Table 1**  
Example sentence stimuli.

Condition	Sentences
Well-formed	The huge house still <u>belongs</u> to my aunt. The huge houses still <u>belong</u> to my aunt.
Grammatical violation	The huge houses still <u>belongs</u> to my aunt. The huge house still <u>belong</u> to my aunt.
Semantic violation	The huge house still <u>listens</u> to my aunt. The huge houses still <u>listen</u> to my aunt.
Double grammatical & semantic violation	The huge houses still <u>listens</u> to my aunt. The huge house still <u>listen</u> to my aunt.

The critical word for ERP averaging is underlined.

grammatical verbs average frequency = 0.59, semantic/double verbs average frequency = 0.56,  $t = 0.441$ ,  $p = 0.659$ ). The calculations accounted for the average frequency of both the base and –s forms of the verbs. Due to an oversight, the two sets of verbs were not matched in word length; verbs in the well-formed/grammatical condition had an average length of 6.23 letters (standard deviation = 1.88) while the verbs in the semantic/double condition had an average length of 5.68 letters (standard deviation = 1.65;  $t = 3.405$ ,  $p < 0.001$ ). Despite this confound, prior work has shown that during sentence reading, differences in word length modulate the latency of ERP components that appear earlier than the N400, not the N400 itself (Hauk and Pulvermüller, 2004; King and Kutas, 1995; Osterhout et al., 1997). Thus, we would not expect these small differences in word length to influence our components of interest, the N400 and P600, during sentence reading.

The eight versions of each sentence were distributed across eight experimental lists, such that each list only had one version of each sentence. There were 15 sentences from each of the eight versions in each list, and thus there were 30 sentences per condition in each list. Each list contained an additional 60 filler sentences, all of which were grammatically correct. In total, each list contained 180 sentences. The sentence order in each list was randomized, and lists were divided into 3 blocks of 60 sentences each. Participants were pseudorandomly assigned one of the lists.

## 2.5. Procedure

Participants took part in three sessions, each of which lasted no more than two hours. An interpreter was present for all sessions with deaf participants unless the participant said they did not need an interpreter. In the first session, participants completed all background questionnaires and the speechreading and reading comprehension tests. ERPs reported here were recorded during either the second or third session. ERPs from an experiment not reported here were recorded in the other session. Half of the participants saw the stimuli reported here in the second session, and half saw the stimuli in the third session. Participants were pseudorandomized to determine the order of their ERP sessions. During ERP recording, participants sat in a comfortable recliner in front of a CRT monitor. Participants were instructed to relax and minimize movements and eye blinks while silently reading the stimuli in their minds. Each sentence trial consisted of the following events: a blank screen for 1000 ms, followed by a fixation cross, followed by a stimulus sentence presented one word at a time. The fixation cross appeared on the screen for 500 ms followed by a 400 ms interstimulus interval (ISI). Each word of the sentence appeared on the screen for 600 ms followed by a 200 ms ISI. The presentation rate used was slower than the rate typically used in ERP studies of first language users; it is standard procedure for ERP studies of a second language to use a slower presentation rate for all participants, both the first and second language users (Foucart and Frenck-Mestre, 2012, 2011; Tanner et al., 2014, 2013). After the final word of the sentence, there was a

1000 ms blank screen, followed by a “yes/no” prompt. Participants were instructed to give a sentence acceptability judgment at the “yes/no” prompt, where “yes” was the response for sentences that were correct in all ways and “no” was the response for sentences that contained any kind of error. Participants were instructed to make their best guess if they were not sure if the sentence contained an error. The “yes/no” prompt remained on the screen until participants responded; as soon as a response was given, presentation of the next sentence began. Participants were pseudorandomly assigned to use either their left or right hand for the “yes” response.

## 2.6. Data acquisition and analysis

Continuous EEG was recorded from 19 tin electrodes attached to an elastic cap (Electro-cap International) in accordance with the 10–20 system (Jasper, 1958). Eye movements and blinks were monitored by two electrodes, one placed beneath the left eye and one placed to the right of the right eye. Electrodes were referenced to an electrode placed over the left mastoid. EEG was also recorded from an electrode placed on the right mastoid to determine if there were experimental effects detectable on the mastoids. No such effects were found. EEG signals were amplified with a bandpass filter of 0.01–40 Hz (–3 dB cutoff) by an SAI bioamplifier system. ERP waveforms were filtered offline below 30 Hz. Impedances at scalp and mastoid electrodes were held below 5 k $\Omega$  and below 15 k $\Omega$  at eye electrodes. Continuous analog-to-digital conversion of the EEG and stimulus trigger codes was performed at a sampling frequency of 200 Hz. ERPs, time-locked to the onset of the critical word in each sentence, were averaged offline for each participant at each electrode site in each condition. Trials characterized by eye blinks, excessive muscle artifact, or amplifier blocking were not included in the averages. ERPs were quantified as mean amplitude within a given time window. All artifact-free trials were included in the ERP analyses; 5.3% of trials from deaf participants and 4.1% of trials from hearing participants were rejected. The rejection rate was not significantly different between the two groups ( $t(334) = 1.728$ ,  $p = 0.085$ ). Within each group, there was no significant difference in the rejection rates of the four different sentence conditions (deaf:  $F(3,164) = 0.541$ ,  $p = 0.655$ ; hearing:  $F(3,164) = 0.500$ ,  $p = 0.683$ ). In accordance with prior literature and visual inspection of the data, the following time windows were chosen for analysis: 300–500 ms (N400), and 500–900 ms (P600), relative to a 100 ms prestimulus baseline. Differences between sentence conditions were analyzed using a repeated-measure ANOVA with two levels of semantic correctness (semantically plausible, semantic violation) and two levels of grammaticality (grammatical, ungrammatical). ANOVA analyses included deaf and hearing participants in the same model, using group as a between-subjects factor, unless otherwise specified. Data from midline (Fz, Cz, Pz), medial (right hemisphere: Fp2, F4, C4, P4, O2; left hemisphere: Fp1, F3, C3, P3, O1), and lateral (right hemisphere: F8, T8, P8; left hemisphere: F7, T7, P7) electrode sites were treated separately in order to identify topographic and hemispheric differences. ANOVAs on mid-

line electrodes included electrode as an additional within-subjects factor (three levels), ANOVAs on medial electrodes included hemisphere (two levels) and electrode pair (five levels) as additional within-subjects factors, and ANOVAs over lateral electrodes included hemisphere (two levels) and electrode pair (three levels) as additional within-subjects factors. The Greenhouse-Geisser correction for inhomogeneity of variance was applied to all repeated measures on ERP data with greater than one degree of freedom in the numerator. In such cases, the corrected  $p$ -value is reported. In instances where follow-up contrasts were required, a familywise Bonferroni correction was used based on the number of follow-up contrasts being performed, and the adjusted alpha level is reported.

Individual differences analyses were also planned, following prior work (Tanner et al., 2014; Tanner and Van Hell, 2014). For all individual differences analyses, the size, or effect magnitude, of a particular ERP response was calculated. An ERP effect magnitude refers to the “size” of a particular response as compared to the relevant control response. The ERP responses of interest were the N400 response to semantic violations (either alone or as part of a double violation), and the P600 response to grammatical violations (again, either alone or as part of a double violation). To compute the effect magnitude of each of these responses, for each of the four sentence conditions, we calculated participants’ mean ERP amplitudes in a central-posterior region of interest (electrodes C3, Cz, C4, P3, Pz, P4, O1, O2) where N400 and P600 effects are typically the largest (Tanner et al., 2014; Tanner and Van Hell, 2014). For sentences with semantic violations alone, semantic violation N400 effect magnitude was calculated as the mean activity in the well-formed minus semantic conditions between 300 and 500 ms. For sentences with grammatical violations alone, grammatical violation P600 effect magnitude was calculated as the mean activity in the grammatical violation minus well-formed conditions between 500 and 900 ms. For sentences with double violations, double violation N400 and P600 effect magnitudes were calculated in similar ways, comparing activity in the double violation and well-formed conditions. Thus, these calculations produced four ERP effect magnitudes: a) semantic violation N400, b) grammatical violation P600, c) double violation N400, and d) double violation P600.

We explored the relationship between these ERP effect magnitudes and participants’ standardized reading comprehension scores (WRMT-III Reading Comprehension scores, described earlier) in two ways. For all of these analyses, separate analyses were used for the deaf and hearing participants. First, we examined the simple correlations between the four ERP effect magnitudes and participants’ standardized reading score. However, since many factors may lead to variation in a person’s reading skill, we employed theoretically-driven multiple regression models to account for some of this variation and clarify the relationship between these ERP measures and participants’ standardized reading skill. As with the simple correlations, separate models were used for deaf and hearing participants, for two reasons. First, language background was included as a predictor for deaf participants, which was not a measure that existed for hearing participants. Second, the relationships between predictors and outcomes was potentially different between the groups. The outcome measure of all multiple regression models was participants’ standardized reading comprehension score. Predictors for each of the multiple regression models were of two kinds: a) background measures known to influence reading skill, and b) ERP measures of interest. Sequential predictor entry was used such that the background predictors were entered into the regression model first, and the ERP measures were entered second, allowing us to determine how much additional variance in standardized reading comprehension score the ERP measures accounted for. The background measures included as predictors for all participants, deaf and hearing, in each model were 1) years of education, and 2) standardized speechreading score. An additional background measure predictor was used in models with deaf participants: 3) score on the 1–7 scale used to assess language use while growing up. The four ERP effect

magnitudes were split into two groups of predictors: 1) semantic violation N400 and grammatical violation P600, and 2) double violation N400 and P600. These two sets of ERP effect magnitudes were used as the final predictors in two separate multiple regression models (each with the same background measure predictors) for each group of participants. The design of these models was motivated by prior ERP language research that has shown that the N400 and P600 index different neural language processing streams (Kim and Osterhout, 2005; Kuperberg, 2007; Tanner, 2013). These planned multiple regression models were used as the basis for power analysis used to determine the number of participants needed for this research study. With the assumption of a large effect size, a multiple regression model with five predictors and an alpha level  $\alpha = 0.05$  requires 42 participants to obtain 80% power (Cohen, 1992). Thus, we recruited 42 deaf and 42 hearing participants for this study.

### 3. Results

#### 3.1. End-of-sentence acceptability judgment task

To analyze responses to the end-of-sentence acceptability judgment task,  $d'$  scores for the judgment task were calculated for each participant for each of the three types of sentence violations: grammatical violations, semantic violations, and double violations, as well as an overall  $d'$  score (Fig. 2). Four  $t$ -tests (corrected alpha  $\alpha = 0.0125$ ) were used to compare deaf and hearing participants’ performance across the three sentence conditions and overall. For all three types of sentence violations, and overall, hearing participants were better at discriminating sentence violations than deaf participants ( $t(82)$  ranged from 4.070 to 5.111, all  $p$ 's < 0.001). For both deaf and hearing participants, there was a significant difference in  $d'$  score between the different sentence conditions (deaf:  $F(2,123) = 23.982$ ,  $p < 0.001$ ; hearing:  $F(2,123) = 16.202$ ,  $p < 0.001$ ). Tukey post-hoc tests determined that in both groups,  $d'$  scores were lower for sentences with grammatical violations than for sentences with semantic or double violations (all  $p$ 's < 0.001). For both groups, there was no significant difference in the  $d'$  scores between sentences with semantic and double violations ( $p = 0.796$ ). In summary, the hearing participants were better at discriminating sentence acceptability than deaf participants across all types of sentence violations, but both groups showed similar  $d'$  patterns for the different types of sentence violations. For both deaf and hearing participants, all  $d'$  scores were significantly positively correlated with participants’ reading comprehension score; as  $d'$  scores increased, so did reading comprehension scores ( $R^2$  values in Table 2). Of particular note

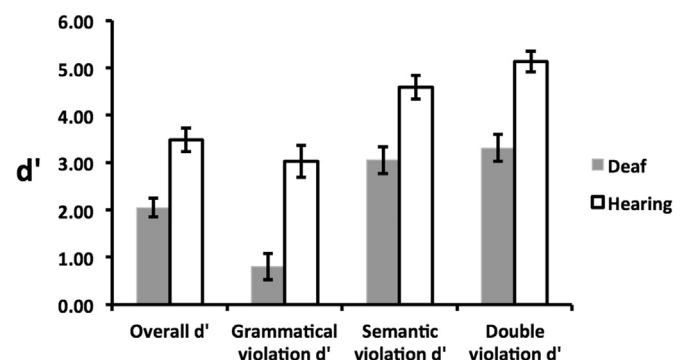


Fig. 2. End-of-sentence acceptability judgment  $d'$  scores for deaf and hearing participants. Overall  $d'$  refers to performance across all types of sentence violations. Hearing participants had higher  $d'$  scores than deaf participants for all types of sentence violations and overall ( $t(82)$  ranged from 4.070 to 5.111, all  $p$ 's < 0.001, corrected alpha  $\alpha = 0.0125$ ). For both deaf and hearing participants,  $d'$  scores differed between conditions (deaf:  $F(2,123) = 23.982$ ,  $p < 0.001$ ; hearing:  $F(2, 123) = 16.202$ ,  $p < 0.001$ ). In both groups of participants,  $d'$  scores were lower for sentences with grammatical violations than for sentences with semantic or double violations (Tukey post-hoc tests, all  $p$ 's < 0.001). Error bars represent one standard error of the mean (SEM).

**Table 2**  
Coefficients of determination ( $R^2$ ) for the relationship between  $d'$  scores and standardized reading comprehension score.

	Standardized reading comprehension score	
	Deaf	Hearing
Overall $d'$	0.50 <sup>***</sup>	0.22 <sup>**</sup>
Grammatical violation $d'$	0.12 <sup>*</sup>	0.32 <sup>***</sup>
Semantic violation $d'$	0.52 <sup>***</sup>	0.16 <sup>**</sup>
Double violation $d'$	0.56 <sup>***</sup>	0.29 <sup>***</sup>

All correlation coefficients ( $r$ ) are positive.

\*  $p < 0.05$ .

\*\*  $p < 0.01$ .

\*\*\*  $p < 0.001$ .

**Table 3**  
Average percent of end-of-sentence acceptability judgments that were answered correctly in each condition.

	Deaf	Hearing
Well-formed sentences	86.4%	91.6%
Grammatical violation sentences	33.0%	73.3%
Semantic violations sentences	82.6%	96.0%
Double violation sentences	86.6%	97.9%

is that for deaf participants, while the grammatical violation  $d'$ , semantic violation  $d'$ , and double violation  $d'$  all explain a significant amount of variance in reading comprehension score, the semantic and double violation  $d'$  scores each explain more than four times as much variance in reading comprehension compared to the grammatical violation  $d'$  score. Finally, the percent correct of end-of-sentence acceptability judgments are given in Table 3.

### 3.2. Grand mean results

For both the deaf and hearing groups, grand mean ERP results for semantic violations can be seen in Fig. 3, for grammatical violations in Fig. 4, and for double violations in Fig. 5. Data from the 300–500 ms (N400) and 500–900 ms (P600) time windows were analyzed separately. All trials (with both correct and incorrect acceptability judgments) were included in the ERP analysis.

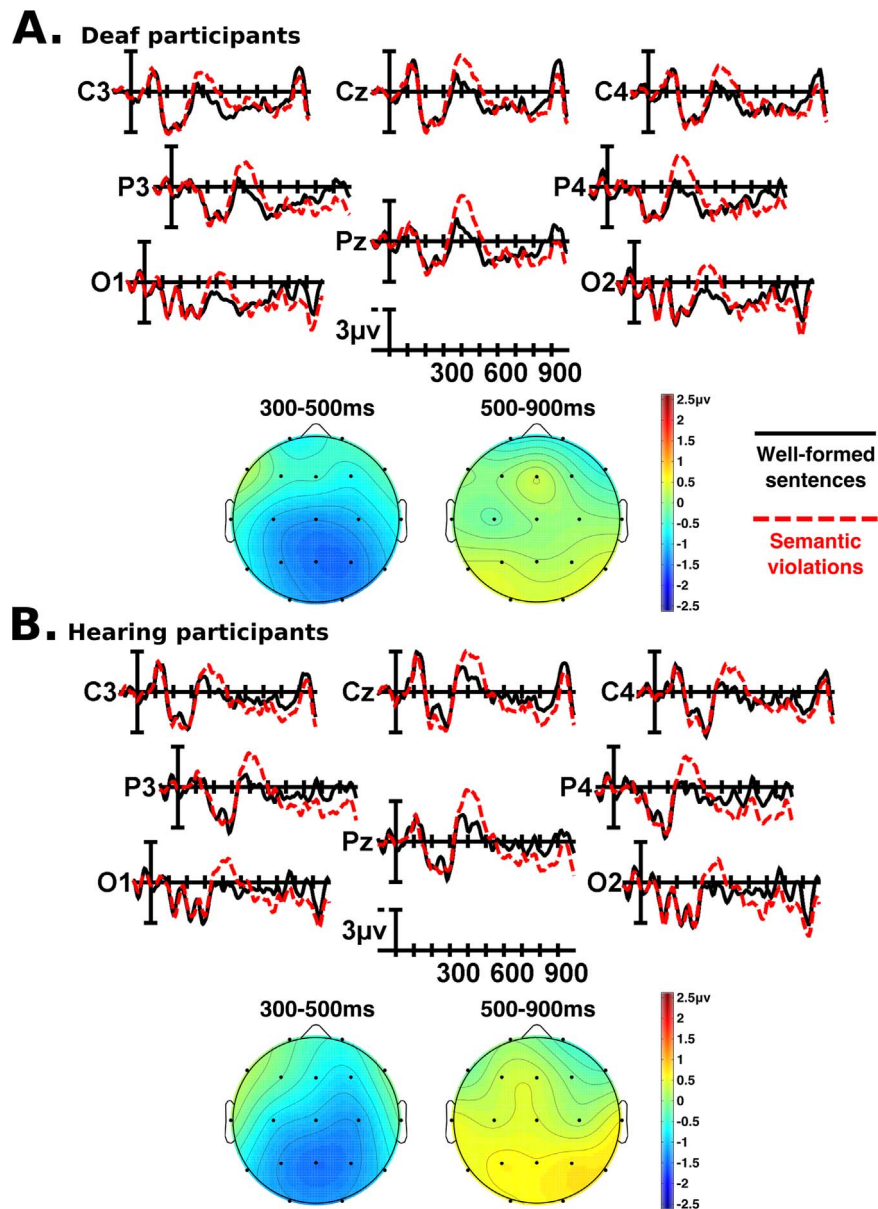
#### 3.2.1. Grand mean results: N400 (300–500 ms) time window

Visual inspection of the grand mean ERP waveforms showed that relative to well-formed sentences, sentences with a semantic violation (Fig. 3) or a double violation (Fig. 5) elicited a widely-distributed negativity with a posterior maximum between approximately 300 and 500 ms (an N400 effect) in both deaf and hearing participants. There did not appear to be differences in this time window between well-formed sentences and sentences with grammatical violations. Statistical analyses confirmed these observations. Reflecting the widespread N400 in response to semantic and double violations, there was a main effect of semantic correctness (midline:  $F(1,82) = 58.026$ ,  $p < 0.001$ ; medial:  $F(1,82) = 55.105$ ,  $p < 0.001$ ; lateral:  $F(1,82) = 21.637$ ,  $p < 0.001$ ). This effect was strongest over posterior electrodes, (semantic correctness  $\times$  electrode interaction: midline:  $F(2,164) = 25.371$ ,  $p < 0.001$ ; medial:  $F(4,328) = 18.338$ ,  $p < 0.001$ ; lateral:  $F(2,164) = 25.269$ ,  $p < 0.001$ ). There was also an interaction between semantic correctness and grammaticality at midline and lateral electrode sites (semantic correctness  $\times$  grammaticality interaction: midline:  $F(1,82) = 4.239$ ,  $p = 0.043$ ; lateral:  $F(1,82) = 4.039$ ,  $p = 0.048$ ). Follow-up contrasts (corrected alpha  $\alpha = 0.025$ ) indicated that this interaction was driven by two factors. First, the interaction was driven by the lack of N400 change elicited by grammatical violations alone, as compared to the much larger N400 elicited by double violations (midline:  $F(1,83) = 45.666$ ,

$p < 0.001$ ; lateral:  $F(1,83) = 28.134$ ,  $p < 0.001$ ). Second, at lateral electrode sites the N400 response to double violations was significantly larger than the N400 response to semantic violations alone (midline:  $F(1,83) = 3.588$ ,  $p = 0.062$ ; lateral:  $F(1,83) = 7.132$ ,  $p = 0.009$ ). Deaf and hearing participants did not differ significantly in any of the analyses reported above. Small hemispheric differences in the N400 response at lateral electrodes were observed (semantic correctness  $\times$  hemisphere: lateral:  $F(1,82) = 4.384$ ,  $p = 0.039$ ), which were driven by an interaction between the groups (semantic correctness  $\times$  hemisphere  $\times$  group interaction: lateral:  $F(1,82) = 4.041$ ,  $p = 0.048$ ). Follow-up ANOVAs with the deaf and hearing participants in separate models (corrected alpha  $\alpha = 0.025$ ) indicated that this interaction was due to deaf participants having larger N400s in the right hemisphere than in the left (deaf participants: semantic correctness  $\times$  hemisphere interaction: lateral:  $F(1,41) = 7.114$ ,  $p = 0.011$ ) while hearing participants did not show any significant hemispheric differences. In summary, in the 300–500 ms time window, both deaf and hearing participants showed large N400 responses to sentences with semantic or double violations, and no significant response within this time window to sentences with grammatical violations.

#### 3.2.2. Grand mean results: P600 (500–900 ms) time window

Visual inspection of the grand mean ERP waveforms showed marked differences between hearing and deaf participants in the 500–900 ms time window. In hearing participants, relative to well-formed sentences, sentences with a grammatical violation (Fig. 4) or a double violation (Fig. 5) elicited a large, widely-distributed positivity beginning around 500 ms with a posterior maximum (a P600 effect). Deaf participants showed only weak evidence of a P600 in response to sentences with grammatical or double violations (Figs. 4 and 5). Both groups of participants showed a small positivity in this time window in response to sentences with semantic violations (Fig. 3). Statistical analyses confirmed these observations. Across all participants, there was a main effect of grammaticality (midline:  $F(1,82) = 10.181$ ,  $p = 0.002$ ; medial:  $F(1,82) = 10.414$ ,  $p = 0.002$ ; lateral:  $F(1,82) = 3.779$ ,  $p = 0.055$ ). The interpretation of this main effect is conditional upon the fact that there was an interaction between group (deaf vs. hearing) and grammaticality (grammaticality  $\times$  group interaction: midline:  $F(1,82) = 10.413$ ,  $p = 0.002$ ; medial:  $F(1,82) = 8.440$ ,  $p = 0.005$ ; lateral:  $F(1,82) = 4.108$ ,  $p = 0.046$ ). Follow-up ANOVAs with the deaf and hearing participants in separate models (corrected alpha  $\alpha = 0.025$ ) showed that for hearing participants, sentences containing a grammatical or double violation elicited a robust P600 relative to well-formed sentences (hearing participants: grammaticality main effect: midline:  $F(1,41) = 14.637$ ,  $p < 0.001$ ; medial:  $F(1,41) = 13.194$ ,  $p = 0.001$ ; lateral:  $F(1,41) = 6.156$ ,  $p = 0.017$ ), whereas there was no main effect of grammaticality for deaf participants alone. Across all participants, the main effect of grammaticality was largest over posterior electrodes (grammaticality  $\times$  electrode interaction: midline:  $F(2,164) = 27.520$ ,  $p < 0.001$ ; medial:  $F(4,328) = 21.806$ ,  $p < 0.001$ ; lateral:  $F(2,164) = 18.514$ ,  $p < 0.001$ ), but there were differences between the two groups in this interaction (grammaticality  $\times$  electrode  $\times$  group interaction: midline:  $F(2,164) = 11.380$ ,  $p < 0.001$ ; medial:  $F(4,328) = 8.087$ ,  $p = 0.001$ ; lateral:  $F(2,164) = 7.200$ ,  $p = 0.005$ ). Follow-up ANOVAs with the deaf and hearing participants in separate models (corrected alpha  $\alpha = 0.025$ ) indicated that in hearing participants, the P600 elicited by sentences with grammatical and double violations was largest at posterior electrodes (hearing participants: grammaticality  $\times$  electrode interaction: midline:  $F(2,82) = 32.445$ ,  $p < 0.001$ ; medial:  $F(4,164) = 22.892$ ,  $p < 0.001$ ; lateral:  $F(2,82) = 20.868$ ,  $p < 0.001$ ), whereas this interaction was not present in the deaf participants alone. Deaf and hearing participants showed different hemispheric effects of grammaticality (grammaticality  $\times$  hemisphere  $\times$  group interaction: medial:  $F(1,82) = 4.423$ ,  $p = 0.039$ ; lateral:  $F(1,82) = 5.110$ ,  $p = 0.026$ ). Follow-up ANOVAs with the deaf and hearing participants in separate models (corrected alpha  $\alpha = 0.025$ ) showed that in deaf participants,



**Fig. 3.** Grand mean ERP waveforms and scalp topographies for sentences with semantic violations alone as compared to well-formed sentences, for deaf (Panel A) and hearing (Panel B) participants. Full ERP waveforms are shown for the central-posterior region of interest (electrodes C3, Cz, C4, P3, Pz, P4, O1, O2). Onset of the critical word in the sentence is indicated by the vertical bar. Calibration bar shows 3  $\mu$ V of activity; each tick mark represents 100 ms of time. Negative voltage is plotted up. Topographic map scale is in microvolts.

there were hemispheric differences in the effect of grammaticality at medial electrodes (deaf participants: grammaticality  $\times$  hemisphere interaction: medial:  $F(1,41) = 6.282, p = 0.016$ ). Hemispheric interactions at lateral electrode sites did not reach significance, and there were no significant hemispheric interactions with grammaticality in the hearing participants. While there was no main effect of semantics in the 500–900 ms time window, all participants had a small P600-like response at posterior medial and lateral electrodes to sentences containing a semantic violation, whether it be a semantic violation alone or a double violation (semantic correctness  $\times$  electrode interaction: medial:  $F(4,328) = 5.735, p = 0.006$ ; lateral:  $F(2,164) = 14.157, p < 0.001$ ).

Finally, there was also an interaction between semantic correctness and grammaticality in all participants (semantic correctness  $\times$  grammaticality interaction: midline:  $F(1,82) = 12.629, p = 0.001$ ; medial:  $F(1,82) = 10.413, p = 0.002$ ; lateral:  $F(1,82) = 9.164, p = 0.003$ ). Follow-up contrasts comparing double violation sentences with semantic and grammatical violation sentences (corrected alpha  $\alpha = 0.025$ ) indicated that this interaction was driven by two factors. First, the P600 elicited

by sentences with grammatical violations alone was significantly larger than the P600 elicited by sentences with double violations (midline:  $F(1,83) = 9.597, p = 0.003$ ; medial:  $F(1,83) = 9.484, p = 0.003$ ; lateral:  $F(1,83) = 5.268, p = 0.024$ ). This is likely due to the N400 response also elicited by double violations, the presence of which pulls the P600 “up” – more negative/less positive. While deaf participants did not display widely-distributed P600s in response to grammatical or double violations, this trend is still present in their responses when comparing the relative voltages of the 500–900 ms time window between the response to grammatical violations alone versus the response to double violations. Second, at posterior electrodes, the P600 elicited by semantic violations alone was smaller than the P600 elicited by double violations, while this trend reversed at anterior electrodes (electrode interaction: midline:  $F(2,166) = 20.109, p < 0.001$ ; medial:  $F(4,332) = 9.582, p < 0.001$ ; lateral:  $F(2,166) = 9.508, p = 0.001$ ). In summary, in the 500–900 ms time window, hearing participants displayed large P600s in response to sentences with grammatical or double violations, while deaf participants did not show evidence of significant P600s as a

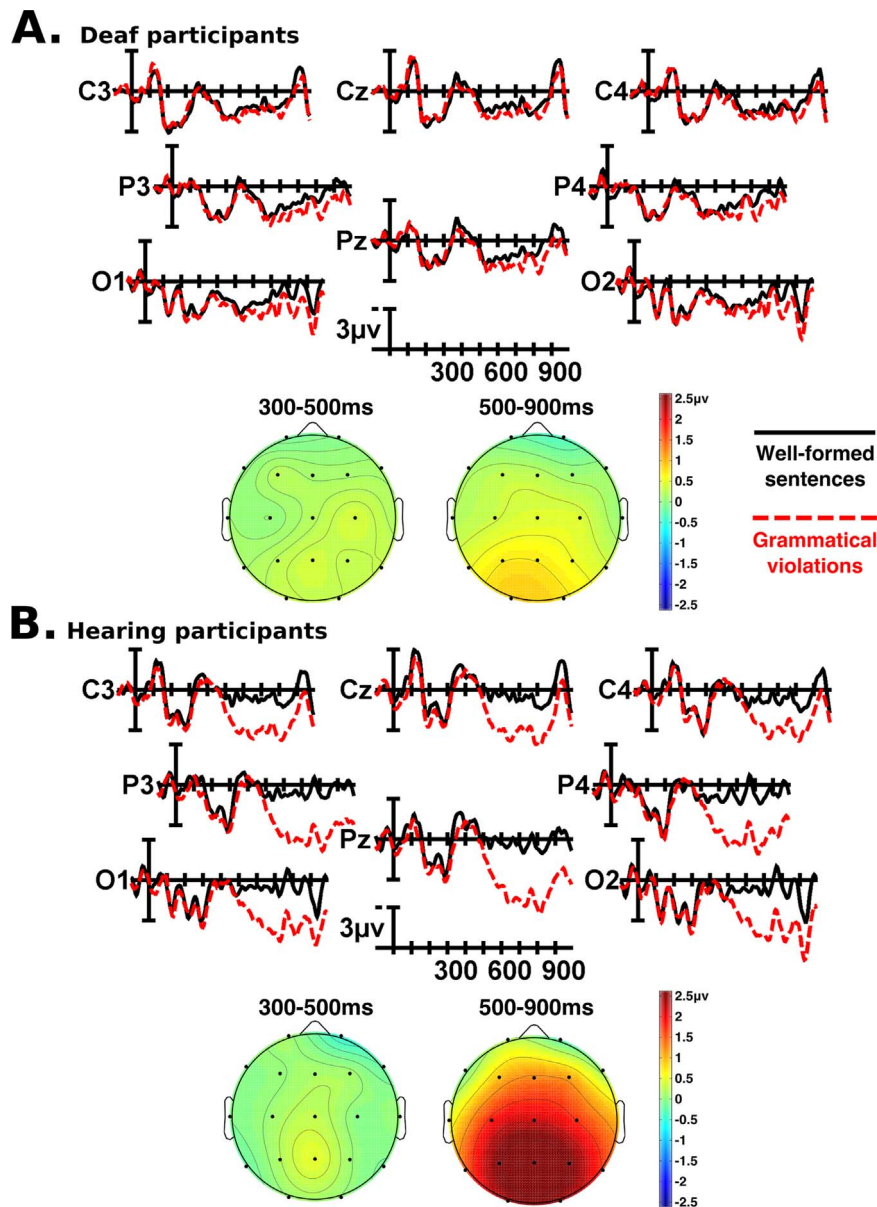


Fig. 4. Grand mean ERP waveforms and scalp topographies for sentences with grammatical violations alone as compared to well-formed sentences, for deaf (Panel A) and hearing (Panel B) participants. Full ERP waveforms are shown for the central-posterior region of interest (electrodes C3, Cz, C4, P3, Pz, P4, O1, O2). Onset of the critical word in the sentence is indicated by the vertical bar. Calibration bar shows 3  $\mu\text{V}$  of activity; each tick mark represents 100 ms of time. Negative voltage is plotted up. Topographic map scale is in microvolts.

group. Both groups of participants displayed small P600-like responses to semantic violations in sentences.

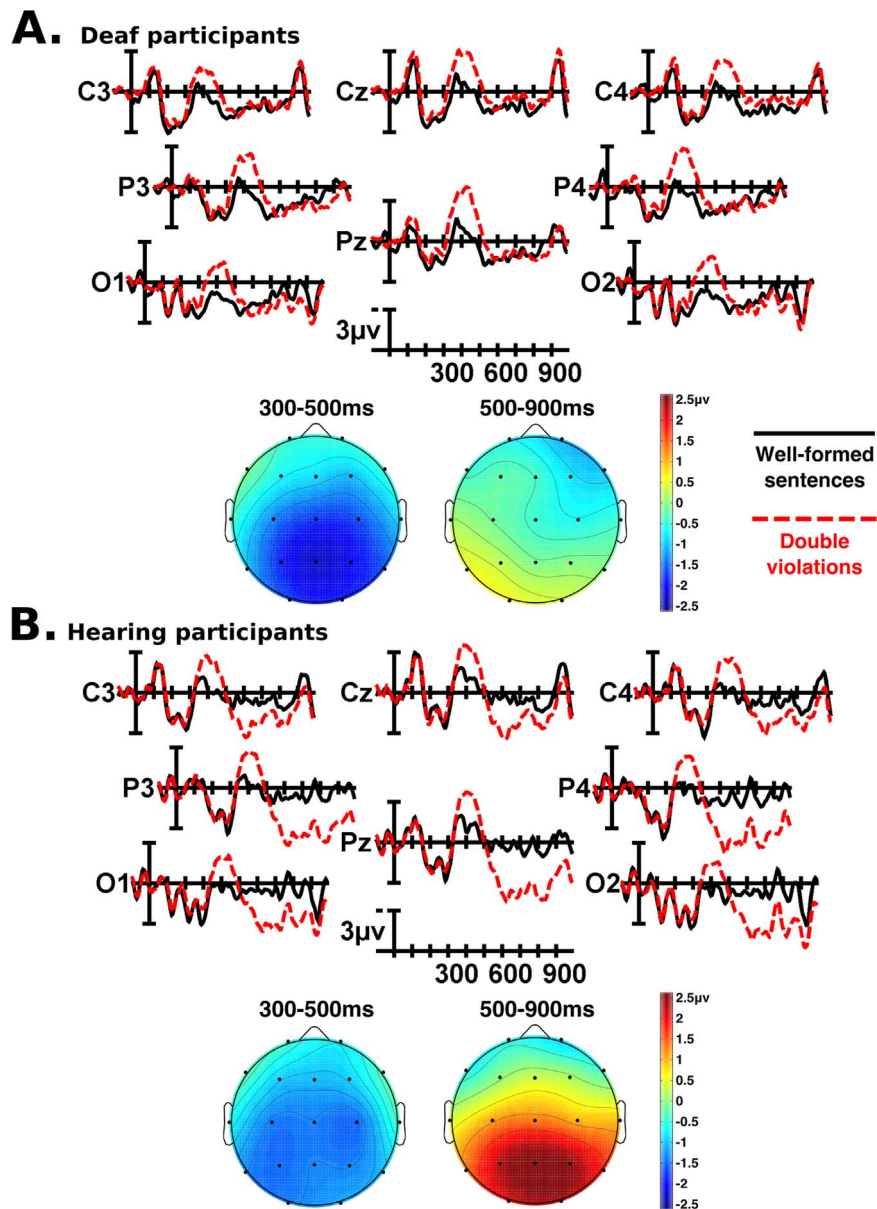
### 3.3. Individual differences analyses

As described in the methods Section 2.6, effect magnitudes of the following ERP responses were calculated for use in individual differences analyses: a) semantic violation N400, b) grammatical violation P600, c) double violation N400, and d) double violation P600. Simple correlations between the four ERP effect magnitudes and participants' standardized reading comprehension scores can be seen in Table 4. For deaf participants, as the magnitude of the N400 elicited by sentences with double violations increased, reading comprehension score also increased. No other ERP effect magnitude was significantly correlated with reading comprehension score for the deaf participants. For hearing participants, however, a notably different set of relationships was found. In the hearing participants, as the magnitude of the P600 elicited by sentences with either a grammatical violation alone or a

double violation increased, reading comprehension score increased.

As was previously described, many factors may lead to variation in a person's reading skill. To better understand the relationship between the ERP effect magnitudes and participants' standardized reading comprehension scores, we used theoretically-driven multiple regression models to account for some of this variation. As described in the methods 2.6, the first set of multiple regression models included as ERP predictors the effect magnitudes of the semantic violation N400 and grammatical violation P600 (deaf participants: Table 5; hearing participants: Table 6). To answer our questions, the most relevant results are the predictive value of each individual predictor, as well as how much additional variance in reading comprehension score the ERP measures account for. For both deaf and hearing participants, an increase in the number of years of education predicted a higher reading comprehension score. For only deaf participants, a higher score on the NTID speechreading test predicted a higher reading comprehension score. Our measure of language use while growing up did not significantly predict reading comprehension score for deaf participants. In terms of





**Fig. 5.** Grand mean ERP waveforms and scalp topographies for sentences with double semantic and grammatical violations as compared to well-formed sentences, for deaf (Panel A) and hearing (Panel B) participants. Full ERP waveforms are shown for the central-posterior region of interest (electrodes C3, Cz, C4, P3, Pz, P4, O1, O2). Onset of the critical word in the sentence is indicated by the vertical bar. Calibration bar shows 3  $\mu$ V of activity; each tick mark represents 100 ms of time. Negative voltage is plotted up. Topographic map scale is in microvolts.

**Table 4**  
Correlation coefficients (*r*) for the relationship between ERP effect magnitudes and standardized reading comprehension score.

	Standardized reading comprehension score	
	Deaf	Hearing
Semantic violation N400	0.161	0.291 <sup>^</sup>
Grammatical violation P600	0.193	0.339 <sup>^</sup>
Double violation N400	0.310 <sup>*</sup>	0.098
Double violation P600	0.151	0.323 <sup>^</sup>

<sup>\*</sup> *p* < 0.05.  
<sup>^</sup> *p* < 0.10.

the ERP predictors, for deaf participants, an increase in the effect magnitude of the N400 elicited by semantic violations alone predicted a higher reading comprehension score. The effect magnitude of the P600 elicited by grammatical violations alone was not a predictor of reading

comprehension score for deaf participants (Table 5; Fig. 6A). Conversely, in hearing participants, an increase in the effect magnitude of the P600 elicited by grammatical violations significantly predicted higher reading comprehension scores. The effect magnitude of the N400 elicited by semantic violations did not significantly predict hearing participants' reading comprehension scores (Table 6; Fig. 6B). The second set of regression models included as ERP predictors the effect magnitudes of the double violation N400 and P600 (deaf participants: Table 7; hearing participants: Table 8). As in the first set of models, for both deaf and hearing participants, an increase in the number of years of education completed predicted a higher reading comprehension score. For only the deaf participants, a higher score on the NTID speechreading test predicted a higher reading comprehension score. Similar relationships between ERP predictors and reading comprehension score observed in the first model were found here. For deaf participants, an increase in the effect magnitude of the N400 elicited by double violations predicted a higher reading comprehension score. The effect magnitude of the P600 elicited by double violations was not a

**Table 5**  
Deaf participant multiple regression model of reading comprehension score with semantic violation N400 and grammatical violation P600 as ERP predictors.

Block 1	$R^2_{change}$	$F_{change}$	$R^2_{total}$	$R^2_{Adj}$	$F_{total}$	
		.64	22.12 (3,38) <sup>***</sup>	.64	.61	22.12 (3,38) <sup>***</sup>
	B	(SE)	$\beta$	t	p	
Years of Education	3.73	(1.07)	0.41	3.50	**	0.001
NTID Speechreading Score	0.47	(0.11)	0.53	4.10	***	< 0.001
Growing Up Language	0.81	(0.99)	0.09	0.82	0.419	
Block 2	$R^2_{change}$	$F_{change}$	$R^2_{total}$	$R^2_{Adj}$	$F_{total}$	
	.05	2.98 (2,36) <sup>^</sup>	.69	.64	15.85 (5,36) <sup>***</sup>	
	B	(SE)	$\beta$	t	p	
Years of Education	4.11	(1.03)	0.45	3.98	***	< 0.001
NTID Speechreading Score	0.44	(0.11)	0.51	4.03	***	< 0.001
Growing Up Language	0.22	(0.98)	0.03	0.23	0.821	
Semantic Violation N400	2.38	(0.98)	0.23	2.42	*	0.021
Grammatical Violation P600	-0.29	(1.03)	-0.03	-0.28	0.778	

B = unstandardized regression coefficient;  $\beta$  = standardized regression coefficient.  
 \*  $p < .05$ .  
 \*\*  $p < .01$ .  
 \*\*\*  $p < .001$ .  
 ^  $p < 0.10$ .

**Table 6**  
Hearing participant multiple regression model of reading comprehension score with semantic violation N400 and grammatical violation P600 as ERP predictors.

Block 1	$R^2_{change}$	$F_{change}$	$R^2_{total}$	$R^2_{Adj}$	$F_{total}$	
		.17	3.99 (2,39) <sup>^</sup>	.17	.13	3.99 (2,39) <sup>^</sup>
	B	(SE)	$\beta$	t	p	
Years of Education	1.83	(0.72)	0.37	2.56	*	0.014
NTID Speechreading Score	0.14	(0.13)	0.16	1.09	0.284	
Block 2	$R^2_{change}$	$F_{change}$	$R^2_{total}$	$R^2_{Adj}$	$F_{total}$	
	.17	4.79 (2,37) <sup>^</sup>	.34	.27	4.78 (4,37) <sup>**</sup>	
	B	(SE)	$\beta$	t	p	
Years of Education	1.47	(0.67)	0.30	2.18	*	0.036
NTID Speechreading Score	0.16	(0.12)	0.18	1.29	0.205	
Semantic Violation N400	1.72	(0.86)	0.28	2.00	*	0.053
Grammatical Violation P600	1.40	(0.56)	0.34	2.53	0.016	

B = unstandardized regression coefficient;  $\beta$  = standardized regression coefficient.  
 \*  $p < .05$ .  
 \*\*  $p < .01$ .  
 ^  $p < 0.10$ .

significant predictor of reading comprehension score for deaf participants (Table 7; Fig. 7A). Conversely, in hearing participants, an increase in the effect magnitude of the P600 elicited by double violations significantly predicted higher reading comprehension scores. The effect magnitude of the N400 elicited by double violations again did not significantly predict hearing participants' reading comprehension scores (Table 8; Fig. 7B).

To determine whether the relationship between N400 effect magni-

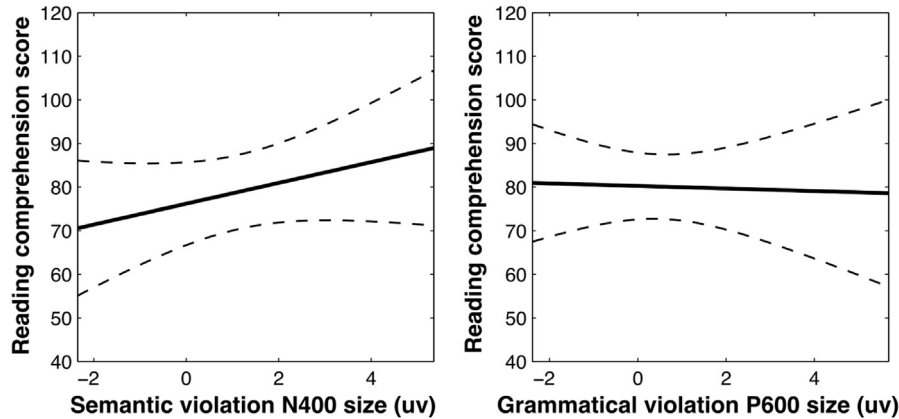
tude and reading skill was driven by the less-proficient deaf readers, we ran a third set of multiple regression models including only deaf participants whose reading comprehension score was within the effective range of the hearing participants (i.e., scores above 80). Twenty-four of the 42 deaf participants fell in this range. The grand mean ERP responses for this subset of higher-skill deaf readers can be seen in Fig. 8. Visual inspection suggested that the grand mean ERP responses for the subset of high-skill deaf readers was similar to the responses seen in the full sample of deaf participants. Statistical analysis of the ERP responses of interest confirmed this. In the 300–500 ms (N400) time window, the subset of higher-skill deaf readers showed a widespread N400 in response to sentences with semantic and double violations (main effect of semantic correctness; midline:  $F(1,23) = 24.359, p < 0.001$ ; medial:  $F(1,23) = 21.194, p < 0.001$ ; lateral:  $F(1,23) = 8.966, p = 0.006$ ). As in the full sample of deaf participants, the N400 effect was strongest at posterior electrodes (semantic correctness  $\times$  electrode interaction: midline:  $F(2,46) = 15.375, p < 0.001$ ; medial:  $F(4,92) = 7.553, p = 0.003$ ; lateral:  $F(2,46) = 9.816, p = 0.002$ ). Also similar to the full sample of deaf participants, the N400 response was larger to double violations than semantic violations (midline:  $F(1,23) = 6.238, p = 0.020$ ; medial:  $F(1,23) = 4.344, p = 0.048$ ; lateral:  $F(1,23) = 4.982, p = 0.036$ ). In the 500–900 ms (P600) time window, the subset of higher-skill deaf readers did not show a main effect of grammatically, similar to the full sample of deaf participants. As with the full sample, the subset did show an interaction between semantic correctness and grammaticality (semantic correctness  $\times$  grammaticality interaction: midline:  $F(1,23) = 7.496, p = 0.012$ ; medial:  $F(1,23) = 4.263, p = 0.050$ ; lateral:  $F(1,23) = 5.198, p = 0.032$ ). Follow up contrasts (corrected alpha  $\alpha = 0.0167$ ) indicated that this interaction was driven by the fact that while the subset of higher-skill deaf readers did not show a significant P600 in response to grammatical violations alone, at posterior electrodes they showed a small P600-like response to semantic and double violations, a trend that reversed at anterior electrodes (semantic violations alone, electrode interaction, medial:  $F(4,92) = 5.114, p = 0.009$ ; double violations, electrode interaction, lateral:  $F(2,46) = 6.095, p = 0.015$ ). This interaction is similar to what was seen in the full sample of deaf participants. In regards to end-of-sentence acceptability judgments, when compared to the remaining lower-skill deaf readers ( $n = 18$ ), the higher-skill deaf readers had higher overall  $d'$  scores, higher semantic violation  $d'$  scores, and higher double violation  $d'$  scores (corrected alpha  $\alpha = 0.0125, t(40)$  ranged from  $-4.826$  to  $-5.368$ , all  $p$ 's  $< 0.001$ ). However, there was no difference between the two groups in the  $d'$  scores for sentences with grammatical errors. In summary, the subset of higher-skilled deaf readers had qualitatively and quantitatively similar ERP grand mean responses as the full sample of deaf participants. The higher-skilled deaf readers were generally better at the sentence acceptability task, except when judging sentences with grammatical errors alone.

For the regression models with the subset of higher-skill deaf readers, given the reduction in statistical power from this reduced sample, many of the regression predictors, both background variables and ERP responses, were no longer significant predictors (Tables 9 and 10). However, across both multiple regression models, the single variable that was still a significant predictor of better reading comprehension score was N400 size in response to double violations in sentences (Table 10); showing the same relationship as the full group of deaf participants.

ERP effect magnitudes were also compared to participants'  $d'$  scores from each type of sentence violation (Table 11). For deaf participants, none of the ERP effect magnitudes were significantly correlated with their corresponding  $d'$  score. For hearing participants, the effect magnitude of the grammatical violation P600 was significantly correlated with grammatical violation  $d'$  score; as the grammatical violation P600 effect magnitude increased, the grammatical violation  $d'$  score also increased.

Overall, these multiple regression models show that for the deaf

### A. Deaf participants



### B. Hearing participants

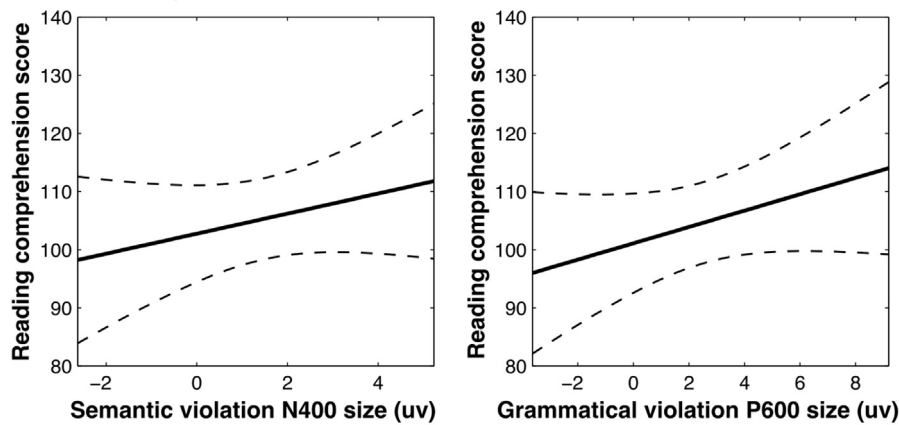


Fig. 6. Effects of the semantic violation N400 predictor (left column) and grammatical violation P600 predictor (right column) while holding all other predictors constant in the semantic N400/grammatical P600 multiple regression models (A: deaf participant model, B: hearing participant model). The solid line shows the relationship between the predictor (on the x-axis) and the outcome, reading comprehension score, while holding all other predictors constant. Dotted lines show 95% confidence bounds. Plots produced with MATLAB's plotSlice() function.

Table 7

Deaf participant multiple regression model of reading comprehension score with double violation N400 and P600 as ERP predictors.

Block 1	$R^2_{change}$	$F_{change}$	$R^2_{total}$	$R^2_{Adj}$	$F_{total}$
	.64	22.12 (3,38)***	.64	.61	22.12 (3,38)***
	B	(SE)	$\beta$	t	p
Years of Education	3.73	(1.07)	0.41	3.50	** 0.001
NTID Speechreading Score	0.47	(0.11)	0.53	4.10	*** < 0.001
Growing Up Language	0.81	(0.99)	0.09	0.82	0.419
Block 2	$R^2_{change}$	$F_{change}$	$R^2_{total}$	$R^2_{Adj}$	$F_{total}$
	.14	10.76 (2,36)***	.77	.74	24.39 (5,36)***
	B	(SE)	$\beta$	t	p
Years of Education	4.17	(0.91)	0.46	4.57	*** < 0.001
NTID Speechreading Score	0.38	(0.10)	0.43	3.98	*** < 0.001
Growing Up Language	-0.01	(0.82)	0.00	-0.01	0.990
Double Violation N400	3.69	(0.89)	0.45	4.16	*** < 0.001
Double Violation P600	1.10	(0.99)	0.13	1.11	0.273

B = unstandardized regression coefficient;  $\beta$  = standardized regression coefficient.

\*\*  $p < .01$ .

\*\*\*  $p < .001$ .

Table 8

Hearing participant multiple regression model of reading comprehension score with double violation N400 and P600 as ERP predictors.

Block 1	$R^2_{change}$	$F_{change}$	$R^2_{total}$	$R^2_{Adj}$	$F_{total}$
	.17	3.99 (2,39)*	.17	.13	3.99 (2,39)*
	B	(SE)	$\beta$	t	p
Years of Education	1.83	(0.72)	0.37	2.56	* 0.014
NTID Speechreading Score	0.14	(0.13)	0.16	1.09	0.284
Block 2	$R^2_{change}$	$F_{change}$	$R^2_{total}$	$R^2_{Adj}$	$F_{total}$
	.14	3.74 (2,37)*	.31	.24	4.15 (4,37)**
	B	(SE)	$\beta$	t	p
Years of Education	1.61	(0.68)	0.33	2.39	* 0.022
NTID Speechreading Score	0.15	(0.13)	0.17	1.18	0.247
Double Violation N400	1.83	(0.99)	0.29	1.84	~ 0.073
Double Violation P600	1.51	(0.58)	0.39	2.59	* 0.014

B = unstandardized regression coefficient;  $\beta$  = standardized regression coefficient.

\*  $p < .05$ .

\*\*  $p < .01$ .

~  $p < 0.10$ .

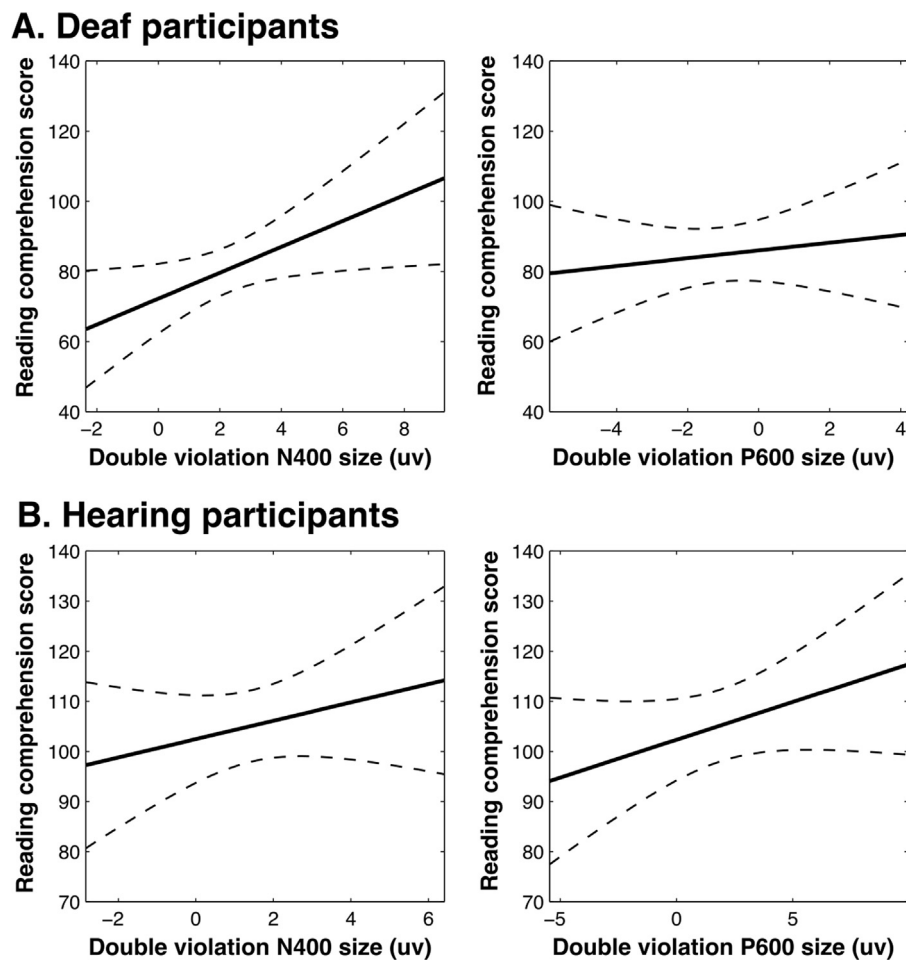


Fig. 7. Effects of the double violation N400 predictor (left column) and double violation P600 predictor (right column) while holding all other predictors constant in the double N400/P600 multiple regression models (A: deaf participant model, B: hearing participant model). The solid line shows the relationship between the predictor (on the x-axis) and the outcome, reading comprehension score, while holding all other predictors constant. Dotted lines show 95% confidence bounds. Plots produced with MATLAB's plotSlice() function.

participants, the magnitude of the N400 response to semantic violations alone or double violations was the best ERP predictor of reading comprehension score, with larger N400s predicting better reading scores. For the hearing readers, the magnitude of the P600 response to grammatical violations alone or double violations was the best ERP predictor of reading comprehension score, with larger P600s predicting better reading. There was no relationship between the magnitude of deaf participants' P600 response and their reading comprehension scores.

#### 4. Discussion

Grand mean analyses showed that deaf and hearing readers had similarly large N400 responses to semantic violations in sentences (alone or in double violations). In contrast, hearing participants showed a P600 response to grammatical and double violations, while deaf participants as a group did not. The N400 is associated with neural processing involved in lexical access (i.e., accessing information about word meanings) and semantic integration (Kutas and Federmeier, 2000; Kutas and Hillyard, 1984, 1980), while the P600 is generally associated with the integration or reanalysis of grammatical information (Kaan et al., 2000; Osterhout et al., 1994; Osterhout and Holcomb, 1992). Thus, at a group level, deaf and hearing participants responded similarly to semantic information during sentence reading, but only hearing readers responded strongly to grammatical information. Individual differences analyses allowed us to better understand the relationship between participants' reading skill and their sensitivity to semantic and grammatical information in text, as indexed by the

magnitude of their ERP responses. An important note is that while the average reading comprehension score of the deaf participants was lower than the average score of the hearing participants, the best deaf and hearing participants had equally high reading scores. Thus, we are able to assess deaf and hearing readers who have attained similar levels of reading proficiency. For deaf readers, the magnitude of the N400 response to semantic violations alone or in double violations was the best ERP predictor of reading comprehension score, with larger N400s predicting better reading scores. These relationships held (and indeed, were strongest) even after controlling for differences in years of education, standardized speechreading score, and language background. For hearing readers, the magnitude of the P600 response to grammatical violations alone or in double violations was the best ERP predictor of reading comprehension score, with larger P600s predicting better reading scores. There was no relationship between deaf participants' P600 size and their reading scores. These results show that the best deaf readers responded most to semantic information in sentences, while the best hearing readers responded most to grammatical information.

An overarching question surrounding deaf literacy is whether deaf children read in similar or different ways than hearing children. The results of this study clearly show that equally proficient hearing and deaf adults rely on different types of linguistic information when reading sentences. Specifically, the best deaf readers seem to rely primarily on semantic information during sentence comprehension, while the best hearing readers rely on both grammatical and semantic information, with a stronger link between grammatical processing and better reading skill. Given this difference between deaf and hearing

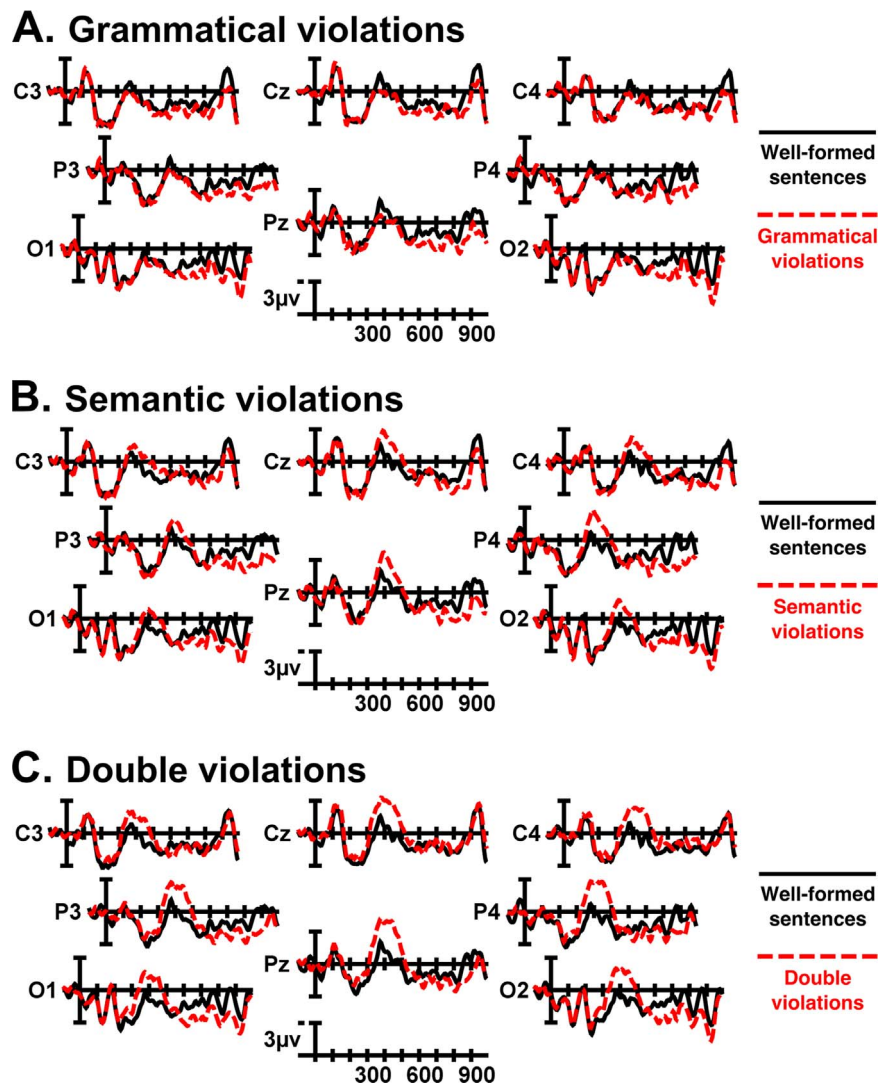


Fig. 8. Grand mean ERP waveforms across all three sentence violation conditions for the 24 deaf participants whose reading comprehension scores were within the effective range of the hearing participants (scores above 80). Onset of the critical word in the sentence is indicated by the vertical bar. Calibration bar shows 3  $\mu$ V of activity; each tick mark represents 100 ms of time. Negative voltage is plotted up.

adults, there is little reason to assume that literacy teaching strategies should necessarily be the same for hearing and deaf children. Though further research with children is needed to fully answer this question, our research gives clear evidence that hearing and deaf adults read at similar levels of proficiency in different ways.

As in the deaf population generally, only ten percent (4 of 42) of the deaf participants in this study were exposed to ASL from an early age – all other deaf participants were exposed to ASL later in life. Thus, our results are primarily applicable to non-native signing deaf adults, rather than to deaf native signers. Prior ERP research has shown that deaf native signing adults and hearing adults reading in their native language (L1) showed similar P600s when reading grammatical violations in sentences, while non-native signing deaf adults had smaller P600s (Skotara et al., 2012, 2011). In that work, all three groups had similar N400 responses to semantic violations in sentences. Though Skotara and colleagues did not analyze individual differences in participants' ERP responses, nor how ERP responses related to reading proficiency, their grand mean results are similar to the results from non-native signing deaf adults in this study. Additionally, though the four native signing deaf participants in this study were too small a sample to analyze statistically, visual inspection of their grand mean ERP responses to grammatical violations suggests they may show more P600-like activity than the group of deaf participants as a whole

(Fig. 9). Together, our results and those of Skotara and colleagues support the idea that non-native signing deaf adults likely process written language differently than deaf native signers and hearing adults. Interestingly, our results with primarily non-native deaf signers differ from what is typically seen in second-language (L2) hearing readers. In general, as hearing individuals become more proficient in their L2, they show larger and/or more robust P600 effects in response to grammatical violations in sentences (McLaughlin et al., 2010; Rossi et al., 2006; Tanner et al., 2013; Tanner and Van Hell, 2014; Weber-Fox and Neville, 1996). However, in this study there was no relationship between deaf participants' reading comprehension scores and the size of their P600s in response to grammatical violations, even though the best deaf readers read as proficiently as the best hearing readers. This suggests that deaf individuals are not simply another group of second language learners, but that their language learning experience, especially in terms of grammatical knowledge, is markedly different from hearing second language learners.<sup>1</sup>

A potential confound would exist if the less-proficient deaf readers

<sup>1</sup> We also examined whether the ERP results for the six deaf readers who became deaf after birth but before age two (and thus had some early oral language exposure) differed from the full sample of deaf readers. Inspection of the grand means revealed similar ERP patterns, although the small sample size prevented statistical analyses.

**Table 9**

Multiple regression model of reading comprehension score with semantic violation N400 and grammatical violation P600 as ERP predictors for the subset of deaf participants whose reading comprehension score was within the effective range of the hearing participants (scores above 80; n = 24).

Block 1	$R^2_{change}$	$F_{change}$	$R^2_{total}$	$R^2_{Adj}$	$F_{total}$
		.19	1.60 (3,20)	.19	.07
	B	(SE)	$\beta$	t	p
Years of Education	0.64	(1.49)	0.09	0.43	0.674
NTID Speechreading Score	0.25	(0.21)	0.35	1.22	0.236
Growing Up Language	-0.53	(1.32)	-0.11	-0.40	0.694
Block 2	$R^2_{change}$	$F_{change}$	$R^2_{total}$	$R^2_{Adj}$	$F_{total}$
	.12	1.55 (2,18)	.31	.12	1.63 (5,18)
	B	(SE)	$\beta$	t	p
Years of Education	0.11	(1.64)	0.02	0.07	0.946
NTID Speechreading Score	0.29	(0.21)	0.40	1.39	0.182
Growing Up Language	-0.13	(1.30)	-0.03	-0.10	0.923
Semantic Violation N400	1.69	(1.16)	0.30	1.46	0.160
Grammatical Violation P600	0.95	(1.24)	0.18	0.77	0.452

B = unstandardized regression coefficient;  $\beta$  = standardized regression coefficient.

**Table 10**

Multiple regression model of reading comprehension score with double violation N400 and P600 as ERP predictors for the subset of deaf participants whose reading comprehension score was within the effective range of the hearing participants (scores above 80; n = 24).

Block 1	$R^2_{change}$	$F_{change}$	$R^2_{total}$	$R^2_{Adj}$	$F_{total}$
		.19	1.60 (3,20)	.19	.07
	B	(SE)	$\beta$	t	p
Years of Education	0.64	(1.49)	0.09	0.43	0.674
NTID Speechreading Score	0.25	(0.21)	0.35	1.22	0.236
Growing Up Language	-0.53	(1.32)	-0.11	-0.40	0.694
Block 2	$R^2_{change}$	$F_{change}$	$R^2_{total}$	$R^2_{Adj}$	$F_{total}$
	.18	2.58 (2,18)	.37	.20	2.14 (5,18)
	B	(SE)	$\beta$	t	p
Years of Education	1.25	(1.43)	0.17	0.87	0.396
NTID Speechreading Score	0.32	(0.20)	0.44	1.63	0.121
Growing Up Language	0.06	(1.25)	0.01	0.05	0.963
Double Violation N400	2.38	(1.06)	0.49	2.24	* 0.038
Double Violation P600	0.72	(1.07)	0.15	0.07	0.510

B = unstandardized regression coefficient;  $\beta$  = standardized regression coefficient.

drove the relationship between larger N400s and better reading skill. However, this was not the case. When analyzing data from only deaf participants whose reading comprehension score was within the effective range of hearing participants' scores, the double violation N400 effect magnitude remained a significant predictor of reading score, while all other variables, including years of education and speechreading score, were no longer significant predictors. Though N400 size to semantic violations alone was not a significant predictor for the subset of high-skill deaf readers, its predictive ability in the full group of participants was less dramatic than the double violation N400,

**Table 11**

Correlation coefficients (r) for the relationship between ERP effect magnitudes and  $d'$  scores.

Correlation	Deaf	Hearing
Semantic violation N400 vs. semantic violation $d'$	0.078	0.297 <sup>~</sup>
Grammatical violation P600 vs. grammatical violation $d'$	0.143	0.389*
Double violation N400 vs. double violation $d'$	0.205	0.234
Double violation P600 vs. double violation $d'$	0.221	0.033

\*  $p < 0.05$ .  
~  $p < 0.1$ .

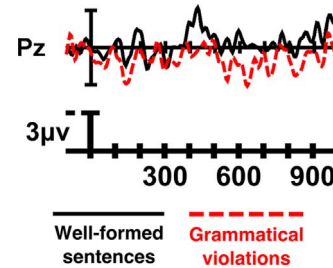


Fig. 9. Grand mean ERP waveform from electrode Pz for sentences with grammatical violations alone as compared to well-formed sentences for deaf participants who were native ASL users (n = 4). Onset of the critical word in the sentence is indicated by the vertical bar. Each tick mark represents 100 ms of time. Negative voltage is plotted up.

and thus likely suffered from the reduction in statistical power.

The  $d'$  results from the end-of-sentence acceptability judgment task also support the conclusion that for deaf participants, responses to sentences with semantic violations (alone or in double violations) are the best real-time predictors of reading skill. While better  $d'$  scores to all types of sentence violations accounted for a significant amount of variance in deaf participants' reading score, semantic and double violation  $d'$  scores each explain over four times as much variance in reading comprehension than do grammatical violation  $d'$  scores. This reinforces the importance of the relationship in deaf participants between reading comprehension and responses to semantic information. For hearing participants, the greatest amount of variance in reading score was accounted for by grammatical violation  $d'$ , supporting the fact that for hearing participants, larger P600 size was the strongest ERP predictor of better reading skill.

The idea that the best deaf readers respond most to semantic information in sentences is seemingly at odds with a standard model of language, in which grammatical proficiency is thought to be the primary driver of sentence comprehension proficiency (Frazier and Rayner, 1982). However, an alternative model, the 'Good-Enough' theory of sentence comprehension, challenges this idea (Ferreira et al., 2002). The 'Good-Enough' theory was developed in an effort to account for observations in sentence reading behavior that did not track with prior theories of grammar-focused sentence processing. In the 'Good-Enough' theory, readers analyze grammatical structure only as much as is necessary for the task at hand. In normal day-to-day communication and reading, oftentimes most necessary information can be obtained from the meanings of words in a sentence, without needing to determine the exact grammatical role served by every word (Ferreira et al., 2002; Ferreira and Patson, 2007). The most successful deaf readers in this study may be using something akin to the 'Good-Enough' approach when reading. They do not appear to be performing much syntactic integration or reanalysis of grammatical information; the best deaf readers are the ones with the largest meaning-associated ERP responses. This is consistent with prior behavioral research showing that deaf adults focus on semantic information when reading (Domínguez et al., 2014; Domínguez and Alegria, 2010) and with online sentence processing studies indicating that deaf readers use semantic information to help them understand grammatically complex

sentences (Traxler et al., 2014).

Given that the best deaf readers respond most to information about meaning in sentences, this suggests that for deaf children learning to read, a focus on teaching and encouraging vocabulary development may be more beneficial than teaching grammatical information in great detail - i.e., a focus on meaning rather than linguistic form (Long and Robinson, 1998). It is crucial that any potential changes in reading education for deaf children be thoroughly tested before implementation. Nevertheless, the results from this study give clear, brain-based data showing that the best deaf readers (who were primarily non-native signers) respond most to semantic information, not grammatical information, in sentences, and that this pattern of responses is markedly different than what is seen in equally proficient hearing readers. This provides crucial new information in the effort to improve literacy education for deaf children.

## Acknowledgements

This work was supported by National Institutes of Health grant F31DC013700 (A.S.M.) and National Institutes of Health grant T32DC005361 (A.S.M). We thank the members of the Cognitive Neuroscience of Language Lab at the University of Washington for their help in collecting data, especially Emma Wampler and Judith McLaughlin. We also thank Eatai Roth for technical assistance with figures.

## References

- Allen, T., 1986. Patterns of academic achievement among hearing-impaired students. In: Schildroth, A., Karchmer, M. (Eds.), *Deaf Children in America*. College Hill Press, San Diego, CA, pp. 161–206.
- Allen, T., Clark, M.D., del Giudice, A., Koo, D., Lieberman, A., Mayberry, R., Miller, P., 2009. Phonology and reading: a response to Wang, Trezek, Luckner, and Paul. *Am. Ann. Deaf* 154, 338–345.
- Baayen, R.H., Piepenbrock, R., Gulikers, L., 1995. The CELEX Lexical Database (CD-ROM).
- Boudewyn, M.A., Long, D.L., Swaab, T.Y., 2015. Graded expectations: predictive processing and the adjustment of expectations during spoken language comprehension. *Cogn. Affect. Behav. Neurosci.* 15, 607–624.
- Cohen, J., 1992. A power primer. *Psychol. Bull.* 112, 155–159.
- Cohen, S.R., 1967. Predictability of deaf and hearing story paraphrases. *J. Verbal Learn. Verbal Behav.* 6, 916–921.
- Domínguez, A.-B., Alegria, J., 2010. Reading mechanisms in orally educated deaf adults. *J. Deaf Stud. Deaf Educ.* 15, 136–148.
- Domínguez, A.-B., Carrillo, M.-S., Pérez, M., del, M., Alegria, J., 2014. Analysis of reading strategies in deaf adults as a function of their language and meta-phonological skills. *Res. Dev. Disabil.* 35, 1439–1456.
- Easterbrooks, S.R., Huston, S.G., 2008. The signed reading fluency of students who are deaf/hard of hearing. *J. Deaf Stud. Deaf Educ.* 13, 37–54.
- Ewoldt, C., 1981. A psycholinguistic description of selected deaf children reading in sign language. *Read. Res. Q.* 17, 58.
- Ferreira, F., Patson, N.D., 2007. The “Good Enough” approach to language comprehension. *Lang. Linguist. Compass* 1, 71–83.
- Ferreira, F., Bailey, K.G.D., Ferraro, V., 2002. Good-enough representations in language comprehension. *Curr. Dir. Psychol. Sci.* 11, 11–15.
- Foucart, A., Frenck-Mestre, C., 2011. Grammatical gender processing in L2: electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Biling. Lang. Cogn.* 14, 379–399.
- Foucart, A., Frenck-Mestre, C., 2012. Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *J. Mem. Lang.* 66, 226–248.
- Frazier, L., Rayner, K., 1982. Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences. *Cogn. Psychol.* 14, 178–210.
- Goldin-Meadow, S., Mayberry, R.L., 2001. How do profoundly deaf children learn to read? *Learn. Disabil. Res. Pract.* 16, 222–229.
- Hanson, V., 1989. Phonology and reading: evidence from profoundly deaf readers. In: Shankweiler, D., Liberman, I. (Eds.), *Phonology and Reading Disability: Solving the Reading Puzzle*. University of Michigan Press, Ann Arbor, pp. 69–89.
- Hauk, O., Pulvermüller, F., 2004. Effects of word length and frequency on the human event-related potential. *Clin. Neurophysiol.* 115, 1090–1103.
- Jasper, H.H., 1958. The ten–twenty system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 371–375.
- Kaan, E., Harris, A., Gibson, E., Holcomb, P.J., 2000. The P600 as an index of syntactic integration difficulty. *Lang. Cogn. Process.* 15, 159–201.
- Kim, A., Osterhout, L., 2005. The independence of combinatory semantic processing: evidence from event-related potentials. *J. Mem. Lang.* 52, 205–225.
- King, J., Kutas, M., 1995. A brain potential whose latency indexes the length and frequency of words. *Newsl. Cent. Res. Lang.* 10, 1–9.
- Kroese, J., Lotz, W., Puffer, C., Osberger, M., 1986. Chapter 10: academic skills, in: *Language Learning Skills of Hearing-Impaired Students*. ASHA Monographs March, vol. (23), pp. 66–76.
- Kuperberg, G.R., 2007. Neural mechanisms of language comprehension: challenges to syntax. *Brain Res.* 1146, 23–49.
- Kutas, M., Hillyard, S., 1980. Reading senseless sentences: brain potentials reflect semantic anomaly. *Science* 80–207, 203–205.
- Kutas, M., Hillyard, S., 1984. Brain potentials during reading reflect word expectancy and semantic associations. *Nature* 307, 161–163.
- Kutas, M., Federmeier, K.D., 2000. Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470.
- Kyle, F.E., Harris, M., 2010. Predictors of reading development in deaf children: a 3-year longitudinal study. *J. Exp. Child Psychol.* 107, 229–243.
- Long, M., Robinson, P., 1998. Focus on form: theory, research and practice. In: Doughty, C., Williams, J. (Eds.), *Focus on Form in Classroom Second Language Acquisition*. Cambridge University Press, Cambridge, England, pp. 15–41.
- Mayberry, R.L., del Giudice, A. a., Lieberman, A.M., 2011. Reading achievement in relation to phonological coding and awareness in deaf readers: a meta-analysis. *J. Deaf Stud. Deaf Educ.* 16, 164–188.
- Mayer, C., Trezek, B.J., 2014. Is reading different for deaf individuals?: reexamining the role of phonology. *Am. Ann. Deaf* 159, 359–371.
- McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., Osterhout, L., 2010. Brain potentials reveal discrete stages of L2 grammatical learning. *Lang. Learn.* 60, 123–150.
- Mohammed, T., Campbell, R., MacSweeney, M., Barry, F., Coleman, M., 2006. Speechreading and its association with reading among deaf, hearing and dyslexic individuals. *Clin. Linguist. Phon.* 20, 621–630.
- Newman, A.J., Tremblay, A., Nichols, E.S., Neville, H.J., Ullman, M.T., 2012. The influence of language proficiency on lexical semantic processing in native and late learners of English. *J. Cogn. Neurosci.* 24, 1205–1223.
- Ojima, S., Matsuba-Kurita, H., Nakamura, N., Hoshino, T., Hagiwara, H., 2011. Age and amount of exposure to a foreign language during childhood: behavioral and ERP data on the semantic comprehension of spoken English by Japanese children. *Neurosci. Res.* 70, 197–205.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Osterhout, L., Holcomb, P., 1992. Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806.
- Osterhout, L., Nicol, J., 1999. On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Lang. Cogn. Process.* 14, 283–317.
- Osterhout, L., Holcomb, P.J., Swinney, D.A., 1994. Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 786–803.
- Osterhout, L., Bersick, M., McKinnon, R., 1997. Brain potentials elicited by words: word length and frequency predict the latency of an early negativity. *Biol. Psychol.* 46, 143–168.
- Pakulak, E., Neville, H.J., 2010. Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *J. Cogn. Neurosci.* 22, 2728–2744.
- Paul, P., Wang, Y., Trezek, B.J., Luckner, J.L., 2009. Phonology is necessary, but not sufficient: a rejoinder. *Am. Ann. Deaf* 154, 346–356.
- Perfetti, C., Sandak, R., 2000. Reading optimally builds on spoken language: implications for deaf readers. *J. Deaf Stud. Deaf Educ.* 5, 32–50.
- Pickering, M.J., Garrod, S., 2007. Do people use language production to make predictions during comprehension? *Trends Cogn. Sci.* 11, 105–110.
- Power, D.J., Quigley, S.P., 1973. Deaf children's acquisition of the passive voice. *J. Speech Lang. Hear. Res.* 16, 5–11.
- Qi, S., Mitchell, R.E., 2012. Large-scale academic achievement testing of deaf and hard-of-hearing students: past, present, and future. *J. Deaf Stud. Deaf Educ.* 17, 1–18.
- Quigley, S.P., Smith, N.L., Wilbur, R.B., 1974. Comprehension of relativized sentences by deaf students. *J. Speech Lang. Hear. Res.* 17, 325–341.
- Rossi, S., Gugler, M.F., Friederici, A.D., Hahne, A., 2006. The impact of proficiency on syntactic second-language processing of German and Italian: evidence from event-related potentials. *J. Cogn. Neurosci.* 18, 2030–2048.
- Russell, W., Quigley, S., Power, D., 1976. *Linguistics and Deaf Children*. Alexander Graham Bell Association for the Deaf, Washington, DC.
- Sarachan-Deily, A.B., 1980. *Deaf Readers' Comprehension of Individual Sentences [microform]/Ann Beth Sarachan-Deily*. Distributed by ERIC Clearinghouse, Washington, D.C.
- Shulman, J., Decker, N., 1980. *Readable English for Hearing-impaired Students: Multi-level Guidelines for Linguistically Controlled Reading Materials*. Multi-level Captioning Project, the Caption Center, WGBH-TV, Boston.
- Sims, D., 2009. *NTID Speechreading: CID Everyday Sentences Test*. Rochester Institute of Technology, Rochester, NY.
- Skotara, N., Kügöw, M., Salden, U., Hänel-Faulhaber, B., Röder, B., 2011. ERP correlates of intramodal and crossmodal L2 acquisition. *BMC Neurosci.* 12.
- Skotara, N., Salden, U., Kügöw, M., Hänel-Faulhaber, B., Röder, B., 2012. The influence of language deprivation in early childhood on L2 processing: an ERP comparison of deaf native signers and deaf signers with a delayed language acquisition. *BMC Neurosci.* 13.
- Spencer, L., Barker, B., Tomblin, J., 2003. Exploring the language and literacy outcomes of pediatric cochlear implant users. *Ear Hear.* 24, 236–247.
- Tanner, D., 2013. Individual differences and streams of processing. *Linguist. Approaches Biling.* 3, 350–356.

- Tanner, D., Van Hell, J.G., 2014. ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia* 56, 289–301.
- Tanner, D., McLaughlin, J., Herschensohn, J., Osterhout, L., 2013. Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Biling. Lang. Cogn.* 16, 367–382.
- Tanner, D., Inoue, K., Osterhout, L., 2014. Brain-based individual differences in on-line L2 grammatical comprehension. *Biling. Lang. Cogn.* 17, 277–293.
- Traxler, C.B., 2000. The Stanford achievement test, 9th Edition: national norming and performance standards for deaf and hard-of-. *J. Deaf Stud. Deaf Educ.* 5, 337–348.
- Traxler, M.J., Corina, D.P., Morford, J.P., Hafer, S., Hoversten, L.J., 2014. Deaf readers' response to syntactic complexity: evidence from self-paced reading. *Mem. Cogn.* 42, 97–111.
- Trybus, R.J., Buchanan, C., 1973. Patterns of achievement test performance. *Studies in achievement testing, hearing-impaired students* (Series D, No 11). Gallaudet College, Office of Demographic Studies, Washington, DC.
- Wang, Y., Williams, C., 2014. Are we hammering square pegs into round holes?: an investigation of the meta-analyses of reading research with students who are d/deaf or hard of hearing and students who are hearing. *Am. Ann. Deaf* 159, 323–345.
- Wang, Y., Trezek, B.J., Luckner, J.L., Paul, P.V., 2008. The role of phonology and phonologically related skills in reading instruction for students who are deaf or hard of hearing. *Am. Ann. Deaf* 153, 396–407.
- Weber-Fox, C., Neville, H.J., 1996. Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *J. Cogn. Neurosci.* 8, 231–256.
- Weber-Fox, C., Davis, L.J., Cuadrado, E., 2003. Event-related brain potential markers of high-language proficiency in adults. *Brain Lang.* 85, 231–244.
- Woodcock, R., 2011. *Woodcock Reading Mastery Tests*, Third edition. Pearson, Bloomington, MN.
- Yurkowski, P., Ewoldt, C., 1986. A case for the semantic processing of the deaf reader. *Am. Ann. Deaf* 131, 243–247.