Technology in Cognitive Research: Methods to Examine Second Language Processing in Study Abroad Research

Mandy Faretta-Stutenberg  
*Northern Illinois University*, mfs@niu.edu

Irene Finestrat  
*Northwestern University*

Kara Morgan-Short  
*University of Illinois at Chicago*

Follow this and additional works at: [https://huskiecommons.lib.niu.edu/allfacultyother-bookschapters](https://huskiecommons.lib.niu.edu/allfacultyother-bookschapters)

Original Citation

This Book is brought to you for free and open access by the Faculty Research, Artistry, & Scholarship at Huskie Commons. It has been accepted for inclusion in Faculty Books & Book Chapters by an authorized administrator of Huskie Commons. For more information, please contact jschumacher@niu.edu.
Chapter 7

Technology in Cognitive Research: Methods to Examine Second Language Processing in Study Abroad Research

Mandy Faretta-Stutenberg*, Irene Finestrat**, and Kara Morgan-Short***

*Northern Illinois University, ** Northwestern University, ***University of Illinois Chicago
1. Introduction

Study Abroad (SA) research to date has largely focused on investigating second language (L2) development by measuring changes in *performance* on language tasks. Less SA research has examined development of language *processing*, that is, the mechanisms involved in real-time language comprehension and production, which underlies task performance (VanPatten & Jegerski, 2010). Given the unique language demands inherent to the SA environment, it is possible that learning in this context is more conducive to the development of efficient L2 processing than learning that takes place in more controlled contexts, such as domestic classroom environments. Thus, examining language processing in addition to language performance can help provide a more holistic view of L2 development as shaped by the SA context.

Our chapter aims to highlight methods for examining language processing by presenting a targeted overview of research instruments and technologies that can provide indices of language processing development for SA contexts (see also Marijuan & Sanz, 2017; for more general overviews of methodological approaches to examining L2 processing, see Grey & Tagarelli, 2018; Sanz & Grey, 2015). We first present behavioral methods that can be employed using tools accessible to almost every researcher: a computer, keyboard, and mouse. We aim to show how these methods can be leveraged to provide powerful information about how learners process language in real time. We then describe eye-tracking and event-related potential techniques, both of which require specialized technology, but have the benefit of providing unique datasets that can offer novel insights into processing. Subsequent to the overview of each method, we briefly summarize empirical research that has employed the method to address
questions about SA. We discuss limitations of each technique, and offer guiding principles for researchers interested in incorporating the method into their own work. Finally, we conclude the chapter by offering specific future research directions that can be addressed with one or more of the methods presented.

2. Behavioral Methods

Information about language processing speed or efficiency can be assessed behaviorally via tasks that include measures of reaction times. Such psycholinguistic tasks record the time between the presentation of a linguistic stimulus and a participant’s behavioral response (for example, a button press). Reaction time data can be examined on their own, or in tandem with behavioral performance data to provide fine-grained details about language processing and learning (e.g., Grey & Tagarelli, 2018; Marijuan & Sanz, 2017; Sanz & Grey, 2015). Here, we discuss a number of behavioral tasks that make use of reaction time to provide a window into L2 processing.

First, some tasks that are commonly used to assess performance can also be leveraged to provide reaction time data. For example, when engaging in language decisions tasks, participants provide judgments about different aspects of language. Whereas researchers have typically examined the accuracy of such responses, the reaction time for the response can also provide useful information. Various decision tasks have been employed to examine L2: (a) in a grammaticality judgment task, participants generally indicate whether a sentence is ‘good’ and ‘bad’ based on its grammaticality or acceptability, (b) in a semantic classification task, participants categorize experimental items by specific semantic properties, (c) in a lexical
decision task, participants indicate whether stimuli are words or non-words in a given language, and finally, (d) in a translation-recognition task, participants judge whether pairs of words in a given set are translations of one another (for an in-depth overview of the use of judgment or decision tasks in L2 research, see Plonsky et al., 2020). Reaction time data, which can be recorded for each response made on any of these decision tasks, provide a measure of processing speed - the length of time a participant needs to make a decision about a stimulus. In addition to measuring processing speed, researchers have indexed processing efficiency by examining relationships between a participant’s mean reaction time and the standard deviation of their reaction time (coefficient of variation, e.g., Segalowitz & Freed, 2004), as well as relationships between reaction time and performance accuracy (e.g., Grey et al., 2015; Sunderman & Kroll, 2009). Essentially, shorter reaction times without (a) increased variability in reaction times or (b) decreases in accuracy are interpreted as indicative of more efficient processing.

Another commonly used L2 performance measure particularly well suited to the collection of reaction time data is a picture selection task. In this task, participants are presented with a sentence (aurally) and instructed to select the picture that best matches the stimulus from two or more images as quickly as possible. Critically, the stimuli are designed to include a (informative) condition that contains linguistic cues that allow participants to potentially differentiate the target from the distractor image(s) before the end of the sentence (e.g., singular vs. plural determiners allow participants to select between images that contain singular or plural objects). Reaction time data recorded based on participant’s image selection thus reveal whether learners process and use the linguistic cue in language comprehension (for additional information about picture selection tasks, see Gerken & Shady, 1996).
Other behavioral methods have been designed to specifically examine language processing through reaction time measures. Self-paced reading and listening tasks are among the most widely used behavioral methods to investigate language processing. In these paradigms, participants are instructed to read or to listen to language stimuli that are presented in segments, which reflect different linguistic conditions (e.g., correct vs. violation morphosyntactic stimuli). Critically, the participant controls the rate of presentation of each segment via a button press such that the time elapsed between button presses can be recorded as a processing time measure for a given segment. Time spent reading target segments, as well as comparisons between segments from different linguistic conditions, can indicate processing difficulties, for example, in segments that contain violations or ambiguities (for more information on self-paced reading and listening tasks, see Jegerski, 2014; Papadopoulou et al., 2014).

The word-monitoring task is another oft-used behavioral method that exploits reaction-time measures to assess L2 processing. In this task, participants are instructed to monitor written or auditory language stimuli for a particular target word and to press a button as soon as the target word is detected. Target words are embedded in a carrier sentence, and reaction times are computed from the target word onset until the participant’s button press. Experimental paradigms may manipulate the position and/or the linguistic properties of the target word (e.g., transitive/intransitive verb) as well as the linguistic condition represented by the stimuli in the carrier sentences (e.g., correct vs. violation sentences). Analyses that compare reaction times across different conditions can reveal information about various aspects of processing (for more information on word-monitoring tasks, see Kilborn & Moss, 1996; Suzuki & DeKeyser, 2015).

Finally, the priming technique can be employed to assess various L2 processing mechanisms. In general, priming experiments make use of reaction time data to explore whether
exposure to particular linguistic stimuli (the “prime”) facilitates (speeds up) or inhibits (slows down) the processing of subsequently presented stimuli (the “target”). In reaction time studies using the priming technique, participants are presented with a prime word that is manipulated based on a linguistic characteristic (e.g., semantic relatedness). Subsequently a target word is presented for which participants are generally asked to make a judgment (e.g., semantic categorization). Reaction times for participants’ decision about the target word are then analyzed based on the different prime conditions in order to probe whether the prime affected processing of the target word (for a review of priming methods in L2 research, see Marsden, 2009; McDonough & Trofimovich, 2009).

In all, researchers can collect and analyze reaction-time measures for responses on various behavioral tasks to gain insights into language processing that performance measures alone may not reveal. Because these methods generally only require a computer and may be able to be administered over the internet, they can be used quite flexibly in different SA contexts. In addition, for some of these behavioral tasks, latency measures other than the reaction time of a response can be collected to examine L2 processing. For example, mouse tracking can be used to obtain detailed information about the time-course of processing as a decision is being made (Freeman, 2018). Because mouse tracking follows similar logic to eye tracking, we provide a brief overview of mouse tracking in that section.

2.1. Study Abroad Research Using Behavioral Methods

Researchers have made use of reaction time data to examine the role of SA in L2 processing using both single-session (Sunderman & Kroll, 2009) and pre-post experimental
designs (Grey et al., 2015, Segalowitz & Freed, 2004; Taguchi, 2008). These SA studies have employed some of the behavioral methods examined here in order to explore grammar processing (Grey et al., 2015), lexical access (Grey et al., 2015; Segalowitz & Freed, 2004; Sunderman & Kroll, 2009), and speed of pragmatic comprehension (Taguchi, 2008).

Sunderman and Kroll (2009) assessed speed of semantic production and comprehension among learners with and without SA experience. The researchers employed picture naming and translation-recognition tasks to measure L2 performance and processing, and also assessed working memory capacity via a reading span task. The two participant groups comprised native speakers of English majoring in Spanish at a U.S. university and studying at a similar curricular level; one group reported no prior SA experience, whereas the other reported an average of 3.8 months (range 2-10 months) of prior SA experience that took place around 12 weeks before testing. Results revealed that learners with prior SA experience exhibited a processing speed advantage, both in production and comprehension, compared to L2 learners who had not had such experience. In addition, better working memory resources predicted better L2 performance for the semantic comprehension task regardless of prior SA experience.

Using a pre-post design, Grey and colleagues (2015) found evidence of L2 processing gains for different aspects of language from the start to the end of a five-week SA experience. The researchers explored language gains in a group of advanced learners of L2 Spanish (first language [L1] English) enrolled in an intensive college-level SA program in Spain. At the beginning and at the end of their SA experience, participant accuracy and reaction times were recorded during the completion of a lexical decision task and a grammaticality judgment task (average reaction time as participants classified letter strings as words and non-words and sentences as grammatical or ungrammatical, respectively). Results revealed a significant reaction
time decrease in both L2 tasks, indicating development of faster processing of L2 semantics and grammar. Participants’ working memory was also assessed through a sentence span task and non-word repetition task, however, working memory was not found to predict the reaction time decrease in either lexical or grammatical processing.

Lexical processing development in an SA context has also been examined among intermediate-level Spanish learners studying abroad in Spain for one semester (L1 English; Segalowitz & Freed, 2004) and among beginning-level learners of L2 English studying abroad at a U.S. university for four months (L1 Japanese; Taguchi, 2008). In both longitudinal studies, a semantic classification task was used to assess speed of lexical access in the L2 (average reaction time as participants classified visually presented targets as ‘living’ or ‘non-living’). Segalowitz and Freed (2004) also included a measure of reaction time variability that controls for overall response speed to assess lexical access efficiency. The inclusion of this additional measure allowed the researchers to distinguish between quantitative changes (resulting from mere performance speed-up) and qualitative changes (resulting from automatization or restructuration of underlying processes). Finally, both studies included a self-report measure of the amount of L2 exposure outside of class. Results revealed significant gains in L2 lexical access, both in terms of a quantitative decrease in reaction times (Segalowitz & Freed, 2004; Taguchi, 2008) and a qualitative improvement in efficiency (Segalowitz & Freed, 2004). Regarding the effect of extracurricular L2 contact, results revealed that, whereas the amount of L2 contact predicted speed development (Segalowitz & Freed, 2004; Taguchi, 2008), it was not correlated with gains in efficiency (Segalowitz & Freed, 2004). The role of learning context in supporting these processing changes, however, remains ambiguous: Segalowitz and Freed (2004) included
analyses with a control, at-home group of learners and did not find evidence that gains were significantly different between learners in the SA and control groups.

In addition to lexical access, Taguchi (2008) also assessed changes in speed of pragmatic processing within the same group of SA learners. Participants completed a test in which they listened to short conversations containing implicatures and answered comprehension questions to assess inference of pragmatic meaning. Again, results revealed a significant decrease in reaction times from beginning to end of SA, indicating gains in speed of pragmatic comprehension over a 4-month immersion period. Amount of L2 contact hours reported (speaking and reading, but not listening) were found to predict gains in processing speed, such that learners who reported more time speaking and reading the L2 experienced greater reaction time decreases, and therefore, faster processing at the end of SA.

Taken together, results from the available research using reaction time data as a window into L2 processing suggests that SA can lead to increased speed of access to grammatical, semantic, and pragmatic knowledge. Specifically, research has found significant decreases in reaction times among L2 learners from beginning to advanced proficiency levels with SA experiences ranging from five weeks to four months. Results are mixed regarding the comparison of processing gains in SA and non-SA learners, with some evidence pointing to advantages for learners with SA (Sunderman & Kroll, 2009) and others pointing to a lack of difference in development (Segalowitz & Freed, 2004). Finally, studies that have examined the contribution of cognitive factors (working memory) and experiential factors (L2 extracurricular exposure) to such processing development have yielded mixed results. In conclusion, more research is needed to understand the effects of SA on the development of different processing
mechanisms involved in language comprehension and use as revealed by the tasks described above, as well as the internal and external factors that play a role in such development.

2.2. Limitations and Guiding Principles of Behavioral Methods

Behavioral measures such as reaction time and accuracy provide a convenient way of assessing L2 processing that is inexpensive, does not require special technical equipment, and can be conveniently administered onsite or remotely. These methods, nonetheless, are not without limitations.

The main constraint of behavioral methods is that they do not provide a direct measure of language processing. Rather, researchers must carefully design or modify experimental designs to isolate a linguistic phenomenon of interest and allow for the analysis of reaction time and accuracy data that purportedly capture the engagement of particular processing mechanisms. Importantly, reaction time measures can be influenced by a number of external factors that may lead to timing effects that do not reflect the manipulation of interest. Indeed, the validity of reaction time measures is dependent on participant’s attention and focus on the task (Jiang, 2012). In addition, variability among speakers in their overall processing efficiency can also impact the timing and the size of effects (Kaan et al., 2015). Regarding stimuli design, linguistic factors such as word length, frequency, or degree of regularity, as well as other subjective factors such as imageability or concreteness can affect speed of word access and, therefore, have an impact on overall sentence processing (Harley, 2013). Another limitation is that some of these behavioral techniques reflect unnatural instances of language processing, both in terms of stimuli design and task performance. Specifically, low ecological validity could be an issue particularly
for tasks that require processing of sentences that are created to be ambiguous or that contain syntactic structures that appear with low frequency in a given language. Furthermore, all the tasks described in this section require some form of response from the participant (e.g., linguistic judgment, picture selection, button press to unmask the following segment of a sentence), which is an unnatural way of comprehending language.

In light of these limitations, we offer some guiding principles for researchers interested in using behavioral methods to capture L2 processing in the SA context. First, regarding experimental design, a critical consideration is to minimize the effect of factors that can distort reaction time measures, such as participant attention to the task. The inclusion of a comprehension question following each (or some) of the experimental items has been shown to increase participants’ engagement throughout the task. Indeed, it is a common practice to analyze only those trials for which the follow-up question response was correct. In order to minimize unintended impacts related to linguistic traits of the stimuli, experimental designs should (a) control for word length, frequency, and regularity of the lexical items, and (b) counterbalance lexical items within critical conditions as well as sentences across subjects and lists (e.g., Marsden, Thompson, & Plonsky, 2018). Additionally, norming studies and pilot testing to determine, for example, that the effect of interest is evidenced among native speakers, are fundamental in establishing the validity of the experimental design. Regarding data analysis and interpretation, tasks that rely on reaction time analyses can yield particularly nuanced and noisy data. Thus, informed decisions need to be made regarding data cleaning procedures, such as the identification, removal, and replacement of extreme and/or missing data points in order to minimize the influence of extraneous variables and maximize the validity and accuracy of the results (for discussion, see Nicklin & Plonsky, 2020). Given the labor and effort associated with
creating experimental designs that isolate various processing mechanisms behaviorally, we encourage researchers to make their instruments, data, and analyses publicly available. Open research practices contribute to replicability, helping to validate experimental methods and enhance methodological transparency and overall, the quality of research (Marsden et al., 2018).

3. Eye tracking

Eye-tracking methods allow researchers to examine the real-time cognitive processing that may occur before a response is made, or even in the absence of a response, through the use of specialized equipment to detect and record an individual’s eye movements as they engage with stimuli (Conklin et al., 2018; Godfroid, 2020). Eye-tracking methods have been increasingly employed by researchers to examine a range of questions about L2 processing and representation (Godfroid & Hui, 2020). Eye tracking is well suited to examine the processing that underlies written and spoken language comprehension but can also be used to examine written production (Marijuan & Sanz, 2017). Two overarching eye-tracking paradigms are utilized in L2 research (Godfroid, 2020): (a) a text-based paradigm, in which written linguistic stimuli (e.g., words, phrases, etc.) are presented to participants, and (b) a visual world paradigm, in which aural linguistic stimuli are presented along with visual images. Both paradigms are generally premised on the idea that the eyes and the mind are linked in such a way that an individual’s eye movements reflect moment-to-moment cognitive processes (Rayner, 1998), with different cognitive models offering more specifications about the influence of higher-level (cognitive and linguistic) vs. lower-level (oculomotor) factors in eye-movement data (for more on the models, see Godfroid, 2020). In both paradigms, eye movement data are recorded as participants interact with the stimuli (e.g., for comprehension or to identify words vs. non-words). Data tied to the critical stimuli for the paradigm are then analyzed based on the particular
research question. Notably, the paradigms can be leveraged to answer different kinds of research questions. For example, text-based paradigms can be used to examine questions about the role of attention in L2 learning, and visual world paradigms can be used to address questions about whether L2 learners are able to engage in prediction processes.

Eye-tracking measures are collected by setting up a participant with an eye-tracking device, often a head-, desk-, or tower-mounted eye tracker that consists minimally of a camera capable of recording eye movements and specialized software to record and analyze the eye-movement data. Some eye trackers are designed to be portable, which allows researchers to collect data at an SA site away from their home institution. Stimuli are often presented to participants on a computer monitor but can also be presented ‘live.’ Importantly, before (and often at one or more points during an experiment), researchers conduct a calibration procedure to ensure that the eye tracker is recording the spatial location of the eye gaze correctly.

Once a participant is set up with the eye tracker and calibration has been performed, the eye-tracking experiment can begin. Researchers will need to have made several decisions about the experimental design involving the stimuli (including its location and size on the computer monitor, for example) and about the spatial areas of interest from which recorded eye-movements will be deemed relevant for analysis (Godfroid, 2020). Researchers will also have to decide which eye-movement measures to analyze. Measures can include (a) fixations on and skips over stimuli (including the count, duration, latency, and location of the fixation), (b) regressions back to particular stimuli, as well as (c) more integrated measures such as heat maps or scanpaths (Godfroid, 2020). Very generally, measures based on the initial or early fixations and skips are interpreted as reflecting more automatic processes whereas later and total fixation and regression measures are interpreted as reflecting more controlled processes. Finally,
researchers will need to establish procedures for data cleaning and analyses. As can be seen, the researcher degrees of freedom, that is, the number of methodological and analytical decisions that a researcher makes when conducting a study, are quite large in eye-tracking research. The ‘correct’ decision will vary widely and will depend on the research questions and experimental paradigm (Godfroid & Hui, 2020), and it is recommended that researchers pre-register their decisions or submit their study as a Registered Report in order to minimize even unintentional bias on the part of the researcher or during the publication decision process (Marsden et al., 2018).

3.1. Study Abroad Research Using Eye tracking

What does eye-tracking research reveal about L2 development in SA contexts? Although some eye-tracking studies have examined L2 (and L1) processing in immersion contexts (e.g., Dussias & Sagarra, 2007; Foucart & Frenck-Mestre, 2011), to our knowledge, to date, only two eye-tracking studies have examined SA specifically (LaBrozzi, 2012; Sagarra & LaBrozzi, 2018), both of which focus on L2 morphosyntax and individual differences in cognitive abilities.

First, LaBrozzi (2012) examined whether studying abroad would affect the extent to which learners with a morphologically poor L1 (English) would attend to morphology when processing a morphologically rich L2 (Spanish), and whether processing would be mediated by inhibitory control. Participants in this study were learners enrolled in seventh- or eighth-semester Spanish courses at a U.S. university who had one semester of SA experience (approximately three months prior to the experiment) or who had not studied abroad (non-SA learners). Eye-movement data were recorded as these learners read correct or violation adverb-verb tense agreement sentences (e.g., Spanish versions of *Yesterday the professor cooked... vs *Yesterday
the professor cooks...), with the adverb positioned either before or after the verb. Learners also completed a measure of inhibitory control. Analyses were conducted based on fixations on and regressions to the target adverbs and verbs as well as the words immediately following the target words. Results based on learners’ first fixation on target words suggested that both SA and non-SA learners were sensitive to tense violations, as they fixated longer on words that followed violations on target adverbs or verbs. Results for total reading time (i.e., the sum of the first fixation duration and the regression durations) suggested that both SA and non-SA learners attended to adverbs more when there was a tense violation in the sentence. Interestingly, only the SA learners showed longer total reading times for the target verb when there was a tense violation. In regard to inhibitory control, it did not account for processing differences in either learner group. LaBrozzi interpreted this pattern of results as indicating that, although both groups were sensitive to tense violations, only the SA group showed evidence of attending to verbal morphology when processing the violations and this difference was not explained by individual differences in cognitive abilities.

Examining a different morphosyntactic structure and cognitive ability, Sagarra and LaBrozzi (2018) explored the effect of SA on L2 processing and asked if processing might be mediated by working memory. Similar to LaBrozzi (2012), the researchers tested L2 learners enrolled in sixth-semester Spanish classes at a U.S. university who either had or had not studied abroad to probe how they processed subject-verb agreement in L2 Spanish (Spanish versions of ...the patients smoke... vs. *...the patient smoke...). SA and non-SA learners read correct and violation sentences as their eye-movements were recorded. Additionally, working memory was assessed with a complex reading span task. Analyses were conducted based on the first fixation duration and the total reading time on the target subject, verb, and the word immediately
following the verb. The results revealed a main effect of violation for total reading time on the verb that was qualified by an interaction between violation and group. Follow-up analyses demonstrated that the SA group was more sensitive to subject-verb violations. For working memory, no statistically significant main effects or interactions were evidenced, but pairwise comparisons suggested that high but not low working memory SA learners were sensitive to the subject-verb agreement violations.

Taken together, these studies suggest that studying abroad may have a facilitative effect on the processing of L2 morphosyntax with some indication that the effect might be mediated by individual differences in cognitive abilities. Importantly, they demonstrate the potential of eye-tracking studies to reveal how SA may affect processing ‘under the hood’ that is not apparent in accuracy, proficiency, or fluency performance measures. Thus, further eye-tracking research might reveal the processes and mechanisms of L2 learning as conditioned by a particular context and how these processes and mechanisms might lead to increased L2 performance.

3.2. Limitations and Guiding Principles of Eye-tracking Methods

Eye-tracking research methods hold several advantages over other techniques for examining real-time processing, including the fact that they provide high-resolution eye-movement data related to relatively natural presentations of written or aural stimuli. There are also limitations associated with this methodology, several of which are issues of practicality. First, the cost of securing the equipment and software required to conduct an eye-tracking study can be prohibitive. However, Godfroid (2020) offers some creative solutions, such as renting an eye-tracker or volunteering in an established eye-tracking lab as a means of conducting initial eye-tracking research to put oneself in a stronger position to obtain an eye-tracker through a
successful grant application or start-up package. Another potential challenge is securing the training to develop the necessary technical skills to design, conduct, and analyze an eye-tracking study. Again, this challenge is not insurmountable, as there are more and more resources available including (a) Godfroid’s (2020) comprehensive book geared toward second language and bilingualism researchers, (b) other publications with recommendations for eye tracking (e.g., Conklin et al., 2018; Fiedler et al. 2019; Godfroid & Hui, 2020), and (c) workshops offered by professional organizations and by companies that sell eye-tracking equipment (a simple internet search for eye-tracking workshops brings up many possible opportunities). In addition, with the movement for replication and transparency in research (e.g., Marsden et al., 2018), researchers should have more access to publicly posted stimuli sets, protocols for data cleaning, and scripts for analyses that will allow for replications in the field and that can more generally serve as standard templates for new research projects.

Conducting eye-tracking research for SA offers its own set of practical challenges, including the ability to collect data at the SA site. However, there are also solutions to this issue: portable eye-trackers are available that can be transported to the site, or investigators can work to develop relationships and collaborations with researchers who have established eye-tracking laboratories at institutions at or near the SA program sites or who have access to portable eye-trackers. Finally, if researchers are not able to collect eye-tracking data at the SA site, they can develop studies to test participants at their home institution before and after the SA program. Such designs could lead to other issues though, as it may be particularly difficult to test and retain enough participants to obtain sufficient statistical power for one’s study. Yet, possible solutions exist. For example, researchers can collect data over multiple semesters, or they can
develop multi-site collaborations with other researchers with similar interests and resources (e.g., Issa et al., 2020; Zalbidea et al., 2020).

Although we do not discourage researchers from pursuing eye-tracking methodologies, if the practical limitations of the methodology are prohibitive, alternate methods to address their research questions can be considered. Mouse tracking (Freeman, 2018) is one method that might allow researchers to approximate some of the eye-tracking paradigms, such as the visual-world paradigm, by providing a means of tracking participants’ mouse trajectories. Like eye tracking, mouse tracking can provide dynamic and detailed information about the timing of decision-making and resolution of competing information. Furthermore, software that allows researchers to set up, run, and analyze a mouse-tracking experiment is freely available (Freeman & Ambady, 2010). Several other behavioral methods can also be informative to questions about real-time L2 processing (see Behavioral Methods section above).

4. Event-Related Potentials

Similar to eye-tracking methods, event-related potentials (ERPs) allow researchers to examine real-time cognitive processing that may occur even before a response is made or in the absence of a response. However, ERPs more directly reflect neurocognitive processing because they are a measure of the electrophysiological brain activity associated with a particular event or stimulus (Luck, 2014). They offer excellent temporal resolution and have been used extensively to examine the real-time neurocognitive processes underlying language comprehension in a wide range of speaker and learner populations. In this technique, researchers make use of electrodes placed at specific locations on a participant’s scalp to detect, amplify, and record the continuous signal of electrical potentials (electroencephalogram, EEG) produced by the human brain. For ERP analyses, EEG data that is time-locked to specific events (e.g., target lexical items or
grammatical structures) is extracted and then averaged across stimuli of the same type. This procedure results in a waveform that represents the neurocognitive processing associated with the particular type of stimuli of interest. Waveforms are characterized by their latency (i.e., their timing relative to the onset of the target stimuli), their polarity (i.e., whether the voltage is positive or negative), their amplitude (i.e., the size of the voltage measured in microvolts), and their spatial distribution (i.e., the location on the scalp where the signal is detected). By comparing ERP waveforms elicited by different conditions (e.g., correct versus violation stimuli), different participant groups (e.g., learners with or without SA experience), or at different time points (e.g., before, during, and after SA), researchers can make inferences about quantitative and qualitative differences in the neural response to cognitive events. For example, differences in latency and amplitude suggest quantitative differences in timing and in the degree to which a particular process is engaged, respectively. Differences in spatial distribution or polarity may reflect the engagement of qualitatively different processes.

The processing signatures associated with different types of linguistic stimuli (e.g., lexical, syntactic, morphosyntactic) are fairly well documented in the ERP literature for both L1 and L2. For example, two important components in language research, both typically detected among electrodes in centro-posterior regions of the scalp, are the N400 and the P600. The N400, often elicited in response to semantic violations, is a negative deflection that occurs around 300-500 ms after the onset of the target word. One prominent interpretation of the N400 component is that it reflects the process of accessing meaning (Kutas & Federmeier, 2011). The P600, often elicited in response to grammatical violations or complexities, is a positive deflection that occurs around 600-900 ms after the onset of the target word. A general interpretation of the P600 component is that it indexes the process of structure building (e.g., Swaab et al., 2012; for an in-
depth review of these ERP language components attested in L1, see Swaab et al., 2012; and for L2, see Morgan-Short, 2014; Morgan-Short et al., 2015).

In L2 research, ERPs are often analyzed at the group level by averaging across individuals. Although any reliable difference between waveforms elicited by different linguistic conditions (e.g., semantically congruous vs. anomalous words) can be meaningfully interpreted based on the design of a study, participant groups are said to exhibit a particular ERP component (e.g., an N400) when the group-averaged waveform is consistent with the timing, polarity, and distribution of that component established through previous ERP research. More recently, researchers have begun to examine individual participant’s ERP processing signatures by calculating the overall size of neural response (e.g., the amplitude difference of the waveforms for congruous vs. anomalous words) within a particular region of interest (e.g., central-posterior group of electrodes to examine N400-magnitudes) and/or by using an index of response dominance (e.g., the negative or positive polarity of the overall difference between congruous vs. anomalous words; for more information about these measures, see Tanner et al., 2014). Although there is ample evidence that L2 processing signatures change with increases in proficiency and performance (e.g., Steinhauer et al., 2009), research has also demonstrated that learners trained under different conditions (explicit, implicit) may evidence different processing signatures even when behavioral performance does not differ (e.g., Morgan-Short et al., 2010; Morgan-Short et al., 2012) and that individual differences (e.g., motivation in Tanner et al., 2014) may play a role in individuals’ processing signatures.

In all, ERPs provide a non-invasive, low cost (in comparison with other neural techniques) method of assessing neurocognitive processing. Although researchers often collect EEG data while participants complete a language task (e.g., acceptability judgment task), ERP
methods do not require participant responses to stimuli, thereby making them an option for a wide range of experimental paradigms and participant populations. Used in both cross-sectional and longitudinal designs, ERP data can offer useful insights for exploring differences both between and within groups and can thus provide a method of examining the neurocognitive processes that underlie language comprehension and development.

4.1. Study Abroad Research Using Event-Related Potentials

Very few ERP studies have directly examined research questions related to SA. However, two general approaches have been used that are informative to questions about L2 processing as affected by SA. One approach to explore the effects of SA on L2 processing is to include prior SA experience in selection criteria for study participants. Utilizing this approach, Bowden and colleagues (2013) collected EEG data during the completion of a grammaticality judgment task to examine processing of semantic and syntactic (word order) violations in Spanish. Participants included a group of advanced L2 learners who all reported one to two semesters of previous SA experience (L2 Advanced), a group of low-intermediate L2 learners without previous SA experience (L2 Low) and a native speaker control group. Results revealed that semantic violations reliably elicited N400 responses in all three groups, suggesting that all participants accessed meaning through similar processes. For syntactic processing, qualitative differences were evidenced between the L2 Low group, on the one hand, and the L2 Advanced and L1 groups, on the other. Specifically, the L2 Advanced group evidenced a biphasic neural response that reflected the pattern evidenced by the L1 group (i.e., an early anterior negativity followed by a P600), whereas the L2 Low group showed an early anterior positivity as well as a broad, later
positivity. These results suggest that L2 syntactic processing can approximate L1 processing patterns for learners with sufficient proficiency and exposure, and perhaps specifically for those with immersion experience that can be gained through SA. Future work that examines processing among learner groups who are matched for proficiency, but differ in terms of SA experience, could utilize post-SA ERP testing to further elucidate the role of SA in L2 processing.

Another way to more directly examine the effects of SA on processing is to utilize a longitudinal design to examine whether (and how) processing may change over the course of an SA experience. Taking this approach, Faretta-Stutenberg and Morgan-Short (2018a, 2018b) explored changes in ERP signatures from pre- to post-SA to examine the processing of (morpho)syntax among L2 learners of Spanish completing grammaticality judgment tasks. For these studies, L2 processing among intermediate-level university learners (studying at the fifth semester level or above) was assessed before and after a semester-long SA experience. Faretta-Stutenberg and Morgan-Short (2018a) examined the processing of grammatical gender agreement on articles and adjectives, exploring relationships between pre-SA proficiency and L2 contact during SA. In group-level analyses of N400- and P600-magnitudes, the researchers reported no statistical changes from pre- to post-SA for either condition. They noted, however, the presence of substantial variability in individual learner processing at each testing session, as well as in how individual processing signatures changed from pre- to post-SA, suggesting a range of processing strategies and efficiency levels within the group. Neither of the individual difference measures (pre-departure proficiency, reported L2 contact) were significantly related to changes in neurocognitive processing signatures for either condition. The researchers did, however, find relationships with N400- and P600-magnitudes at post-SA testing. Specifically, a significant correlation between reported L2 contact and post-SA article processing indicated that
learners who reported more time speaking Spanish while abroad evidenced larger, more positive neural responses in the P600 time window (and smaller, more positive neural responses in the N400 time window), suggesting more reliance on structure building, as opposed to meaning-based, processing.

Faretta-Stutenberg and Morgan-Short (2018b) examined syntactic processing (word order violations) and relationships with cognitive abilities (declarative and procedural learning ability, working memory). Participants exhibited changes in neurocognitive processing signatures from pre- to post-SA, with analyses for the full group of participants revealing an anterior positivity to syntactic violations at pre-testing and an N400 at post-testing. A negative correlation between individual N400- and P600-magnitudes suggested that individual learners primarily exhibited either a negativity- or positivity-dominant neural response. Subgroups based on learner response dominance (negativity or positivity) were thus created in order to explore whether processing differences present in subgroups were being obscured in full group-level analyses. Indeed, analyses of these subgroups showed (a) clear N400s at pre- and post-testing for negativity-dominant learners, indicative of reliance on meaning-based processing strategies, possibly chunking, (b) an anterior positivity at pre-testing, and (c) a clear P600 at post-testing for positivity-dominant learners, indicative of reliance on structure building processes. In terms of relationships with cognitive abilities and changes in individual ERP signatures, analyses demonstrated that working memory and procedural learning ability were significant predictors of an increase in overall response magnitude (regardless of response type) from pre- to post-SA. This relationship suggests that individual differences in cognitive abilities may have facilitated learner processing of L2 input during SA, leading to a greater neural sensitivity to syntactic violations.
Taken together, the extant research that has employed ERP methods with SA learners indicates that neurocognitive processing signatures (both size and type of response) are likely to change with SA experiences. Importantly, the limited existing research reveals considerable individual variability in processing strategies, and suggests that individual differences in experiential, cognitive, and affective variables are likely to play a role in accounting for this variability.

4.2. Limitations and Guiding Principles of Event-related Potential Methods

The potential insights into L2 acquisition and processing afforded by ERPs are substantial, particularly when utilized within a multidimensional approach. The technique, however, is not without its challenges and limitations. As with eye-tracking research, there are several practical challenges involved in collecting data using specialized technology and equipment at an SA site. The suggestions for creative solutions outlined in Section 3.2 are also applicable here, particularly in regard to securing portable EEG equipment, and perhaps more practically, establishing collaborations with other researchers that have shared interests by complementary resources. However, researchers should also be aware of and account for a number of challenges and limitations that are specific to ERP methods.

First, a number of issues such as participant muscle and eye movement or fatigue can lead to noise and artifacts in the EEG data. In order to combat potential data loss and to obtain sufficient power in general, ERP experimental designs require the inclusion of many trials (thus, experiments can be quite long) and written target stimuli must be presented one word at a time in the center of the computer screen (to limit horizontal eye movement). To lessen participant fatigue, trials can be divided into smaller blocks, with brief break periods between blocks.
Additionally, it is helpful to allow participants to advance to the next trial at their own pace, signaling their readiness via mouse click or other input device. This allows the participant to take an eye rest as needed and provides a convenient opportunity for the researcher to correct any technological or electrode issues that may arise over the course of the experiment.

Another limitation of ERP methods is the challenge of arriving at a clear understanding about the functional interpretation and the spatial resolution of ERP results: An averaged waveform is likely to reflect multiple psychological or neural processes, making it difficult to identify a specific underlying process that might be represented or the specific brain area(s) that generate an ERP effect (Kappenman & Luck, 2012). In order to make claims about the functional significance of a particular component or response, one must rely on previous research in which components are repeatedly identified in association with particular cognitive processes and remain abreast of current discussions in the literature regarding the functional significance of the components (e.g., Kutas & Federmeier, 2011; Osterhout et al., 2012; Swaab et al., 2012). Although there are limitations and challenges to ERP research, a number of excellent resources are available to help guide researchers interested in utilizing ERP methods in their work (e.g., Luck, 2014).

5. Future Directions

Although there is limited empirical work that has examined L2 processing for SA learners, the extant literature makes clear the potential for real-time processing data (whether collected via behavioral, eye-tracking, or ERP methods) to provide insights into how the processes involved in L2 comprehension and development may be influenced by learning
context. In order to better understand the role of SA in the development of L2 processing mechanisms, future research should examine not only additional linguistic targets, proficiency levels, SA durations, and language pairings, but also the internal and external factors that may contribute to individual differences in the development of language processing over the course of various SA experiences. Indeed, psycholinguistic research has provided mounting evidence that variability in cognitive and linguistic resources, as well as language experience, influence the real-time engagement of language processing assessed using the techniques included in this chapter (e.g., Faretta-Stutenberg & Morgan-Short, 2018a, 2018b; Finestrat, 2021; Sagarra & LaBrozzi, 2018; Segalowitz & Freed, 2004). As such, further research should be conducted to understand how various individual differences interplay with learning context both to help understand the development of language processing.

Of the studies that have utilized behavioral, eye-tracking, and ERP methods to examine the impact of SA on L2 processing, researchers have either compared processing among groups of learners with and without prior SA experience or have made use of pre-post designs that allow for the measurement of changes from the beginning to the end of a sojourn. Thus, an important future direction for longitudinal research is to incorporate multiple testing sessions during SA to obtain more fine-grained insights into how processing might change and develop over the course of time abroad. Furthermore, in order to understand the durability of processing strategies learners may develop while immersed, researchers are encouraged to include delayed testing sessions after the completion of SA.

A benefit of the methods for measuring processing described in this chapter is that they are often used together with L2 performance measures, streamlining experimental designs and allowing for the triangulation of data. Moreover, the combination of multiple sources of
processing data (e.g., both reaction time and eye-tracking data) along with performance data in future research will advance our understanding of the underlying processes that contribute to L2 comprehension and development. Such experimental paradigms should be feasible given that a number of the studies presented here made use of very similar experimental (violation) paradigms but utilized different techniques to assess L2 (morpho)syntactic processing in SA (e.g., reaction time: Grey et al., 2015; eye tracking: Sagarra & LaBrozzi, 2018; ERPs: Bowden et al., 2013). Although some stimuli and parameters would need to be adjusted for particular techniques, it would be possible to collect within-subject data using two or more processing measures in addition to examining performance data. Such an approach could elucidate the complex dynamics that contribute to the development of L2 processing and knowledge during, and as a result of, SA experiences.

A final suggestion for future research is to leverage the processing measures described in this chapter to examine the effects of SA on L1 processing (e.g., Baus et al., 2013; Linck et al., 2009). In addition to serving as a relevant research question in their own right, changes in L1 processing, including flexibility and inhibition, may have a facilitative effect on L2 acquisition and processing in a number of domains (e.g., Bogulski et al., 2019).

As researchers address the suggestions above, there are a number of ethical considerations they should bear in mind when designing and conducting future research. First is the inclusion of diverse participant groups, including a variety of language backgrounds and experiences (e.g., bilinguals, multilinguals, heritage speakers of the target language, learners with a variety of L1s). Importantly, researchers should consider the potential impact of language background and experience on experimental task performance. The use of oral stimuli and tasks can help mediate these impacts and ensure that researchers are able to examine language
processing among participants with more experience in the oral/aural rather than written domains. Additionally, physical characteristics of participants such as eye size or hairstyle can pose challenges for collection of usable eye-tracking and ERP data. Researchers should continue to appeal to companies to provide equipment that allows investigators to obtain valid data across the largest possible spectrum of participants.

Another important consideration is the adoption of open research practices, which serve to improve methodological rigor by increasing transparency and replicability. Such practices include making instruments, data, and analyses available to other researchers (Marsden et al., 2018). These efforts are critical to the advancement of ethical methodology in SA research that includes processing measures.

6. Conclusion

This chapter presented an overview of methods and technologies that can be used to examine the development of L2 processing for SA learners. We also presented evidence based on these technologies that the SA experience results in changes to semantic, (morpho)syntactic, and pragmatic processing, in terms of speed of processing, attention to syntactic cues, and neurocognitive processing signatures. This processing benefit of SA would have remained obscure if researchers had exclusively measured L2 fluency, accuracy, and knowledge via performance measures. As this chapter highlighted, many of the techniques available to capture real-time L2 processing do not require the development of new independent experimental paradigms, but rather, are implemented through the modification of existing tasks to allow for the collection of reaction time, eye-tracking, or ERP data. Given the theoretical relevance of exploring the benefits of SA on L2 processing and the practicality of the available techniques, we
conclude by encouraging researchers to consider incorporating the use of such techniques in their research, possibly in conjunction with other performance measures. Such research will provide a holistic view of L2 development and the effects of SA experience.
References


https://doi.org/10.1111/j.1467-9922.2009.00554.x


https://doi.org/10.1162/jocn_a_00119


https://doi.org/10.1017/S0267190520000057


https://doi.org/10.1017/CBO9781139029377.025


handbook study abroad research and practice (pp. 149–164). Routledge.

https://doi.org/10.4324/9781315639970


https://doi.org/10.1017/S0142716408090048


https://doi.org/10.1111/lang.12138


https://doi.org/10.1093/oxfordhb/9780195374148.013.0197

