‘I knew that you were there, so I was talking to you’: the use of screen-recording videos in online language learning research

Wing Yee Jenifer Ho
Department of English, City University of Hong Kong, Hong Kong

Abstract
As language learning has become increasingly globalised, mobile and online, instances of language learning of significant value cannot be obtained by using conventional means such as on-site observation or video recording in classrooms. In this article, I present a new approach to collecting data in the online language learning context with an aim to capture the multimodal and embodied nature of language learning. Screen-recording as a research tool is an under-explored area; this article discusses some methodological and practical issues that researchers would encounter when using this approach, and outlines the considerations researchers need to make when collecting data in such kind of contexts. The article argues that screen-recording is an innovative data collection tool in the research of language learning, and it should be included in the repertoire of mobile methods to study (im)mobilities of language learners, teachers and knowledge.

Keywords
Online learning, language learning, video, multimodality

Introduction
The use of the Internet in qualitative research is an emerging phenomenon as the world is becoming more and more connected. At the same time, researching the Internet is different from researching a physical environment as it is virtual and fluid. Internet users are now afforded with multimodal ways of meaning-making (Kress, 2010), which has brought to the fore the need for researchers to obtain data that truly reflect the complexity, fluidity and mobility of communication and diverse ways of meaning-making with the use of the Internet.
In this paper I am concerned with the learning of Chinese by multilingual adult learners through the use of an online language learning platform. This paper reports on a larger study by Ho (2018b) on self-directed Chinese learning in which 11 multilingual adult learners were observed; all of them were located in different parts of the world at the time of the research. All data collection was done virtually, which created unique challenges to the researcher. Specifically, the research question that I seek to answer in this paper is how can screen-recording be used as a method to understand how multilingual learners mobilise linguistic and semiotic resources in online language learning contexts?

Online language learning platforms offer a virtual learning space where learners can decide their own learning goals by utilising resources of their own choice, including linguistic and semiotic means of making-meaning, different life and work experience, as well as their personal history. This kind of learning is relatively informal, and learners have a high degree of control over various learning decisions such as what kind of content to learn, how much time is devoted to learning and where to learn. Online language learning refers to diverse learning settings, and this study focuses on language learning using online language learning platforms, or what Reinhardt (2019) calls ‘Social network-enhanced commercial CALL sites and services’ (SNECSs), in combination with other resources that learners had, such as textbooks, digital gaming, socialisation with speakers of the language or their prior experience of language learning. Studies focusing on online language learning platforms, or SNECSs abound (e.g. boyd and Ellison, 2008; Harrison and Thomas, 2009; Stevenson and Liu, 2010), but few studies have focused on the mobilisation of resources by learners as they used these platforms. The mobilisation of resources in online language learning has to be explored in greater detail, which requires the use of innovative research methods.

Whilst online learning of English has been widely discussed in the literature (e.g. Sockett, 2013), there is an increased interest in the online learning of other languages. Looking at how learners used Duolingo for ab initio learning of Turkish, Isbell et al. (2017) found that learners’ persistence in learning varied across timescales, and motivation was hard to sustain due to the quality of materials offered by the platform, which resulted in limited language learning outcomes. Online language learning also includes informal language learning, which makes use of online resources that can be found in online communities. For instance, Isbell (2018) observed how learners of Korean used a Reddit forum to gain metalinguistic knowledge about Korean, and revealed that only little target language was used in interactions. In relation to Chinese learning, there has been a strong interest in researching how technology assists Chinese teaching and learning both in and outside of the classroom. For instance, Qian et al. (2018) conducted a study to understand the learning strategies used by distance learners when using a mobile application to learn Chinese characters. They found the emergence of new learning strategies that were not salient in ‘traditional’ learning contexts. In light of this new development, White and Zheng (2018) suggested a list of priorities for future research, one of which is a call for researchers to identify new research questions, and to address them by creating new research tools and methodologies.

The need for new research tools and methodologies is particularly evident in online language learning settings in which there is only limited understanding of how this kind of learning can be best captured and recorded, as researchers are often remote from the
site of learning. In view of the research questions of the larger study, one specific question that is relevant is: how can researchers capture this type of learning for analysis? Based on the research context of online Chinese learning, this paper focuses on exploring the use of screen-recording to capture the goings-on when learners were using a particular online language learning platform.

One method that has been used to capture this multifaceted world of online learning and meaning-making is by using virtual ethnography. Hine (2000) commented that virtual ethnography can provide a detailed understanding of how technology is experienced when it is being used. Observation is central to any kind of ethnographic research, regardless of whether the researcher opts to ‘doing ethnography’, ‘adopting an ethnographic perspective’ or ‘using ethnographic tools’ (Green and Bloome, 1997). While observation in the physical world has been done for decades, observation on the Internet is still at its infancy, primarily due to the methodological challenges faced by researchers in obtaining observation data from the Internet.

As the social and physical worlds are changing, there is a need for researchers to find new ways of inquiry to understand these worlds. Law and Urry (2004) argue for a need for the social sciences to ‘re-imagine themselves, their methods, and their “worlds” if they are to work productively in the twenty-first century where social relations appear increasingly complex, elusive, ephemeral, and unpredictable’ (p. 390). They also argue for the use of ‘messy’ methods that may not sit comfortably within the existing ‘academic habits’. This paper is structured as follows. I begin by explaining how screen-recording videos work as an effective tool to observe and capture online language learning, and then I present an example of how the study is conducted and the practical considerations of using this method. I then discuss the issue of ‘observer’s paradox’ in the context of screen-mediated observation, and conclude by stressing the importance of the need for researchers to seek new ways to capture the complexity and fluidity of online language learning.

**Screen-recording videos as observation tool**

Videos have long been used as a data source to observe human interactions in different settings (see Bezemer, 2008; Flewitt, 2006; Heath and Hindmarsh, 2002). They provide a valuable source of data and offer the possibility of preserving a record of interactions that can be viewed repeatedly for later analysis. It often requires setting up a video camera in a physical site. However, when it comes to virtual interaction, a considerable number of studies rely on researchers being ‘present’ when the online interaction occurs to observe the goings-on during the chat, or rely on transcripts generated retrospectively from online chatrooms or forums as their main source of data. While these are possible ways of capturing the essence of the online interactions, as communication is increasingly online, there is a need to capture the interactions on the screen as they happen, and thus this paper suggests that screen-recording could be used as an observation tool to understand the goings-on on the screen, while also taking into account the multimodal and embodied nature of interactions. Supplemented by thinking-aloud protocol and semi-structured interviews, thick descriptions can also be made in line with the ethnographic orientation of the study.
Thus far, the use of screen-recording as a research tool has been used in the field of Computer Science and Business, such as usability testing for websites or computer games. For instance, screen-recording is used by Salinger et al. (2008) to investigate how graduate students do pair programming while sharing the same computer. Increasingly, screen-recording is taken up by other disciplines such as Psychology. Meredith (2016) discussed the process of developing a transcription system with screen captures of online chat data. As regards the use of screen-recording technology in language teaching and learning, Develotte et al. (2010) explored how teacher trainees learnt to teach French through Skype. The authors observed that teacher trainees faced various challenges teaching in webcam-mediated context, some of which included the ephemeral nature of the webcam image, and also the restrictive view of the image. However, they also argued that the use of webcam enhances learner’s motivation because a visual link is created between the teacher trainees and the learners. They identified five degrees of utilisation of webcam by teacher trainees, ranging from not appearing on the video at all, to looking straight into the webcam, directly to the learners. One interesting point the authors raised was that ‘[t]he video window can be compared to a theatre stage that the teacher trainees use to enact their role: they learn to adapt their gestures to the size of the stage’ (p. 309), which points to the performative nature promoted by the use of webcam. In the field of online language learning and teaching, webcam was often used as part of the research, but there is a lack of research on using it as an observation and data-gathering tool.

Screen-recording technology as a data collection method has raised issues related to privacy. Tang et al. (2006) conducted a study using screen-recording to better understand how team members accomplish work collaboratively. The authors reflected on the use of screen-recording as a data collection method, the sensitive nature of recording participants’ screens and how being observed had changed participants’ behaviour. Whilst the authors appreciated the advantage of the ‘unobtrusive’ nature of using screen-recording to collect data on online interactions, they also acknowledged its ‘invasive’ nature. As they noted, participants expressed concerns about exposing data related to third parties without their informed consent. Nevertheless, the authors argued that the advantages of obtaining rich, empirical data outweigh the privacy concerns provided that researchers communicate with the participants clearly about how the data would be used and safeguarded. Guichon (2017) also provided useful guidance to help researchers deal with ethical issues associated with collecting multimodal data, especially in situations where participants’ faces are shown.

The use of screen-recording to research language learning, in which much of the data generated is multimodal and show embodied actions, is still scarce due to various methodological and practical constraints. More studies focusing on the learners’ perspective of interacting with online platforms are needed to understand the multimodal, multisensory and embodied nature of language learning. The use of new technology as an observation tool is also crucial in understanding the individual trajectory taken by different learners as they learn a language in online platforms. As Reinders and White (2011) commented,

A reconceptualization of language education as the provision of a collection of affordances that start from the learners as individuals, and include classrooms, materials, native speakers, teachers, assessment, other learners, the workplace, and so on, has been made more practically
feasible, and methodologically easier to investigate, through the pervasive use of technology (p. 2)

Table 1 summarises the similarities and differences of screen-mediated observation and physical observation. With the help of Camtasia, the screen-recording software used in this study, it is relatively easy to observe the screen, facial expression and verbal commentary simultaneously on the same timeline, whereas this kind of triangulation is hard to do had the researcher only carried out observation physically. Using Camtasia as an observation tool is also less obtrusive to the participants as the researcher need not be physically present in the same room with the participant. However, the drawback of using screen-mediated observation is the loss of contextual detail, which has to be obtained through other means.

This study adopts an ethnographic approach, and therefore only obtaining screen-recording data is not enough to build a ‘thick description’ of the learning process. In this study, screen-recording videos were the dominant data source, complemented by thinking-aloud method and semi-structured interviews. These data sources complemented each other.

**Thinking-aloud protocol**

The thinking-aloud protocol is a kind of introspective method that is commonly used in psycholinguistics. The key characteristic of the think-aloud method is that learners are asked to ‘verbalise their thought processes’ (Nunan, 1992: 117). This is usually done while the task is ongoing so that researchers can access short-term happenings in the learner’s working memory, which differs from the talk-aloud protocol in which verbalisation is done retrospectively (Ericsson, 2003).

Drawing from a Vygotskian perspective of learning, second language learning is seen as a mediated process. In particular, Lantolf (2000) mentioned the importance of self-mediation in language learning whereby the use of self-directed speech plays an important role in learning. Nevertheless, this kind of self-directed speech is not easy to obtain, especially from adult learners, as verbalisation during task is normally suppressed as learners get older. The thinking-aloud method could help to elicit this self-directed speech data from participants, allowing the researchers to understand how they solved problems through languaging (Swain, 2006). The design of the present study provided minimal prompting in the verbalisation process as Ericsson and Simon (1984) pointed out that ‘differences in

| Similarities | Observation through screen recording | Observation in physical setting |
|--------------|-------------------------------------|--------------------------------|
| **Observation** | The focus on the screen content | Hard to observe both the screen content and facial expression at the same time |
| **Differences** | Relatively easy to observe both screen content and facial expression with the help of screen-recording software such as Camtasia | Presence of contextual detail |
| Absence of contextual detail | ‘Retrospective observation’ | ‘Real time observation’ |
| Less obtrusive to participants | | More obtrusive to participants |
performance were induced by telling the subject how to verbalise’ (p. 107). In this study, learners were asked to use this technique while they were going through the different lessons to let me understand their selection and mobilisation of resources. Whilst there are advantages of using the thinking-aloud method in the current research study, such as its ability to elicit real-time narration of the learning process, as opposed to the risk of declined accuracy in recall in retrospective methods, the drawback of using such a method is that the data obtained would be messy and unstructured. The unfamiliarity of the thinking-aloud task by participants may also affect the depth and clarity of the narration (Cowan, 2017).

Semi-structured interviews
Semi-structured interviews were conducted primarily through Skype as most of the participants in the study were located in different countries. Email interviews were occasionally conducted when clarifications were needed. In line with an ethnographic approach, the purpose of these interviews was to obtain deep understanding about the context in which online learning took place, and how participants brought with them their life experiences and histories in the learning context. The interviews were informed by a brief interview guide but were structured loosely so as to encourage deep and open conversations between the participants and the researcher.

Online language learning in action
As my main research question for the larger project (see Ho, 2018b) was to understand how multilingual learners mobilise their linguistic and semiotic resources to learn Chinese, it was crucial for me to gain access to the learners’ engagement with the platform by using screen-recording as an observation tool. In the larger study, participants were recruited through various online channels. As mentioned in Tang et al. (2006), because of the ‘invasive’ nature of this data collection method, it was difficult to recruit participants to join the study, and the recruitment process took almost 1 year, with repeated ‘call for participants’ circulated in various online channels throughout the year. Information of the research was circulated through social media, and through the official blog of the online language learning platform concerned. Once participants signed up and expressed interest to participate in the research, they were invited to attend a Skype meeting so that more details of the research were explained to them, including instructions on how to install Camtasia. They were clearly notified that they had full control of what information would be recorded, and advanced notice was given to participants to be careful not to include any third-party online interactions, such as emails and chats, while they were recording their screen. They had also been assured that if any third-party materials were recorded by accident, they would be disregarded and data segments containing third-party interactions would be deleted. After explaining the functions of the screen-recording software, informed consent was sought, and the data collection process started.

The data collection lasted for 4 weeks. During these 4 weeks, participants were asked to learn Chinese using the online platform at their leisure. Some recorded their learning at home, and some at their workplace. They were encouraged to upload one ‘learning session’ to the designated secure, password-protected Dropbox each week. The duration
and frequency of these ‘learning sessions’ were decided by the participants so as to allow flexibility, and to reflect the normal routine of the learners. The duration of these ‘learning sessions’ ranged from 5 min to 1 h.

By asking recruited learners to record their screens whenever they wished without the researcher’s presence, it was one way to imitate the setting of out-of-class learning in real life, where learners used the platforms at their leisure without any kind of supervision. Nonetheless, there are limitations to using screen-recording data as well. Similar to video recording, screen-recording only provides a partial representation of reality. It cannot capture happenings outside of the screen that might offer rich contextual, situational cues, which may affect learners’ behaviour. Therefore, the use of screen-recording data alone is not enough; it has to be used with other data collection methods to obtain a more holistic description of the interactions, and in the present study, they include the use of thinking-aloud method and semi-structured interviews. Limitations on recording without researcher’s presence also include an increased level in self-consciousness and performativity (Stenström et al. 2002, cited in Macaulay, 2009), which is addressed later in this paper.

**Analysis of screen-recording data**

This study adopts a qualitative approach to data analysis in order to understand how multilingual learners mobilise resources in online platforms for Chinese learning. The use of screen-recording videos as an observation tool in the study had generated more than 18 hours of screen-recording videos in total. This is a large amount of multimodal data that needs to be analysed.

The 18 hours of recording were viewed repeatedly by the researcher. Videos were coded and multimodal analysis was carried out to the selected segments. The selection of segments for further analysis was informed by Moment Analysis, which has its roots in Interpretative Phenomenological Analysis (Eatough and Smith, 2008; Smith and Osborn, 2008; Smith et al., 2009). Moment Analysis guides the researcher to focus on ‘critical and creative moments of individuals’ actions’ (Li, 2011: 1224), which is in line with the socio-linguistic focus of the study to understand how linguistic and semiotic resources in the learners’ repertoire are ‘brought about’ through multilingual practices (Li, 2011: 1224).

As informed by my research question, a social semiotic approach to multimodality (Kress, 2010; Kress and van Leeuwen, 2006) was used as an overarching analytical framework to understand what resources are mobilised in the learners’ repertoire to learn Chinese in the online platform. The social semiotic approach to multimodality challenges the ideological preoccupation of language as the dominant mode of communication. It takes into account how a range of modes make meaning in a multimodal ensemble, that is, how modes such as speech, writing, images orchestrate to make meaning. This framework also has a focus on sign-makers’ agency as it emphasises on how signs are made rather than used (Kress, 2010). This framework is particularly relevant to the case of online learning in which learners are constantly traversing between online and offline resources, orienting to the screen and adjusting their bodies to facilitate their learning (Ho, 2018a; Ho and Li, forthcoming). At a micro level, a social semiotic approach to multimodality also allows the fine-grained analysis of how learners interact with the platform, and how they make use of the affordances of the platform as they learn Chinese.
Let us now look at an example to demonstrate how screen-recording videos were created by recruited learners. Figure 1 shows the interface of Camtasia as seen from the researcher’s computer. Recruited learners were given instructions to install the free trial version of the software on their computer, and to turn it on when they started language learning on the online platform. It recorded the screen that the learner was looking at, and the frontal camera recorded the face of the learner. The internal microphone also recorded the thinking-aloud of learners while the recording was in progress. All these sources of data are situated on the same timeline. It is therefore much easier to synchronise the goings-on on the screen as well as the learners’ facial expression and speech.

As can be seen from Figure 1, Camtasia offers the researcher access to three types of data simultaneously: (a) the screen that the learner was looking at, (b) learners’ facial expression (including modes of gaze, posture, gesture), and (c) learners’ thinking-aloud commentary. These three sources of data were then transcribed multimodally for subsequent analysis.

One of the challenges faced by researchers in analysing multimodal data is about transcription. The issue of representing multimodal data has been widely discussed in the literature (e.g. Bezemer and Mavers, 2011; Dooly and Helm, 2017; Mavers, 2009; Meredith, 2016; Mondada, 2007; Norris, 2004).

In this research, I designed a multimodal transcript that contains the following components: speech and cursor movement (obtained from thinking-aloud and the screen), screenshot (obtained from the screen), typing (obtained from the screen), and facial expressions and other contextual information (obtained from the webcam of the participants’ computer). An example of the layout of the transcription is shown in Figure 2.

Figure 2 shows a snapshot of a learner learning a Chinese character using an online platform. He was shown completing an online quiz to type in the box the English translation of the given Chinese character. This transcript is by no means a representation of all the modes used in the interaction. As with all types of transcriptions, no matter how detailed they are, they only present a partial view of the dynamic process of learning (Flewitt et al., 2014). This transcript shows how the learner mobilises and orchestrates different resources at a given point in time. In his speech, we know how he came up with the English translation of the given Chinese character, which is by reading the clue, in small font underneath the Chinese character, which says ‘adverb’. With this knowledge, he felt ‘perhaps’ could be the answer for this exercise. Nevertheless, from his facial expression, gaze and tone of his voice we can see a degree of uncertainty; even though he has the correct answer, it was not a confident one; it was more of an educated guess. This insight could not be obtained had screen-recording and thinking-aloud were not used. We would not be able to establish his uncertainty by just looking at the final ‘product’ – a correct answer.

Reflections on the use of screen-recording videos as the observation tool

The use of screen-recording videos as observation and data-gathering tool is a novel attempt in the field of language learning, and therefore a pilot study was done to understand the affordances and constraints of the method in order to better inform the main study. The insights from the pilot study had changed the research design. For example, in the pilot
Figure 1. The interface of Camtasia as seen from the researcher’s computer (Ho, 2018b).
| Speech and cursor movement | Screenshot | Typing | Facial expressions and other relevant happenings |
|---------------------------|------------|--------|--------------------------------------------------|
| Now, the only reason I’m writing (2) ‘perhaps’ is because it says adverb underneath, and | ![Screenshot](image) | perhaps | Uncertain whether it is ‘perhaps’ or not. The pause before uttering the word ‘perhaps,’ the uncertain gaze of his eyes, and the rising tone. |

| | | Staring at the screen intently. Body leaning close to the screen. |

**Figure 2.** Extract of multimodal transcription (from Li and Ho, 2018).
study, stimulated recall technique was used in an attempt to clarify specific actions that occurred in the recording. As my research question requires understanding of not only what the learner produces, but also the process of learning, it was thought that a retrospective stimulated recall would work well as an introspective method. However, after reviewing the literature, and after carrying out retrospective stimulated recall with a participant in the pilot study, it was then deemed unsuitable for this study because of the time gap between the learning session and the retrospective stimulated recall session. Due to practical constraints, such as the widespread geographical location of the participants, and the busy schedules of both the participants and the researcher, it was difficult to conduct the stimulated recall interviews shortly after the learning sessions, and therefore resulting in a decrease in accuracy of the recall (Bloom, 1954; Cotton and Gresty, 2006; Yinger, 1986).

As a result, the thinking-aloud method was used instead to replace retrospective stimulated-recall technique in the main study. The thinking-aloud protocol was used alongside screen-recording. The primary purpose of adopting this protocol is to gain access to the self-directed, inner speech of learners in order to obtain an emic understanding of the learners’ interaction with the platform, and to understand the learning process. Participants were instructed by the researcher to verbalise their thoughts in order to make explicit the resources that they drew on when learning Chinese. They were not instructed to speak to the researcher directly, and therefore it could very likely be the case that their speech was directed to themselves. The thinking-aloud method is particularly in line with the concept of languaging, which involves the use of language to solve problems (Swain, 2006).

Throughout the data collection process, especially in the pilot study, every care was taken to minimise my effect in the screen recording process with the intention to preserve the authenticity of the data, although it is not an achievable goal. Nevertheless, I made an effort to imitate their usual learning environment as best as possible, such as asking the recruited learners to install Camtasia and record their lessons at their ‘usual’ learning place so they could do their lessons in a comfortable and natural setting, instead of doing at a designated place. I gave minimal instructions on what topics they should learn, and I stayed out of the recording process as much as possible, giving them autonomy by letting them make learning decisions by themselves. However, reflecting on the process at a practical level, it has to be admitted that my perceived ‘absence’ in the process did not help much in terms of putting learners at ease and enabling them to behave ‘naturally’. For instance, one of the learners mentioned in the follow-up semi-structured interview that ‘I knew that you were there, so I was talking to you’ (Learner 1, semi-structured interview). It is clear that whatever she did, she was aware that I would be watching, although not in real time, and she would adjust her behaviour because of this. In another case, a female learner admitted in the follow-up interview that she had purposefully ‘made herself presentable’ such as by wearing make-up and dressing up in beautiful clothes because she knew that she would be on camera. Inevitably, it can be seen that the use of screen-recording as an observation tool leads to the ‘observer’s paradox’ (Labov, 1972), as well as the concern that this is actually a ‘performance of learning’, which are discussed later on.

The pilot study had helped me anticipate some of the technical difficulties faced by the participants in the main study. For instance, some participants needed help with the installation of the software. As the participants come from different backgrounds and
belong to different age groups, computer knowledge cannot be assumed. Care must be taken to make sure that all learners felt comfortable participating in a study that required the use of a software which was unfamiliar to them. As mentioned before, learners were given the autonomy to decide how much they wanted to learn in each ‘learning session’, and therefore it was difficult to estimate how long the videos would be. Some ‘learning sessions’ only lasted for 5 minutes, whereas some lasted for almost an hour. The large file size in some of these videos created a problem during file transfer. Some learners recalled spending hours trying to upload the videos to the designated Dropbox. This was unavoidable due to the large file size of the recording. It can only be hoped that future technological advancement could solve this problem. Some of the learners also expressed privacy concerns, and it was resolved by allowing learners to have full control of the data through the secured Dropbox folder. As the data were generated by the learners, they were able to decide whether they would like to share the videos with the researcher or not. This is one way to safeguard the privacy of participants.

The use of the thinking-aloud technique was successful in the study. As I watched the recording, I felt that I existed in the same space with the learners and that they were directly talking to me, although they were not instructed to talk to me. They treated me as an imaginary audience. Not only did I gain remote access to the physical environment in which the recording was made, I also gained access to the inner thoughts of the learners when they were languaging, observing how they orchestrated resources of different temporal scales, and how they used their bodies to adjust and adapt to the learning environment (see Ho and Li, forthcoming).

Some practical issues had to be aware of when carrying out a study involving the use of screen-recording videos. Some of these issues were also informed by the pilot study. For example, I realised that explicit instructions had to be given to participants on how to use Camtasia to record the screen and the face, and therefore in the main study I created a simple guide to help learners with the use of the software (see Appendix). In the pilot study, some learners only recorded the screen, or only their face, but not both. Furthermore, while watching the videos, I realised the constraints of screen-recording videos. When learners pointed at the screen with their fingers, it was not captured in the camera, as those actions fall outside of the scope of the camera. Therefore, after the pilot study, I encouraged participants instead of pointing at the screen with their fingers or using deictic expressions such as ‘here’ and ‘there’, it would be better to also use the cursor to indicate the part that they were referring to, because the pointing action and the deictic expressions lost meaning if they were not captured in the screen recording.

The virtual and remote nature of the research requires the researcher to pay special attention to the participants. For example, regular check-ins were required, as participants were located in different parts of the world. I kept in touch with participants on a weekly basis to allow them a chance to ask questions regarding what they had to do and to help them solve technical problems remotely. I speculate that if it had not been done, the attrition rate of this kind of study could be high, especially if learners felt they were being left without help. In the present study, 20 participants were originally in the study, and 9 of them withdrew either because of their lack of time to commit in the study, or because of loss of contact. It is indeed not surprising as it was hard to build up trust at the very beginning when the participants did not know the researcher well.
Reflexivity and the observer’s paradox

Hine (2000) argues that ethnographers can only present a selective account of the reality that has been observed, which is determined by the perspective taken by each individual researcher. Therefore, it has to be acknowledged that my role as a researcher shaped the data in different ways. To be reflexive as a researcher means recognising ‘the dominant epistemological position from which ethnographic work should be undertaken’ (Dicks et al., 2005: 32), and recognising the fact that the description of the social by the researcher is never a neutral one.

My role as a researcher and as a Chinese speaker makes a difference in my relationship with the participants. Whilst all the communications related to the research was done in English, when it came to the screen-recording, thinking-aloud and semi-structured interview, Chinese was used occasionally when learners were trying to explain to me what they had learnt. They were sometimes shy to say the Chinese characters out loud because they were conscious of their pronunciation, and they were afraid of being judged. There was no instruction on the language used in the thinking-aloud, but all participants understood that they had to use English because of me. Some learners did not say a word in the thinking-aloud, possibly because English was a foreign language to them, and it could be challenging for them to think-aloud using English. In this instance, it can be seen that the research is shaped by the linguistic repertoires that we share. My role as a researcher is also significant. As the participants had granted their permission for the data to be shown in academic conferences and publications, they knew that they would potentially be speaking to the global academic community, and therefore they only shared with me their ‘best’ videos.

The kind of observation that I chose to conduct, that is, to be more of an ‘observer’ rather than a ‘participant’ reflected my orientation to the recruited learners. Whiteman (2012) pointed out that unobtrusive observation prevents the researcher from ‘muddying the waters’.

Unobtrusive observation is an ideal, but it is not always achievable. In practical terms, the ‘observer’s paradox’ is an inescapable outcome of any research that involves observation and recording. This term was first used by Labov (1972) who had a sociolinguistic interest to find out the way how people talk. The paradox is outlined as follows: ‘the aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain these data by systematic observation’ (Labov, 1972: 209).

Although the focus of the current study is not on how people talk, rather on what they talk about, the effect of ‘observer’s paradox’ can still be felt. As argued by Law and Urry (2004), research methods are performative in that they enact realities. In this study, screen-recording with the use of thinking-aloud protocols is inevitably susceptible to ‘observer’s paradox’. Whilst effort was made to minimise its effects in the recording process, as outlined in the previous discussions, I also argued that it was difficult, if not impossible, to collect ‘naturalistic data’, a view also echoed by Meyerhoff et al. (2011). It has to be acknowledged that this can actually be seen as a case of ‘performance of learning’, which was done for the sake of the research, but differs from studies that place participants under experimental conditions. Relatively speaking, this ‘performance of learning’ allowed participants to have more space to express their ‘usual’ learning trajectory as best they could. However, it must be admitted that due to the constraints of the screen-recording method, participants had to change some of their behaviours, and they had to be acknowledged in the analysis. For example, in addition to making themselves...
presentable, some learners admitted that they normally used mobile apps for learning, but they had to switch to the desktop site of the platform so that they could record their learning with the software; by the time of this research, screen-recording could only work reliably on a computer. It also changed the time and place where learners used the platform. Some learners reported that they had to make an effort to find a quiet place and time so that they could record themselves without other people around, which can be seen as a self-initiated effort to protect the privacy of their families and friends. Self-generated or self-gathered data can therefore be seen as a way to ethically collect data involving recording and observation in which participants have full control.

There are indeed limitations to the use of the thinking-aloud technique, which include learners having to split their attention between the task and the narrative, as well as the performative nature of learners, as they had to do something that they were not familiar with (Cowan, 2017; Ericsson and Simon, 1984). Nonetheless, in this study, the benefits of thinking-aloud outweigh its limitations. Unlike natural sciences, the aim of studies informed by ethnographic methods is not to achieve absolute objectivity. Using Duranti’s (1997) words, ‘the problems with the term “objectivity” arise from its identification with a form of positivistic writing that was meant to exclude the observer’s subjective stance’ (p. 85), which is ‘a questionable goal’. I agree with Duranti’s view that the effect of the researcher or artefacts are often times overestimated. The purpose of the study is to understand learners’ mobilisation of resources while taking into account the context in which the recording was undertaken as well as the effect of the technology. As an applied linguist with an orientation to ethnographic methods, a balance of objectivity and subjectivity has to be maintained (Duranti, 1997; Schilling, 2013).

**Conclusions**

The current language learning landscape has changed dramatically and it has become increasingly mobile with a low degree of presupposability (Blommaert and Backus, 2013). As applied linguists, this rapidly changing environment calls for innovative ways to capture how language learning is done by individual learners who are highly mobile. I have presented in this article a novel way to obtain such data with the use of screen-recording videos complemented by the thinking-aloud technique and semi-structured interviews. I have also addressed in detail how the methods I chose were relevant to the research question. This article also discussed the issue of ‘observer’s paradox’ associated with the use of observation and recording in an online setting.

In recent years, there has been a call by social researchers to develop more mobile methods of studying (im)mobilities, such as the exploration of imaginative and virtual mobilities of people through analysing their digital footprints, recovering people’s memories using photographs or letters or following participants from one place to another (Büscher and Urry, 2009). Screen-recording can be added to the repertoire of mobile methods to understand how transnational learners with multilingual, multimodal, multisensory and multisemiotic resources learn in online platforms. The use of screen-recording videos for data collection is still at its infancy, and more research has to be done to explore the affordances and constraints of using such method, as well as ways to analyse the vast amount of data generated, such as how to transcribe multimodal data obtained from screen-recording. As communication becomes more global and mobile, it is crucial for us to step
out of our comfort zones and keep on searching for innovative ways to capture the complexity of language learning ‘on the move’.

Acknowledgements

Thanks must be given to the volunteers who participated in this study, and to the anonymous reviewers who have greatly improved the quality of this article.

Disclosure

The author(s) report no conflict of interest. The authors alone are responsible for the content and writing of the paper.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The data from this research is from the Author’s PhD study which is funded by the Bloomsbury Colleges Research Studentship.

ORCID iD

Wing Yee Jenifer Ho https://orcid.org/0000-0002-9725-9494

References

Bezemer J (2008) Displaying orientation in the classroom: students’ multimodal responses to teacher instructions. *Linguistics and Education* 19(2): 166–178.

Bezemer J and Mavers D (2011) Multimodal transcription as academic practice: a social semiotic perspective. *International Journal of Social Research Methodology* 14(3): 191–206.

Blommaert J and Backus A (2013) Superdiverse repertoires and the individual. In: De Saint-Georges I and Weber J-J (eds) *Multilingualism and Multimodality: Current Challenges for Educational Studies*. Rotterdam: Sense Publishers, 11–32.

Bloom B (1954) The thought process of students in discussion. In: French SJ (ed.) *Accent on Teaching: Experiments in General Education*. New York: Harper, 23–46.

boyd DM and Ellison NB (2008) Social network sites: definition, history, and scholarship. *Journal of Computer-Mediated Communication* 13(1): 210–230.

Büscher M and Urry J (2009) Mobile methods and the empirical. *European Journal of Social Theory* 12(1): 99–116.

Cotton D and Gresty K (2006) Reflecting on the think-aloud method for evaluating e-learning. *British Journal of Educational Technology* 37(1): 45–54.

Cowan J (2017) The potential of cognitive think-aloud protocols for educational action-research. *Active Learning in Higher Education* 20(3): 219–232. DOI: 10.1177/1469787417735614.

Develotte C, Guichon N and Vincent C (2010) The use of the webcam for teaching a foreign language in a desktop videoconferencing environment. *ReCALL* 22(3): 293–312. Run Run Shaw Library, City University of Hong Kong. DOI: 10.1017/s0958344010000170.

Dicks B, Mason B, Coffey A, et al. (2005) *Qualitative Research and Hypermedia: Ethnography for the Digital Age*. London: SAGE.

Dooly M and Helm F (2017) Challenges in transcribing multimodal data: a case study. *Language Learning & Technology* 21(1): 166–185.

Duranti A (1997) *Linguistic Anthropology*. Cambridge: Cambridge University Press.

Eatough V and Smith JA (2008) The SAGE handbook of qualitative research in psychology. In: Willig C and Stainton-Rogers W (eds) *The SAGE Handbook of Qualitative Research in Psychology*. London: SAGE Publications, 179–194.
Ericsson KA (2003) Valid and non-reactive verbalization of thoughts during performance of tasks. *Journal of Consciousness Studies* 10(9–10): 1–18.

Ericsson KA and Simon HA (1984) *Protocol Analysis: Verbal Reports as Data*. London: MIT Press.

Flewitt R (2006) Using video to investigate preschool classroom interaction: education research assumptions and methodological practices. *Visual Communication* 5(1): 25–50.

Flewitt R, Hampel R, Hauck M, et al. (2014) What are multimodal data and transcription? In: Jewitt C (ed.) *The Routledge Handbook of Multimodal Analysis*. 2nd edn. London: Routledge, 44–59.

Green J and Bloome D (1997) Ethnography and ethnographers of and in education: a situated perspective. In: *Handbook of Research on Teaching Literacy Through the Communicative and Visual Arts*. New York: Macmillan Publishers, 181–202.

Guichon N (2017) Sharing a multimodal corpus to study webcam-mediated language teaching. *Language Learning & Technology* 21(1): 56–75. Available at: http://llt.msu.edu/issues/february2017/guichon.pdf.

Harrison R and Thomas M (2009) Identity in online communities: social networking sites and language learning. *International Journal of Emerging Technologies & Society* 7(2): 109–124.

Heath C and Hindmarsh J (2002) Analysing interaction: video, ethnography and situated conduct. In: May T (ed.) *Qualitative Research in Action*. London: SAGE Publications, 99–121.

Hine C (2000) *Virtual Ethnography*. London: SAGE Publications.

Ho WY (2018a) Mobility and language learning: A case study on the use of an online platform to learn Chinese as a foreign language. *London Review of Education* 16(2): 239–249.

Ho WY (2018b) ‘Translanguaging in online language learning: Case studies of self-directed Chinese learning of multilingual adults’. *Unpublished PhD Thesis*. Institute of Education, University College London.

Ho WY and Li W (forthcoming) Mobilising learning: a translanguaging view. *Chinese Semiotic Studies* 15(4): 1–26.

Isbell DR (2018) Online informal language learning: insights from a Korean learning community. *Language Learning & Technology* 22(3): 82–102.

Isbell DR, Rawal H, Oh R, et al. (2017) Narrative perspectives on self-directed foreign language learning in a computer- and mobile-assisted language learning context. *Languages* 2(4): 1–20.

Kress G (2010) *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London: Routledge.

Kress G and van Leeuwen T (2006) *Reading Images: The Grammar of Visual Design*. 2nd edn. London: Routledge.

Law J and Urry J (2004) Enacting the social. *Economy and Society* 33(3): 390–410.

Labov W (1972) *Sociolinguisitc Patterns*. Philadelphia: University of Pennsylvania Press.

Lantolf JP (2000) Second language learning as a mediated process Survey article. *Language Teaching* 33(2): 79–96.

Li W (2011) Moment Analysis and translanguging space: discursive construction of identities by multilingual Chinese youth in Britain. *Journal of Pragmatics* 43(5): 1222–1235.

Li W and Ho WY (2018) Language learning *Sans Frontieres*: a translanguaging view. *Annual Review of Applied Linguistics* 38: 33–59. DOI: 10.1017/S0267190518000053.

Macaulay R (2009) *Quantitative Methods in Sociolinguistics*. London: Palgrave Macmillan.

Mavers D (2009) Student text-making as semiotic work. *Journal of Early Childhood Literacy* 9(2): 141–155.

Meredith J (2016) Transcribing screen-capture data: the process of developing a transcription system for multi-modal text-based data. *International Journal of Social Research Methodology* 19(6): 663–676.
Meyerhoff M, Adachi C, Nanbakhsh G, et al. (2011) Sociolinguistic fieldwork. In: Thieberger N (ed.) The Oxford Handbook of Linguistic Fieldwork. Oxford: Oxford University Press, 121–146.

Mondada L (2007) Commentary: transcript variations and the indexicality of transcribing practices. Discourse Studies 9(6): 809–821.

Norris S (2004) Analyzing Multimodal Interaction: A Methodological Framework. London: Routledge.

Nunan D (1992) Research Methods in Language Learning. Cambridge: Cambridge University Press.

Qian K, Owen N and Bax S (2018) Researching mobile-assisted Chinese-character learning strategies among adult distance learners. Innovation in Language Learning and Teaching 12(1): 56–71.

Reinders H and White C (2011) Special issue commentary: learner autonomy and new learning environments. Language Learning and Technology 15(3): 1–3. Available at: http://llt.msu.edu/issues/october2011/.

Reinhardt J (2019) Social media in second and foreign language teaching and learning: blogs, wikis, and social networking. Language Teaching 52(1): 1–39.

Salinger S, Plonka L and Prechelt L (2008) A coding scheme development methodology using grounded theory for qualitative analysis of pair programming. Human Technology 4(1): 9–25. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.142.7634&amp;rep=rep1&amp;type=pdf.

Schilling N (2013) Sociolinguistic Fieldwork. Cambridge: Cambridge University Press.

Smith JA, Flowers P and Larkin M (2009) Interpretative Phenomenological Analysis: Theory, Method and Research. London: SAGE Publications.

Smith JA and Osborn M (2008) Interpretative phenomenological analysis. In: Smith JA (ed.) Qualitative Psychology: A Practical Guide to Research Method. 2nd edn. London: SAGE Publications, 53–80.

Sockett G (2013) Understanding the online informal learning of English as a complex dynamic system: an emic approach. ReCALL 25(1): 48–62.

Stenström A-B, Andersen G and Hasund IK (2002) Trends in Teenage Talk: Corpus Compilation, Analysis, and Findings. Amsterdam/Philadelphia: John Benjamins Publishing.

Stevenson MP and Liu M (2010) Learning a Language with Web 2.0: exploring the use of social networking features of foreign language learning websites. CALICO Journal 27(2): 233–259.

Swain M (2006) Languaging, agency and collaboration in advanced second language proficiency. In: Byrnes H (ed.) Advanced Language Learning: The Contribution of Halliday and Vygotsky. London: Continuum, 95–108.

Tang JC, Liu SB, Muller M, et al. (2006) Unobtrusive but invasive: using screen recording to collect field data on computer-mediated interaction. In: Computer Supported Cooperative Work, 2006, Alberta, Canada, 479–482. Available at: https://jameslin.name/research/recent_shortcuts/screen_recording-cscw2006.pdf.

White C and Zheng Y (2018) Editorial. Innovation in Language Learning and Teaching 12(1): 1–5.

Whiteman N (2012) Undoing Ethics: Rethinking Practice in Online Research. New York; London: Springer.

Yinger RJ (1986) Examining thought in action: a theoretical and methodological critique of research on interactive teaching. Teaching and Teacher Education 2(3): 263–282.

Author biography

Wing Yee Jenifer Ho is assistant professor at the Department of English, City University of Hong Kong. Her research interests lie in the broad area of mobile and self-directed language learning. She is interested in the use of multimodal methods to collect online learning data and to understand the complexities of learning in virtual spaces.
Appendix: A guide on how to use Camtasia for screen-recording, which was sent to participants

This guide is made using Camtasia 2 for Mac, but it works in a similar way in Camtasia Studio for Windows

1. Click the red recording button to start the recording menu.
2. In the recent region drop-down menu, you can select to record a customised part of the screen. Make sure the three green buttons are lit up and your face is clearly visible on the screen. Press the red ‘REC’ button to start recording.
3. Press ‘Stop Recording’ on the top right hand corner of your screen when you want to stop recording.