Think-aloud protocols

Lawrence Jun Zhang and Donglan Zhang

Introduction

The use of verbal report data involving the think-aloud or thinking-aloud method has aroused substantial controversies that have led to many discussions not only in the fields of psychology, educational psychology, and cognitive science (Ericsson, 2002; Ericsson & Simon, 1980, 1993; Jack & Roepstorff, 2002; Pressley & Afflerbach, 2012), but also in the field of second or foreign language studies, or applied linguistics broadly defined (Bowles, 2010; Cohen, 2013; Deschambault, 2017; Leow & Morgan-Short, 2004; Matsumoto, 1994). The existing literature shows that the use of the think-aloud method for gathering information about human information processing is the work of scholars in the field of psychology and cognitive science (e.g., Ericsson, 2003; Ericsson & Simon, 1980, 1993; Jack & Roepstorff, 2002). The centrality of the issue surrounding the use of this method is about the validity, reactivity, and veridicality (assertion of the truth) of the data collected through this method. A number of reviews of this method are documented in the field of second language studies (e.g., Bowles, 2010; Deschambault, 2018; Goo, 2010; Hu & Gao, 2017; Leow & Morgan-Short, 2004; Matsumoto, 1993). In such scholarly debates, the focus has always been on the issue of whether these data reflect the psychological reality of the participants in the experiments.

Understandably, consensus is not always easy to reach among scholars regarding whether such data speak truthfully of the cognitive processes of the participants (e.g., Barkaoui, 2011; Yang & Zhang, 2015; Yang, Zhang, & Parr, 2019). We posit that, despite some concerns (e.g., Deschambault, 2018; Yang & Zhang, 2015), it is not so much about the validity, reactivity, and veridicality of this research method that are significant to second language researchers (e.g., Cohen, 1996, 2013); rather, it is about how researchers make an informed decision to embrace it wisely for getting the useful data to gain insights into the learning process that really matters (see e.g., Cowan, 2017; see Fox, Ericsson, & Best, 2011, for a meta-analysis). In other words, due diligence needs to be taken in interpreting the data, and this needs to be done with caution. It will soon become clear in this chapter that all the research reported in the literature, and the related discussions on the utility of the think-aloud method for pedagogical purposes and the effects of employing the think aloud method in educational intervention – especially the think-aloud protocols collected through this method as evidence of the cognitive activity and
processes of the human mind – point to the tremendous interest in and the value of this research tool, even to the present day (e.g., Cowan, 2017; Gu, 2014; Pressley & Afflerbach, 2012).

**What is a think-aloud protocol?**

A think-aloud protocol is a recording or record of what is revealed in the participants’ voluntary reporting of what they are going through while they are engaged with a particular real problem-solving or learning task (Cowan, 2017; Ericsson & Simon, 1993; see also Fox et al., 2011). Thinking-aloud is widely accepted as a research method for collecting data through participants’ verbalisation of “what is going through their minds as they are solving a problem or performing a task” (Gass & Mackey, 2000, p. 13). The data collected in such a way, particularly those from using concurrent thinking-aloud, are usually regarded as truthful representations of participants’ mental activity or processes (Ericsson & Fox, 2011; Ericsson & Simon, 1993; Fox et al., 2011), namely, what is happening “at the time the language learning or use events are taking place” when such a procedure is used for collecting information about how language learners go through the learning process (Cohen, 1996, p. 7). They are therefore treated by the majority of L2 researchers as evidence of direct access to learners’ short-term mental states, as represented in their working memory, or what language learners are doing at the moment when such data are collected (Bowles, 2010; Cohen, 2013; Gu, 2014; Kasper, 1998; Polio & Friedman, 2017). There are two kinds of thinking-aloud, as explained next.

**Concurrent think-aloud vs retrospective think-aloud**

Ericsson and Simon (1993) clearly distinguish concurrent thinking-aloud from retrospective verbalisations. Concurrent thinking-aloud is what the informant or the participant is reporting, at the time when something is happening, while they are engaged in an activity or a learning task. Retrospective verbalisation is what the informant or the participant is asked to think aloud about an activity or a task that has been just completed. The crucial difference between the two verbal report forms of thinking-aloud is the time-lag that is present in the latter, which is often regarded as having influence on the accuracy and completeness of the data.

In their recent meta-analysis and recommendations for best reporting methods regarding the procedures for verbal reporting of human thinking processes, however, Fox et al. (2011) conclude that in thinking-aloud “concurrent verbalisation does not influence the accuracy of performance and, by implication, does not alter the cognitive processes mediating task performance” (p. 335). To many researchers who have confidence in, and have been using, this method, such a statement is, in effect, reassurance that collecting think-aloud protocols as a research method can justifiably offer useful insights into the cognitive process of the human mind. Nonetheless, given that the scholarship on second and foreign language studies surrounding the issues of veridicality and reactivity has still been very active (see e.g., Bowles, 2010; Deschambault, 2018; Polio & Friedman, 2017; Yang & Zhang, 2015), we think it is necessary that some essential information on the debate be provided so that readers who are interested in using this method for data collection are well-informed and fully aware of the debates, should they want to use it at all.

**Approaches to understanding think-aloud**

Various approaches to understanding the think-aloud method and its related data are well-documented in the literature. Some scholars state that thinking-aloud is so far “the best available
measure of conscious, easily verbalisable thoughts” despite having some limitations, as other data collection methods (Wilson, 1994, p. 249; see also Payne, 1994). Others question its validity or, at best, caution researchers on the potential challenges of using this method for collecting useful data (Bowles, 2010; Yang, Hu, & Zhang, 2014). Deschambault (2018) broadly categorises the think-aloud data collection methods into three approaches: cognitive, sociocognitive, and sociocultural and discursive. Given that the real purpose of using thinking-aloud is to uncover the cognitive processes of human information processing, we confine our discussion of this research method to this cognitive perspective in this chapter. Readers interested in a more elaborate and critical appraisal of the think-aloud method from sociocognitive, and sociocultural and discursive perspectives are strongly recommended that they refer to scholarly debate by Sasaki (2008), Potter (2010), Pressley and Afferbach (2012), and Deschambault (2018).

**Think-aloud primarily as cognitive activity**

We concur with psychological views on thinking-aloud that, as a research tool, thinking-aloud involves bringing thoughts into consciousness, coding the thoughts verbally if needed, and then verbalising them (Ericsson & Simon, 1993). Such a view is most dominant in the work of Ericsson and Simon (1980, 1993), who distinguish three types of verbalisation related to the use of thinking-aloud (also see Bowles, 2010; Yang et al., 2014, for elaborations). In Type 1 verbalisation, participants’ think-aloud protocols do not require verbal recoding, as their thinking processes are readily discernible in the think-aloud protocols. Type 2 verbalisation, however, requires researchers’ to recode these thoughts, which are not in any explicit verbal code. In Type 3 verbalisation, participants are asked to offer specific information that the researchers want, usually in the form of some kind of reasons or explanations. Such verbalisation is often retrospective (Yang et al., 2014; Yang & Zhang, 2015; Zhang, 2016a), and reactivity might be likely, which presents itself as an issue. Ericsson and Simon predicted and found, consistently, empirical evidence that Type 1 and Type 2 verbalisations are fairly accurate representations of the nature of the cognitive processes of the human mind, despite such reporting causing participants to spend slightly more time on the task, and that Type 3 verbalisation, which is metacognitive in nature (Bowles, 2010; Yang et al., 2014), would change the order in which the heeded information is generated in the think-aloud protocols. They also found that Type 3 verbalisation increased the time participants spent on completing or performing the task.

Researchers’ prompts for participants to explain, justify, or describe in detail what is on their minds, or in their thought processes, all possibly contribute to delays in information processing. This is much related to the ideas discussed in relation to L2 studies in the works of Faerch and Kasper (1987), Kasper (1998), Matsumoto (1994), Goh (1998), Leow and Morgan-Short (2004), and especially the works by Sasaki (2008), and Deschambault (2012, 2018), where participants’ cultural knowledge and patterned thinking are, to a great extent, consciously brought to the fore. This phenomenon is also discussed and reviewed in Smagorinsky (1998, 2001). Consequently, reactivity would become an issue, and if it is really serious, the validity of such data is put at risk. This is a point we discuss in some detail in the next section.

**Concerns over using thinking-aloud for data collection**

There appears to be consensus among researchers in the fields of psychology, cognitive science, and education, including second or foreign language studies, or applied linguistics broadly defined, that concurrent thinking-aloud (verbalisation Type 1, as mentioned earlier) is a major source of useful and reliable data (Bowles, 2010; Cohen, 2013; Cowan, 2017; Fox et al., 2011;
Zhang, Gu, & Hu, 2008). In researching participants’ cognitive processes in completing tasks in first- and second-language contexts, researchers collect think-aloud protocols from informants, code them, and then analyse them to find out otherwise unobservable behaviours or processes that relate to their informants’ performance in doing or completing the tasks (e.g., Gu, Hu, & Zhang, 2005; Polio & Friedman, 2017). Like what has been discussed in psychology and cognitive science, as previously explained, in second language studies, if participants verbalise their thought processes without any justification or explanation, such verbalisation is treated as non-metalinguistic think-aloud (NMTA) (Type 1 and Type 2) and if verbalisations with explanations and justifications are part and parcel of the process, then they are regarded as metalinguistic think-aloud (MTA) (Type 3) (Bowles, 2010; Bowles & Leow, 2005). Bowles (2010) also refers to them, respectively, as non-metacognitive and metacognitive verbal reports. Yang et al. (2014, p. 60) further make a distinction between succeeding metacognitive think-aloud (SMTA) and preceding metacognitive think-aloud (PMTA). The former refers to reason-reporting initiated after a writing move and targeted at what has been written, and the latter refers to reason-reporting initiated before a writing move and directed at what is to be written.

In the field of applied linguistics, many studies using think-aloud as the main data collection method have been conducted to understand the processes of how language learners perform various tasks in language learning and language testing situations. Research along this line has examined the cognitive processes of second or foreign language learners’ learning of reading (Gu et al., 2005; L. Zhang, 2016; Zhang et al., 2008), listening (Goh, 1998), lexis (Deschambault, 2012; Hu & Nassaji, 2014), speech production (Duff & Kormos, 1998), interlanguage knowledge acquisition (Færch & Kasper, 1987), the use of learning strategies (Cohen, 1996; Hu & Gao, 2017; Rao, Gu, Zhang, & Hu, 2007); problem-solving and strategy-deployment in doing translation (Krings, 1986), assessment of students’ language skills (Raimes, 1985; J. Zhang, 2016), and strategies for feedback provision in writing (Cohen, 1991; Raimes, 1985). Relatively recently, scholarly interest appears to be expanding. For example, second language writers’ writing processes have been explored in relation to how research methodological choices can be fully leveraged for achieving the research objectives (Hyland, 2016; Polio & Friedman, 2017; Yang et al., 2014). However, in the existing discussions on using think-aloud as a method for data collection, particularly in the fields of second language studies or applied linguistics, its validity is scrutinised, especially in relation to the issue of reactivity and veridicality, as shown next.

**Reactivity**

The centrality of the question is about whether reporting the thinking process at the same time when the informant is required to perform the task could impose additional cognitive load so much so that the informant is not reporting what he is actually doing. If yes, then the additional requirement could alter the very thinking process it is supposed to represent. Despite the fact that large numbers of studies in cognitive psychology have been conducted that use the think-aloud method successfully and that the majority of scholars show trust in concurrent think-aloud protocols as mental representations of the psychological reality when the participants report the process in their first language (Fox et al., 2011), second language researchers seem to be more intensely interested in reactivity than psychologists in recent years (for reviews, see Bowles, 2010; Deschambault, 2018; Polio & Friedman, 2016). Bowles’ (2010) meta-analysis of 14 studies into reactivity shows that research on reactivity in the field of second language studies mainly used reading tasks; Sachs and Polio’s (2007) study examined how teacher reformulations in revision could induce reactivity effects. Given the challenging nature of writing in a language other than one’s first language, Yang et al.
Lawrence Jun Zhang and Donglan Zhang

(2014) investigated if there were reactivity effects in students writing in English as a foreign language (EFL); if yes, how seriously the reactivity effects presented themselves in affecting the think-aloud quality and the veridicality of the data. Interestingly, they indeed found reactivity effects of thinking-aloud while writing, but the significant reactive effects seem to touch upon rather minor and easy-to-be-unattended aspects of writing performance such as fluency and syntactic variety, but are unable to impinge on those core formal concerns in measuring L2 writing, i.e., syntactic complexity, lexical variety and accuracy, and above all, on the overall quality of writing. In other words, the degree of severity of reactivity effects is rather minor.

Some researchers have also investigated if participants who were thinking-aloud and those who completed the tasks silently would show performance differences in accuracy in reporting the cognitive process and latency as shown in participants’ increases in reaction time. The findings so far are varied and inconclusive. This suggests that reactive effects of thinking-aloud are moderated and modulated by variables such as the type of reporting (e.g., concurrent/non-metacognitive vs retrospective/metacognitive) (see Fox et al., 2011; Yang et al., 2014), language of reporting, learning tasks, and the age of participants. Such results are also reviewed in detail in Bowles (2010), who suggests, as other studies have shown, too, that reactive effects are more protruding while thinking aloud during reading and when the type of reporting is retrospective/metacognitive. Accordingly, Greene and Azevedo (2009) warn researchers of reactive effects of think-aloud on participants’ reading processes in self-regulated reading research, as participants are expected to engage in metacognitive processes (see Hu & Gao, 2017, for a fuller review).

**Veridicality**

Debates on the veridicality of think-aloud data in the field of psychology and cognitive science have been going on for a long time; only until recently did Fox et al.’s (2011) meta-analysis show clearly that it is not much of an issue. Nonetheless, in studies that involved second or foreign language learners as participants, the language of reporting became a confounding variable due to the fact that the participants were required to report on the cognitive processes of how they complete a second or foreign language learning task. It is noteworthy to mention that many second language researchers do allow their participants to report the cognitive process in the language(s) of their choice, either in their first language or the second/foreign language, or a combination of the two, allowing code-switching and co-mixing. Even so, some researchers have questioned the veridicality of the think-aloud data. In other words, it is not absolutely clear how accurately thinking-aloud reflects what is really happening in the human brain (Rogers, Revesz, & Rebuschat, 2016; Yang & Zhang, 2015). Potential inaccuracy and incompleteness of data from thinking-aloud are two major issues raised by scholars (Barkaoui, 2011; Wilson, 1994). Nonetheless, research on veridicality is far too infrequent when compared with studies into the reactivity effects of thinking aloud. The sparseness of research on the veridicality of think-aloud data probably has much to do with the trust that researchers have put in concurrent think-aloud, which is believed to be more truthfully reflecting the psychological reality, as compared with retrospective think-aloud (see Hu & Gao, 2017). This might be the reason why Fox et al.’s (2011) meta-analysis appears to provide evidence that there is no need to worry too much about reactivity and veridicality.

Those endorsing think-aloud are of the view that the verbalisation that occurs while the participant is performing the task is, in effect, a real reflection of the cognitive processes.
Think-aloud protocols

For example, Robinson (2001) argues that “[during] concurrent reports (CRs), or thinking aloud or talking aloud . . . processing and verbalisation occur at the same time and, therefore, no thought, feeling, or action would be omitted because the participant has not had time to forget” (p. 211). Bowles and Leow (2005, p. 417) also lend their support to this view, positing that although some evidence for nonveridicality has been found for retrospective protocols, which can ‘yield substantial forgetting or fabrication in all tasks’ . . . concurrent protocols, on the other hand, are not subject to the same critique, as they are collected during the task.

Of course, in addition to the issue of time, researchers’ use of clear instruction to participants for completing the think-aloud task is crucial to getting high-quality and valid data. Therefore, Hu and Gao (2017) remind us of the insufficient discussions of the veridicality of think-aloud in relation to research on self-regulated reading and suggest that more research be carried out in the future.

Effective use of think-aloud as a research method

Despite discussions on the drawbacks of using think-aloud as a method for data collection, the dominant view is that it is a trustworthy methodological choice for researchers who probe into learners’ cognitive processes (Bowles, 2010). The real challenge is probably not about the psychological limitations of this method, but rather the actual challenges in its administration, which is in most cases time-consuming and labour-intensive (Barkaoui, 2011). As discussed earlier in this chapter, there are also concerns about reactivity and veridicality of using think aloud. We suggest that as long as care is taken and researchers do not over-interpret the findings, think-aloud data can offer rich information about the cognitive processes of human think. Given the challenges mentioned in this chapter, we think it is important that training be provided to participants so that they know how to carry out concurrent thinking-aloud while completing a task. Concurrent think-aloud should be preferred to retrospective think-aloud, as the former can substantially minimise reactivity effects and allows the researcher to collect valid and veridical data.

Informant training and prompting

Many researchers have used this method and recommended that training be provided to informants/participants so that they become familiar with this method and are capable of performing the tasks that are to be assigned for them to complete. In fact, all the available research in the literature appears to have documented this procedure quite carefully, particularly in explaining the steps for providing training in explicit terms (see e.g., Gu et al., 20015; Rao et al., 2007; Zhang et al., 2008). There are reports on the benefit of providing informant training, as the validity of the think-aloud data is enhanced. This appears to be a justifiable recommendation, as not all informants know how to report (Afflerbach & Johnston, 1984; Bowles, 2010; Ericsson & Simon, 1993). As Barkaoui (2011) states, any lack of clarity in researchers’ instruction on how to complete thinking aloud tasks would lead to incomplete verbalisation. Therefore, it is necessary to give participants clear rationales and explicit purposes for doing the thinking-aloud and to tell them the ways in which they are expected to think aloud (Bowles, 2010). Pressley and Afflerbach (1995, 2012) stress the importance of asking the participants to report as accurately and completely as possible what they are reading. Some useful prompting expressions can be used in the data collection process, whether prior to data collection or while data collection is in progress. These expressions include: “tell me whatever you
are thinking about while reading” (e.g., Bråten & Strømsø, 2003) and “tell me whatever comes to your mind while you are doing the task” (Gu et al., 2005; Zhang et al., 2008). In their review of 23 studies using thinking-aloud protocols, Hu and Gao (2017) point out that at least two of them (i.e., Kendeou & van den Broek, 2007; de Milliano, van Gelderen, & Sleegers, 2016) did not appear to be consciously aware of the reactivity effects because their participants were asked to explain why they did the reading the way they did it, instead of using the aforementioned words or expressions as prompts.

Teachers or researchers with experience in thinking aloud can show, orchestrate, or illustrate to their informants how to think aloud using concrete things as objects for the description of the mental processes. This kind of technique has been reported as a successful way of collecting data, even from children as young as 10 years old, who were shown by the researchers how to think aloud while the researchers were playing with toys to practise thinking aloud (see Gu et al., 2005; Zhang et al., 2008). In de Milliano et al.’s (2016) study video tapes or demo clips were used to indicate the process of thinking aloud while reading. Researchers’ demonstration and informants’ practice must go hand in hand in order for the informants to know how to think aloud so that researchers can collect accurate, complete, and reliable data. Informants should be provided with ample opportunities to ask any questions regarding the thinking-aloud technique as well as other related matters. In Gu et al.’s (2005) case, the tasks for practising thinking-aloud were deliberated among the research team members before a simple task was chosen (see also Rao et al., 2007; Zhang et al., 2008). Bowles (2010) summarises that a practice task “should be carefully chosen so that it does not prime the participants for the target structure being investigated in the study” (p. 117). She also advises that think-aloud tasks for demonstration and practice purposes should not be the same as those to be used in the real study, and that the same illustrative examples should be used to minimise any potential bias (see e.g., Denton et al., 2015). These are practical and useful strategies for collecting reliable and valuable think-aloud data.

From our review and discussion of issues involved in the use of the think-aloud method, we can see that there is some consensus that as long as care is taken, thinking-aloud can be a useful method for collecting data on human cognitive processes. Researchers need to be meticulous in the very process of data collection, however. For the think-aloud method to be successful, reminding informants – either through verbal prompts orally presented to them by researchers or by means of any visible markings inserted in the reading text or indicated on the writing pad – to report their thoughts, is a highly recommended strategy (Bowles, 2010). It needs to be pointed out that prompting questions posed to the informants could also be a source of interference with their cognitive processing of the tasks to be completed (Pressley & Afflerbach, 1995). Any researcher intending to use the think-aloud method should be aware of the advantages and the potential challenges of using prompting questions. To avoid putting the researchers’ words into the participants’ mouths, the very neutral questions should be preferably something like these: “what are you thinking right now?” “please say what you are thinking,” “don’t forget to read out loud,” “say what you are looking for now,” and “why are you clicking there,” among others (see Hu & Gao, 2017). Placing colourful stickers in the texts (especially after each chunk when the researcher thinks there is a need to think aloud at that point) as a prompting technique was found to be a good way to reduce the interference of verbal prompting (see Færch & Kasper, 1987; see recent uses of this modified technique as reported in Scott, 2008).

**Future directions in using the think-aloud method**

Thinking-aloud is a well-established research method and has been widely and successfully used in numerous studies in the fields of psychology, cognitive science, education, and applied linguistics.
Yet, there are still scholars who are concerned about its use, especially in relation to the reactivity effects and the veridicality of the data collected through this method. The debates have not prevented researchers from embracing this method for researching second or foreign language teaching and learning (Cohen, 2013). Controversies are detailed in a number of studies both in the fields of psychology and L2 studies, but Fox et al.’s (2011) recent meta-analysis of 94 studies that compared performance between groups doing concurrent verbalisations and groups without verbalisation is illustrative. Based on nearly 3,500 participants, their analysis shows that the “thinking-aloud” effect size is indistinguishable from zero ($r = -0.03$) and that this procedure remains non-reactive even after statistically controlling additional factors such as task type (primarily visual or nonvisual). Indeed, as their meta-analysis shows, the effects of verbalisation on performance are determined in part by verbal reporting procedures, as we have discussed in this chapter. These findings lend support to Ericsson and Simon’s (1980, 1993) theory that think-aloud verbalisation results in little or no reliable difference in performance across verbal report and silent conditions.

In response to what other researchers have raised as a concern, Fox et al. (2011) conclude that the type of instructions used, type of reminders employed, interactions between informants and researchers, and differences between the verbalising condition and the silent control condition are possible contributors to reactivity. In other words, “only by developing methods of data collection with precision and reliability is it possible to attain reproducible data on thinking processes” (p. 339). As a further reinforcement of their meta-analysis results and assurance of the utility of the think-aloud method, Ericsson and Fox’s (2011) reply to Scholler’s (2011) query clearly and convincingly shows that “thinking aloud is not a form of introspection but a qualitatively different methodology” (p. 351). Collecting data through think-aloud gives researchers qualitatively rich information about participants’ thinking or cognitive processes. Nonetheless, like any other research methods for data collection, think-aloud as a data-collection tool has its shortcomings, too, as discussed in this chapter. For an in-depth synthesis and discussion of the issues involving the use of think-aloud data in second language research, see Deschambault (2018), which is a valuable resource. Fox et al. (2011) is another very comprehensive resource.

As a way of compensating for the evident drawbacks of using one single method, we would like to recommend leveraging methodological triangulation, which can be a very useful strategy for overcoming these drawbacks. Additionally, with the increasing use of modern technologies and equipment, thinking-aloud can be complemented with data collected through resort to modern techniques such as eye tracking (see Pellicer-Sánchez & Conklin, this volume), event-related episodes (ERP), and magnetic resonance imaging (MRI). Understandably, these technologies also have their own limitations. It is exactly because of their limited use that a combination of several methods in one single study should be regarded as an ideal solution to the methodological dilemmas facing researchers in social sciences – using multiple methods for collecting the data to corroborate the findings is a worthy endeavour.

Evidently, pedagogical uses of thinking-aloud still have an appeal to classroom teachers as well as classroom-based researchers for improving pedagogical practice, especially those who are keen to do action research. Such uses have been widely reported in various studies. For example, Matsumoto (1996) reported his use of thinking-aloud to help L2 learners to reflect on language learning in the classroom. Goh (1998) also reported on the value of using thinking-aloud for promoting student use of listening strategies for enhancing the learning of L2 listening.

To close, we recommend that readers consult a recent study by Cowan (2017), which provides a good summative review in addition to reporting research findings using thinking-aloud as a research method.
References


Think-aloud protocols


