Investigating L2 Speech Production through Verbal Reports: A Systematic Review of Recent Studies

Yin Sheun Chong[a]
yinsheunchong@gmail.com
Center for Research in Language and Linguistics,
Faculty of Social Sciences and Humanities,
Universiti Kebangsaan Malaysia, Malaysia

Bahiyah Abdul Hamid
bahiyahabdulhamid@gmail.com
Center for Research in Language and Linguistics,
Faculty of Social Sciences and Humanities,
Universiti Kebangsaan Malaysia, Malaysia

Normalis Amzah
normalis@ukm.edu.my
Center for Research in Language and Linguistics,
Faculty of Social Sciences and Humanities,
Universiti Kebangsaan Malaysia, Malaysia

Khazriyati Salehuddin[b]
khazudin@ukm.edu.my
Center for Research in Language and Linguistics,
Faculty of Social Sciences and Humanities,
Universiti Kebangsaan Malaysia, Malaysia

ABSTRACT

Introspective verbal reports have been incorporated into many studies on L2 reading and writing. In contrary, they are less utilised in studies on L2 listening and speaking. The objective of this review is to describe the use of verbal reports in recent studies on L2 speech production. The studies were identified through a comprehensive search across two (2) online databases: the Education Resources Information Center (ERIC) database and the Language and Speech (LAS) journal database. In particular, the search was restricted to studies published in the past ten (10) years (January 2012 to December 2021). A total of six (6) studies on different aspects of L2 speech production that met the predetermined selection criteria were identified and reviewed. All of the studies are set within the ESL/EFL context. The review resulted in two (2) major findings. Firstly, the verbal reports in recent studies on L2 speech production are all retrospective reports or stimulated recalls. Secondly, the verbal reports mostly played the role of triangulation to confirm and explain the data from other research instruments about the main construct being studied. Moreover, verbal reports revealed the psycholinguistic processes behind L2 learners' speech including processes associated with bilingual speech production. Based on the review's results, suggestions related to the use of verbal reports in future studies on L2 speech production were made.

Keywords: L2 speech production; verbal reports; psycholinguistic processes; bilingual speech production; ESL/EFL

INTRODUCTION

Introspective verbal reports were first used in philosophy and psychology before they gained popularity in language research. Within a research context, introspection refers to participants' own reflection of the mental processes behind their actions (Mackey & Gass, 2016). The fundamental idea behind introspection is that a person's mind is separate from a person's body and as such, the workings of a person's consciousness can also be observed. However, there are various procedures related to the use of introspection in research that affect the validity of the reports and the validity of the corresponding research findings. For example, introspection can occur simultaneously with a task (concurrent reporting), or it can be carried out retrospectively. Over the years, verbal reports have become an important part of first language (L1) and second language (L2) research. In L2 research especially, verbal reports are widely used in studies on reading and writing (Abas & Noor Hashima, 2016; Abdulkhaleq, 2020; Barkaoui, 2015; Hazita et al., 2017; Kim & Yoon, 2014; Noorizah, 2010; Nur Ainil et al., 2020; Nurul Adila et al., 2017; Zhao & Brown, 2014). In most of these studies, verbal reports were collected from L2 learners based on their reading or writing activities. The verbal reports then served to uncover the mental processes underlying the learners' language use. This has led to discoveries about L2 learners' knowledge types, knowledge structures, and their strategies when it comes to language use and language learning. Apart from studies on reading and writing, verbal reports have also been used in studies on L2 lexical organisation, L2 vocabulary acquisition, and studies on the role of awareness and attention in language learning (see Bowles, 2019, for a review).

On the other hand, verbal reports are less used in studies on L2 listening (Rossa, 2016; Rukthong & Brunfaut, 2019; Siricord & Melor, 2016). This is partly due to the fact that only certain types of verbal reports are suitable with listening tasks. For example, concurrent reporting is not suitable for listening tasks as it may be confusing for participants to listen and report simultaneously. Moreover, concurrent reporting may slow down the listening task and affect task performance (Gass & Mackey, 2017). Similarly, concurrent reporting is not suitable for tasks that involve speaking. Nevertheless, in an earlier review, Kormos (1998) noted that the use of verbal reports in studies on L2 speaking are able to provide important information about the process of L2 speech production. As a result, Kormos concluded her review with a proposal for the use of verbal reports to be expanded "far more extensively" (Kormos, 1998, p. 357) than what was available at the time. As more than 20 years have gone by following Kormos' proposal, it is timely for us to take a look at what has been achieved in this regard. To capture the progress that has been made, this paper systematically reviews some of the recent studies on L2 speech production that incorporated verbal reports as part of their methodology. Specifically, this review aims to describe the use of verbal reports in these recent studies on L2 speech production. Thus, this review is guided by the following review questions:

1. What are the main procedures surrounding the use of verbal reports in recent studies on L2 speech production?
2. What are the significant findings produced via verbal reports in recent studies on L2 speech production?

Accordingly, the review begins with a brief discussion of some of the basic concepts related to the use of verbal reports as a research instrument. Following that, the full methodology for the review is outlined. Next, selected recent studies on L2 speech production that incorporated verbal reports as part of their methodology are reviewed and the findings are presented. Finally, the limitations of the review and the implications from the review are delineated.

VERBAL REPORTS AS A RESEARCH INSTRUMENT

The use of verbal reports as a research instrument is primarily based on a seminal framework developed by Ericsson and Simon (1993). Drawing on information processing theories, Ericsson and Simon argued that verbal reports could reveal the information 'heeded' in one's short-term memory while he or she is performing a task. In other words, verbal reports can provide access to the mental processes behind a person's actions. However, there are different types of verbal reports or verbalisations. Level 1 verbalisation is the direct verbalisation of information available in a person's short-term memory while Level 2 verbalisation requires non-verbal information (such as images) to be encoded into the verbal form before it can be reported. Level 3 verbalisation is the verbalisation of information that is not available in a person's short-term memory (i.e., the information is stored in a person's long-term memory). Therefore, during reporting sessions, researchers should ask questions that can lead to Level 1 or Level 2 verbalisations. For example, instead of asking participants to reflect on a task which may induce them to refer to their general knowledge or long-term memory, researchers should ask questions that target specific events within the task.

In general, there are two (2) aspects of verbal reports that must be considered in the instrumentation process namely, reactivity and veridicality. Reactivity refers to when the 'primary' process, that is, the construct being studied, is altered due to the reporting process (Gass & Mackey, 2017). For example, in a study on teaching strategies, learners may be given a post-test on the target structures following a reporting session. It has been suggested that the reporting session itself could have strengthened learner perception of the target structures leading to a better performance during the post-test. This would result in conclusions being made about the strategies used that may not be entirely valid. On the other hand, veridicality refers to the "correspondence" (Gass & Mackey, 2017, p. 125) between verbal reports and the primary process. This is because people tend to offer explanations for their behaviour when questioned (Ormerod & Ball, 2017). Thus, verbal reports may not be a true representation of their mental activities. In addition to reactivity and veridicality, another important aspect of verbal reports that must be considered in the instrumentation process is the different types of knowledge possessed by people. Procedural knowledge is the tacit and intuitive knowledge of how to do things that people have whereas declarative knowledge is a person's factual knowledge of the world (Ormerod & Ball, 2017). For example, L1 knowledge is mostly procedural in nature but L2 users who consciously learned a language (either formally or informally) typically possess declarative knowledge. Verbal reports are said to be more suited for studying declarative knowledge as people usually have limited access to their procedural knowledge. This is why verbal reports are widely used in L2 research especially in studies on reading and writing.

To overcome these issues, various procedures have been developed concerning the use of verbal reports as a research instrument. As mentioned earlier, verbal reports can be concurrent to a task or carried out retrospectively. Concurrent reporting or 'think-aloud' ensures fewer issues with veridicality as the reporting process occurs simultaneously with the task. Interestingly,

concurrent reporting is also mostly non-reactive (Bowles, 2019). However, concurrent reporting does increase the overall length of time spent on a task. In contrary, retrospective reporting either takes place immediately after the task has been completed or it takes place after a certain amount of delay. To improve on the veridicality of reports, retrospective reporting is often supported by some form of stimulus to help participants with their reflection such as a video or audio recording of them performing the task. As a result, retrospective reporting is usually referred to as 'stimulated recall'. Retrospective reporting generally has no issues with reactivity unless there is further data collection following the reporting session including the aforementioned post-test. Nevertheless, retrospective reporting should ideally occur within 72 hours of the task (Gass & Mackey, 2017). In cases where an extended delay is unavoidable due to scheduling conflicts or technical issues, an additional stimulus such as a transcript of the video/audio recordings may be necessary. Apart from that, a specific form of retrospective reporting known as 'immediate recall' was introduced in a study on recasts and noticing by Egi (2004). Rather than carrying out the reporting process at the end of a task, immediate recall takes place right after a target utterance *during* the task. This entails that there are multiple points throughout the task whereby a learner's speech is interrupted to make way for recalls. In this particular study, the points were signalled using double knocking sounds performed by the researcher. After a recall, the task is resumed until it has been completed. Although the issue of veridicality has undoubtedly been mitigated in this way, the technique was criticised for the unnatural production experience and for being reactive on the learner's subsequent language use (Mackey & Gass, 2016). However, Egi (2004) found no significant difference in the post-test results of learners who carried out immediate recalls and those who were engaged in the more conventional stimulated recalls, suggesting that the reactivity effect is minimal. Finally, retrospective reporting can be prompted by the researcher or it can be self-initiated by participants (Mackey & Gass, 2016). In the event of researcher-initiated reporting, the prompts used must be prepared and trialled beforehand to ensure that they are able to elicit the required information from the participants. For example, Gass and Mackey suggested questions such as "Can you remember what you were thinking when…?" (Gass & Mackey, 2017, p. 55). As mentioned earlier, questions that target specific events within a task allow the information heeded in participants' short-term memory to be successfully elicited (Level 1 or Level 2 verbalisations).

In short, verbal reports allow for the mental processes behind a person's actions to be revealed, recorded, and analysed. However, in incorporating verbal reports into a study, the issues of reactivity and veridicality must be addressed through proper research design to ensure the validity of the reports. In general, the procedures surrounding the use of verbal reports must be planned out along three (3) 'dimensions': the timing of the reports, the availability, type and amount of support provided, and whether the reporting process is initiated by the researcher or the participants themselves. Each of these dimensions has its own effects on the reactivity and veridicality of the verbal reports. In addition, researchers must also consider how the different dimensions work together. For example, reports that are collected after a slight delay may have a lower level of veridicality but it can be subsequently strengthened with a strong stimulus (such as a video stimulus) and a combination of researcher and participant initiation of reporting. By first approaching the dimensions individually and then collectively, the requirements for validity can be balanced out with the more practical aspects of data collection, leading to the most effective use of verbal reports in a study.

METHODOLOGY OF REVIEW

This review aims to describe the use of verbal reports in recent studies on L2 speech production. The methodology of the review is adapted from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). Although the PRISMA statement was originally created for systematic reviews in the medical field, it is also widely referred to in social sciences (Chapman, 2021). This section outlines the full methodology of the review including the selection process and the data analysis process.

PHASE 1: IDENTIFICATION

In the first stage of the selection process, the key phrase for the review was identified as "L2 speech production" based on the objective of the review. Phrases similar to the key phrase were also identified. Following that, a full list of search terms was developed (Table 1). The search terms were keyed into two (2) databases namely, the Education Resources Information Center (ERIC) database and the Language and Speech (LAS) journal database. The ERIC database was chosen as it is one of the world's largest databases for educational research including language education while the LAS journal is a leading journal dedicated to research on human languages and speech that has been in publication since 1958. Both databases publish primarily in English.

TABLE 1. The search terms used

Database	Search Terms
ERIC/LAS	• "L2 speech production" • "L2 oral production" • "second language speech production" • "second language oral production"

The search was conducted in December 2021. The search was conducted by keying in the search terms (with the quotation marks) one at a time using the general search function in the two (2) databases. The search was restricted to articles published in the past ten (10) years (January 2012 to December 2021). The search resulted in a collection of 47 articles (42 articles from the ERIC database and five (5) articles from the LAS journal database).

PHASE 2: SCREENING

In the second stage of the selection process, all 47 articles were screened based on the following inclusion and exclusion criteria (Table 2).

TABLE 2. The inclusion and exclusion criteria

Criterion	Inclusion	Exclusion
Type of article	Journal (research article)	Journal (reviews), book, book chapter, conference proceeding
Research topic	Different aspects of L2 speech production	Other than L2 speech production
Research instruments	Must include both production task and verbal reports	Without production task and/or verbal reports

For the screening process, the titles and abstracts of the articles were carefully read. One (1) article was removed as it is a book review and not a research article. 12 articles that were not directly related to L2 speech production and 27 articles that did not incorporate production task and/or verbal reports as part of their studies were also removed. In total, 40 articles were removed in the screening process.

PHASE 3: ELIGIBILITY

In the third stage of the selection process, the seven (7) remaining articles were further assessed for their final eligibility. The full texts of the articles were carefully read to ensure that they meet all of the inclusion criteria. One (1) article was removed as the full text did not provide excerpts of the verbal reports collected during the study. Finally, six (6) articles were selected for the review. The flow diagram below (Figure 1) illustrates the complete selection process. The template of the diagram was adapted from the PRISMA 2020 statement (Page et al., 2021).
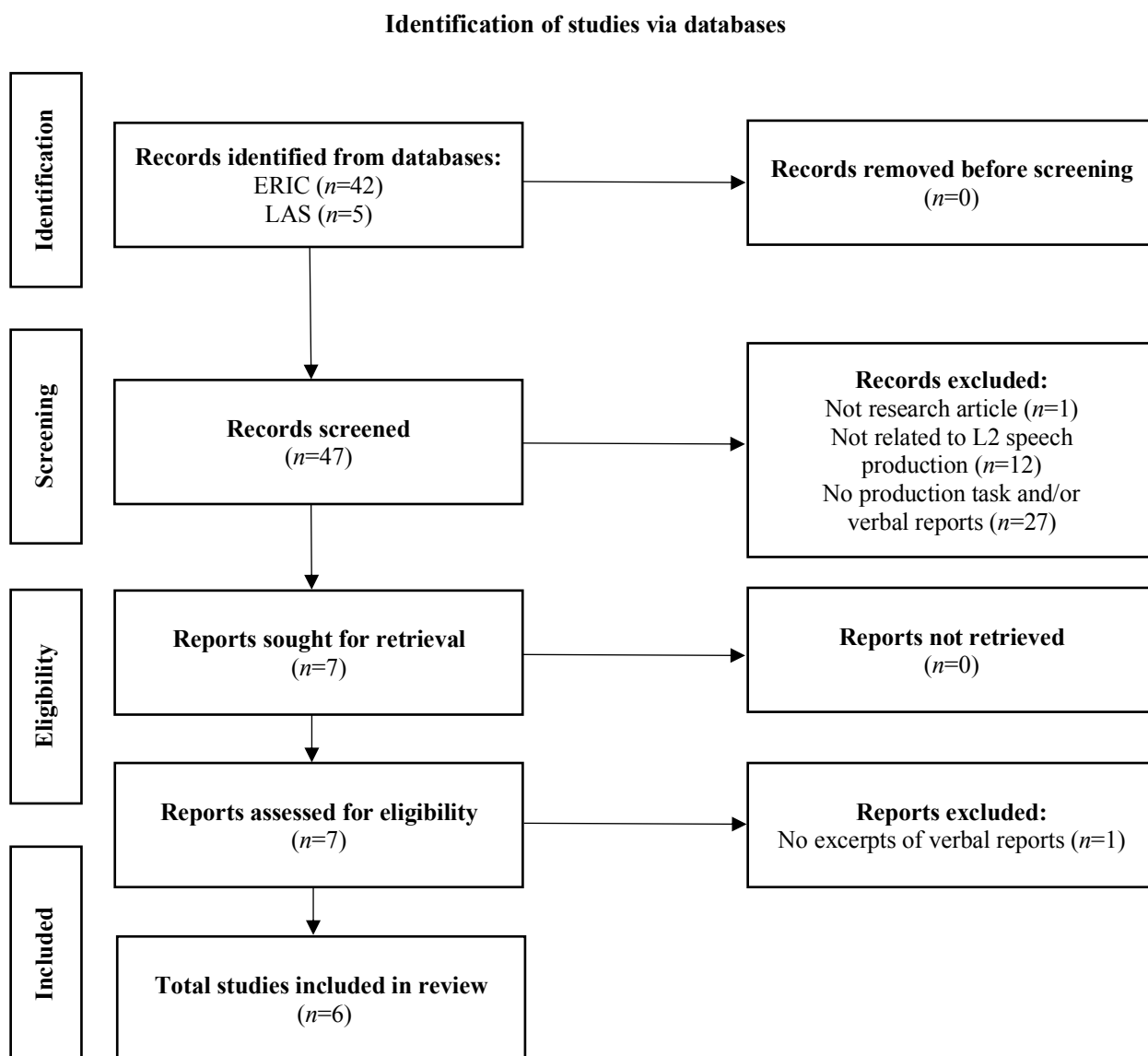
Identification of studies via databases

FIGURE 1. Flow diagram of the selection process (adapted from Page et al., 2021)

PHASE 4: DATA ANALYSIS

For the process of data analysis, several types of information were first extracted from the six (6) articles using a table that has been prepared specifically for this review (Table 3). Based on the objective of the review, the information extracted includes basic information about the studies such as the author(s), year of publication, the objectives, the sample, and the research instruments used. Information related to the verbal reports in the studies was also extracted. The studies were then reviewed based on the available information to answer the two (2) review questions. The findings from the review are presented in the following section.

TABLE 3. Summary of L2 speech production studies reviewed

Author(s)/Year of Publication	Objectives of the Study	Sample	Research Instruments	Verbal Reports
Azizullah & Najmeh (2012)	To probe into the constraints underpinning fluent and non-fluent speakers' use of PSMs (problem-solving mechanisms) in L2 oral communication	60 fluent and 60 non-fluent L2 English speakers from several Iranian universities	• 2 tasks in monologic and dialogic conditions • questionnaire • retrospective interviews (20 participants randomly selected from each group of students)	• stimulus in the form of audio recordings
Mohammad Javad et al. (2012)	To shed more light on how the degree of task difficulty, operationalised as the existence of a loose or tight storyline structure, affects the number and types of self-repair behaviour that EFL learners effectuate in their L2 oral speech	30 17- to 21-year-old lower-intermediate EFL learners from a private language centre in Iran	• 2 narrative tasks with different video prompts (1 structured and 1 unstructured) • stimulated recalls	• performed in the participants' L1 • 3-4-minute interval after the main task • prompting based on disfluency markers • participant-initiated reporting allowed • stimulus in the form of audio recordings
Kahng (2014)	To investigate how L1 and L2 speakers' fluency differs in terms of utterance and cognitive fluency	31 Korean learners of English and 15 English native speakers from a university in the United States	• spontaneous speech task • questionnaire on L2 learning background (all Korean learners) • stimulated recalls (9 low proficiency and 8 high proficiency Korean learners) *The English native speakers only took part in the spontaneous speech task and their performance was compared with the performance of the Korean learners.	• performed in the participants' L1 • participant-initiated reporting allowed • stimulus in the form of audio recordings

Fukuta (2016)	To examine the effects of task repetition on learners' attention orientation by comparing data of attention shifts during repeated task engagement	28 Japanese undergraduate and postgraduate students with upper-intermediate proficiency in English. The participants were randomly assigned into an experimental group and a comparison group.	• 2 narrative tasks • retrospective interviews (experimental group) *In the first week, both groups of students performed the same narrative task. 1 week later, the experimental group was given the same task as they had performed in the first trial whereas the comparison group was given a task with the same structure but different content.	• performed in the participants' L1 • prompting based on disfluency markers • stimulus in the form of audio recordings
Qiu (2019)	To investigate the influence of two (2) types of tasks on lower- and higher-proficient EFL learners' speaking performance from the perspectives of complexity, accuracy, and fluency (CAF)	60 first year undergraduate students from a university in China	• English proficiency test • 2 narrative tasks • stimulated recall interviews (9 low proficiency and 8 high proficiency learners)	• performed in the participants' L1 • prompting based on disfluency markers • questions asked by the researcher include "What were you thinking when you paused?" and "Why did you think so?" • participant-initiated reporting allowed • stimulus in the form of video recordings
Páez (2020)	To investigate the impact of the pushed output hypothesis and its three (3) functions on EFL students' L2 oral production	16 7th grade (11 to 12 years old) intermediate English learners in a private school in Colombia. The researcher randomly assigned the participants to output and non-output groups.	• pre- and post-test • 'pushed output' activities • stimulated recalls (output group) • interviews *For 5 weeks, the researcher asked the output group to perform different activities that 'pushed' them to produce oral L2 output whereas the non-output group was only exposed to comprehension activities.	• performed in the participants' L1 • stimulus in the form of audio recordings

FINDINGS AND DISCUSSION

GENERAL FINDINGS

Firstly, the review found that all of the recent studies on L2 speech production that incorporated verbal reports as part of their methodology were based on the English learning context: either English as a Second Language (ESL) or English as a Foreign Language (EFL). Nevertheless, the studies focused on different aspects of L2 speech production within the ESL/EFL context including problem-solving mechanisms (Azizullah & Najmeh, 2012), the effects of task difficulty (Mohammad Javad et al., 2012), the relationship between utterance and cognitive fluency (Kahng, 2014), the effects of task repetition (Fukuta, 2015), the effects of task types (Qiu, 2019), and the 'pushed output' hypothesis (Páez, 2020). It can be seen that the studies covered a wide range of topics and were not limited to any particular aspect of L2 speech production. The L2 learners in the studies consisted of Korean, Japanese, Chinese, Spanish, and Persian speakers. Interestingly, all of the participants were university students except for Páez (2020), whose study was conducted with secondary school students. This suggests that studies on L2 speech production that incorporated verbal reports tend to be conducted with older participants. However, none of the studies had a substantial sample. Only Azizullah and Najmeh (2012) had more than 100 participants in their study. For most of the studies, verbal reports were collected from a subset of the main sample. For Azizullah and Najmeh (2012), one-third of their participants was randomly selected to take part in the reporting process. In addition to verbal reports, most of the studies also had a minimum of two (2) other research instruments. Verbal reports were utilised together with instruments such as proficiency tests, pre- and post-tests, questionnaires, and interviews. All of the studies also had at least one (1) oral production task in the form of picture description, pair discussion, narrative tasks based on video prompts, or storytelling. In Mohammad Javad et al. (2012) and Fukuta (2016), verbal reports served as the main research instrument apart from the oral production task. The main findings from the review are presented below according to the two (2) review questions.

MAIN FINDINGS

WHAT ARE THE MAIN PROCEDURES SURROUNDING THE USE OF VERBAL REPORTS IN RECENT STUDIES ON L2 SPEECH PRODUCTION?

Although all of the recent studies on L2 speech production in this review incorporated verbal reports as part of their methodology, the verbal reports differ slightly in terms of their procedures. As mentioned earlier, there are three (3) separate but connected dimensions in the procedures for verbal reports.

Timing of the Reports

All of the studies in this review utilised retrospective reports or stimulated recalls as concurrent reporting is not possible with tasks involving speaking. As such, verbal reports were also occasionally referred to as 'retrospective interviews' (Azizullah & Najmeh, 2012; Fukuta, 2016). However, there was a lack of information as to when the reporting process was carried out except for Mohammad Javad et al. who noted that it took place after "a 3–4 minute time interval" (Mohammad Javad et al., 2012, p. 320). According to the researchers, the time was used to instruct their participants for the reporting session. In other words, the reporting process essentially took

place immediately after the production task. Similar procedures have been observed in previous studies on L2 reading and writing that incorporated retrospective interviews (Kim & Yoon, 2014; Nur Ainil et al., 2020). According to Gass and Mackey (2017), prompt reporting ensures the veridicality of the reports.

Availability and Type of Support Provided

The stimulus provided during the reporting process in most of the studies was in the form of audio recordings. Only one (1) study had video recordings as the stimulus (Qiu, 2019). This is in contrary to studies on L2 reading, writing, and listening whereby most stimulated recalls were accompanied by videos (Barkaoui, 2015; Hazita et al., 2017; Rukthong & Brunfaut, 2019). Videos are generally viewed as being more effective when it comes to eliciting recalls (Gass & Mackey, 2017). Nevertheless, this trend in the current review may simply be the result of practicality. Since video recording and playback often require extra equipment and preparation time, researchers could have made the decision to rely on audio recordings to ease the process of data collection.

Initiation of Reporting

All of the studies utilised researcher-initiated reporting. Disfluency markers in learners' speech such as pauses and false starts served as a guide for the researchers to prompt their participants during the reporting process. In Mohammad Javad et al. (2012), Kahng, (2014), and Qiu (2019), the participants were also encouraged to speak up whenever they have a comment that they would like to make. However, these studies involved high proficiency L2 learners who may be more confident in commenting on their L2 speech production compared to the low proficiency learners. Similarly, not much information was available as to the specific prompts used by the researchers, except for Qiu who asked questions such as "What were you thinking when you paused?" (Qiu, 2019, p. 10). According to Gass and Mackey (2017), this manner of questioning allows participants to clearly recall their thoughts during an activity. It should also be noted that most of the verbal reports were in the learners' L1. Thus, most of the excerpts found in the articles (and quoted here) are translated versions of the learners' original comments. Previous studies have revealed that reporting in the L1 prevents participants from not being able to respond due to proficiency issues (Abdulkhaleq, 2020; Nurul Adila et al., 2017; Rossa, 2016; Rukthong & Brunfaut, 2019).

WHAT ARE THE SIGNIFICANT FINDINGS PRODUCED VIA VERBAL REPORTS IN RECENT STUDIES ON L2 SPEECH PRODUCTION?

The studies included in this review focused on different aspects of L2 speech production such as problem-solving mechanisms, the effects of task difficulty, the relationship between utterance and cognitive fluency, the effects of task repetition, the effects of task types, and the 'pushed output' hypothesis. The studies can generally be divided into two (2) categories namely, studies on the effects of different manifestations of learning tasks including task difficulty and task repetition aimed at improving the teaching of L2 speaking skills (Fukuta, 2016; Mohammad Javad et al., 2012; Páez, 2020; Qiu, 2019) and studies on more general aspects of L2 speech production such as problem-solving mechanisms (Azizullah & Najmeh, 2012; Kahng, 2014). The findings produced via the verbal reports in the studies depended on the main research focus of the studies. Accordingly, there are three (3) main themes in the following discussion including an additional theme related to findings on the psycholinguistic processes behind L2 speech production.

Findings on Learners' Performance across Different Types of Tasks

Studies on learning tasks often compare learners' performance across different types of tasks. Verbal reports were then used to explain the differences (or similarities) in learners' performance. Two (2) groups of learners were compared in the study on task repetition by Fukuta (2016). The first group performed the exact same task twice whereas the second group performed two (2) tasks with the same structure but different content. The first group of learners showed improvements in both their level of accuracy and the variety in their vocabulary when they repeated the same task for a second time. Learners' verbal reports eventually revealed that they were able to pay more attention to their language use during the second attempt as they were already familiar with the content of the task, leading to a better performance. Based on these findings, the study proposed task repetition as a viable approach in teaching specific sub-skills in L2 speaking. Similar procedures were reported in Barkaoui (2015). Verbal reports were used to investigate L2 learners' writing process in an 'independent' task (an essay task with a writing prompt) and an 'integrated' task (an essay task with a textual stimulus). This suggests that the use of verbal reports to find out more about learners' performance across different types of tasks is highly relevant in studies of L2 productive skills.

Findings on Different Aspects of L2 Speech Production

On the contrary, studies on more general aspects of L2 speech production typically do not involve comparison of learners' performance. As such, verbal reports were used to explain the range of learners' L2 use in relation to the main construct. In Azizullah and Najmeh (2012), disfluency markers in the production data were taken as a sign of production difficulties and possible problem-solving activities.

- **Production data**:   People should use both … … should use only good programs.
  Verbal report: At first, I mistakenly wanted to say people should use both good and bad programs but then replaced it with good programs.

In this example, the extended pause in the production data (represented by a double ellipsis in the transcription) was interpreted as the learner having trouble with his or her speech. The learner's subsequent production was then linked to the problem-solving strategy of replacement. The learner's verbal report later confirmed this analysis. Moreover, the learner's verbal report revealed that the source of trouble lies in the message that he or she wanted to convey. This piece of information was not available in the learner's production data. As a result, the researchers were able to establish the connection between disfluency markers in the learner's speech with the underlying cause and the final problem-solving strategy, leading to a better understanding of the problem-solving mechanisms among L2 learners. This particular effect of verbal reports has been noted previously in a study on reading strategies among L2 learners (Abdulkhaleq, 2020). Interviews successfully elicited the reading strategies commonly used by the participants. However, verbal reports during the reading task revealed that high, medium, and low proficiency students actually adapt the same reading strategies in different ways according to their own needs. Thus, it can be concluded that verbal reports are able to produce additional findings related to the main construct of a study that are unavailable elsewhere.

Findings on the Psycholinguistic Processes behind L2 Speech Production

A number of the studies reviewed also made references to the psycholinguistic processes behind speech production based on L2 learners' verbal reports. According to a seminal psycholinguistic model developed by Levelt (1989), speech production consists of three (3) main stages: conceptualising, formulating, and articulating. The conceptualiser first creates a 'preverbal message' through macroplanning and microplanning. Next, the preverbal message is forwarded to the formulator for lexical choice. At this point, all of the relevant lemmas (i.e., lexical items that are related to a particular concept stored in a person's mental lexicon (Tokowicz & Degani, 2013)) are activated. From this collection of lemmas, the lemma which is the most appropriate for the message at hand is selected. This in turn, activates the lemma's syntactic and morpho-phonological information. Syntactic and phonological encoding result in a 'surface structure' which is then verbalised through the articulator. Finally, speech production also involves the process of monitoring whereby a speaker checks on his or her own pre- and post-production utterances for mistakes.

A number of the studies reviewed made references to these psycholinguistic processes based on L2 learners' verbal reports. For example, in Fukuta (2016), the following report was attributed to the process of lexical choice.

- **Production data**:   At the airport, err… a man, err is waiting for a bus.
  Verbal report: Maybe the man was waiting for a bus, but I was wondering whether the word 'wait' is really appropriate.

Based on the learner's comment, Fukuta (2016) asserted that he or she was involved in the process of selecting the most suitable word for his or her message which is part of the formulating stage of speech production. Below are reports in the same study that are related to syntactic encoding and phonological encoding which are also part of the formulating stage of speech production.

- **Production data**:   There was a stranger ... who ... followed him.
  Verbal report: At that time, I was wondering whether I should say 'There was a stranger who followed him' or 'There was a stranger who was following him,' grammatically. (*syntactic encoding*)
- **Production data**:   One day, a man saw a cherry bros... cherry blossoms.
  Verbal report: I was confused whether the appropriate pronunciation is 'blossoms' or 'brossoms'. (*phonological encoding*)

Qiu (2019) on the other hand, found reports related to lexical choice and syntactic encoding. (The production data was not included in the article).

- **Verbal report**:   I had some other substitutive words in mind which could be used without affecting the general meaning of the story, but I couldn't use them because none of them was shown on the picture.
- **Verbal report**:   I was more cautious about grammatical mistakes in the short speech task. I knew I mistakenly used 'he' to replace 'she', so I corrected it. I was more aware of the tense. For the picture task, I made mistakes, but I was not aware of them.

Finally, Mohammad Javad et al. (2012) focused on the monitoring stage of speech production. In this example, it is clear that the learner was involved in the process of pre-production monitoring.

- **Production data**:   There were musicians and a [**pause**] <u>beautiful little girl</u>
  Verbal report: First I wanted to say pretty girl but then I thought 'beautiful little girl' is better.

A comparable use of verbal reports can be seen in a study on cognitive strategies employed by L2 learners during listening (Rukthong & Brunfaut, 2019). In this study, verbal reports were used to investigate the cognitive and metacognitive processes that L2 learners go through during a listening task including selective/directed attention, monitoring and so on. Taken together, these examples showed that verbal reports can lead to important findings about the psycholinguistic aspect of L2 speech production.

This observation then prompted the review to look at the verbal reports in the other studies (Azizullah & Najmeh, 2012; Kahng, 2014; Páez, 2020). Upon closer inspection, it was discovered that it is also possible to draw connections between the verbal reports in some of these studies and the psycholinguistic processes. For example, in Kahng (2014), there were reports that are related to the process of lexical choice. (Numbers with brackets in the production data represent the length of pauses while numbers without brackets represent the length of fillers, both measured in milliseconds.)

- **Production data**:   I enjoy :: (1504) <u>um1090 {(in a whisper) watching seeing looking (288) reading} (333)</u> reading books //
  Verbal report: I got confused among watching, seeing, looking, and reading. I learned that with "books", I should say "reading".

Similar to the example from Fukuta (2016), the learner's comment described the process of selecting the most suitable word for his or her message. However, unlike the earlier example, the event whereby several lexical items (watching, seeing, looking, reading) that are conceptually related (books) became activated can clearly be observed in this example. Furthermore, there are reports in Kahng's study related to the co-activation of languages. The co-activation of languages is an added stage in the process of speech production for L2 speakers or bilinguals. When bilinguals try to formulate a message, the corresponding lemmas from both languages are activated (Costa et al., 2013). The co-activation of languages is what makes it possible for bilinguals to switch between different languages when they speak.

- **Production data**:   I <u>um594 (529)</u> accept my mom {advise} advice //
  Verbal report: I was asking myself "what is 'padadeulida' ('accept' in Korean) in English?"

Based on the learner's comment, in the process of coming up with the English word 'accept', the L1 Korean equivalent '*padadeulida*' was also active in the learner's mind. This suggests that the languages were co-activated for the learner even though he or she was engaged in an L2 task. Interestingly, comments in this capacity were not found in any of the other studies including the ones that specifically referred to the psycholinguistic aspect of L2 speech production.

This may be due to the fact that Kahng's study was the only study conducted in an immersive ESL/EFL setting whereby the participants were learning English while living in an English-speaking country.

CONCLUSION

The review above showed that the use of verbal reports in studies on different aspects of L2 speech production has been consistent for the past ten (10) years. In particular, the review found that verbal reports have relatively similar procedures across all of the studies. The review also found that verbal reports were primarily used to triangulate and explain the data obtained from other research instruments in the studies. Nevertheless, the review has some limitations. For example, all of the studies reviewed were based on the ESL/EFL context. In future reviews, studies from other language teaching contexts including non-Western languages such as Mandarin or Japanese should be included to create a more diverse context for the study of verbal reports in L2 speech production research. Moreover, most of the studies reviewed were qualitative studies in which verbal reports were used to complement other qualitative instruments. Future reviews should include more studies where verbal reports are paired with quantitative instruments (such as Azizullah and Najmeh (2012)). Alternatively, future reviews can consider studies where verbal reports are analysed quantitatively via analytical methods like content analysis.

Based on the review, there are also some implications for the use of verbal reports in future studies on L2 speech production. Firstly, the procedures surrounding verbal reports are important to ensure the reports' validity. Thus, future studies on L2 speech production must strive for a sound and rigorous methodology when it comes to verbal reports. For example, efforts should be made to carry out retrospective reporting as soon as possible following a task (Gass & Mackey, 2017). In comparison to the other studies reviewed, Mohammad Javad et al. (2012) collected verbal reports from their participants within minutes from the end of the main production task. This enabled their participants to recall the mental processes behind their actions with greater ease. However, it is not always possible to achieve such a small gap between the main task and the verbal reports. Thus, what is important is that any delays in the reporting process must be explicitly stated in subsequent publications. In addition to that, careful planning is also required for the stimulus that will be provided to help participants with their recollection of the task during retrospective reporting. Audio recording is the preferred stimulus among recent studies on L2 speech production that incorporated verbal reports due to the relatively simple technical requirements. However, the foremost consideration in the selection of stimulus should be the extent to which it can help to elicit information from participants. According to Gass and Mackey (2017), apart from audio and video recordings, the stimulus for verbal reports can also be in the form of samples of written work, field notes, and computer or software generated data. Once a relevant stimulus has been chosen based on the objectives of a study, researchers must then ensure that they are able to gather or produce the stimulus. Following that, they must also ensure that the stimulus can be presented to the participants in a suitable manner. Finally, researchers should ask questions that will allow participants to explicitly recall their mental processes (Gass & Mackey, 2017). For example, Qiu (2019) asked questions that targeted specific events within the main production task. This method of questioning will push participants to recall their mental processes more accurately. Moreover, allowances should be made for participant-initiated reporting such as those demonstrated in Mohammad Javad et al. (2012) and Kahng (2014). Participant-initiated reporting minimises researcher interference and encourages participants to verbalise the entirety

of their mental processes. However, researchers must take into account factors such as participants' age and their proficiency levels for this type of reporting. Where possible, a combination of both researcher and participant initiation should be practiced for a better reporting effect.

Furthermore, the use of verbal reports should be expanded to include more studies on interactive L2 speech. Out of all the studies reviewed, only Azizullah and Najmeh (2012) had an interactive task (pair discussion) in their study. Using verbal reports to complement interactive tasks will allow for a more detailed comparison of learners' L2 use under different conditions. For example, in Azizullah and Najmeh's study, the discussion task revealed problem-solving activities associated with interlocutor speech that were not available via the non-interactive task. Verbal reports allowed the researchers to compare the problem-solving mechanisms used by learners under dialogic and monologic conditions. In this way, using verbal reports along with interactive tasks may lead to new insights on L2 speech production that can inform the teaching of L2 speaking skills. On the other hand, it may be worthwhile to consider verbal reports for studies on bilingual speech production given that verbal reports can provide access to the psycholinguistic aspect of L2 speech production. Research on bilingual speech production from the psycholinguistic perspective including code switching behaviours is traditionally based on quantitative methods. Verbal reports will be able to lend qualitative support to the findings in this area. For example, the co-activation of languages in bilingual speech production is often studied via naming experiments involving cognates (Costa et al., 2013). The studies found that cognates are named faster compared to non-cognates. These findings were then taken as a sign of the co-activation of languages among bilinguals. In Kahng's study, a similar phenomenon was successfully described by an L2 learner via verbal reporting. This signals the capacity of verbal reports in revealing the psycholinguistic processes behind bilingual speech production in a qualitative manner.

In conclusion, this review has shown that the use of verbal reports in L2 speech production research has indeed diversified considerably following Kormos' proposal more than two (2) decades ago (Kormos, 1998). Verbal reports have continued to provide important information about the process of L2 speech production. However, there is room for improvement still in terms of methodology and research topics in order to fully maximise the potential of this research instrument in L2 speech production research.

ACKNOWLEDGEMENT

Note
1. All production data and verbal reports quoted in this article are reproduced in the same format as the original articles including the use of punctuation, bold font and underline.

REFERENCES

Abdulkhaleq A. Al Qahtani. (2020). Exploring reading strategies through introspective and retrospective think-aloud protocol. *TESOL International Journal*, *15*(4), 37–63.

Azizullah Mirzaei, & Najmeh Heidari. (2012). Exploring L2 speech production management from a cognitive perspective: A focus on fluency. *Iranian Journal of Applied Linguistics*, *15*(2), 33–60.

Barkaoui, K. (2015). *Test Takers' Writing Activities During the TOEFL iBT Writing Tasks: A Stimulated Recall Study* (Report No. RR–15-04). TOEFL Committee of Examiners.

Bowles, M. (2019). Verbal reports in instructed SLA research. In R. Leow (Ed.), *The Routledge Handbook of Second Language Research in Classroom Learning* (pp. 31–43). Abington: Routledge.

Chapman, K. (2021). Characteristics of systematic reviews in the social sciences. *Journal of Academic Librarianship*, *47*(5), 1–9.

Costa, A., Sadat, J., & Martin, C. (2013). Lexical access and selection in bilingual speech production. In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Acquisition* (pp. 382–386). New York: Routledge.

Egi, T. (2004). Verbal reports, noticing, and SLA research. *Language Awareness*, *13*(4), 243–264.

Ericsson, A., & Simon, H. (1993). *Protocol Analysis: Verbal Reports as Data*. Massachusetts: Bradford Books.

Fukuta, J. (2015). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research*, *20*(3), 321–340.

Gass, S., & Mackey, A. (2017). *Stimulated Recall Methodology in Applied Linguistics and L2 Research* (2nd ed.). New York: Routledge.

Hazita Azman, Sepideh Mirzaeifard, & Zaini Amir. (2017). Hypermedia literacy: An insight into English as a foreign language online reading processes. *Akademika*, *87*(1), 207–220.

Imelda Hermilinda Abas, & Noor Hashima Abd Aziz. (2016). Exploring the writing process of Indonesian EFL students: The effectiveness of think-aloud protocol. *Advances in Language and Literary Studies*, *7*(2), 171–178.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809–854.

Kim, Y., & Yoon, H. (2014). The use of L1 as a writing strategy in L2 writing tasks. *GEMA Online Journal of Language Studies*, *14*(3), 33–50.

Kormos, J. (1998). Verbal reports in L2 speech production research. *TESOL Quarterly*, *32*(2), 353–357.

Levelt, W. (1989). *Speaking: From Intention to Articulation*. Massachusetts: The MIT Press.

Mackey, A., & Gass, S. (2016). *Second Language Research: Methodology and Design* (2nd ed.). New York: Routledge.

Mohammad Javad Ahmadian, Parisa Abdolrezapour, & Saeed Ketabi. (2012). Task difficulty and self-repair behavior in second language oral production. *International Journal of Applied Linguistics*, *22*(3), 310–330.

Noorizah Mohd Noor. (2010). ESL learners' reading approaches of an academic expository text. *3L: Language, Linguistics and Literature*, *16*(2), 19–46.

Nur Ainil Sulaiman, Khazriyati Salehuddin, & Rozainee Khairuddin. (2020). Reading English academic texts: Evidence from ESL undergraduates' eye movement data. *3L: Language, Linguistics, Literature*, *26*(1), 60–78.

Nurul Adila Hamdan, Maslawati Mohamad, & Shahizan Shaharuddin. (2017). Hypermedia reading materials: Undergraduate perceptions and features affecting their reading comprehension. *The Electronic Journal of E-Learning*, *15*(2), 116–125.

Ormerod, T., & Ball, L. (2017). Cognitive psychology. In C. Willig & W. Stainton-Rogers (Eds.), *The SAGE Handbook of Qualitative Research in Psychology* (2nd ed., pp. 554–575). Los Angeles: Sage.

Páez, K. L. (2020). The impact of oral pushed output on intermediate students' L2 oral production. *GIST – Education and Learning Research Journal*, *20*, 85–108.

Page, M., McKenzie, J., Bossuyt, P., Boutron, I., Hoffmann, T., Mulrow, C., Shamseer, L., Tetzlaff, J., Akl, E., Brennan, S., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L., Stewart, L., Thomas, J., Tricco, A., Welch, V., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*, 1–11.

Qiu, X. Y. (2019). Picture or non-picture? The influence of narrative task types on lower- and higher-proficiency EFL learners' oral production. *IRAL - International Review of Applied Linguistics in Language Teaching*, 1–27.

Rossa, H. (2016). The cognitive processes elicited by L2 listening test tasks: A validation study. In A. Lenzing & M. Liebner (Eds.), *Developing, Modelling and Assessing Second Languages* (pp. 207–237). Amsterdam: John Benjamins.

Rukthong, A., & Brunfaut, T. (2019). Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks. *Language Testing*, *37*(1), 31–53.

Siricord, T., & Melor Md Yunus. (2016). Learners' perceptions on the effectiveness of videoscribe on improving listening and speaking in rural school of Sarawak. *Asian EFL Journal*, *6*, 50–61.

Tokowicz, N., & Degani, T. (2013). Lexical concepts. In P. Robinson (Ed.), *The Routledge Encyclopaedia of Second Language Acquisition* (pp. 386–389). New York: Routledge.

Zhao, Y., & Brown, P. (2014). Building agentive identity through second language (L2) creative writing: A sociocultural perspective on L2 writers' cognitive processes in creative composition. *Asian EFL Journal*, *16*(3), 116–154.

ABOUT THE AUTHORS

Chong Yin Sheun is an English teacher from Malaysia who is currently on study leave to pursue her doctorate in English Language Studies at the Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia. Her PhD study is on the interactional and psycholinguistic factors behind L2 teachers' use of code switching in class.

Bahiyah Dato' Hj. Abdul Hamid (PhD) was an Associate Professor at the Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia (UKM). Her research interests include language and gender, gender issues, identity construction, code alternation and code choice, and discourse and semiotics analysis.

Normalis Amzah is a Japanese Language Lecturer at the Center for Research in Language and Linguistics, Universiti Kebangsaan Malaysia. Her expertise is in Japanese language and Japanese studies, translation and foreign language acquisition. Her research interests include advertising and marketing communication, foreign language teaching and learning, translation, and Japanese political economy and management.

Khazriyati Salehuddin (Ph.D) is an Associate Professor and a Psycholinguist at the Faculty of Social Sciences & Humanities, Universiti Kebangsaan Malaysia. She received her PhD in Psycholinguistics in 2010 from Western Sydney University. Khazriyati has led several research groups related to the area, including two national-level grants, namely Qur'anic Memorisation Techniques: A Psycholinguistic Module for Non-Arabic-Speaking Malay Speakers (FRGS) and Exploring the cognitive and perceptual processes in reading among Malaysian readers (ERGS). Khazriyati uses various methods in her research and has published several journal articles, books, and book chapters on the area. One of her works is published in South and Southeast Asian Psycholinguistics by Cambridge University Press and her book, Psikolinguistik: Penerokaan Minda Berlandaskan Bahasa, is published by UKM Press.. For more: goo.gl/rskBFV