

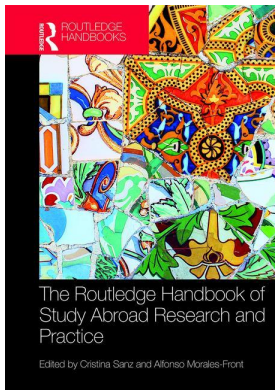
This article was downloaded by: 10.3.98.93

On: 22 Sep 2019

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



The Routledge Handbook of Study Abroad Research and Practice

Cristina Sanz, Alfonso Morales-Front

Quantitative Approaches for Study Abroad Research

Publication details

<https://www.routledgehandbooks.com/doi/10.4324/9781315639970-3>

Sarah Grey

Published online on: 26 Jun 2018

How to cite :- Sarah Grey. 26 Jun 2018, *Quantitative Approaches for Study Abroad Research from: The Routledge Handbook of Study Abroad Research and Practice* Routledge

Accessed on: 22 Sep 2019

<https://www.routledgehandbooks.com/doi/10.4324/9781315639970-3>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Quantitative Approaches for Study Abroad Research

Sarah Grey

Introduction

Stakeholders in the study abroad (SA) field—students, international program offices, program leaders, parents, and teachers—often want to “see” the results of the extensive time and resources that are invested in the experience. From a linguistic perspective, evidence of the effectiveness of studying abroad is generally related to quantifiable development in target foreign language abilities.

This chapter discusses current quantitative approaches in SA research. Such approaches allow researchers to measure particular aspects of linguistic knowledge and development: for example, in oral fluency, speed, and accuracy of accessing words in a foreign language or sensitivity to target language grammar structures. Thus, these approaches are able to provide measurable insights into the effects of studying abroad on foreign language abilities.

The chapter begins with a review of the main SA research design options and a discussion of their advantages and limitations. Following this, I provide a brief review of global quantitative methods, that is, methods that provide broadly measured evidence of the effects of SA. Then, I discuss specific quantitative approaches, such as measures of response time (RT) and brain wave activity, which reveal more detailed information about the linguistic underpinnings of SA.

Designs in Study Abroad Research

One approach for measuring the efficacy of SA is providing evidence that it confers linguistic benefits that are not otherwise realized in a matched classroom setting. In this vein, researchers employ between-subjects designs that compare an SA group with an at-home (AH) group (e.g., Isabelli-García, 2010; Segalowitz & Freed, 2004). This approach operates on the assumption that AH serves as the experimental equivalent of a control group. Intuitively, AH vs. SA designs make sense. Because the two contexts are so different from one another—for example, in the amount, type, and frequency of target language comprehension and production opportunities—any

differences in language outcomes between the two groups seem reasonably due to the quality/quantity of foreign language exposure and use in SA. The quality/quantity differences in AH/SA are highlighted in descriptions of the two contexts (e.g., Collentine & Freed, 2004) and often underscored in explanations for observed gains reported for SA compared to AH (for a review, see Llanes, 2011).

However, methodologically, between-subjects AH vs. SA designs introduce a number of confounding learner-level variables. Although some researchers attempt to limit preexisting differences between AH/SA groups by administering preprogram measures (e.g., Isabelli-García, 2010), other studies administer no preprogram measures at all, and learners may be tested several months post-SA (e.g., LaBrozzi, 2012; Sunderman & Kroll, 2009). The core limitation in AH vs. SA designs is that students who elect to study abroad are likely to differ from their AH counterparts in a number of important factors—such as motivation, aptitude, or attitude—and this, compounded with self-selection bias for SA, means that it is not just the SA *context* that differs from AH but the SA *learner* as well. Currently, AH vs. SA designs cannot avoid these potentially confounding factors, even with preprogram measures. This makes considering AH a reliable comparison group for SA difficult (for related discussion, see Rees & Klapper, 2008).

An alternative SA design employs within-subjects comparisons of pre- and post-program SA but without comparison to an AH group (e.g., Grey, Cox, Serafini, & Sanz, 2015). This allows researchers to enhance knowledge about the efficacy of studying abroad without introducing confounding factors of an AH comparison. From a within-subjects perspective, preprogram measures serve as an experimental baseline and postprogram measures reveal whether SA learners advanced from their own baselines. The aim of this design is to closely characterize the effects of the SA context on linguistic development, rather than compare different contexts.

Regardless of whether researchers choose a within-subjects SA focus or a between-subjects AH/SA approach, the next consideration is *when* learners are tested. SA research may be conducted in situ: for example, just after learners arrive to the SA setting and again before they leave (e.g., Grey et al., 2015; Isabelli-García, 2010; Segalowitz & Freed, 2004). This allows researchers to control for non-SA target language exposure or practice that might occur if learners are tested weeks/months before beginning SA or after returning from SA. In situ within-subjects designs are arguably the experimental ideal for SA research, but they are not always feasible due to scheduling or other programmatic factors. Additionally, researchers may not have the necessary testing equipment or materials in the SA setting.

Rather than conduct the study in situ, researchers may opt to test learners prior to their SA departure and after their SA return (e.g., Farett-Stutenberg & Morgan-Short, 2017) or focus on testing learners after their SA experience (e.g., LaBrozzi, 2012; Sunderman & Kroll, 2009). The approach of testing learners after their SA return is subject to at least two critical limitations. First, learners are immersed in their native language environment at time-of-testing. This is qualitatively and quantitatively different from the target language immersive context of SA and therefore is likely to affect the study's outcomes. Second, the time-lapse between SA and testing (which in some studies has been up to three months), paired with potential target language exposure/practice in the interim, makes it very difficult to directly relate the study's outcomes with SA.

Overall, each of these SA research designs has advantages and limitations. The specific design employed in any given study will depend on a number of factors, including the research questions, access to equipment/materials, and program structure and length. Critically, researchers must be cognizant of these design limitations when motivating their study and especially when interpreting SA outcomes.

Global Quantitative Measures

Global measures of language development are valuable in providing a broad metric for the effects of SA. Additionally, due to the standardized nature of many global measures used in SA research, the outcomes can be effectively interpreted across studies to better understand similarities or differences in results. This section discusses two main global measures, one production-based and the other survey-based, both of which have been fruitful in SA research.

Oral Proficiency Interviews

Due to increased opportunities to interact in the target language, SA is often considered to be a catalyst for promoting development in language production abilities. Indeed, across SA research, its most consistent effect has been found to be broad gains in oral fluency and proficiency (Llanes, 2011).

Language production abilities are reliably elicited with *Oral Proficiency Interviews* (OPIs), which are a popular tool in SA research (e.g., Hernández, 2010; Isabelli-García, 2010). They can be administered via in-person interviews, a computer avatar, or a (prerecorded) simulated interview (SOPI) composed of pictures and situational prompts. (Avatar-delivered OPIs and SOPIs have the advantage of controlling for potential interviewer bias.) (S)OPIs are useful because they can provide a number of data points from which to assess oral abilities. First, (S)OPIs produce generalizable proficiency ratings that are interpretable across different studies. Moreover, the acquired oral data can be coded for a number of additional variables: for example, conversational turn-taking behavior and length, fluency (i.e., speech rate, filled pauses, syllables per minute, longest fluent run), lexical and grammatical accuracy, speech complexity (T-units or C-units), and creativity (i.e., lexical diversity). These additional variables for oral production data can also be gathered with other measures, such as storytelling, picture description, and role-play (e.g., Allen & Herron, 2003; Arnett, 2013). This makes (S)OPIs and related oral production measures rich sources of data for quantitatively assessing oral abilities related to SA.

Language Contact Profile

SA is generally believed to provide students with greater exposure to language, for both production and comprehension, as well as in different modalities via print media, television or radio, social interactions, etc. To gather focused information on the various language exposure/interaction opportunities that characterize individual learners' SA experiences (and thus may influence SA's linguistic outcomes), Freed, Dewey, Segalowitz, and Halter (2004) developed the *Language Contact Profile* (LCP) survey.

The LCP collects information on the frequency and types of language contact that occur during SA. Researchers often administer this survey at pre-/post-SA

time points in situ (e.g., Isabelli-García, 2010; Pérez-Vidal & Juan-Garau, 2011) to assess the trajectory of potential contact-changes that take place during SA. The LCP gathers an array of language use and exposure details. For instance, it surveys where students live during SA, i.e., whether they lived with a host family or in a student dormitory, and gathers subsequent details on these living arrangements, i.e., whether the host family spoke the student's native language or whether there was a dormitory roommate who spoke the target language. It also examines students' self-assessments of how many hours-per-day and days-per-week they spend (or seek out) speaking with native speakers of the target language, reading it in various contexts (e.g., magazines, schedules, menus), listening to the language in different settings (e.g., television, songs, conversations), and writing different types of products (e.g., e-mails, homework). The LCP also assesses these activities for students' native language, which enables researchers to compare profiles of native and foreign language behavior during SA. Similar SA-based survey measures have recently been developed: for example, Mitchell, Tracy-Ventura, and McManus's (2017) *Language Engagement Questionnaire*, which may reduce the memory demands placed on learners via the LCP.

This survey-based information (as well as OPIs) provides quantifiable information on the effects of SA on language behavior. However, because global measures are, by their nature, broad, they cannot capture more detailed information on how SA impacts language knowledge and processing. To gain this insight, researchers utilize measures that are able to examine specific aspects of language.

Specific Quantitative Measures

To reveal precise linguistic information on the effectiveness of SA—for example, to study lexical processing or the development of particular grammatical structures—researchers often employ psycholinguistics-based methods. These methods are effective in elucidating the cognitive processes that are involved in language comprehension and production. They are therefore very useful in revealing the effects of SA on language processing and development. This section reviews several psycholinguistic methods that help inform interests in SA research. It begins with behavioral measures and ends with a review of a neuroscientific measure: event-related potentials (ERPs).

Decision-Elicitation Measures

Many psycholinguistic tasks elicit a decision (in response to a stimulus) from participants. Participants' responses on these tasks are typically examined in terms of ratios, raw accuracy, or discrimination ability (i.e., *A* or *d-prime* values, which help control for participant response biases; Wickens, 2002; Zhang & Mueller, 2005). From participants' responses, researchers are able to infer underlying linguistic representations or processes. For instance, a phoneme discrimination task elicits a decision about whether a pair of speech sounds are the same or different, and can be employed in SA research to examine the influence of SA on learners' phonological representations (e.g., Mora, 2008). Decision-elicitation tasks are able to measure processing across all levels of language, are well attested in the field of psycholinguistics (Traxler & Gernsbacher, 2011), and can be conveniently administered within the SA setting.

For lexical processing, decision-elicitation involves asking participants to make a decision about individual words. The accuracy (and speed; see the Latency Measures section) of participants' decision provides insight into the organization of and access to the interlanguage lexicon. For SA research, these tasks tap into whether access to words in the target language becomes more accurate with SA experience. Sunderman and Kroll (2009), for example, administered a *translation recognition task* to English native speakers who either did or did not have prior Spanish SA experience. The goal of this task is to elicit a decision on whether two words are translation equivalents of each other, i.e., whether the Spanish word 'cara' (face) is a translation equivalent of the English word 'card.' In an earlier study, Segalowitz and Freed (2004) employed a *semantic classification task* with English native speakers studying Spanish in an AH context or completing SA in Spain. In the task, participants were asked to decide if a word was living (e.g., the boy) or nonliving (e.g., a boat) to help probe speed and efficiency of lexical access due to SA. Another common decision-elicitation task for lexical processing is the *lexical decision task*. Grey et al. (2015) used this task with English native speakers completing a short-term SA program in Spain. Participants were presented with letter strings that constituted real Spanish words (e.g., 'ventana'; window) or nonwords, which are letter strings that follow the phonotactic and orthographic constraints of the language but are not real words (e.g., 'ventapa'). The aim is to decide whether each string is a word or not. Accuracy in correctly accepting words and rejecting nonwords is understood to reflect aspects of participants' lexical competence and lexical knowledge.

A widely used sentence-level decision task in psycholinguistics and Second Language Acquisition (SLA) research elicits grammaticality judgments. In *grammaticality judgment tasks* (GJTs), participants decide whether a sentence is grammatically acceptable, and sentences are designed to be grammatically well formed or not, as in (1). Using this task, researchers examine learners' sensitivity to target language grammar information.

- 1 a El lago es tranquilo por la mañana (grammatically well-formed sentence)
 - b El lago es tranquila por la mañana (error in grammatical gender, bolded)
- 'The lake is tranquila in the morning' (example from Bowden, Steinhauer, Sanz, & Ullman, 2013).

GJTs have particular promise for advancing SA research because researchers can design their sentences to examine learners' knowledge of specific linguistic structures, such as grammatical gender agreement (Isabelli-García, 2010) and syntactic word order (Grey et al., 2015). This provides precise information on the impact of SA in promoting specific linguistic development.

Elicited Imitation

Another method for examining grammatical abilities is the *elicited imitation* (EI) task (e.g., Yan, Maeda, Lv, & Ginther, 2016). EI is a very simple task: Participants listen to a sentence and are asked to repeat it verbatim. The psycholinguistic assumption behind EI data is that if participants can repeat the sentence quickly and accurately, they possess the linguistic knowledge contained in the sentence. EI can serve as a measure of global proficiency (e.g., Wu & Ortega, 2013), but it can also assess specific linguistic structures (e.g., Rassaei, Moïnzadeh, & Youhannaee, 2012).

Notably, EI is considered a valid assessment of implicit (i.e., unconscious, automatic) linguistic knowledge (e.g., Erlam, 2006; Serafini & Sanz, 2016; Spada, Shiu, & Tomita, 2015). Therefore, applying EI in SA research could elucidate not only linguistic abilities for specific grammatical structures but also help researchers ascertain the effects of SA on the development of implicit knowledge of those structures. To date, there is almost no SA research that has utilized EI (but note Mitchell et al., 2017). However, it seems a promising tool for the field, and it can easily be administered with in situ SA designs, which is an important methodological advantage.

Latency Measures

The decision-elicitation tasks described previously are often employed in tandem with experimental recording of participant RTs, the time (in milliseconds) that it takes participants to respond to an external stimulus, such as a sound, word, or sentence. When coupled with decision-elicitation, RT essentially indexes how long it takes to make and execute the decision. And when measuring both accuracy and RT, research can assess processing efficiency by examining whether there is a speed-accuracy tradeoff, that is, whether increased speed (shorter latency) is accompanied by decreased accuracy. If faster RTs are observed with no concomitant decreases in accuracy, there is no speed-accuracy tradeoff and processing is deemed efficient. Examination of speed-accuracy details can inform the effects of SA for specific linguistic structures in sentence contexts (Grey et al., 2015) and speed of lexical access (Sunderman & Kroll, 2009).

Another method for determining processing efficiency is to calculate a *coefficient of variation* (CV) from the RT data (Segalowitz & Segalowitz, 1993). In brief, the CV “reflects the relative noisiness of the processes underlying a person’s response time” (Segalowitz & Freed, 2004, p. 177). From this perspective, lower CV (and a positive RT-CV correlation) indexes change in the underlying processes, and this is interpreted as higher processing efficiency and stability (for discussions, see Hulstijn, Van Gelderen, & Schoonen, 2009; Lim & Godfroid, 2015).

Although it has not yet been applied to SA research, *mouse-tracking* is a recently developed latency measure that provides continuous measurement of participants’ decision trajectories as they make a response among multiple options on a screen. It does this by sampling the movement of the computer mouse many dozen times a second (Freeman & Ambady, 2010; Hehman, Stolier, & Freeman, 2015). This technique allows researchers to gather precise information on the onset and timing of an unfolding decision and observe online how competition among items (e.g., words or pictures) is resolved during decision-making. Mouse-tracking has recently been used to study language processing, for example morphological complexity (Blazej & Cohen-Goldberg, 2015), pragmatic intent (Roche, Peters, & Dale, 2015), and lexical competition (Bartolotti & Marian, 2012). Thus, it should be highly informative in testing pertinent questions in SA research. Furthermore, the experimental software for mouse-tracking is freely available (*MouseTracker*, Freeman & Ambady, 2010), and with a testing laptop and the appropriate software, mouse-tracking data are relatively easy to collect remotely. In fact, latency data in general are suitable for remote data collection, making these measures convenient and reliable for in situ SA designs.

Tracking Eye Movement

Eye-tracking measures naturally occurring eye behavior as it unfolds online during language processing: for example, as participants read sentences (Dussias, 2010) or look at visual scenes (note that this chapter does not discuss this visual world paradigm; for a thorough review, see Huettig, Rommers, & Meyer, 2011). Tracking eye movement provides three main dynamic measures of processing: fixations, saccades, and proportion-of-looks to regions of interest. Fixations refer to the amount of time spent on a location (e.g., a word in a sentence) and include both early and later measurements. First-fixation duration and gaze duration are considered early measurements, whereas total time spent in a location is a later measurement. Saccades are quick eye movements from one location to another. They are typically discussed in terms of forward saccades (i.e., forward movements in reading) and regressive saccades (or regressions; returns to a location). Using these measurements, researchers can test questions about the processing of specific lexical items and grammatical structures during sentence reading. There is currently little research that has applied eye-tracking to research questions for SA, likely due to its higher cost and relative immobility, that is, it cannot easily be transported to the SA setting for in situ testing. Nonetheless, as reviewed earlier, researchers may opt to test participants upon their return from SA, as in LaBrozzi (2012), who used eye-tracking to investigate the extent to which English native speakers would use morphological and lexical cues to process verb morphology during Spanish sentence reading after an SA experience.

Brain Processing: Event-Related Potentials

In the last 15 years, research in SLA has begun to use the ERP technique in order to investigate questions about adult language learning and processing from a neural perspective (for a review, see Morgan-Short, 2014). ERPs are derived through amplifying and averaging naturally occurring electroencephalogram data, which consist of changes in the brain's electrical activity recorded from electrodes placed on the scalp. Through the study of these changes in the brain's electrical activity, ERP researchers investigate neurocognitive processing with very high temporal precision (see Luck, 2014, for elaboration).

ERPs reflect brain activity elicited in response to a time-locked external event, such as the words 'tranquilo' compared to 'tranquila' in example (1). Language studies using ERPs often use "violation paradigms"—for instance, within the context of a GJT, which contains both well-formed sentences and sentences with grammar errors—in order to study the time-course of language processing. Using this paradigm, ERP language processing research has revealed a set of well-studied brain wave patterns, or ERP effects, that are understood to reflect distinct neurocognitive processes. For example, the N400 effect has been reliably linked to lexical/semantic processing and the P600 effect has been reliably tied to grammatical processing (Kutas & Federmeier, 2011; Swaab, Ledoux, Camblin, & Boudewyn, 2012).

In general, SLA research shows that at lower proficiency, learners show N400s while processing target foreign language grammar (e.g., McLaughlin et al., 2010), which suggests that they rely on lexical/semantic information. At higher proficiency, learners are more likely to show P600s while they process target grammar (e.g., Steinhauer, White, & Drury, 2009). This indicates that the process of adult language

learning involves a neurocognitive progression from using lexical/semantic information to using structural, grammatical information while processing foreign language grammar (see also Tanner, Inoue, & Osterhout, 2014).

These field-wide insights can be exploited for SA research. Using well-studied ERP effects, researchers can examine language processing during or as a result of an SA experience. A recent study by Faretta-Stutenberg and Morgan-Short (2017) applied the ERP technique to SA research by testing sentence processing of an AH and SA group in a pre-/post-SA design. By gathering ERPs, the study could assess the potential impact of SA for not only behavioral performance but also for the online neurocognitive processing of target language structures (for related details, see also Chapter 27, this volume). This seems to be a promising area for future SA research. Note, however, that ERP-SA designs, like eye-tracking, will predominantly rely on measuring processing after a return from SA since the equipment and setup for ERP research is not usually mobile. In lieu of measuring processing post-SA, researchers may seek out in-country collaborator institutions that have the facilities needed to collect this type of data (including for eye-tracking). While ERP equipment can be quite expensive, there are excellent lower cost ERP equipment options on the market that are, importantly, also suitable for mobile research (e.g., ActiCHamp from Brain Products, Germany), and these could be useful for in situ SA studies.

Conclusion

This chapter has reviewed quantitative approaches for measuring the effects of SA on foreign language behavior and processing. I have discussed common SA design options and the global as well as specific measures that researchers can utilize within those designs to elucidate the effects of studying abroad on foreign language knowledge and development. The chapter did not discuss all available methodological approaches but rather focused on currently employed methods (OPIs, GJTs) as well as additional interdisciplinary methods (EI, mouse-tracking) that SA research might also benefit from.

With many of the technologically advanced methods reviewed in the chapter (ERPs, eye-tracking), SA researchers stand at the forefront of language learning and processing research. Pairing these methods with interests in SA research, and the overall application of quantitative methods, enables researchers to reveal compelling information on the measurable positive effects of studying abroad.

Key Terms

Oral proficiency	Event-related potentials
Grammaticality judgment	Elicited imitation
Response time	Lexicon, grammar
Eye-tracking	

Further Reading

- Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, 18(3), 384–401. (This paper describes in detail the analysis techniques for mouse-tracking data. It also describes the basic mouse-tracking paradigm and its theoretical orientation.)
- Morgan-Short, K. (2014). Electrophysiological approaches to understanding second language acquisition: A field reaching its potential. *Annual Review of Applied Linguistics*, 34, 15–36.

(This paper reviews the ERP technique and the background of its use in SLA research. It also summarizes conclusions drawn from the existing corpus of SLA-ERP research.)

Taguchi, N. (2012). *Context, individual differences and pragmatic competence* (Vol. 62). Bristol, UK: Multilingual Matters. (This book discusses a longitudinal quantitative study on pragmatic development [appropriateness, accuracy, and processing speed] in a second language immersion education context.)

References

- Allen, H. W., & Herron, C. (2003). A mixed-methodology investigation of the linguistic and affective outcomes of summer study abroad. *Foreign Language Annals*, 36(3), 370–385.
- Arnett, C. (2013). Syntactic gains in short-term study abroad. *Foreign Language Annals*, 46(4), 705–712.
- Bartolotti, J., & Marian, V. (2012). Language learning and control in monolinguals and bilinguals. *Cognitive Science*, 36(6), 1129–1147.
- Blazej, L. J., & Cohen-Goldberg, A. M. (2015). Can we hear morphological complexity before words are complex? *Journal of Experimental Psychology: Human Perception and Performance*, 41(1), 50.
- Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia*, 51(13), 2492–2511.
- Collentine, J., & Freed, B. F. (2004). Learning context and its effects on second language acquisition: Introduction. *Studies in Second Language Acquisition*, 26(2), 153–171.
- Dussias, P. E. (2010). Uses of eye-tracking data in second language sentence processing research. *Annual Review of Applied Linguistics*, 30(1), 149–166.
- Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, 27(3), 464–491.
- Faretta-Stutenberg, M., & Morgan-Short, K. (2017). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research*. doi:10.1177/0267658316684903
- Freed, B. F., Dewey, D. P., Segalowitz, N., & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition*, 26(2), 349–356.
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42(1), 226–241.
- Grey, S., Cox, J. C., Serafini, E. J., & Sanz, C. (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during short-term intensive language exposure. *The Modern Language Journal*, 99(1), 137–157.
- Helman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, 18(3), 384–401.
- Hernández, T. A. (2010). The relationship among motivation, interaction, and the development of second language oral proficiency in a study-abroad context. *The Modern Language Journal*, 94(4), 600–617.
- Huetting, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171.
- Hulstijn, J. H., Van Gelderen, A., & Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us? *Applied Psycholinguistics*, 30(4), 555–582.
- Isabelli-García, C. (2010). Acquisition of Spanish gender agreement in two learning contexts: Study abroad and at home. *Foreign Language Annals*, 43(2), 289–303.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential. *Annual Review of Psychology*, 62, 621–647.
- LaBrozzi, R. M. (2012). *The role of study abroad and inhibitory control on processing redundant cues*. Paper presented at the 14 Hispanic Linguistics Symposium. Indiana University, Indianapolis, IN.
- Lim, H., & Godfroid, A. (2015). Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, Van Gelderen, and Schoonen's 2009 study. *Applied Psycholinguistics*, 36(5), 1247–1282.

- Llanes, À. (2011). The many faces of study abroad: An update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism*, 8(3), 189–215.
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- 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. *Language Learning*, 60(s2), 123–150.
- Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *Anglophone students abroad: Identity, social relationships, and language learning*. New York, NY: Routledge.
- Mora, J. C. (2008). Learning context effects on the acquisition of a second language phonology. In A. B. Gaya (Ed.), *A portrait of the young in the new multilingual Spain* (pp. 241–263). Clevedon, UK: Multilingual Matters.
- Morgan-Short, K. (2014). Electrophysiological approaches to understanding second language acquisition: A field reaching its potential. *Annual Review of Applied Linguistics*, 34, 15–36.
- Pérez-Vidal, C., & Juan-Garau, M. (2011). The effect of context and input conditions on oral and written development: A study abroad perspective. *IRAL-International Review of Applied Linguistics in Language Teaching*, 49(2), 157–185.
- Rassaei, E., Moinzadeh, A., & Youhannaee, M. (2012). Effects of recasts and metalinguistic corrective feedback on the acquisition of implicit and explicit L2 knowledge. *The Journal of Language Teaching and Learning*, 2(1), 59–75.
- Rees, A. J., & Klapper, J. (2008). Issues in the quantitative longitudinal measurement of second language progress in the study abroad context. In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 89–105). New York, NY: Routledge.
- Roche, J. M., Peters, B., & Dale, R. (2015). “Your tone says it all”: The processing and interpretation of affective language. *Speech Communication*, 66, 47–64.
- Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26(2), 173–199.
- Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, 14, 369–369.
- Serafini, E. J., & Sanz, C. (2016). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, 38(4), 607–646.
- Spada, N., Shiu, J. L. J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, 65(3), 723–751.
- Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research*, 25(1), 13–41.
- Sunderman, G., & Kroll, J. F. (2009). When study abroad fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics*, 30, 79–99.
- Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components. In S. Luck & E. S. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp. 397–440). New York, NY: Oxford University Press.
- Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in on-line L2 grammatical comprehension. *Bilingualism: Language and Cognition*, 17, 277–293.
- Traxler, M., & Gernsbacher, M. A. (2011). *Handbook of psycholinguistics*. New York, NY: Academic Press.
- Wickens, T. (2002). *Elementary signal detection theory*. Oxford, UK: Oxford University Press.
- Wu, S. L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals*, 46(4), 680–704.
- Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*, 33(4), 497–528.
- Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika*, 70(1), 203–212.