ISELT 146
Interdisciplinary Studies in English Language Teaching

Questioning and Responding in Target Language Use Situation Tasks: A Think-aloud Study

Mina Madadi[1] & Reza Rezvani[1,*]

[1] Yasouj University, Yasouj, Iran[†]

10.22080/iselt.2021.22007.1018

Received
April 23, 2023
Accepted
July 25, 2023
Available online
August 25, 2023

Keywords
IELTS, Questioning, Responding, TLUS, Think-aloud

Abstract

The present study reports a qualitative study aiming to explore whether cognitive processes underlying responding and questioning in English are cognitively analogous. It further sought to discover the latent differences between responding and questioning cognitive processes in Target Language Use Situation tasks. To this end, 20 Iranian general IELTS applicants from two different institutes in Shiraz, with two different language proficiency levels (intermediate and advanced) participated in the study. They were administered a normal oral IELTS responding and a reverse questioning task. Articulated Thoughts in a Simulated Situation (ATSS) paradigm as a think-aloud approach was used to collect qualitative data during task completion. The applicants' recorded voices during task completion were transcribed and analyzed to examine the potential differences between responding and questioning cognitive processes. The analysis of the qualitative data through ATSS paradigm, in general, indicated that the cognitive processes underlying these two processes are not exactly parallel. The applicants tended to be more accurate and fluent during responding tasks. More specifically, they had fewer and shorter pauses and generated more intelligible and comprehensible productions by committing fewer grammatical errors in both proficiency levels. The paper discusses the findings and the implications for second language learners, teachers, and test developers.

1. INTRODUCTION

Assessment plays a key role in any educational system. It is defined as "Any systematic method of obtaining information from tests and other sources, used to draw inferences about characteristics of people, objects, or programs" (AERA, 1999, p. 172). Essentially, a language test is a procedure for eliciting implicit knowledge through what we can observe, and from which we can infer the amount and type of language knowledge which we cannot directly observe (Douglas, 2010). Assessments developed and used locally are likely to hold lower stakes than large-scale tests which

* **Corresponding Author:** Reza Rezvani, Department of English Language and Literature, Faculty of Humanities, Yasouj University, Yasouj, Iran, **Email**: rezvanireza@gmail.com

† This article is based on the M.A thesis of "A Psycholinguistic Study of Iranian IELTS Applicants' Responding and Questioning Processes in Target Language Use Situation Tasks: ATSS paradigm"

are “conducted on a regional, national, or international scale involving large student populations” (Simon, Ercikan, & Rousseau, 2012, p.1)

The key to large-scale assessment is uniformity in development and administration (Kunnan, 2009; Wendler & Walker, 2006). This uniformity and systematicity is both an asset (Kunnan, 2009) and may lead to imprecision and potential misinterpretation (Fox, 2008; Read, 2009) when used across time, regions, administrations, and examinees. Developers of large-scale tests are, thus, accountable for their products, and test-users are held responsible to ensure that appropriate interpretation and use of test scores are made (Chalhoub-Deville & Turner, 2000). The higher the stakes a test takes, the more considerable the demand for its validation.

Large-scale language tests have also attained growing importance “in many parts of the world in school, college, and university contexts" (Kunnan, 2009, p. 135). The International English Language Testing System (IELTS), Test of English as a Foreign Language (TOEFL), and Pearson Tests of English (PTE) are, for example, widely used all around the world. The increasing numbers of international students around the world and the importance of English as a means of communication in an international society have led to a considerably large population of university applicants. The scores obtained from these tests are used to make critical decisions that affect test-takers' life and prospective career. Standardized academic language proficiency test scores are frequently used for several purposes, including admissions of international students to degree programs and identification of students' post-entry language support requirements (Ockey& Gokturk, 2019). Such admission instruments are different in terms of length, format, and test content (Chalhoub-Deville & Turner, 2000) though they might be used for generically similar uses and decisions. Although TOEFL and PTE as admission tests are as valuable as IELTS, the focus of the present study is on the IELTS examination.

IELTS examination as one of the most popular tests, administered by the University of Cambridge Local Examinations Syndicate (UCLES) used for university admission purposes, predicts the extent to which a candidate will be able to begin studying through the medium of English (O’loughlin & Arkoudis, 2009). IELTS scores are required for students from particular countries to gain their visas to be admitted to English-medium universities. Through this test both academic and general English language proficiencies can be assessed.

Given the increased value placed on interactive communication in the classroom, a question arises as to whether the existing English tests for entrance to tertiary education can adequately target relevant interactional skills. Douglas (2014) argues that language use for sheer display is at best unnatural and at worst a distortion. If the performance tests elicit is in some way abnormal, the inferences made about the ability that produced the performance will stand a good chance of being wrong. Therefore, the lack of question-raising tasks in classroom interactions and language assessments raises the issue of whether the test inferences are made logically?

Gibbons (2003) suggested that students require various forms of interaction to build their four skills while direct instruction (i.e., which is characterized by highly didactic curriculum, students’ receptive role, and imitation and repetition activities) remains a dominant approach in L2 education (Pufahl & Rhodes, 2011). Although language instructors express preferences for more classroom interactions, they have not always been successful at encouraging students’ interactive engagement (Pianta et al., 2012). The teacher-student question-answer interaction usually deprives students of opportunities to express diverse thoughts and practice the language. Therefore, as Tan (2007) claimed, there is a vital need for more decentralized ways of teaching to promote students’ conversation. Accordingly, Tocalli-Beller and Swain (2005) advocated peer-to-peer questioning in the classroom which is in line with the results of many studies suggesting that classroom interaction could be fostered through student-generated questioning (Song et al., 2017).

It follows then that if students' questioning skill is essential in interactions in or out of class, it must also be incorporated and reflected in a test that is intended to assess target language use situations (TLUS). However, IELTS applicants are rarely given a chance to reciprocally ask questions. It is assumed that their responding is parallel to questioning and in effect they are analogous. When the applicants answer questions, their questioning skill is also reflected. This assumption motivated the current research to see how these two skills compare cognitively.

2. LITERATURE REVIEW

Language assessment

The last few years, nevertheless, have seen the introduction of 'assessment' terminology into the language evaluation research discussion, signaling not merely a semantic change but a thoughtful theoretical one, with 'assessment' perceived to be a principal term used to refer to 'all methods and approaches to testing and evaluation whether in research studies or educational contexts' (Kunnan, 2004, p. 1). However, this conceptual shift goes beyond notions of alternative assessment (or alternatives in assessment, Brown & Hudson, 1998), perceiving the language evaluation process as a socially constructed activity embedded in the local context with teachers, students, and other community members recognized as meaningful assessment partners (Leung, 2004; Lynch, 2001; Lynch & Shaw, 2005; McNamara & Roever, 2006). While some available courses have reserved their language testing orientation and deal mostly with issues about the design and use of tests as the means for determining language proficiency (Bailey & Brown, 1995, and see Brown & Bailey, this issue), others have integrated additional assessment components, as well as an examination of the social roles of tests and testers in the assessment process (Kleinsasser, 2005). In language assessments, as we cannot observe the criterion directly, we use the test to make inferences about the candidates' subsequent performances. We make a distinction between the criterion which is the most relevant communicative behavior in the target situation and the test, which is often designed according to the test developer's understanding of the characteristics of the criterion (McNamara, 1997).

Speaking assessment

Assessing speaking in admission English proficiency tests revolves around the idea of whether the admitted students can cope with the language requirements of their studies. From a testing point of view, speaking is a productive skill that is interactive in nature and has to be measured directly in live interaction (Karim & Hag, 2014). In speaking tests, not only the knowledge of the language but also the ability to use it during task completion is assessed (Luoma, 2004). Therefore, linguistic knowledge areas such as grammar, vocabulary, pronunciation along with communicational, functional, and sociolinguistic knowledge are assessed as essential components in interactions. Concerning the purpose of the assessment and test interpretation, test designers decide to use different types of talks and language functions as the main focus of the assessment tasks. Various ways of speaking assessment have been proposed by many researchers (e.g. Alderson et al., 1995; Bachman & Palmer, 1996; Lynch & Davidson, 1994). Hughes (2003) suggested some standards for assessing oral ability emphasizing that to assess learners' speaking ability firstly an appropriate task to elicit information is needed and secondly validity and reliability of elicited sample and its scoring are to be ensured.

While there might be no truly authentic situation for EFL learners to practice language use and to be assessed, attempts can be made to improvise TLUS and interactions among students and their teachers (Madadi & Rezvani, 2020). Language use is related to a sociocultural view of learning as a social process framed within broader contextual practices (Vygotsky, 1986). In a Vygotskian sense, active learners do not take sole responsibility for their learning processes and for discovering

meanings. The teacher in the Vygotskian classroom carefully designs learning environments that enable learners to use languages in meaningful ways (Brevik, 2015). Likewise, a multilingual perspective on language learning considers classrooms as fundamentally social contexts in which learners use their languages as they engage in various classroom practices (Cenoz & Gorter, 2017; Garcıa & Li Wei, 2014). By doing so, students are shaped by their use of languages in communicative interaction in the social context of the classroom (Cenoz & Gorter, 2014, 2017; Park & De Costa, 2015). The interactions that learners make are part of the process of language learning and can affect their language development.

As classroom activities are based on communication and interaction, assessments must be interactive and proficiency-based as well. The learners must be put in target language use situations where they can hear and react to real uses of language. Therefore, not only they must learn to respond but also, they should learn how to pose a question to initiate or continue a communication. The appropriateness of interviews in assessing speaking and interactional competence has raised questions and concerns in language assessment (Karim & Hag, 2014) as the interactive nature of communication and more specifically TLUS calls for the need to design reciprocal tasks assessing both questions raising and responding abilities. Assessing learners through such reciprocal tasks assumes instruction on its development, though in practice teachers seldom get students to initiate or to ask questions (Graesser & Person, 1994; Rezvani & Sayyadi, 2016; Willis and Willis, 2007). Despite the instructors' remarkable resort to questioning to instruct and manage classes, they were observed to neglect and fail to encourage their learners to develop practically such a critical skill (Rezvani & Sayyadi, 2015).

Questioning in IELTS

Even though an interview as a gentle conversation between an interviewer and an interviewee is the most widely utilized task for testing speaking skills such as IELTS tests, it has been argued to have its shortcomings. In interviews, the interviewer remains dominant because he is in charge of taking all the initiatives, while the candidate or interviewee is merely supposed to respond to the questions she has been asked. Subsequently, only one style of speech is prompted, and many functions such as an inquiry for information are not characterized in the candidate's performance. According to Hughes (2003), the relationship between the interviewer and the candidate is typically such that the candidate speaks to a predominant person and he is reluctant to take the initiative. If the interviewees were given chances to pop up the questions which might cross their minds, it would, on the one hand, help them build up their confidence and be at more ease and, on the other, would aid the interviewer in assessing applicants' questioning skill (Karim & Hag, 2014).

In one of the few studies concering questioning and responding in IELTS examinations it was found that applicants who are more familiar with the format of the responding tasks and typical topics about everyday life which are commonly used in the test and practiced in IELTS preparation courses demand less information processing load (Madadi & Rezvani, 2019). To complete responding tasks the applicants had to use some ready-made chunks requiring them to use less of the cognitive mechanism resulting in short pauses and fewer hesitations (Madadi & Rezvani, 2019). To put it another way, familiar tasks imposed less information load and less demanding conditions and cognition. The question here arises as to what cognitive processes might take place when applicants face unfamiliar tasks of questioning. To this end, this paper attempts to explore whether questioning and responding as the main constituents of an authentic dialogue or discussion have analogous cognitive processes in TLUS as represented in IELTS speaking modules.

Research Question

1. How do questioning and responding in TLUS IELTS tasks compare cognitively?
2. What are the main differences between questioning and responding regarding their underlying cognitive processes?

3. METHOD

This paper described the qualitative findings of a larger project involving both qualitative and quantitative approaches. The study aimed to point out whether questioning and responding have parallel underlying cognitive processes or not. The articulated thoughts in simulated situations (ATSS) paradigm which is a think-aloud approach to cognitive assessment was used to get the participants to verbalize their thoughts while engaged in given tasks.

Participants

The sample of the study included 20 general IELTS applicants selected purposively from among applicants of two institutes in Shiraz, Iran. To see whether the proficiency of the participants might also impact the results we included both intermediate (5 male and 5 female) and advanced (5 female and 5 male) applicants. Their age ranged from 27 to 42.

Instruments

ATSS was employed to collect qualitative data in this study. The think-aloud approach is particularly useful in understanding the processes of cognition. As think-aloud methods usually assess cognitions concurrently with their occurrence, they may be better suited to tap actual thought content than other modes do (Davison, Vogel, & Coffman, 1997). In a standard think-aloud method, such as ATSS, researchers ask participants to verbalize their thoughts while performing the given tasks, and the verbalizations are recorded for subsequent analysis. The ATSS paradigm is a measure that prompts respondents' immediate cognitive and affective reactions to specific situations that just have been presented to them (Zanov & Davison, 2010). In particular, the model we used in this study is Ericsson and Simon's model (1993) known as non-metacognitive verbalization reflecting cognitive process accurately, since the applicants were not asked to explain or bring any justifications for their thoughts. As Ellis (2004) asserted this method would appear to be the most valid measure of a learner's explicit knowledge.

Concurrent think-aloud was the main method of inquiry in this study since the applicants had to think aloud while doing the tasks, though there were some unintentional reports of retrospective think-aloud as well. In this study, we were interested in thoughts the applicants had when they were completing the tasks. Often, when people are completing some tasks, they have a kind of internal monologue going through their heads, a continual stream of thoughts or feelings which mirror their reactions to something which is happening (Davison et al., 1983). The reason we opted for ATSS was that it resorts to simulated situations exactly aligned with what we intended to look at, that is, the participants' thinking in doing TLUS tasks. The tasks as ATSS prompts were questions adopted from actual general IELTS oral interviews, and respective answers representing typical questions suggested and approved by three experienced IELTS preparation course teachers. While completing the tasks, all the applicants were asked to think and talk aloud whereby their talks were recorded. We avoided interrupting the applicants but we occasionally reminded them to think aloud in case they forgot to verbalize their thoughts.

Data collection procedures

Authentic IELTS speaking tasks were given to the applicants in a counterbalancing order, half starting with questions as the first tasks and responses as the second tasks. The other group of candidates did the same tasks in reverse order to enhance the interval validity of the study. When

a sample question from the actual test was given, the applicants were asked to respond to it and when an analogous statement was prompted, they were supposed to build the original question. They were also asked to think aloud whilst they were doing the tasks. The applicants' verbal data were recorded for further analysis. The anonymity of participants' responses was emphasized to encourage open thought articulation.

Data analysis

The articulated thoughts verbalized during task completion were transcribed line by line. Then, an attempt was made to extract what the main intentions or ideas were that the participants were expressing. As a result, the researchers divided up the transcribed data into smaller parts, namely "ideas units". As Green (1998) suggested an "ideas unit" may include a single or several utterances with a single aspect of the events as the focus. The next step as suggested by Davison et al. (1983) involved grouping the ideas units into categories depending on the aim of the study and research questions. The process of identifying ideas units and categorizing was recursive and continuous as the researchers must return to raw data to re-do and re-think the transcribed data until they faced reasonable ideas unit. In accord with Glaser's (1978) recommendation every time a new idea or theme was found from the data, the researchers made notes and these memos were included in the analysis as well. To make sure whether the ideas units were consistent and reliable the inter-coder reliability of the data was examined using Cohen's kappa as an index of the agreement for categorical and sub-categorical codes. The Kappa value was 0.93 indicating substantial agreement for the coding (see Landis & Koch, 1977 for Kappa index interpretations).

4. RESULTS AND DISCUSSION

The first set of associated questions and answers concerned the participants' favorite food and TV programs; ten participants answered an actual IELTS question (What's your favorite flavor?) and the rest tried to build a question for a typical answer to it (My favorite flavor is sweet.). All the applicants had a clear expression of their opinion although they had some pauses (silent and filled) thinking about what to answer. The given question seemed rather challenging since most of the applicants didn't know what their real flavors were. A point worthy of note in most verbal reports was that they first identified the keywords of the question. This was followed by thoughts of how it could be structured in the answer. Having identified the keywords semantically or grammatically as the first step, they moved on to structure them into simple statements as the next step. Then, challenges aroused as most of them could not decide what their favorite flavors were. As the following example suggested while one of the intermediate applicants (extract A) recognized the appropriate structures to answer the question he had a silent pause since he was not sure about his favorite flavor. The given response by the applicant manifested that the pause couldn't be because of a lack of grammatical knowledge as the answer was grammatically and semantically correct, thus the applicant had doubts about the flavor. This might be admitted as the applicant had used the word "think" which is a sign of doubt.

- Extract A (intermediate)

Question: What is your favorite flavor?

Responding Verbatim think-aloud: I (1) think spicy is my favorite flavor.

*numbers in parentheses indicate the pause length.

The other advanced applicants (B & C) answered the same question with a filled pause (um), though the pause wasn't suggestive of a lack of grammar or lexical knowledge since the applicants responded accurately and used more lexis to complete the question in longer statements. It was observed that filled pauses were used by advanced applicants in comparison to intermediate applicants, who on the contrary had more tendency to pause silently.

- Extract B (advanced)

Question: What is your favorite flavor?

Responding Verbatim think-aloud: I like (um,1) mint and lemon.

- Extract C (advanced)

Question: What is your favorite flavor?

Responding verbatim think-aloud: I prefer (um,1) chocolates and candies (1) which their flavor is sweet.

Based on the data analysis “pause length" was identified as the first "ideas unit". Completing the responding task, all the applicants seemed to reduce the length of pauses while giving the responses. The applicants' short pauses were natural as native speakers also might normally pause while answering such questions. Consequently, effective pauses not only were beneficial for comprehension but also, were the manifestation of more natural and native-like speech. This is analogous with the finding that pausing is natural and necessary for breathing needs and for pragmatic use during speech (Ling, 2006), though long and unnecessary pauses might have negative effects on speaking fluency.

Fluency was considered as the second “ideas unit”. Levelt (1989) emphasized that speaking involves the processing of a considerable amount of data in a limited period, that is, two or three words are produced per second in natural speech. He claimed that fulfilling this great task requires automaticity, not conscious monitoring as human capacity is too limited to focus consciously on the information (Segalowitz & Hulstijn, 2005). Other researchers also claimed that a pivotal difference between fluent and non-fluent L2 speakers is the extent to which lexical processing is automatized. Different factors such as speech rate, repairs, amount and frequency of hesitation, location of pauses, and length of runs of fluent speech between pauses are associated with the psycholinguistic facets of performance and production (e.g. Lennon, 1990; Möhle, 1984; Towell et al., 1996). Levelt (1989) also maintained that the automaticity of language production has generated uninterrupted fluent speech, therefore short pauses in the responding tasks in addition to the correct form of responses might lead to the fluency of the applicants in responding tasks. As such, and as fluency and connected speech are interrelated, it can be inferred from the applicants' verbal reports that both intermediate and advanced applicants had used more connected speech in their given responses and were considered as more fluent.

This might suggest that information load or cognitive processing demand (Krivokapić, 2007) is an effective factor regarding the length of pauses. Information load is likely to be minimal if the learners are familiar with the tasks they have to perform. It would also be of a minimum load if the language learners are familiar with the topics they are going to talk about. Higher information load, in turn, makes cognition more careful and slower (Rabbitt, 1968; Robinson, 2001). This is in agreement with other studies concluding that learners that were more familiar with the format of the responding tasks and typical topics about everyday life which are commonly used in the test and practiced in IELTS preparation courses very often demand little information processing load (Madadi & Rezvani, 2019).

In questioning tasks on the other hand (extract D) borderlines between lexical words were more clearly identified through long pauses resulting in less natural speaking. To complete questioning tasks the applicants had to construct questions from scratch, requiring them to use more of the cognitive mechanism resulting in a longer pause as an indicator of cognitive processing (Madadi & Rezvani, 2019). In other words, unfamiliar tasks imposed extra information load and more demanding conditions and cognition. The long pauses would break up the questions into smaller parts leading to less connected and natural speech.

- Extract D (advanced):

Response: Yes, I think holidays are becoming more and more important.

Questioning verbatim think-aloud (advanced): (Uh,3) Do you think (2) we should have a (2) program (1) or plan for (2) on holidays, (1) for our holidays

It was observed that the pause length in these tasks increased to 4-6 seconds and self-correction strategies were used a lot to construct or recast an appropriate question as manifested in extract E and F.

- Extract E (advanced):

Response: I want to travel to the places which are unique and interesting.

Questioning verbatim think-aloud (advanced): (Uh,4) which/ what kind of places do you want to go?

- Extract F (intermediate):

Response: I prefer traveling in a group rather than traveling alone.

Questioning verbatim think-aloud (advanced): (Uh,3) Do you prefer to go to trip (1) with (→ with omitted) alone or (2) with somebody/ with someone.

The analysis and comparison of the applicants' verbal reports revealed that IELTS applicants tended to be more fluent and native-like in responding tasks than questioning tasks irrespective of their proficiency levels. Similarly, it was found that connected speech resulting in more fluency contributed toward naturalness and intelligibility. Intelligibility has to do with 'speech clarity' or the proportion of a speaker's output that a listener can readily understand. Hence, more fluent applicants were more intelligible as well.

Focusing on verbal reports and transcriptions, we inferred that the applicants had good knowledge of grammar, as far as the responding tasks were concerned, since all the applicants regardless of their proficiency levels were able to answer the questions correctly. More specifically, their verbal reports also manifested that in addition to knowing the grammatical rules theoretically, the applicants were also able to operationalize the rules to make correct answers to the questions. So, both explicit (verbal reports) and implicit (correct answers to the questions) knowledge of grammar were available among all the applicants while doing the responding tasks. As a result, "knowledge of grammar" was identified as the "ideas unit" closely associated with accuracy. The following exemplary excerpts evidence the findings.

- Extract G (intermediate):

Question: What's your favorite flavor?

Responding Verbatim think-aloud 1 (intermediate): "this is a WH question [!]", "so I should initiate the answer like this" "my favorite flavor is (1) let's say dark, my favorite flavor is dark."

Responding Verbatim think-aloud 2 (advanced): (1), "a WH question, with favorite as the focus", ok. "My favorite flavor is lemon and mint."

In questioning tasks, the same ideas unit was identified in their verbal reports, though the applicants could not accomplish accuracy. As the following sample extracts exemplify, the applicants of both levels (intermediate: H & I, advanced: J & K) tried to focus on using grammatically correct structures to make related questions, but they were not quite successful. The applicants centered their attention on the correct forms in making questions as the recorded verbal reports of the applicants' articulated thoughts exhibit.

- Extract H (intermediate):

Response: I may visit my family or have a short trip in holiday.

Questioning Verbatim think-aloud: ok, I should use "do", (1) do you may visit your family? Or let's say it without "May", (2) do you visit in holiday?

- Extract I (intermediate)

Response: I want to travel to places which are unique and interesting.

Questioning Verbatim think-aloud: where, (2) (should begin with where and want). Where do you want, no, no, no, like to travel to? Where do you like (1) to travel in future?

- Extract J (advanced)

Response: I may visit my family or have a short trip in holidays.

Questioning Verbatim think-aloud 1: (1) "you want me to make a question?" (2), "I should use part of the sentence as a part of question," (uh, 2), about holidays (2), "which kind of holiday you visited your parents?"

Questioning Verbatim think-aloud 2: a WH question about holidays/ (3) it might be a general question (um, 2) When did you (1) when did you visit your family?

- Extract K (advanced)

Response: Yes, I think holidays are becoming more and more important?

Questioning Verbatim think-aloud: "It needs a yes/no question, (um, 2) which holidays do you know", (2) "and you prefer it's important for you?"

In the sample extracts (H & I) the intermediate applicants were verbalizing the correct structural rules to accomplish the questioning tasks, though they couldn't use them to build up questions. Other sample extracts (J & K) also pointed to the same problem among advanced applicants. To conclude, most applicants irrespective of their proficiency level couldn't apply the verbalized and well-known rules of structures while performing the questioning tasks.

Another point worthy of note is the applicants' conscious and unconscious thoughts and strategies to smooth task completion (for a discussion see Cohen, Ledford, & Spreitzer, 1996) and to personalize the task accomplishment procedure. In both tasks completion the applications of planning, monitoring, and evaluating as metacognitive strategies can be identified. Metacognition refers to awareness of one's knowledge, what one does and does not know, and one's ability to understand, control, and manipulate one's cognitive processes (Meichenbaum, 1985). It includes knowing when and where to use particular strategies for learning and problem solving as well as how and why to use specific strategies. Chamot (2005) maintains that strategic language learners hold metacognitive knowledge about their thinking and learning approaches, a good understanding of what a task entails, and the ability to organize the strategies that best suit both the task demands and their strengths. The language learners do not think about these strategies while performing them but, if they are asked what they are doing, they can usually accurately describe their metacognitive processes and strategies, therefore the applicants of this study were asked to think and talk aloud while completing the tasks to pinpoint what metacognitive strategies they had used during task completion. As an example, the articulated thoughts of an applicant (extract J) are provided below. it can be seen that the applicant was thinking about preparation for task execution, identifying procedures and requirements of a task to plan it, by asking the question that "you want me to make a question?"

- Extract J (advanced)

Response: I may visit my family or have a short trip in holidays.

Questioning Verbatim think-aloud 1: (1) "you want me to make a question?" (2), "I should use part of the sentence as a part of question," (uh, 2), about holidays (2), "which kind of holiday you visited your parents?"

There were also generated thoughts that represented the recognition of a problem (grammatical errors, inappropriate lexical choice) followed by no further attempt to correct it (extract L) or evaluations that lead to improvements or revisions through self-correction (extract M). These examples point to the monitoring and evaluation processes. Such examples of strategies use led to the identification of the metacognitive strategies as another "ideas unit".

- Extract L (intermediate)

Response: I prefer traveling in a group rather than traveling alone.

Questioning Verbatim think-aloud: (4) how do you like travel? I think, I'm not sure. Something is missing.

- Extract M (advanced)

Response: I prefer traveling in a group rather than traveling alone.

Questioning Verbatim think-aloud: (Uh, 3) Who (1) do you prefer to go on a trip? (1) On a trip with (self-correction)?

In both tasks, metacognitive strategies were frequently used by the applicants. However, in responding tasks, the operation of the metacognitive strategies such as self-correction yielded almost completely correct responses, while in questioning tasks the applicants did evaluate the questions, but very often they failed to self-correct even if they could identify the problem. The following is a telling example that shows that the applicant identified the problem, self-corrected but the cognitive demands and loads disrupted the memory of the questioning task.

- Extract N (advanced)

Response: Yes, I think holidays are becoming more and more important.

Questioning verbatim think-aloud: (Uh, 3) do you think we should have a (2) program or plan for on holidays (self-correction), (1) for our holidays? Would you please repeat the sentence?

Overall, we found more errors in questioning tasks though the lexis in both questions and responses produced were of a similar size. The errors for measuring accuracy included lexical, morphological, and syntactic errors and omissions of article, verb, and subject (see Michel, Kuiken, & Vedder, 2007).

It was also noted that the applicants used more ready-made chunks to respond to questions. This also reduced the pause length and the number of errors. It can be inferred that as the participants had practiced similar tasks they remembered or were coached to provide quicker and more formulaic responses. This also minimizes the cognitive load and frees more of the cognitive capacity for analysis as it is argued that in producing routines speakers do not necessarily implement all stages of speech production (Pickering & Garrod, 2004). Kormos (2011), comparing routines and creatively constructed elements, also contended that routines follow the same processing stages faster, and with less conscious effort, resulting in shorter pauses and less cognitive load. When one needs to create novel plans or creative responses, consciousness is necessary for short- and long-term planning (Mandler, 1975).

In view of speech production, the control of the articulatory system might be unconscious but planning what to say might be conscious, particularly if one is expressing some new ideas, or expressing some old ideas in a novel way. According to Bock (1982), syntactic planning by skilled speakers is also relatively automatic and outside conscious voluntary control. This provides an account of why the IELTS applicants known as skilled speakers, had rather more automatic

syntactic planning or unconscious/subconscious process leading to faster-responding task completion because of the greater role of task familiarity and ready-made chunks. In the following task, the applicant could easily employ a recurrent and familiar statement.

- Extract O

Question: What is your favorite flavor?

Response verbatim think-aloud 1 (intermediate): my favorite flavor is sweet.

Response verbatim think-aloud 1 (advanced): my favorite flavor is dark, (1) like dark chocolate.

In question construction tasks, on the other hand, a completely novel task must be done. Planning what to say, finding relevant lexis, and putting them into meaningful linguistic forms, however, require more cognitive efforts. Jackson (1982) suggested that the amount of planning required depends on whether the speech is "new" speech or "old" speech. Old speech (well-known phrases) requires little planning and is relatively continuous. New speech demands planning and is characterized by hesitation pauses as indicated in this study. In Skehan's words (1998) and in short, cognitive familiarity and cognitive processing are two effective factors concerning cognitive complexity. Madadi and Rezvani (2019) further argued that applicants' unfamiliarity with questioning as an untrodden path augments the task difficulty and results in more and longer pauses, and turn, in less natural/ native-like speech.

To sum up, the following ideas units were identified as the researchers organized the transcriptions for easy retrieval. The pause and length of the pauses, the applicant's ability to operationalize their knowledge of grammar, metacognitive strategies such as self-evaluation and self-correction, and the use of readymade linguistic chunks were categorized as "ideas units".

Under the emerging codes, the following results have been noted and discussed. Table 2 summarizes the findings

Table 1: Ideas Units from Stimulated Recall Protocols

1. Pause length
2. fluency
3. Knowledge of grammar
4. Metacognitive strategies
5. Ready-made chunks

Table 2: Main Differences between Questioning and Responding Tasks

questioning	responding
less accurate	more accurate
less fluent	more fluent
more pauses/longer pauses	fewer pause/shorter pauses
less connected speech	more connected speech
less use of metacognitive strategies	more use of metacognitive strategies
less natural	more natural

5. CONCLUSION AND IMPLICATIONS

The overall findings emerging from the qualitative analysis of the verbal data indicated that responding tasks were considered as familiar tasks leading the applicants to more accurate and fluent language production, hence the applicants had fewer grammatical errors while completing the responding tasks. More self-corrections were found in responding tasks by using metacognitive strategies. More use was also made of routine linguistic chunks in responding tasks, while in questioning tasks, the applicants attempted to put more novel words together making the task, even more demanding with longer pauses. The findings generally support Plato's account of Socrates stating that, contrary to the general opinion, it is more difficult to ask questions than to answer. From a psycholinguistic point of view, questioning and responding though assumed to be parallel, differ to a noticeable extent in terms of the underlying cognitive processes. The more noticeable accuracy and fluency in responding can be linked to the applicants' familiarity with the frequent responding role students take in education. This reduces the cognitive load and efforts making the differences.

The results of the study might have implications for second language learners, teachers, and test developers. Tests can have effects on learners' and teacher's behaviors in the classroom, and "impact" more widely on teaching materials, educational systems, and even society (Taylor, 2005; Weir & Milanovic, 2003). Indeed, washback can be considered to be one aspect of impact (Taylor, 2005), the former known as micro-level and the latter as macro-level effects (Weir & Milanovic, 2003). The influence of tests on students' learning and future is irrefutable given particularly the high stakes of IELTS examinations. As the study demonstrated there are cognitive differences between questioning and responding underlying processes that merit attention and inclusion in language assessment. The current responding focus, for learners indeed, induces them to hence focus on the development of responding strategies and linguistics forms and lexis.

As Alderson and Wall (1993) argue examinations are expected to influence teachers' content of teaching and how they teach or in the words of Cheng et al. (2015) teachers play an important catalyst role in carrying positive and negative washback effects of tests. Thus, it is advisable to give a more active part to the students and their questioning in L2 classes. We think that this will pay off and active participation and question making will keep or generate new communications, essential in L2 learning.

Negative washback occurs in situations where there may be a discrepancy between the desired goals of instruction and the focus of assessment. An implication of the results of this study is that test developers, classroom or high stakes, should include space for the test-takers initiation of dialogues and questions by incorporating more reciprocal tasks instead of only requiring them to rather passively answer teachers' or examiners' questions. This can lead to a more positive washback for both L2 learners and teachers as discussed above.

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. American Educational Research Association.

Alderson, C. J., Alderson, J. C., Clapham, C., Wall, D., & Swan, M. (1995). *Language test construction and evaluation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511732980

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115-129. https://doi.org/10.1093/applin/14.2.115.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

Brown, J. D., & Bailey, K. M. (1995). Language testing courses: What are they in 2007? *Language Testing, 25*(3), 349-383. https://doi.org/10.1177/0265532208090157

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*(1), 1-30. https://doi.org/10.1037/0033-295X.89.1.1

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly, 32*(4), 653-675. https://doi.org/10.2307/3587999

Cenoz, J., & Gorter, D. (2014). *Focus on multilingualism as an approach in educational contexts*. Springer. https://doi.org/10.1007/978-3-319-02261-6

Cenoz, J., & Gorter, D. (2017). Minority languages and sustainable translanguaging: Threat or opportunity? *Journal of Multilingual and Multicultural Development, 38*(10), 901-912. https://doi.org/10.1080/01434632.2016.1263162.

Chalhoub-Deville, M., & Turner, C. E. (2000). What to look for in ESL admission tests: Cambridge certificate exams, IELTS, and TOEFL. *System, 28*(4), 523-539. https://doi.org/10.1016/S0346-251X(00)00033-5

Chamot, A. U. (2005). Language learning strategy instruction: Current issues and research. *Annual Review of Applied Linguistics, 25*(1), 112-130. https://doi.org/10.1017/S0267190505000061

Cheng, L., Sun, Y., & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching, 48*(4), 436-470. https://doi.org/10.1017/S0261444815000233.

Cohen, S. G., Ledford Jr, G. E., & Spreitzer, G. M. (1996). A predictive model of self-managing work team effectiveness. *Human Relations, 49*(5), 643-676. https://doi.org/10.1177/001872679604900506

Davison, G. C., Robins, C., & Johnson, M. K. (1983). Articulated thoughts during simulated situations: A paradigm for studying cognition in emotion and behavior. *Cognitive Therapy and Research, 7*(1), 17-39. https://doi.org/10.1007/BF01173421

Davison, G. C., Vogel, R. S., & Coffman, S. G. (1997). Think-aloud approaches to cognitive assessment and the articulated thoughts in simulated situations paradigm. *Journal of Consulting and Clinical Psychology, 65*(6), 950-958. https://doi.org/10.1037/0022-006X.65.6.950.

Douglas, D. (2014). *Understanding language testing*. Routledge. https://doi.org/10.4324/9781315889771

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning, 54*(2), 227-275. https://doi.org/10.1111/j.1467-9922.2004.00255.x.

Ericsson, K. A., & Simon, H. A. (Eds.). (1993). *Protocol analysis: Verbal reports as data*. Cambridge University Press.

Fox, J. (2008). Alternative assessment. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education. Language testing and assessment* (pp. 97-109). Springer. https://doi.org/10.1007/978-3-319-02261-6_8

García, O., & Wei, L. (2014). Language, bilingualism, and education. In *Translanguaging: Language, bilingualism, and education* (pp. 46-62). Palgrave Pivot. https://doi.org/10.1057/9781137385765

Gibbons, P. (2003). Mediating language learning: Teacher interactions with ESL students in a content-based classroom. *TESOL Quarterly, 37*(2), 247-273. https://doi.org/10.2307/3588502

Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. University of California.

Graesser, A. C., Baggett, W., & Williams, K. (1996). Question-driven explanatory reasoning. *Applied Cognitive Psychology, 10*(7), 17-31. https://doi.org/10.1002/(SICI)1099-0720(199603)10:1<17::AID-ACP368>3.0.CO;2-V

Graesser, A., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*(1), 104-137. https://doi.org/10.3102/00028312031001104

Green, A. (1998). *Verbal protocol analysis in language testing research: A handbook*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667329

Hughes, A. (2003). *Testing for language teachers*. Cambridge University Press.

Jackson, G. A. (1982). Public efficiency and private choice in higher education. *Educational Evaluation and Policy Analysis, 4*(2), 237-247. https://doi.org/10.3102/01623737004002237

Karim, S., & Haq, N. (2014). An assessment of IELTS speaking test. *International Journal of Evaluation and Research in Education, 3*(3), 152-157.

Kleinsasser, R. C. (2005). Transforming a postgraduate level assessment course: A second language teacher educator's narrative. *Prospect, 20*(1), 77-102.

Kormos, J. (2011). Speech production and the Cognition Hypothesis. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 39-60). John Benjamins Publishing Company. https://doi.org/10.1075/tblt.2.04kor.

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics, 35*(2), 162-179. https://doi.org/10.1016/j.wocn.2006.01.002

Kunnan, A. J. (2004). Regarding language assessment. *Language Assessment Quarterly, 1*(1), 1-4. https://doi.org/10.1207/s15434311laq0101_1

Kunnan, A. J. (2009). Testing for citizenship: The US naturalization test. *Language Assessment Quarterly, 6*(1), 89-97. https://doi.org/10.1080/15434300802606897.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 33(1), 363-374. https://doi.org/10.2307/2529786

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), 387-417. https://doi.org/10.1111/j.1467-1770.1990.tb00669.x

Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice, and change. *Language Assessment Quarterly, 1*(1), 19-41. https://doi.org/10.1207/s15434311laq0101_3

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524605.

Ling, H. (2006). Long pauses in Chinese EFL learners' speech production. *Interlingüística, 17*(1), 606-616.

Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.

Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing, 18*(4), 351-372. https://doi.org/10.1177/026553220101800402

Lynch, B., & Shaw, P. (2005). Portfolios, power, and ethics. *TESOL Quarterly, 39*(2), 263-297. https://doi.org/10.2307/3588311

Lynch, B. K., & Davidson, F. (1994). Criterion-referenced language test development: Linking curricula, teachers, and tests. *TESOL Quarterly, 28*(4), 727-743. https://doi.org/10.2307/3587553

Madadi, M., & Rezvani, R. (2019). Trodden and untrodden paths: A study of cognitive processes in oral responding and questioning. *Journal of Teaching Language Skills, 38*(3), 159-187.

Mandler, J. M., & Day, J. (1975). Memory for orientation of forms as a function of their meaningfulness and complexity. *Journal of Experimental Child Psychology, 20*(3), 430-443. https://doi.org/10.1016/0022-0965(75)90088-6

McNamara, T. F. (1997). Interaction in second language performance assessment: Whose performance? *Applied Linguistics, 18*(4), 446-466. https://doi.org/10.1093/applin/18.4.446.

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Oxford University Press.

Meichenbaum, D. (1985). Teaching thinking: A cognitive-behavioral perspective. In J. W. Segal, S. F. Chipman, & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 2, pp. 407-426). Routledge.

Michel, M., Kuiken, F., & Vedder, I. (2007). The interaction of task condition and task complexity in the oral performance of Turkish and Moroccan learners of Dutch. In *2nd International Conference on Task-based Language Teaching, Hawaii, USA*. https://doi.org/10.1075/tblt.2.09mic

Moehle, J. P. (1984). Seismic response of vertically irregular structures. *Journal of Structural Engineering, 110*(9), 2002-2014. https://doi.org/10.1061/(ASCE)0733-9445(1984)110:9(2002)(

Ockey, G. J., & Gokturk, N. (2019). Standardized language proficiency tests in higher education. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 1-16). Springer. https://doi.org/10.1007/978-3-030-02899-2.

O'Loughlin, K., & Arkoudis, S. (2009). Investigating IELTS exit score gains in higher education. *International English Language Testing System (IELTS), 10*(1), 1-86.

Park, J. H., & De Costa, P. (2015). Reframing graduate student writing strategies from an activity theory perspective. *Language and Sociocultural Theory, 2*(1), 25-50. https://doi.org/10.1558/lst.v2i1.26546

Pianta, R. C., Hamre, B. K., & Allen, J. P. (2012). Teacher-student relationships and engagement: Conceptualizing, measuring, and improving the capacity of classroom interactions. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 365-386). Springer. https://doi.org/10.1007/978-1-4614-2018-7_17

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-190. https://doi.org/10.1017/S0140525X04000056

Pufahl, I., & Rhodes, N. C. (2011). Foreign language instruction in US schools: Results of a national survey of elementary and secondary schools. *Foreign Language Annals, 44*(2), 258-288. https://doi.org/10.1111/j.1944-9720.2011.01130.x.

Rabbitt, P. M. (1968). Channel-capacity, intelligibility, and immediate memory. *The Quarterly Journal of Experimental Psychology, 20*(3), 241-248. https://doi.org/10.1080/14640746808400158

Read, J. (2009). Identifying academic language needs through diagnostic assessment. *Journal of English for Academic Purposes, 8*(1), 180-190. https://doi.org/10.1016/j.jeap.2008.10.002.

Rezvani, R., & Sayyadi, A. (2015). Instructors and learners' questioning: A case of EFL classroom discourse in Iran. *Journal of Teaching Language Skills, 34*(3), 141-164.

Rezvani, R., & Sayyadi, A. (2016). Washback effects of the new Iranian TEFL Ph.D. program entrance exam on EFL instructors' teaching methodology, class assessment, and syllabus design: A qualitative scrutiny. *Journal of Instruction and Evaluation, 9*(33), 155-176.

Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). Cambridge University Press. https://doi.org/10.1017/CBO9781139524780.012

Seedhouse, P., & Egbert, M. (2006). The interactional organization of the IELTS speaking test. *International English Language Testing System (IELTS), 6*(1), 35-52. https://doi.org/10.1016/j.specom.2007.12.012.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371-388). Oxford University Press.

Simon, M., Ercikan, K., & Rousseau, M. (Eds.). (2012). *Improving large-scale assessment in education: Theory, issues, and practice*. Routledge. https://doi.org/10.4324/9780203801006

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press. https://doi.org/10.1017/CBO9780511734953

Song, D., Oh, E. Y., & Glazewski, K. (2017). Student-generated questioning activity in second language courses using a customized personal response system: A case study. *Educational Technology Research and Development, 65*(6), 1425-1449. https://doi.org/10.1007/s11423-017-9536-1

Tan, Z. (2007). Questioning in Chinese university EL classrooms: What lies beyond it? *RELC Journal, 38*(1), 87-103. https://doi.org/10.1177/0033688206076169

Taylor, L. (2005). Washback and impact. *ELT Journal, 59*(2), 154-155. https://doi.org/10.1093/eltj/cci030

Tocalli-Beller, A., & Swain, M. (2005). Reformulation: The cognitive conflict and L2 learning it generates. *International Journal of Applied Linguistics, 15*(1), 5-28. https://doi.org/10.1111/j.1473-4192.2005.00079.x

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17*(1), 84-119. https://doi.org/10.1093/applin/17.1.84.

Weir, C. J., & Milanovic, M. (Eds.). (2003). *Continuity and innovation: Revising the Cambridge Proficiency in English Examination*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667398

Wendler, C. L., & Walker, M. E. (2006). Practical issues in designing and maintaining multiple test forms for large-scale programs. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 445-467). Routledge. https://doi.org/10.4324/9780203874772

Willis, D., & Willis, J. (2007). *Doing task-based teaching*. Oxford University Press.

Zanov, M. V., & Davison, G. C. (2010). A conceptual and empirical review of 25 years of cognitive assessment using the articulated thoughts in simulated situations (ATSS) think-aloud paradigm. *Cognitive Therapy and Research, 34*(3), 282-291. https://doi.org/10.1007/s10608-009-9255-9