



LUND UNIVERSITY

Reviewing the potential of the Experience Sampling Method (ESM) for capturing second language exposure and use

Arndt, Henriette; Granfelt, Jonas; Gullberg, Marianne

Published in:
Second Language Research

DOI:
[10.1177/02676583211020055](https://doi.org/10.1177/02676583211020055)

2023

Document Version:
Peer reviewed version (aka post-print)

[Link to publication](#)

Citation for published version (APA):
Arndt, H., Granfelt, J., & Gullberg, M. (2023). Reviewing the potential of the Experience Sampling Method (ESM) for capturing second language exposure and use. *Second Language Research*, 39(1), 39-58.
<https://doi.org/10.1177/02676583211020055>

Total number of authors:
3

Creative Commons License:
CC BY-NC-ND

General rights

Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

This is the peer reviewed version of the paper “Reviewing the potential of the Experience Sampling Method (ESM) for capturing second language exposure and use” which has been accepted for publication in the journal Second Language Research. Reuse is restricted to non-commercial and no derivative uses.

Reviewing the potential of the Experience Sampling Method (ESM) for capturing second language exposure and use

**Henriette L. Arndt, Jonas Granfeldt, Marianne Gullberg
Lund University**

Frequent language exposure and use are among the most important conditions for successful language learning whether in classrooms, during study abroad, or in other informal contexts. Research probing exposure and usage often relies on one-off self-report questionnaires in which participants estimate their typical level of language exposure over extended periods of time, often long after it occurred. This may negatively affect the validity of the resulting data. This paper instead explores the potential of methods used in medical and psychological research, variably known as the ‘Experience Sampling Method’ (ESM; Csíkszentimihályi and Larson, 1987), ‘Ecological Momentary Assessment’ (EMA; Stone and Shiffman, 1994), or ‘diary methods’ (Bolger et al., 2003). These methods are often combined with electronic and mobile survey applications to elicit self-report assessments at frequent, sometimes randomised intervals. We consider the possibilities of these methods for strengthening research into language exposure and use, second language acquisition more broadly, and study abroad research specifically. The methods have the potential to drastically reduce biases associated with summative recall. Additionally, they enable researchers to collect richer data about how individuals engage with language differently over time, and the contexts in which they do so, thus ultimately contributing to our understanding of individual differences in language acquisition.

Introduction

Exposure to language input, that is, spoken, gestured, and written discourse in the target language, is an essential component of second language acquisition (SLA). According to many current theories of SLA, language learning involves making inferences about pronunciation, vocabulary, grammar and pragmatics from samples of the language to which learners are exposed in the classroom and beyond (Ellis, 2009). Thus, finding out how much, and what kind of second language (L2) exposure is necessary for successful learning has long been among the central issues in SLA research. The answers are of importance not only to our theoretical understanding of the processes which underly language acquisition, but also to the practices of language teachers and learners, as well as policies related to language learning more broadly. Current research, however, is far from being able to address these issues with any precision. This is in large part due to the difficulty of operationalising and measuring language exposure. This paper will review previous attempts, present a methodological option used in other disciplines – the Experience Sampling Method (ESM) – and discuss its advantages and disadvantages for the language sciences in general and for SLA in particular, to study exposure in informal contexts, or ‘in the wild’, such as Study Abroad (SA) settings.

Theoretical background

Language learning can be defined as a process where an individual develops a language in interaction with the environment (Clark, 2003; Tomasello, 2003). However, in the field of SLA, there is currently no uniform and generally accepted term to describe the learner’s exchange with the environment. Depending on the theoretical approach, the terms ‘input’, ‘exposure’, and ‘experience’ are in use, reflecting both different understandings of ‘interaction with the environment’, and different research foci.

In cognitively oriented SLA focusing on learner internal factors, such as the generative approach, the term ‘input’ is widely used to describe the outcome of learners’ engagement with the environment. In this orientation, input is often viewed within the computer metaphor as ‘data passively entering a system’. Krashen (1981) thought of ‘comprehensible input’ as abstract and decontextualized information about the target language feeding automatically into largely implicit processes of language acquisition. Carroll (2017) labels this view ‘input-to-the-language-learning mechanism’ but also takes it a step further by making a distinction between the ‘signal’ and ‘input’. She points to the fact that “there is no direct evidence for input in the signal” (Carroll 2017: 3) and constructs such as ‘verbs’, ‘gender’ or ‘inflection’, which are central to solving the “learning problems”, have to be imputed to learners and their linguistic representations. This line of research has led to a vast amount of research on input processing — the cognitive processes through which input becomes intake — and particularly the role of consciousness and ‘noticing’ in SLA (Badger, 2018; Cosgun Ögeyik, 2018; Truscott and Sharwood Smith, 2011).

In cognitively oriented approaches focusing more on learner external factors, such as the Usage-Based Approach (ULB; Bybee, 2008; Ellis, 2009), ‘exposure’ is a recurrent term. According to this theory, language acquisition results from the application of general cognitive and learning mechanisms to linguistic input, which allow learners to process and acquire lexical and grammatical patterns (Tomasello, 2003; Wulff and Ellis, 2018). In ULB, ‘exposure to language’ means exposure to discourse produced by other speakers, writers or signers of the target

language. Based on this exposure, the task of the language learner is to work out the regularities of the target language, despite partial evidence and considerable individual variation. Therefore, language learning is, at least partly, (implicit) statistical learning (Christiansen, 2019; Rebuschat and Williams, 2012; Saffran, 2003), or “estimation from sample” (Ellis, 2009: 139).

In socio-cultural (Lantolf and Thorne 2006) or socio-cognitive (Atkinson 2011) approaches to SLA, ‘interaction with the environment’ is understood in a fundamentally relational and integrative way. The emphasis here is on the ‘experiences’ in and with the target language, defined by the actions that make it up. Atkinson (2011: 156) stresses that “the concept of input is misleading if it steers us away from understanding the fundamentally ecological nature of human existence and activity”. Activities lead to events that can be defined as ‘linguistic affordances’ (Van Lier, 2000) in the sense of ‘actions in potential’ emerging when the language learner interacts with the physical and social world (Van Lier, 2004: 92). Van Lier argues that learners will pick up the linguistic information they need for their activities and projects as long as they have access to affordances. Thus, linguistic affordances are socially mediated, and an important empirical question should be how, when and where learners gain access to them.

Beyond the theoretical differences briefly outlined above, it seems evident that in the real world the three terms are intrinsically linked. An ‘experience’ involving (verbal) communication of some kind, can lead to ‘exposure’ to a particular language, which can ultimately serve as ‘input’ to the language learning processes. However, the empirical links between ‘experience’, ‘exposure’, and ‘input’ are currently poorly understood and they are not easy to investigate with current methodologies. This situation can at least partly be attributed to a poor understanding of the details about when, how, and where learners encounter the L2 outside of the classroom. In SLA, with its historically strong focus on instructed settings, the issue has not attracted much interest apart from studies within Study Abroad (see below) and research on informal language learning in out-of-school contexts (e.g., Dressman and Sadler, 2020; Sockett, 2014; Sundqvist and Sylvén, 2016).

However, in research on early bilingualism, there is an increasingly rich literature on both the role of the environment for language development and its potential measurement. The interest should be seen against the backdrop of findings in recent years pointing to important individual variation among simultaneous and successive bilingual children, some of which can be explained by differences in the children’s language environments. This literature (see e.g., Grüter and Paradis, 2014) suggests that there is no simple relationship between experiences, exposure and input, let alone between these notions and linguistic development, but there are important insights that should apply also to SLA. Unsworth (2013) points to the cumulative effects of exposure in the two languages. The exposure balance between the two languages typically changes several times over a child’s lifespan, which means that reliance on measuring a child’s input situation at a single given point in time will be misleading. Unsworth argues that it is the accumulated exposure over time that best accounts for the observed variation in linguistic development.

In SLA and in situations like SA in particular, there are several reasons to believe that the experience and exposure situations are even more dynamic. Learners’ social networks develop over time (McManus et al., 2014) and as the learner progresses in linguistic proficiency, more and different activities in the target language become interesting and feasible. Moreover, a perceived successful interaction event may lead to an increased inclination to engage in yet

another event (or, conversely, learners who experience difficulties might be less inclined to seek out further interactions in future). Finally, not only do the amount of language exposure and use matter for learning, but also the context in which it occurs (Unsworth, 2015). Pertinent factors may include whether conversational partners are native or non-native speakers, the number of different interlocutors, and their degree of familiarity with the learner (Dewey, 2017).

In sum, to gauge how an individual engages with the environment it is necessary to employ dynamic, detailed and repeated assessments of L2 learners' exposure and language use over time. While it is certainly true that there have been several previous attempts to achieve such assessments, also in the SA context (cf. review below), we will argue that, since the ESM approach has been designed precisely for this type of assessment, it holds great potential of moving the field forward.

Previous approaches to measuring language exposure and use

Current methodologies for quantifying the amount and type of language exposure and use during SA primarily rely on retrospective self-report questionnaires, which require participants to recall and summarise their past experiences with the second language. Perhaps the most widely used and adapted example is the *Language Contact Profile* (LCP; Freed et al., 2004), which is intended for data collection at the end of the SA period. The LCP includes questions about how frequently (i.e., how many days per week) and for how long (i.e., how many hours per day) learners engaged in different types of language use, including speaking, reading, listening, and writing activities in various settings, such as:

On average, how much time did you spend speaking, in *Spanish*, outside of class with native or fluent speakers of Spanish during this semester?

Typically, how many *days per week*?

0 1 2 3 4 5 6 7

On those days, typically how many *hours per day*?

0-1 1-2 2-3 3-4 4-5 more than 5

(Freed et al., 2004: 354)

The *Language Engagement Questionnaire* (LEQ; McManus et al., 2014) is an example of another variant of self-report questionnaires used in SA research, in which respondents are asked to indicate how frequently they 'typically' engage in a number of different second language activities, rather than reflecting on a specified time period. Activities covered in the LEQ include making small talk or having long casual conversations, instant messaging, watching TV and films, or listening to music. Response options include: 'every day, several times a week, a few times a week, rarely, or never' (McManus et al., 2014: 104).

The core strength of instruments such as the LCP and LEQ is that they offer a relatively quick and simple way to gather data from large participant samples, allowing researchers to discover trends in language exposure and use and to answer questions regarding, for example, the extent to which learners use the community language (vs. English or their first language) during SA, and the circumstances in which target language exposure is most likely to occur. Furthermore, some studies have successfully used self-report instruments to demonstrate a positive relationship between the amount of language exposure and developments in proficiency, although these findings are not consistent. Several critical discussions (e.g., Briggs

Baffoe-Djan and Zhou, 2020; Dewey, 2017; Fernández and Gates Tapia, 2016) have cited this lack of consistency as evidence that these popular tools are not capturing language exposure with sufficient accuracy, although variations in results will likely also be linked to differences between various adaptations of the instruments, how they are administered, and the populations under study. Altogether, these critical reviews have indicated three major shortcomings of these self-report instruments, which put the validity of the resulting data into question.

First, it has been noted that, while easy to administer, the great length of the LCP (which includes 40 items about language use, such as the one quoted above, in addition to demographic questions) and other such questionnaires may fatigue participants and discourage them from responding carefully, resulting in poor quality data (Dewey, 2017).

Second, despite their length, the commonly used questionnaires have been criticised for being severely limited in the aspects of language exposure and use which they capture. For example, the original LCP (Freed et al., 2004) does not cover online L2 practices, such as social media use, although they are included in some more recent adaptations (e.g., Bracke and Aguerre, 2015; Martínez-Arbelaiz et al., 2016). More importantly, these questionnaires only target the frequency and quantity, but not the *quality* of language exposure, including the extent to which learners concentrate on, understand, or enjoy the language being used (Briggs Baffoe-Djan and Zhou, 2020). Finally, since these surveys are often administered only once, at the end of the SA period, they cannot capture fluctuations in language exposure and use over time (Fernández and Gates Tapia, 2016).

The third, and most fundamental, criticism levied against these instruments is that they generally deliver inaccurate and inconsistent estimates of language exposure and use (Dewey, 2017; Fernández and Gates Tapia, 2016). Since questionnaires are often administered a considerable time after the events in question occurred (e.g., several weeks after the end of the SA period), the accuracy of the collected data is likely to suffer from recall errors, which can occur even after relatively short intervals (Bradburn et al., 1987). Furthermore, some individuals are simply better at remembering past events, and will thus be able to reply more accurately to questions about language exposure and use. Such individual differences can introduce further variability and put the validity of interpersonal comparisons into question. Moreover, research shows that recall error constitutes not just interpersonal ('random') error but is also affected by systematic biases: There is a general tendency for people to overestimate the frequency or duration of past events or behaviours (Shiffman et al., 2008; Wearden, 2008), except in case of short, frequently occurring phenomena, which tend to be underestimated (Araujo et al., 2017; Droit-Volet et al., 2018). In the specific case of language use studies, Fernández and Gates Tapia (2016: 33) suggest that participants may purposefully report higher-than-accurate quantities of language contact because they might "feel embarrassed, and perhaps disillusioned, in how little they actually used the L2 while abroad".

The general tendency to over-estimate in self-reports becomes even more problematic when measures of language exposure and use are summed or averaged across time and/or various activities, due to the large number of items in the questionnaires. Dewey (2017) notes that the LCP in particular was intended only to capture the extent to which learners engage in different types of L2 activities, not the quantity of their overall exposure. It has nevertheless been used by many researchers to derive sum-totals. One of the reasons for why the LCP may yield exaggerated, and even unrealistic, estimates of total language exposure and use may be that the

items are not always clear to respondents. For example, they may interpret several activity categories as overlapping (e.g., “speaking to strangers” and “speaking to service personnel”) and report the same interactions multiple times, leading to inflated sum totals (Dewey, 2017; Fernández and Gates Tapia, 2016). Moreover, how learners judge their own language exposure and use seems to vary vastly depending on the frame of reference provided in the questionnaire. For example, Fernández and Gates Tapia (2016) found that participants who completed the LCP reported spending more time speaking to a variety of different interlocutors in their free time than the total overall time they reported speaking in their L2. In other words, when learners are asked to provide retrospective estimates for specific activities, without reference to any total amount of L2 exposure, they are prone to over-estimation (Dewey, 2017).

Employing a think-aloud procedure, Fernández and Gates-Tapia (2016) also showed that learners found it difficult to respond to the LCP questions because they felt that their language use varied significantly across the SA period. As a consequence, it will be challenging for participants to provide valid estimates of average amounts of language use over a certain period of time. What is more, studies on autobiographical memory and survey methods show that various systematic biases result from the heuristic cognitive processes involved in summarising past experiences. People tend to give disproportionate weight to events that are more easily remembered, for example because they are more recent (recency bias) or connected to more intense emotional experiences (saliency bias) (Burt, 2008). Similarly, the way that experiences are recalled and aggregated has been shown to be affected by current attitudes, moods, and circumstances (state-congruent bias), as well as by social scripts (what is perceived as ‘typical’ within a given cultural context) (Brunec et al., 2017; Teitler et al., 2006). Past thoughts and feelings are arguably even more difficult to recall than events or behaviours, and indeed, the biases have even stronger effects on reports of motivation, attitudes, and other cognitive or emotional variables (Thigpen, 2019). This suggests that the use of retrospective self-report questionnaires is an even greater threat to the validity of research which tries to capture not just the quantity but also the quality of second language exposure and use, as discussed previously.

Many researchers have recognised weaknesses in the available instruments and have modified them in order to address their shortcomings. Such strategies have included directing participants’ attention towards the feasibility of individuals time estimates by displaying running totals of daily L2 use times alongside the online questionnaire (Dewey et al., 2012); focusing on the relative proportion of time spent on different activities instead of sum totals (Baker-Smemoe et al., 2012); and eschewing time estimates altogether in favour of ‘how true of me’ rating scales (Briggs, 2015).

Other researchers have employed daily diaries, journals or logs (these terms are used interchangeably throughout the literature, although they connote different levels of open-endedness, according to Rose and McKinley, 2020) as an alternative to one-off surveys. Data collected longitudinally via these methods reflect the dynamic and complex nature of language exposure and can help researchers to discover the significance of individual instances or specific types of language use (Briggs Baffoe-Djan and Zhou, 2020; Magnan and Back, 2017, Ranta and Meckelborg, 2013). For example, García-Amaya (2017) designed the *Daily Linguistic Questionnaire* (DLQ), which prompts respondents via a mobile phone app to outline their schedule during the preceding day (e.g., waking, sleeping, and mealtimes) and to estimate how much time they spent speaking, listening, reading, and writing in their native and second

language during each one of four time blocks (sleeping and waking up, morning, afternoon, dinnertime and late evening). This design significantly reduces the delay between instances of language use themselves and the time of reporting, as well as the intervals across which respondents are required to summarise their experiences (i.e., part of one day instead of several weeks). Nevertheless, diaries such as the DLQ, still rely on aggregated recall, and therefore suffer from the same biases as other retrospective surveys, albeit on a smaller scale. There are, however, alternative methodological approaches for investigating everyday experiences ‘in real time’ (prospectively rather than retrospectively), which will be introduced in the next section.

The Experience Sampling Method

In other fields of research, alternative measurement techniques have been gaining in popularity which involve prompting participants to complete brief surveys several times per day. Different variants of this approach exist, whereby participants may either be instructed to complete it every time a certain trigger occurs (event-based sampling) or signalled at certain time points to complete the questionnaire (signal-based sampling). Whereas researchers initially relied on programmable watches or pagers to signal participants, who would then fill in pen-and-paper questionnaires, smartphones have recently become the primary means of both signalling and data collection. The surveys usually include questions about current behaviours, mood, thoughts, feelings, and features of the physical or social context (e.g., location, interlocutors). Because data are often collected several times per day over a period of multiple days, or even weeks, the questionnaires are typically short and easy to complete, within a minute or two, to reduce participant burden.

A variety of terms have been used in the literature to describe such intensive repeated measurement designs. ‘Ecological Momentary Assessment’ (EMA; Stone and Shiffman, 1994) and ‘diary methods’ (Bolger et al., 2003) are both umbrella terms, which include signal- as well as event-based approaches. The former term is more widely used in clinical psychology and behavioural medicine, whereas the latter appears more frequently in educational and social psychology studies. Furthermore, the term ‘Experience Sampling Method’ (ESM; Csíkszentmihályi and Larson, 1987) is used in many other sub-disciplines of psychology as well as computer science research. Initially, ESM referred specifically to signal-based measurement techniques which employed random signalling schedules (Hektner et al., 2007). In practice, however, the different terms are mostly used interchangeably nowadays (van Berkel et al., 2017).

To our knowledge, the ESM methodology has not been applied to the study of L2 exposure in SA contexts yet. However, we believe that this method holds the potential to strengthen the study of everyday language use and exposure by providing more accurate information than other current, recall-based methods about how frequently, and in which contexts, learners come into contact with their L2. In support of this claim, we will continue with a discussion the strengths and limitations of the Experience Sampling Method, particularly in comparison to previous approaches to measuring language exposure and use.

ESM and the study of second language exposure and use

The two greatest strengths of the Experience Sampling Method are (a) the immediacy and high ecological validity, which increase the reliability and accuracy of the collected data, and (b) that

data are collected repeatedly, in various everyday situations, which facilitates the study of dynamic change across time and contexts. Therefore, this method directly addresses a number of the shortcomings of popular language contact questionnaires introduced above, including the biases associated with summative recall, the lack of insight into individual instances of language use, and the inability to capture longitudinal changes in second language exposure.

The ESM was developed to capture individual actions, thoughts, and perceptions in their “natural setting and [...] to minimize the delay between the event and the time it is recorded” (Krishnamurty, 2008: 197). Whereas instruments such as the LCP and LEQ require participants to recall and/or summarise past experiences across some period of time (ranging from part of one day, in the case of language logs, to several weeks in retrospective questionnaires such as the LCP), ESM largely focuses on current (or very recent) events. When a large amount of ESM data points is collected from the same person, they can be aggregated to derive estimates of a person’s ‘typical’ behaviour (e.g., the proportion of time spent on certain L2 activities). These estimates should be more reliable than those collected via retrospective self-reports, because they have less potential to be affected by recall errors and reporting biases and are more ecologically valid, because they are collected during the natural flow of everyday life.

Split-half analyses (comparisons of one half of the data to the other) indicate that aggregated ESM data indeed have high internal reliability (Chen et al., 2015; Larson et al., 2002). Because of the larger number of data points per participants, mathematical adjustments can even be applied to correct for random error resulting from different interpretations and use of response scales by the respondents (which is a potential problem in all questionnaires, whether they are implemented as part of an ESM or retrospective data collection design).

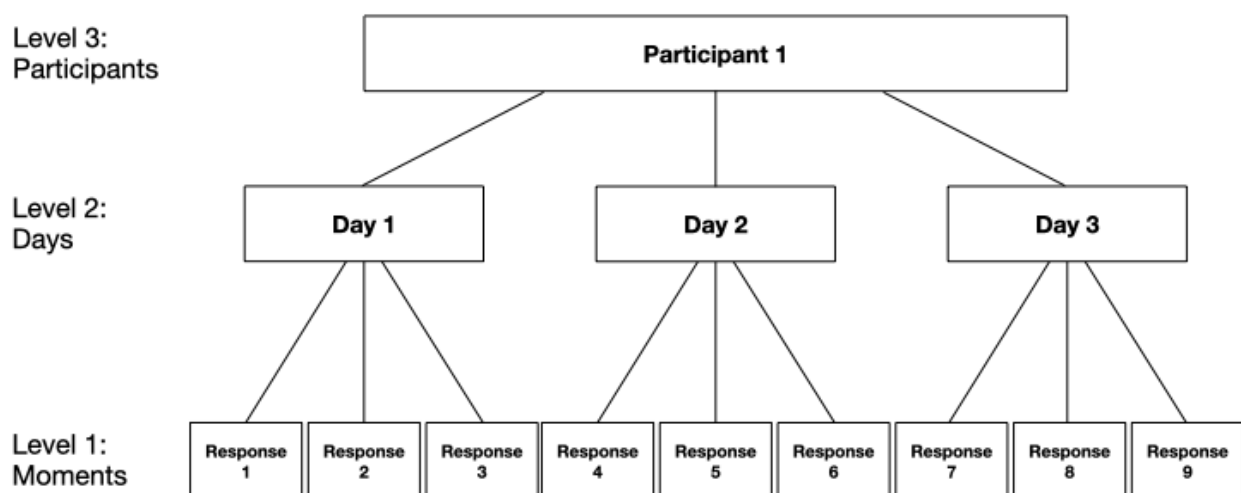
Meanwhile, the high validity of these data is supported by research which has compared ESM measures and recall-based data (see reviews in Ellison et al., 2020; Hektner et al., 2007; Shiffman et al., 2008). Many studies report no more than moderate correlations between aggregated ESM and retrospective survey data (Fazeli and Turan, 2019; Shrier et al., 2005; Turan et al., 2016), which indicates that these methods provide substantially different pictures of an individual’s typical behaviour and experiences. Analyses of the specific differences between ESM and retrospective data appear to suggest that ESM instruments are more sensitive and better able to “capture moments that are otherwise deemed [too] insignificant to remember or report” (Hektner et al., 2007: 111). The discrepancy between ESM and recall-based data has also been found to increase in concordance with other variables that are known to affect retrospective reports in particular, such as the time between the event itself and the moment of reporting, and the amount of variability in the phenomenon under study (Fazeli and Turan, 2019; Jobe, 2003; Stone et al., 2005). This is commonly interpreted as evidence that ESM indeed provides a more accurate picture of a person’s actual thoughts, feelings, and actions, although recall-based surveys are still regarded as valuable for capturing retrospective impressions or global beliefs (Ellison et al., 2020; Shiffman et al., 2008).

In addition to more accurate data on the quantity and frequency of L2 exposure and use (i.e., when it took place, what activities the learners were engaged in, and how long they lasted), researchers employing the ESM can also ask their participants questions about the context in which learners were exposed to or used language, with whom, and even what they were thinking and feeling at the time. Therein, the ESM has the potential to address a second shortcoming of extant measurement approaches, which, as noted previously, have been criticised for being too

limited in their focus on *quantity* of language use over *quality* (Briggs Baffoe-Djan and Zhou, 2020).

So far, our discussion has focused on the advantages of aggregated ESM data when compared to estimates derived from one-off retrospective questionnaires. However, the unique structure of ESM data also enables researchers to conduct different types of analyses, targeting different research questions, than what is possible with traditional methods. Because the same participants are surveyed repeatedly over a set period of time, ESM yields data that are ‘nested’ or ‘clustered’ at three levels (see Figure 1): Individual survey responses are nested within days, which are nested within participants. This structure reflects the assumption that measurements are likely to be more similar if they stem from the same individual than across different participants, and if they are collected on the same day versus different days.

Figure 1: Three level nested structure of ESM data



The nested structure of ESM data allows researchers to study not only how phenomena (such as language exposure and use) vary between participants, but also variations which occur *within* participants, day-by-day and from moment to moment. ESM data can thus be used to study changes in *individual* learners’ patterns of language exposure and use as they unfold in real time. Shiffman et al. (2008: 10) provided the illustrative metaphor of ESM data as “a movie, in which dynamic relationships emerge over time, whereas global or recall measures are analogous to a still photograph, a single static snapshot”. Whereas recall-based questionnaires work on the assumption that personal habits and characteristics are relatively stable over time (traits), ESM responses capture a series of *current states*. This rich data can then be used to investigate intrapersonal changes in language use across time, or to study the co-evolution between language exposure and a host of other psychological and environmental factors, such as the learners’ developing language proficiency, their shifting social networks or, in the case of students studying abroad in 2020, changing COVID-19 related restrictions regarding education and social life in the host country.

ESM data can be analysed in a variety of ways, including qualitatively or by simply plotting changes in the targeted variables over time (see e.g., Carter and Emsley, 2019; Hektner et al., 2007; Shiffman et al., 2008; Walls et al., 2006 for guidance on different types of qualitative and

quantitative approaches to ESM data analysis). Statistical analysis of variations in ESM data over time, however, requires the application of advanced techniques such as random effects regression or multilevel structural equation modelling, since the nested structure of ESM data violates the assumption of independence which underlies more traditional methods such as t-tests and linear regression models. Multilevel modelling makes it possible to study not only concurrent relationships (e.g., whether the number of participants in a conversation correlates to perceptions of difficulty in L2 use), but also to conduct temporal or time-lagged analyses, which evaluate the effect that an observed variable has on another variable at a later point in time (Carter and Elmsley, 2019; Walls et al., 2006). These analyses can address questions such as whether and how a learner's perceptions of different instances of language exposure and use (e.g., how challenging and enjoyable these experiences were) affect the quantity and quality of subsequent L2 exposure and use. Time-lagged analysis seems particularly valuable in an SA context, because it could offer insight into the extent to which learners' self-assessed experience of previous exposure events lead them either to avoid similar events in the (near) future or to replicate them.

Taken together, these arguments suggest to us that ESM holds considerable promise for research on the study of L2 exposure and use. Its implementation in SLA research could help us move forward and improve our understanding of the role of L2 exposure and use, whether in SA or other contexts.

Practical considerations in the implementation of ESM

Of course, ESM is not without its limitations, which need to be taken into consideration by researchers who wish to incorporate this approach into their study design. One commonly cited concern is that completing surveys multiple times per day may lead to reactivity, that is, to participants consciously or subconsciously changing their habits, or the ways in which they experience everyday life, while taking part in the study. This has indeed been observed in the initial stages of some ESM studies, but the effects have been found to diminish quickly as participants habituate to the procedure (Hektner et al., 2007). In return, proponents of ESM have argued that participants would be more likely to give responses which they believe will be viewed favourably by others (desirability bias) when the researchers, and perhaps other participants, are physically present, such as in lab- or classroom-based data collection (Hektner et al., 2007; Xie et al., 2019). In past ESM studies, respondents have certainly not shied away from reporting private or embarrassing events and, when asked, the vast majority (between 80 and 90%) indicated that they lived their lives 'normally' while the study was going on, suggesting that they did not change their behaviour as a function of ESM or fabricated false replies (Csíkszentimihályi and Larson, 1987; Shiffman et al., 2008; Stone and Shiffman, 1994).

Another frequent criticism is that ESM can be overly time-consuming and demanding for the participants, especially if the surveys are long and must be completed at inopportune times. Since not everyone will be willing to commit this amount of time and effort, nor to commit to the potential disruption of their daily life, selection biases may occur. Indeed, ESM studies tend to have lower volunteer rates and higher rates of attrition than studies employing less intensive methods conducted with the same population and recruitment schemes (Hektner et al., 2007). ESM studies also report lower 'signal response rates' (proportion of surveys which are actually completed) than other methods, often ranging between 65–85% when studies last four to seven

days and include five to seven surveys per day (Rintala et al., 2019). What is more, certain groups of people are often overrepresented in ESM samples. For example, women and individuals who are more organised, diligent, and conscientious (in education, this may mean the higher achieving learners) are more likely both to participate in ESM studies and to complete a larger proportion of assessments (Larson et al., 2002; Mulligan et al., 2000; Rintala et al., 2019). Of course, such sampling biases can also occur in studies which employ one-off self-report questionnaires. Nevertheless, the findings above show that ESM researchers in particular need be aware of issues of sampling bias and compliance with protocol in their own studies, and carefully reflect on the representativeness of their sample.

Furthermore, there are several strategies to effectively reduce participant burden, and thus increase participation rates, in ESM studies. First and foremost, researchers must carefully consider their sampling approach, including the choice between an event- or signal-based design. Event-based designs, wherein participants are to complete a survey every time a certain trigger event occurs, require that the phenomena under study are well defined and easily recognisable to the respondents. Researchers must also consider that, if the targeted events occur too frequently, having to submit a report every time may be too much to ask of the participants. Consequently, increases in missing data and participant attrition may be expected (Bolger et al., 2003). Therefore, in cases where researchers are interested in studying events which occur often, or in capturing a cross-section of the entirety of a person's daily experiences (e.g., time use), a signal-based approach may be more appropriate. For example, researchers interested in studying language exposure and use might prompt their participants several times per day to report in which activities they were most recently engaged and for how long, whether these were conducted in their first or second language, and even their concurrent thoughts and feelings.

In signal-based studies, reports may be elicited either at fixed or randomised time points. Whereas a fixed schedule is easier to implement, the predictability of the signals may increase the risk of reactivity, or participants changing their behaviours in anticipation of the next signal (Verhagen et al., 2016). Researchers may also run the risk of participants being engaged in the same activities at the same time every day (e.g., work, schooling). By contrast, random sampling reduces the likelihood of reactivity (since participants cannot predict when the next signal will come) and offers the opportunity to "capture maximum variation of [the] construct being studied" (Rose et al., 2020: 138). Based on the assumption that a large number of reports, collected at random times throughout the day, will together provide a representative picture of participants' overall experiences in everyday life, it has been suggested that randomised assessment schedules are the most appropriate when the researchers' goal is to estimate accurately "how much time people spend doing various activities during their waking hours" (Hektner et al., 2007: 11). Since second language exposure in everyday life occurs at many micro-moments, a random signal-based ESM design seems most suitable for capturing a cross-section of learner's daily L2 contact without placing undue burden on participants (although an event-based approach was successfully employed by Arndt (2019) in a study of informal L2 use among secondary school students in Germany). Researchers should note, however, that a truly randomised schedule runs the risk of occasionally sending signals at intervals that are either very short or very long, which may be confusing or annoying to the participants. For this reason, ESM researchers often prefer to adopt a stratified random sampling approach, whereby the day is

divided into fixed blocks and signals are randomised within each time window, or to impose constraints on the minimum/maximum time between signals.

Once they have settled on a sampling strategy, researchers must carefully consider the number of days across which data are to be collected, and how many times per day participants will be signalled. Research shows that participants are significantly less likely to answer signals in the late evening or at night, as well as during hours spent at school or work (van Berkel et al., 2019). Many researchers therefore choose not to signal participants during certain time periods (e.g., between 10pm and 8am). Naturally, this increases the risk that certain activities will not be recorded, such as language exposure from watching TV in the late evening or socialising with other L2 speakers in a nightclub. In making all these decisions, the need for a richer dataset will have to be balanced against the amount of effort that an intensive assessment schedule requires from the participants.

In addition to the sampling approach, the surveys themselves need to be carefully constructed, taking into consideration the time and effort needed to complete each questionnaire. This is true for all survey studies, obviously, but since ESM respondents are required to complete the same questionnaire many times, survey design becomes potentially even more important. In general, researchers should strive to use simple language, keep the number of questions low, and prioritise formats which are easy to answer (e.g., multiple choice or Likert scale items rather than open ended questions). Extensive piloting of the instruments and training of participants, for example on when to complete records (in fixed schedule designs) or on which events to report, can also improve the data quality (Hektner et al., 2007; Rose and McKinley, 2020).

Finally, it is well documented that both the number of participants and the number of survey responses per participant tend to decline as participants gradually lose interest in the study ('response fatigue'; Naughton et al., 2015; van Berkel et al., 2019). However, Hsieh and colleagues (2009) demonstrated that participant engagement and compliance over time can be improved by providing easy-to-understand visualisations of previous survey responses. Of course, participants may simply miss some signals because they were not paying attention or silenced their signalling devices (phones) during other important activities (e.g., school, work). This can be accounted for by giving participants some leeway for when to complete assessments (e.g., within one hour of being signalled), or sending reminders if responses are not submitted within a certain time period. This has been found to significantly increase response rates in past studies (Litt et al., 1998; Naughton et al., 2015).

ESM and mobile phone technology

Another important factor to consider when implementing an ESM approach is the method for signalling participants and collecting survey responses. Given the effort involved in managing signalling devices and pen-and-paper questionnaires, previous educational studies implementing ESM have largely been limited to classroom settings, where researchers can be in regular contact with the participants, help troubleshoot, or answer any arising questions (Xie et al., 2019). However, the now widespread use of smartphones and their advanced technological capabilities have the potential to make ESM more accessible than ever before, for example by removing geographical barriers, allowing researchers to conduct studies in multiple places at once, and

facilitating the recruitment of larger numbers of participants, which in turn provides greater statistical power for quantitative analysis.

We are not the first to recognize the utility of mobile phone technology for collecting timely, ecologically valid data on target language use during SA. García-Amaya (2017), for example, noted that her participants often chose to complete the Daily Linguistic Questionnaire on their mobile devices. Similarly, participants in Seibert Hanson and Dracos' (2019) research used the mobile course management system Canvas to complete daily logs of their engagement with digital media in their first and second languages. Dewey (2017) also describes an unpublished pilot study which employs an approach not unlike ESM, where SA students were prompted via text messages to complete browser-based surveys about the activities in which they had recently engaged.

While these studies draw on existing technologies to collect data on language exposure and use, a purpose-built tool could introduce several new possibilities for research in this area, which might be difficult to achieve with other research instruments. We ourselves are currently developing the LANG-TRACK-APP, a free iOS and Android native mobile application (and a companion webservice for data management) intended to facilitate the application of ESM methodology to the study of L2 exposure and use.

First and foremost, the app will allow researchers to upload custom electronic questionnaires and automatically assign them to (groups of) participants according to specified (fixed or random) schedules. In this way, large sets of real time data on daily language exposure and use can be collected with minimal continual effort from the research team. Participants are notified about new surveys via push notifications and record their answers on their own mobile phones, which will immediately transmit the data to the university servers. This is convenient both for the participants, who can respond to questions 'on the move' without having to keep track of their responses, and for the researchers, since data are delivered electronically, ready to be analysed.

For experience sampling to be valid, adherence to the chosen assessment schedule is essential and back-filling must be avoided. To ensure that surveys are completed in a timely manner, the app can be programmed to allow participants to send their responses only within a certain time window (e.g., within one hour of being signalled). In addition, data are time stamped, which can help researchers to monitor potential problems, and even follow up with participants who have missed surveys assigned to them. The same data can also be used to generate statistics which show the participants themselves how many surveys they have completed, a strategy which has been shown to increase participation rates (Hsieh et al., 2009).

In addition to timestamping, the technology could be adapted to automatically record other types of data, such as information about smartphone use (e.g., time spent using various apps) or GPS-information, which can be used to contextualise self-reported patterns of language use. This information could be further supplemented with audio recordings, allowing researchers to directly study the second language input to which the learners are exposed outside of formal educational or research contexts (see e.g., Mehl and Robbins, 2012; Ganek and Eriks-Brophy, 2018).

ESM in study abroad: a sample study

In this final section, we will sketch a sample study on the relationship between foreign language anxiety (FLA) and foreign language use in order to illustrate how the ESM approach could be applied to research in a study abroad setting.

Broadly, FLA has most commonly been defined as the worry and negative emotional reactions perceived when learning or using – in particular speaking – a foreign language (MacIntyre, 1999). In a comprehensive research overview, MacIntyre (2017) noted a number of negative consequences of FLA, including increasing avoidance motivation, declining perceptions of competence and lower willingness to communicate (see also Dewaele, 2013). FLA has been widely studied in the foreign language classroom context, but much less is known about the effects FLA during study abroad. In one of the few existing studies, Allen and Herron (2003: 382) looked at 25 college-level learners of French during a summer SA in France and found “significant mean decreases [...] in classroom and non-classroom language anxiety after SA”. Their research used a pretest-posttest design and could only observe that there was a significant change between the two measurements. Their research design did not allow for a study of successive and gradual change and a deeper understanding about what specific aspects of the SA experience were driving decreasing levels of FLA.

Since previous research has found an association between FLA and willingness to communicate, it is reasonable to hypothesize that an increased number of perceived successful communicative events will lead to lower levels of FLA. Moreover, if a learner perceives a particular communicative event during SA as successful, it is reasonable to believe that they are more likely to engage in another at a later point in time and so on. This leads us to the following research questions for our hypothetical study:

- 1) To what extent does a perceived successful communicative experience increase the likelihood that the learner will engage in other communicative experiences in the near future?
- 2) Is there a relation between the number of perceived successful communicative experiences and a subsequent decrease in FLA?

A study design which incorporates the ESM would be ideally suited to investigating these questions more closely. Employing a random signal-based design, participants could be prompted several times per day, via the aforementioned mobile app, to report on the nature of their most recent communicative experiences (e.g., when they last spoke to someone in the L2, with whom, and for how long) and their perception of these experiences (in terms of successfulness, degree of anxiety involved, etc.). To avoid collecting duplicate data, it should be specified that these should be experiences that happened after the participants submitted their last report. In this way, the interplay between FLA, communicative experiences, and their perceived successfulness could be investigated in the short term, across multiple consecutive days of ESM data collection, but also in the longer term, by scheduling multiple data collection periods (e.g., during the beginning, middle, and end of the SA period).

The moment-level ESM data (cf. Figure 1) can then be aggregated to derive estimates of learners’ daily or ‘typical’ engagement in communicative experiences. Additional useful measures could also be collected at the other two ‘levels’: Participant-level data, such as demographic information or other suspected confounding trait variables (e.g., proficiency,

personality traits), could be collected at the beginning of the SA period. Day-level measurements may include an abridged Foreign Language Anxiety Scale (Horwitz et al., 1986; adapted to focus on communication *outside of* the classroom), which the participants would be asked to complete once per day. Multilevel models can then be fitted to evaluate, for example, the extent to which the perceived success of previous communicative experiences predicts the quantity of communicative experiences during the subsequent days or weeks (RQ1), and whether daily/weekly levels of foreign language anxiety can be predicted by the number of perceived successful exchanges in which the learners engaged during the preceding day/weeks (RQ2). Given a sufficient sample size, additional analyses could be conducted to investigate whether these patterns differ between (groups of) participants, for example learners with greater L2 proficiency or those with a more extroverted personality.

Concluding remarks

We have argued in this paper that the Experience Sampling Method, particularly when implemented using smartphone technology, has the potential to significantly strengthen research into language exposure and use in study abroad contexts or second language acquisition more broadly. Ultimately, of course, the relative quality of the data delivered by ESM, as compared to other prominent methods, remains a subject for future empirical research. Studies directly comparing ESM and recall-based questionnaires will be able to reveal the extent to which the rich data which can be obtained via this new method complements and expands upon prior findings.

Acknowledgements

We gratefully acknowledge financial support from the Marcus and Amalia Wallenberg Foundation (grant number MAW 2018.0025). We thank multiple reviewers for very helpful comments on previous versions of the manuscript, and the Systems Developers Stephan Björck and Josef Granqvist for their work on the LANG-TRACK-APP. We also express our thanks for the fruitful and stimulating discussions we had while working on the Study Abroad Research in European Perspectives (SAREP) COST Action 15130 led by Prof Martin Howard, University of Cork. We gratefully acknowledge support from the Lund University Humanities Lab.

References

- Arndt HL (2019) *Informal second language learning: The role of engagement, proficiency, attitudes, and motivation*. PhD Thesis, University of Oxford, UK.
- Allen HW and Herron C (2003) A mixed-methodology investigation of the linguistic and affective outcomes of summer study abroad. *Foreign Language Annals* 36(3): 370–385.
- Araujo T, Wonneberger A, Neijens P and de Vreese C (2017) How much time do you spend online? Understanding and improving the accuracy of self-reported measures of internet use. *Communication Methods and Measures* 11(3): 173–190.
- Atkinson D (2011) *Alternative Approaches to Second Language Acquisition*. Abingdon: Routledge.
- Badger R (2018) From input to intake: Researching learner cognition. *TESOL Quarterly* 52(4): 1073–1084.
- Baker-Smemoe W, Dewey DP, Bown, J and Martinsen RA (2014) Variables affecting L2 gains during study abroad. *Foreign Language Annals* 47(3): 464–486.
- Bolger N, Davis A and Rafaeli E (2003) Diary methods: Capturing life as it is lived. *Annual Review of Psychology* 54: 579–616.
- Bracke A and Aguerre S (2015) Erasmus students: Joining communities of practice to learn French? In: Mitchell R, Tracy-Ventura N and McManus K (eds) *Social interaction, identity and language learning during residence abroad*. European Second Language Association (EUROSLA), pp. 139–168.
- Bradburn N, Rips L and Shevell S (1987) Answering autobiographical questions: The impact of memory and inference on surveys. *Science* 236: 157–161.
- Briggs JG (2015) A context-specific research tool to probe the out-of-class vocabulary-related strategies of study-abroad learners. *International Journal of Applied Linguistics* 25(3): 291–314.
- Briggs Baffoe-Djan JG and Zhou S (2020) Close encounters of the third kind: Quantity, type and quality of language contact during study abroad. In: Howard M (ed) *Study Abroad and the Second Language Learner: Expectations, Experiences and Development*. London: Bloomsbury, pp. 69–89.
- Brunec II, Ozubko JD, Barense MD and Moscovitch M (2017) Recollection-dependent memory for event duration in large-scale spatial navigation. *Learning and Memory* 24(3): 104–114.
- Burt CDB (2008) Time, language, and autobiographical memory. *Language Learning* 58(S1): 123–141.
- Bybee J (2008) Usage-based grammar and second language acquisition. In: Robison P and Ellis NC (eds) *Handbook of Cognitive Linguistics and Second Language Acquisition*. Abingdon: Routledge, pp. 216–236.
- Carroll, SE (2017) Exposure and input in bilingual development. *Bilingualism: Language and Cognition* 20(1): 3–16.

- Carter LA and Emsley R (2019) The analysis of experience sampling data. In: Palmier-Claus J, Haddock G and Varese F (eds) *Experience Sampling in Mental Health Research*. Abingdon: Routledge, pp. 18–39.
- Chen YW, Cordier R and Brown N (2015) A preliminary study on the reliability and validity of using experience sampling method in children with autism spectrum disorders. *Developmental Neurorehabilitation* 18(6): 383–389.
- Christiansen MH (2019) Implicit statistical learning: a tale of two literatures. *Topics in Cognitive Science* 11(3): 468–481.
- Clark EV (2003) *First language acquisition*. Cambridge: Cambridge University Press.
- Cosgun Ögeyik M (2018) The comparative effectiveness of noticing in language learning. *International Review of Applied Linguistics in Language Teaching* 56(4): 377–400.
- Csikszentmihályi M and Larson RW (1987) Validity and reliability of the experience sampling method. *Journal of Nervous and Mental Disease* 175: 526–536.
- Dewaele JM (2013) The link between foreign language classroom anxiety and psychoticism, extraversion, and neuroticism among adult bi- and multilinguals. *Modern Language Journal* 97(3): 670–684.
- Dewey DP (2017) Measuring social interaction during study abroad: Quantitative methods and challenges. *System* 71: 49–59.
- Dewey DP, Bown J and Eggett D (2012) Japanese language proficiency, social networking, and language use during study abroad: Learners' perspectives. *Canadian Modern Language Review/La Revue Canadienne des langues vivantes* 68(2): 111–137.
- Dressman M and Sadler RW (2020) *The Handbook of Informal Language Learning*. Hoboken, NJ: Wiley Blackwell.
- Droit-Volet S, Monceau S, Berthon M, Trahanias P and Maniadakis M (2018) The explicit judgment of long durations of several minutes in everyday life: Conscious retrospective memory judgment and the role of affects? *PLoS ONE* 13(4): e0195397.
- Ellis NC (2009) Optimizing the input: Frequency and sampling in usage-based and form-focused learning. In: Long MH and Doughty CJ (eds) *The Handbook of Language Teaching*. Oxford: Blackwell Publishing Ltd.
- Ellison, WD, Trahan AC, Pinzon JC, Gillespie ME, Simmons LM and King KY (2020) For whom, and for what, is experience sampling more accurate than retrospective report? *Personality and Individual Differences* 163(110071).
- Fazeli PL, Turan B (2019) Experience Sampling Method versus questionnaire measurement of HIV stigma: Psychosocial predictors of response discrepancies and associations with HIV outcomes. *Stigma and Health* 4(4): 487–494.
- Fernández J and Gates Tapia AN (2016) An appraisal of the language contact profile as a tool to research local engagement in study abroad. *Study Abroad Research in Second Language Acquisition and International Education* 1(2): 248–276.
- Freed BF, Dewey DP, Segalowitz N and Halter R (2004) The language contact profile. *Studies in Second Language Acquisition* 26: 349–356.
- Ganek H and Eriks-Brophy A (2018) Language Environment analysis (LENA) system investigation of day long recordings in children: A literature review. *Journal of Communication Disorders* 72: 77–85.

- García-Amaya L (2017) Detailing L1 and L2 use in study-abroad research: Data from the daily linguistic questionnaire. *System* 71: 60–72.
- Grüter T and Paradis J (2014) *Input and Experience in Bilingual Development*. Amsterdam: John Benjamins.
- Hektner JM, Schmidt JA and Csíkszentimihályi M (2007) *Experience Sampling Method: Measuring the Quality of Everyday Life*. London: Sage.
- Horwitz EK, Horwitz MB and Cope J (1986) Foreign language classroom anxiety. *The Modern Language Journal* 70(2): 125–132.
- Hsieh G, Li I, Dey A, Forlizzi J and Hudson SE (2009) Using visualizations to increase compliance in experience sampling. *Proceedings of the 10th International Conference on Ubiquitous Computing*, 164–167.
- Jobe J (2003) Cognitive psychology and self-reports: Models and methods. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation* 12: 219–227.
- Krashen, SD (1985) *The Input Hypothesis: Issues and Implications*. London & New York: Longman.
- Krishnamurty P (2008) Diary. In: Lavrakas PJ (ed) *Encyclopedia of survey research methods*. London: Sage, pp. 197–198.
- Lantolf JP and Thorne SL (2006) *Sociocultural Theory and the Genesis of Second Language Development*. Oxford: Oxford University Press.
- Larson RW, Moneta G, Richards MH and Wilson S (2002) Continuity, stability, and change in daily emotional experience across adolescence. *Child Development* 73(4): 1151–1165.
- Litt MD, Cooney NL and Morse P (1998) Ecological momentary assessment (EMA) with treated alcoholics: methodological problems and potential solutions. *Health Psychology* 17: 48–52.
- MacIntyre PD (1999) Language anxiety: A review of the research for language teachers. In: Young DJ (ed) *Affect in Foreign Language and Second Language Learning: A Practical Guide to Creating a Low-Anxiety Classroom Atmosphere*. New York, MacGraw-Hill Education, pp. 24–45.
- MacIntyre PD (2017) An overview of language anxiety research and trends in its development. In: Gkonou C, Daubney M, Dewaele JM (eds) *New Insights into Language Anxiety: Theory, Research and Educational Implications*. Bristol: Multilingual Matters, pp. 11–30.
- Magnan SS and Back M (2007) Social interaction and linguistic gain during study abroad. *Foreign Language Annals* 40(1): 43–61.
- Martínez-Arbelaiz A, Areizaga E and Camps C (2017) An update on the study abroad experience: Language choices and social media abroad. *International Journal of Multilingualism* 14: 350–365.
- Mehl MR and Robbins ML (2012) Naturalistic observation sampling: The electronically activated recorder (EAR). In: Mehl MR and Conner TS (eds) *Handbook of research methods for studying daily life*. New York: Guildford Press, pp. 176–192.
- McManus K, Mitchell R and Tracy-Ventura N (2014) Understanding insertion and integration in a study abroad context: The case of English-speaking sojourners in France. *Revue Française de Linguistique Appliquée* 2(XIX): 97–116.
- Mulligan CB, Schneider B and Wolfe R (2000). Time use and population representation in the Sloan Study of Adolescents. In: *NBER Working Paper No. t0265*.

- Naughton F, Riaz M and Sutton S (2015) Response parameters for SMS text message assessment among pregnant and general smokers participating in SMS cessation trials. *Nicotine & Tobacco Research* 18(5): 1210–1214.
- Ranta L and Meckelborg A (2013) How much exposure to English do international graduate students really get? Measuring language use in a naturalistic setting. *Canadian Modern Language Review* 69(1), 1–33.
- Rebuschat P and Williams JN (2012) Introduction: statistical learning and language acquisition. In: Rebuschat P and Williams JN (eds) *Statistical learning and language acquisition*. Berlin: Mouton de Gruyter, pp. 1-12.
- Rintala A, Wampers M, Myin-Germeys I and Viechtbauer W (2019) Response compliance and predictors thereof in studies using the Experience Sampling Method. *Psychological Assessment* 31(2): 226–235.
- Rose H and McKinley J (2020) Diaries and journals: Collecting insider perspective in second language research. In: Rose H and McKinley J (eds) *The Routledge Handbook of Research Methods in Applied Linguistics*. London: Routledge, pp. 348–359.
- Rose H, McKinley J and Briggs Baffoe-Djan J (2020) Diaries, journals and logs. In: Rose H, McKinley J and Briggs Baffoe-Djan J (eds) *Data Collection Research Methods in Applied Linguistics*. London: Bloomsbury Academic, pp. 133–152.
- Saffran JR (2003) Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science* 12: 110-114.
- Seibert Hanson AE and Dracos M (2019) The digital dilemma: L1 and L2 technology use, language learning, and motivation among US university students studying abroad. *Study Abroad Research in Second Language Acquisition and International Education* 4(2): 224–251.
- Shiffman S, Stone AA and Hufford M (2008) Ecological momentary assessment. *Annual Review of Clinical Psychology* 4: 1–32.
- Shrier LA, Shih MC and Beardslee WR (2005) Affect and sexual behavior in adolescents: A review of the literature and comparison of momentary sampling with diary and retrospective self-report methods of measurement. *Pediatrics* 115: 573–581.
- Sockett G (2014) *The Online Informal Learning of English*. London: Palgrave MacMillan.
- Stone AA, Schwartz JE, Broderick JE and Shiffman S (2005) Variability of momentary pain predicts recall of weekly pain: A consequence of the peak (or salience) memory heuristic. *Personality and Social Psychology Bulletin* 10: 1340–1346.
- Stone AA and Shiffman S (1994) Ecological momentary assessment in behavioral medicine. *Annals of Behavioral Medicine* 16(3): 199–202.
- Sundqvist P and Sylvén LK (2016). *Extramural English in Teaching and Learning*. London: Palgrave MacMillan.
- Teitler JO, Reichman NE and Koball H (2006) Contemporaneous versus retrospective reports of cohabitation in the fragile families survey. *Journal of Marriage and Family* 68: 469–477.
- Thigpen C (2019) Measurement validity of retrospective survey questions of bicycling use, attitude, and skill. *Transportation Research Part F* 60: 453–461.
- Tomasello M (2003) *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

- Truscott J and Sharwood Smith M (2011) Input, intake, and consciousness. *Studies in Second Language Acquisition* 33: 497–528.
- Turan B, Fazeli PL, Raper JL, Mugavero MJ and Johnson MO (2016) Social support and moment-to-moment changes in treatment self-efficacy in men living with HIV: Psychological moderators and clinical outcomes. *Health Psychology* 35: 1126–1134.
- Unsworth S (2013) Assessing the role of current and CUMULATIVE exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition* 16(1): 86–110.
- Unsworth S (2015) Quantity and quality of language input in bilingual language development. In: Nicoladis E and Montanari S (eds) *Lifespan perspectives on bilingualism*. Berlin: Mouton de Gruyter, pp. 136–196.
- van Berkel N, Ferreira D and Kostakos V (2017) The experience sampling method on mobile devices. *ACM Computing Surveys* 50(6): 40 pages.
- van Berkel N, Goncalves J, Lovén L, Ferreira D, Hosio S and Kostakos V (2019) Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports. *International Journal of Human-Computer Studies* 125: 118–128.
- Van Lier L (2000) From input to affordance: Social-interactive learning from an ecological perspective. In: Lantolf JP (ed) *Sociocultural theory and second language learning: Recent Advances*. Oxford: Oxford University Press, pp. 245–259.
- Verhagen SJW, Hasmi L, Drukker M, van Os J and Delespaul PAEG (2016) Use of the experience sampling method in the context of clinical trials. *Evidence Based Mental Health* 19(3): 86–89.
- Walls TD, Jung H and Schwartz JE (2006) Multilevel models for intensive longitudinal data. In: Walls TD and Schafer JL (eds) *Models for Intensive Longitudinal Data*. Oxford: Oxford University Press, pp. 3–37.
- Wearden JH (2008) The perception of time: Basic research and some potential links to the study of language. *Language Learning* 58(Suppl. 1): 149–171.
- Wulff S and Ellis NC (2018) Usage-based approaches to second language acquisition. In: Miller D, Bayram F, Rothman J and Serratrice L (eds) *Bilingual Cognition and Language. The state of the science across its subfields*. Amsterdam: John Benjamins, pp. 37–56.
- Xie J, Heddy BC and Greene BA (2019) Affordances of using mobile technology to support experience-sampling method in examining college students' engagement. *Computers & Education* 128: 183–198.