Revista Científica da Universidade Eduardo Mondlane, Série: Letras e Ciências Sociais ISSN 2307-3918

Artigo original

DISSECTING THINK ALOUD METHODS (TAM) (PART I): VALIDITY, REACTIVITY, VERIDICALITY AND RELIABILITY: THE CONCEPT, ADVANTAGES AND THE LOOPHOLES

Manuel Cabinda

Eduardo Mondlane University (UEM), Mozambique

ABSTRACT: This paper delves into one of the most prominent research methodologies, the Think Aloud Methods that has evolved in the field of first language research and branched onto second and foreign language and reading comprehension. Essentially, reviews and revisit the Think Aloud Methods – TAM, on its main aspects and applies TAM in a study and test its validity in an EFL multilingual context in Mozambique. The present paper thus intends to provide a review of the method and additional ground to the understanding of Think Aloud Methods (TAM) and its use for a variety of purposes, exploring the history of the concept and look into the major hurdles that one can have when using TAM, mainly reactivity and veridicality and discuss possible definitions for *reactivity* and *veridicality,* concepts that lack clear and straightforward definitions in the TAM literature so far and thus useful to understand TAM and its use in FL. The ways EFL participants vary in their linguistic competence, their background knowledge relative to a target text and others, and their specific individual experiences in the interpretation of texts is of paramount importance in TAM studies (Smith and King, 2013) and of invaluable importance for us to comprehend the field of reading in EFL.

Keywords: Foreign language learning, reactivity, reliability, second language reading, Second Language Acquisition (SLA), Think Aloud, validity, veridicality.

EXAMINANDO A METODOLOGIA *THINK ALOUD* (PENSAR EM VOZ ALTA) (PARTE I): VALIDADE, REACTIVIDADE, VERACIDADE E FIABILIDADE: O CONCEITO, AS VANTAGENS E LACUNAS EXISTENTES

RESUMO: Este artigo, analisa uma das mais proeminentes metodologias de investigação; os métodos *Think Aloud* (Pensar em Voz Alta) que se desenvolveram na área de investigação de Língua Primeira (L1) e se ramificaram para a Língua Segunda (L2) e Língua Estrangeira (LE) e ainda para a área da leitura e compreensão. Essencialmente revisitam-se os métodos *Think Aloud* e os seus principais pressupostos aplicam-se os referidos métodos num estudo, testando a sua validade num contexto multilingue de Ensino de Inglês como Língua Estrangeira (EFL), nomeadamente Moçambique, explorando a história do conceito e a análise dos principais desafios com que se pode deparar quem se usa os TAM, principalmente no que diz respeito a reactividade e veracidade. No trabalho, também se discute as possíveis definições para o conceito de 'reactividade' e 'veracidade', conceitos que hoje em dia não apresentam definições claras e objectivas na literatura dos TAM e que seriam úteis para compreender melhor os métodos e seu LE. A forma como os aprendentes de Inglês como Língua Estrangeira variam em termos de sua competência linguística relativamente a um texto-alvo e outros, e às suas experiências individuais específicas na interpretação dos textos é de extrema importância nos estudos sobre os TAM (Simth e King, 2013) e de valiosa importância para nós podermos compreender o campo da leitura em Inglês como LE.

Palavras-chave: Aprendizagem de língua estrangeira, reactividade, veracidade, Leitura em Língua Segunda, *Think Aloud,* Aquisição de Língua Segunda (ALS), validade, fiabilidade.

Correspondência para: (correspondence to:) manuel.cabinda@uem.ac.mz or manuel.cabinda66@uem.mz.

INTRODUCTION

The study of Second Language Acquisition (SLA) has over the past decades moved towards new dimensions, from the simple notion of verb understanding and drills to more complex reading comprehension skills and strategies. The works of Sheorey and Mohktari (2001), Grabe (2009) and Bernhardt (2011) are a few examples of such evolution: they bring light to reading strategies and text comprehension in foreign language which have evolved from studies in L1. Further, models which describe the intricacies of reading in foreign language are put forth and they do assist in explaining what are the compensatory aspects of reading in a foreign language. Smith and King (2013) in particular have written about the re-conceptualization of second language acquisition (SLA) in a paper that covers very sensitive aspects of the said evolution such as an acknowledgement of the interaction between cognitively-based theories and socially-oriented approaches and their impact on language learning. The emerging awareness of what they call collective social within a cognitive whole, is claimed to be the one largely responsible for the increasingly modified view of language learners and for re-envisioning the latter as a "national asset" (CASTEK *et al*., 2007), but there are still aspects which bring about doubts as to whether this more expansive and inclusive perspective is as evident as it should be in second language research conducted with certain methodologies, like TAM, which I discuss in the present paper.

The understanding of reading comprehension and the use of Think Aloud Methods, namely verbal protocols (TAM) has trod the same path, from Huey (1804) to Frank Smith (1971) to newer trends explained in works by Ericsson and Simon (1984; 1993), Afflerbach (1990) and more recently Bernhardt (2011) and Smith and King (2013). Several authors have comprehensively reviewed these trends showing how TAM operates and what barriers need to be overcome. I shall resort to this literature as a base and to justify some of my queries and or propositions. The main aim of the present paper is thus to revisit some of the aspects about Think Aloud Methods for a better comprehension and to add to what is already known about this research tool by sharing my insights about it and its use in FL and prepare the ground for a presentation of a pilot study in part II.

The insights, hopefully, may shed light to issues related to the validity of TAM. Ericson and Simon (1984) thoroughly discussed this aspect and issues of reactivity and veridicality of TAM were clarified. Below, these aspects are discussed further. In so many words a verbal protocol is *reactive* if verbalization changes the primary process and this is usually considered consequential because it can invalidate the theoretical conclusions the data were designed to provide.

More recently, verbal protocols have gained increased prominence as a tool for understanding reading processes and have been described as being flexible methodological tools. Yet, it is early days for one to assert that the tool is without any faults. Considering this, as asserted by Smith and King, there is a need for a paradigm shift in the use and interpretation of verbal protocols, generally, and specifically within SLA reading research.

Furthermore, we are trying to look at the tool from two perspectives: the first as described above, essentially to further understanding of its use in FL reading comprehension research in an under-researched context and secondly, in part II, a follow-up paper, the validity of the tool as indicated by a pilot study.

Think Aloud Methods (TAM): an overview

Various terms have been used in the literature to refer to essentially the same research methodology process, the Think

Aloud Methods (TAM) or the Think Aloud Verbal Protocols (TAVP). These terms include 'verbal reports' (AFFLERBACH and JOHNSTON, 1984; CRUTCHER, 1994), 'protocol analysis' (ERICSSON and SIMON, 1980; AFFLERBACH, 2000), 'verbal protocols' (ERICSSON and SIMON, 1980; AUSTIN and DELANEY, 1998), 'think-aloud verbal protocols' (COHEN, 1996), 'thinking-aloud protocols' (ERICSSON and SIMON, 1979), 'think alouds' (DAVEY, 1983; KIBBY, 1997; BLOCK and ISRAEL, 2004) and 'think aloud methods' (ERICSSON, 2002B; JOHNSTONE, BOTTSFORD-MILLER and THOMPSON, 2006; YOSHIDA, 2008). The present study will use the term Think Aloud Methods (TAM), except where otherwise specifically mentioned, as part of a specific study's methodology.

The concept explained

TAM constitutes a rigorous method for eliciting concurrent verbalization of an individual's internal cognitive processes, and to structure the verbalization process so that the verbalization can be utilized as data (ERICSSON, 2002A; ERICSSON and SIMON, 1983). They are also described as constituting a methodology for eliciting verbal reports of thought sequences as a valid source of data on thinking. This methodology has been extensively employed in the fields of psychology and cognitive science as a verbal-report method of producing concurrent verbalization of thought sequences (YOSHIDA, 2008).

The think aloud methods draw on thoughts in the short-term memory of subjects because all cognitive processes that generate verbalizations are a subset of the cognitive processes that generate behaviour or action and travel through short-term memory (Ericsson and Simon, 1993). So the conscious thoughts of the subject can be reported [*concurrently*] at the time they are processed and these verbalizations are claimed to be representative of an individual's cognitive processes at that time (YOSHIDA, 2008).

The premise of the think aloud methods is that individuals may not have conscious access to all of their cognitive processes involved in performing a particular task, and as such no attempt is made to gain access to individuals' internal cognitive processes but rather to elicit verbalizations that are representative of cognitive processes of these individuals that take part in the elicitation process (ERICSSON, 2000, 2002a; YOSHIDA, 2008). TAM requires participants to tell researchers what they are thinking and doing while performing a task (YOSHIDA, 2008).

This process is explained through a simple model of the human cognitive system (VAN SOMEREN, BARNARD and SANDBERG, 1994) which is broken down into three parts: i) the sensory system, "that transforms information from the environment into an internal form;" ii) the long-term memory, where knowledge is stored more or less permanently; and iii) the working memory, where the currently 'active' information resides (VAN SOMEREN *et al*., 1994,p. 20). Van Someren *et al.* (1994) claim that the contents of the sensory system and of long-term memory cannot be verbalized unless these contents are retrieved in some form and stored temporarily in the working memory. Therefore, only contents of working memory can be verbalized through the think-aloud methods. When providing an individual with a specific task, he/she is instructed to say anything and everything that crosses his or her mind, speaking constantly, without consciously filtering what is being said (in so far as that is possible). In this manner, the individual should (introspectively) articulate the appropriate cognitive process(es) involved in performing the given task (COAKSEY, 2000, p. 86).

TAM can also be used after the task has been performed. This process is termed

retrospective verbalization, in which "a subject is asked about cognitive processes that occurred at an earlier point in time" (ERICSSON and SIMON, 1980, p. 218). Retrospective verbalizations are used because think aloud utterances are sometimes incoherent at the moment of task resolution (ERICSSON and SIMON, 1993). So if such incoherence is observed, post-TAM interviews (retrospective TAM), which take place right after the think aloud protocol is completed, or within established intervals, can yield more articulate responses.

To collect verbalizations (introspectively and retrospectively), TAM require participants to tell researchers what they are thinking and doing while performing a task (introspection), and because of this particular aspect, participants are usually directed to keep thinking aloud, and act as if they are alone and speaking to themselves (ERICSSON and SIMON, 1993; ERICSSON, 2002; YOSHIDA, 2008; JOHNSTONE *et al*. 2006). The verbalizations are recorded via a tape-recorder, or videotaped, and then transcribed for content analysis. During the analysis process the data is often coded according to a specific classification, i.e. a set of categories developed by the researcher.

METHODOLOGY

It is rather hard to describe a method when the main purpose of a paper is to review other studies and gather evidence to validate a given issue, i.e. a research methodology. However, one can say that the basis behind the present paper is a review of a collection of papers, studies that have used TAM for various purposes and first, in a Part I, review them to build up a sound basis that support its use in research activities, a tool that can be deemed valid, dependable and one which provides data leading to sound conclusions. Thus a considerable number of papers extending from a wide period were reviewed and evidence gathered to explain its origins, the concept, the type of data collected, the major concerns and above its veridicality, validity and reliability in the field of SLA-FL research.

RESULTS AND DISCUTION

Background on the concept of TAM

The TAM formally came into being in the early 1990s with a book entitled "Protocol Analysis: Verbal Reports as Data" (ERICSSON and SIMON, 1993) which clearly posited think aloud data collection as being a valid method for researching cognitive processes. However, the starting point of this endeavour dates back to the emergence of psychology as a scientific discipline, towards the end of the 19th century, which in turn stimulated interest in issues related to consciousness. At this time, psychologists sought to examine the structure and elements of [the individual's] thoughts and subjective experiences through introspective analysis (ERICSSON, 2002). The use of verbalizations as indicators of cognition is a decades-old data collection technique. Psychologist Karl Duncker (1945) originally described think aloud verbalizations as productive thinking and a way to understand his subjects' development of thought in response to stimuli exercises. With the advent of computational programs a new, rather renewed impetus emerged a shift in the way cognitive research was carried out (ERICSSON, 2006) resulted in a trend which moved from a psychological inquiry focused on observable responses to stimuli to a cognitive focus on the processing involved. This shift provided grounds that assisted in validating data from cognitive research whose methods had been questionable, i.e. the fraternity about the validity of data collected through analytic introspection as a scientific method. For example, reports of changed sequences of thoughts due to the need to explain the process as put by Ericsson (2006), or the criticism around the validity and accuracy

of the retrospective/introspective verbal reports where arguments pointed to be unreliable reports and reports showing reliance of participants on rules that were inconsistent with their observed selection behaviour (VERPLANCK, 1962 apud ERICSSON, 2006) or provision of inconsistent explanations *vis-à-vis* their observed behaviour (NISBETT and WILSON, 1977). As such initial demise on the method prompted developments that led to better methodology to instruct participants to elicit 'consistently valid non-reactive reports of their thoughts' (ERICSSON, 2006, p. 227), (ERICSSON and SIMON, 1993, 2002b, 2006; KUCAN and BECK, 1997).

Experimental psychologists developed standardized tests with stimuli and instructions where the same pattern of performance could be replicated under controlled conditions: they observed and asked the individuals to 'think aloud and give immediate verbal expression to their thoughts while they were engaged in problem solving' (ERICSSON and SIMON, 2006, p. 224).

Because of the considerable controversy largely related to experts' lack of capacity to explain the nature and structure of individuals' performance, the validity of the data was questioned. For example, inconsistent descriptions of the same issue by different experts threatened the validity of results from the studies when, in rare cases, verification of the strategy used by a participant during the performance of the task and/or TAM was allowed and this resulted in different descriptions of the actual action and thought verbalization reporting strategy usage as well as of the observations (ERICSSON and SIMON, 2006). Thus recourse to newer robust methods were necessary and computer-developed methods with more sophisticated programs was used to investigate the performance of challenging cognitive tasks.

With this newly reinvented and purportedly more rigorous and accurate research tools, "think aloud" techniques were redeveloped by Ericsson and Simon (1983), who showed that it is possible to instruct participants to verbalize their thoughts in a manner that does not alter the sequence and content of thoughts mediating the completion of a task and therefore participants should reflect on or verbalize immediately available information during thinking. This new approach to collecting various types of verbal reports of thinking has since become the core method of protocol analysis.

Today, the verbal protocol analysis or, simply, TAM, also known as the Think Aloud Protocols (TAP), is considered a rigorous methodology for eliciting valid verbal reports of thought sequences The group of Think Aloud Methods has evolved into one of the main methods for studying thinking in Cognitive Psychology (CRUTCHER, 1994), Cognitive Science (SIMON and KAPLAN, 1989), and Behaviour Analysis (AUSTIN and DELANEY, 1998). Think Aloud Methods also play a major role in applied settings such as the designing of surveys and interviews (SUDMAN, BRADBURN and SCHWARZ, 1996) through the evaluation of computer designed programs which compare several surveys and perform computational data analysis in the testing of computer software (HENDERSON, SMITH, PODD and VARELA-ALVAREZ, 1995). Ultimately the method has undergone several interesting metamorphoses, for example from a tool in psychology to seek understandings and examine the structure and elements of individuals' thoughts and subjective experiences, to a means of investigating higher order cognitive processes, to its adaptation to suit the study of text comprehension (PRESSLEY and AFFLERBACH, 1995), L1 and L2 and FL reading comprehension and analysis (BLOCK, 1992; BLOCK and ISRAEL, 2004; YOSHIDA, 2008), and test taking

(ALDERSON, 1990). Other fields using TAM are mental translation processes and translation studies (KERN, 1994), triage studies (VAN SOMEREN, BARNARD and SANDBERG, 1994; POMERANTZ, 2004), evaluation of on-line resources for nursing students and education (RENKL, 1997). Other examples of the use of TAM are the recent studies on the effect of computer-based read-aloud methodology on test performance of high school students with learning disabilities (DOLAN *et al*., 2005) and on the issue of reactivity on L2 acquisition. Reactivity is defined as the possible changes triggered in learners' cognitive processes by the act of thinking aloud while they are performing the task (LEOW and MORGAN-SHORT, 2004; BOWLES and LEOW, 2005; YOSHIDA, 2008). The latter studies are significant for FL language learning and reading because they have helped researchers in SLA to observe the cognitive processes involved and have been used in reading, writing (reactivity to TAM), testing, language acquisition, discourse analysis, as well as issues related to attention and awareness in the writing process.

Think aloud methods have given SLA researchers information about the types of strategies learners apply in L2 tasks, for example Alanen (1995), Leow, (2001b), Rott (1999) on discourse analysis, and Leow (1998a, 1998b, 2000, 2001a) and Rosa and O'Neill (1999) for problem-solving tasks. However, such studies have not to date dealt empirically with reactivity, an issue of particular relevance to research and which I will discuss in part II. I have delved into this and other validity issues, i.e. key features of TAM, to further expand on how these methods (may) contribute to our understanding of the less visible variables in the language learning and reading process in SLA and FL.

Introspective and retrospective verbalizations

A description of two possible relationships between cognitive processes and verbalizations is given above: introspective or concurrent verbalization, and retrospective verbalization. As defined earlier, *introspective* verbalized data should be understood as information verbalized at the time the subject is attending to a task, while *retrospective* verbalized data should be understood as data collected after the task has been completed and usually in the absence of concurrent data. In this process "a subject is asked about cognitive processes that occurred at an earlier point in time" (ERICSSON and SIMON, 1980, p. 218).

In addition to categorizing verbal reports as introspective or concurrent, Ericsson and Simon (1984, 1993) made a distinction between reports that ask participants to verbalize their thoughts only and those that ask participants to verbalize additional information such as explanations and justifications for their thoughts. Following Bowles and Leow (2005), I will refer to the verbalization of thoughts per se as *non-metalinguistic*, and verbalization of explanations or justifications as *metalinguistic*.

From the study by Johnstone *et al*. (2006) I inferred an additional type of data which has not been explicitly mentioned or discussed in any detail in most of the literature on TAM. These data are different from the post think aloud verbalizations yielded retrospectively by participants in a study, for they are processed by the participants themselves in *written* form. For example, when the students in the Johnstone *et al.* (2006) study completed an item, they were asked non-scripted follow-up questions based on events that arose during the think aloud verbalization for clarification. Participants were asked process questions such as "How did you solve that?" (when the student did not adequately verbalize) and questions or prompts such as "Was there anything that confused you?" (When a student spent several minutes on a sub-

section of an item) and they were required to produce metalinguistic responses in written form showing how the problem was solved. There is some resemblance with retrospective data collection (ERICSSON and SIMON, 1979, 1980, 1984, 1993; PRESSLEY and AFFLERBACH, 1995; ERICSSON, 2006) but there is no mention of written material by these authors. However, these process questions can help the researcher to collect data that could otherwise be lost forever from participants such as occurred with those in the Johnstone *et al*. (2006) study. Nevertheless, there is also the danger of collecting biased data given that participants may remember to add and/or may omit information or even over-describe the process given that they have the freedom to write about the process. I deal with this issue when I talk about advantages and disadvantages of TAM below. Table 1 summarises the three kinds of TAM.

TABLE 1: Types of data yielded from TAM

Type of Data	Example
Introspective/ concurrent	Student [subject or participant] thoughts as they attempted to solve items
Retrospective	Student [subject or participant] perceptions of solving items after they were completed
Process	Student [subject or participant] written material that demonstrates problem-solving process

(adapted from JOHNSTONE; BOTTSFORD-MILLER and THOMPSON, 2006)

Advantages and disadvantages of TAM in SLA and FL Reading Research

Amidst so many aspects, theories, facts that have been written about this tool, the advantages of TAM as a research method to collect data in SLA and FL reading research can be explained through four main aspects, as follows:

First, a key advantage claimed for concurrent verbalizations is that neither participants' thought processes nor their task performance are changed (ERICSSON and SIMON, 1993; ERICSSON, 2006, p. 228). The level of accuracy of performance is not altered during think aloud methods even when the performance is compared to that of other individuals who complete the same tasks silently (ERICSSON and SIMON, 1993).

A second advantage is that most non-visible and audible processes can be verbalized concurrently and/or retrospectively by participants when performing a task. Data collected according to this method is deemed a true and immediate representation of individual cognitive processes. In this context TAM have provided language acquisition researchers with information as to the types of strategies employed by learners when interacting with L2 tasks (YOSHIDA, 2008). In particular, verbalizations have generated insights as to what types of input induce most noticing, and what types of cognitive processes can be accessed by particular types of verbal reports. As reading is normally a silent, hidden process, and researchers cannot determine with any accuracy what is happening in cognitive terms by simple observation or by product-based assessment (YOSHIDA, 2008), asking readers to provide verbal reports or protocols on their reading process, whether retrospectively or concurrently (DOMINORSKY, 1998; ERICSSON and SIMON, 1980, 1993), becomes the most direct and suitable way to access this process. Furthermore, Think

Aloud Methods provide detailed descriptions of task-induced reader behaviours and complexity in reader's thoughts (AFFLERBACH, 2000).

A third advantage that is claimed for the use of TAM in cognitive strategy research (when compared to interviews) is that they cancel out the time gap between processing and reporting, i.e. readers can report their thoughts while simultaneously being involved in the target task (Yoshida, 2008) and, as has been mentioned, it is claimed that this does not affect task performance when they are engaged in concurrent verbalization (ERICSSON AND SIMON, 1993);

A fourth advantage of TAM is the sample size. In TAM the data can come from a small sample of participants. Additionally, unlike large questionnaire or psychometric research projects, TAM samples are not selected randomly; they are purposive and representative of particular subsets deemed important to the project (KOPRIVA, 2001). Moreover, unlike other methods such as strategy questionnaires, interviews (structured and semi-structured with open and closed ended questions), eye-movement indices, and oral reading, TAM gives almost total freedom to the participant to verbalize his/her thoughts and she/he is only restricted when long pauses are observed and the researcher provides prompts for more verbalizations without posing a direct question. The richness of language generated in this process (or lack thereof) of verbalization 'are the greatest assets and liabilities of the verbal reporting methodology' (PRESSLEY and AFFLERBACH, 1995, p. 2).

Within reading research, TAM have been used to study reading processes to find out how readers engage in a variety of literacy activities, how readers of varying abilities adjust to different types of text (PRESSLEY and AFFLERBACH, 1995), for evaluating test design and its effects on student test-taking processes, student understanding of constructs, student skill level, relevance of items to student life experience, and relevance of items to content taught (KOPRIVA, 2001). Another set of studies investigated construct fidelity, potential bias, possibilities for accommodation, comprehensibility of instructions, general comprehensibility, readability, and legibility of items (THOMPSON, JOHNSTONE, and THURLOW, 2002) to aid test producers in understanding how test design affects student performance of tasks in reading exercises. Results from these studies and those by Ericsson and Simon (1981, 1983, 1984, 2002) have helped in identifying a set of setbacks, for example, level of data accuracy, that need to be understood to validate TAM as a research tool.

The disadvantages of TAM in terms of the level of accuracy of data can be grouped into six main areas: time on task, researcher effect, access to short term memory, cognitive load, recording and transcribing time, and lack of clear steps to transcribing data. Since some of the above issues overlap with aspects linked to veridicality and reactivity, issues of veridicality and reactivity are dealt with separately in the next section.

First, a few recent studies have shown that in contrast to earlier assertions in research by entities such as Ericson and Simon, Smith and King, for example, mention some participants who think aloud to 'take somewhat longer to complete the tasks – presumably due to the additional time required to produce the overt verbalization of the thoughts' (ERICSSON, 2006, p. 228). As a matter of fact TAM verbalizations are time-consuming and labour-intensive when participants work for an hour or hours to verbalize their thought processes and this may slow down the process of task completion. It should however be stressed here as do Ericsson and Simon (2002) that the act of verbalizing subjects' thought processes does not change the sequence of the thoughts per se, and

being this the case, no subjects' task performance should change when TAM is used. Ericsson and Simon (1993) comprehensively reviewed a dozen studies and found no evidence that the sequences of thoughts (accuracy of performance) were changed when subjects thought aloud as they completed the tasks, compared to subjects who completed the same tasks silently.

Second, in relation to researcher effect, there is a potential danger in terms of data accuracy or value in both concurrent and retrospective verbalizations in the instructions given by the researcher to a participant to explain the reasons behind the resolution of a problem and description of the content of thought. These additional instructions and/ or questions (***Wh***-questions) are reliably associated with changes in the accuracy of observed performance (ERICSSON and SIMON, 1993).

Third, here the major concern relates to accessing short-term rather than long-term memory and to cognitive and linguistic loads. Although it is not easy to collect data from the short-term memory, this is seen as preferable because thoughts generated from the long-term memory are often affected by participants' perceptions. Ericsson and Simon (1993) argue that once information enters the long-term memory, participants may incorrectly describe the processes they actually used at the time to respond to a task while verbalizations that take place concurrently with cognitive processes are to a large extent free from interpretation by participants (VAN SOMEREN, BARNARD and SANDBERG, 1994). However, obtaining data in real-time can be a dilemma for the researcher due to incoherent utterances (ERICSSON and SIMON, 1993). Because informants know more than they can tell (NISBET and WILSON, 1977), it is crucial to exercise care when concurrent thought verbalizations are being recorded: the researcher must make sure that all incoherent concurrent verbalizations are noted and that attempts at eliciting better verbalizations are made through retrospective data collection. In addition, the researcher must at all costs avoid interrupting and/or asking many questions, attempting instead to prompt participants using neutral cues. More articulate responses can generally be drawn from interviews which take place after the think aloud protocol is completed, i.e. retrospective data collection.

Fourth, a major concern is that the cognitive load of problem solving and speaking simultaneously may be too great for some subjects (BRANCH, 2000). The use of retrospective data collection can mitigate the impact of this problem, and the use of post-process questions with such participants can also provide valuable information, which may facilitate the interpretation and understanding of the data (BRANCH, 2000).

In view of the above, a two-step TAM process appears to be a practical one to handle the data collection: researchers may first collect data in real time, probing participants as infrequently as possible to avoid distraction during problem-solving activities (Ericsson and Simon, 1993). When faced with moments of silence that last for several seconds (considerable enough to be deemed long), the researcher may then prompt the participant to "keep on talking" without any direct or indirect questions such as, "what are you thinking?" or "tell me about X or Z". The purpose of neutral prompts is to encourage the participant to continue verbalizing aloud his/her thoughts and not, for example, to add ideas external to the thought processes of the participant. Researchers can pose follow-up questions once the thought verbalization is finished. The answers to these questions supplement any unclear data but are not necessarily deemed to be the primary data source (BRANCH 2000).

One other issue linked to data treatment is the time that recording and transcribing verbal protocols entails and the lack of clear steps to follow concerning the transcribed data to be analyzed (WHITNEY and BUDD, 1996, p. 344). Failure to electronically record the data may result in the invalidity of the data: a time lapse may corrupt the evident actual knowledge, assertions, and observations noted at the time of the TAM and result in several inadequate or inaccurate interpretations. An additional potential danger in terms of distorting the data is that data from one researcher, if not video or audio-taped, cannot be considered valid if used by another researcher. A number of researchers' individual coding, marking, etc. leaves this type of data open to endless possibilities of interpretation - one the few loopholes and points of concern in TAM.

TAM and the lingering points of contention

Although TAM as been largely discussed, there are still some points of contention and these are related to the retention of full description of the cognitive process that occurs, how conscious and automated the process is and the issues regarding the oral capacity and competence to express one's thought in a different language without major gaps.

A key point of contention when using TAM relates to the extent to which verbal protocols provide a full picture of cognitive processing (NISBET and WILSON, 1977). This appears to be a particular threat when the text being read is "so easy that reading activities are automatic and inaccessible to verbalization" (YOSHIDA, 2008, p. 200). A related problem noted by Leighton (2004), cited by Johnstone *et al*., (2006), is the difficulty of obtaining meaningful data from items that are too challenging for participants. So both overly simple and overly complex texts and tasks can in their own way hamper the provision of what could have been the real construction of cognitive process of the studied individual. A further criticism of TAM is that processes observed with the use of thinking-aloud are limited to conscious and automatized processes (YOSHIDA, 2008; SMITH and KING, 2013). There are some processes that readers may not be aware of or do not attend to while thinking aloud and thus cannot be reported. Consequently, data resulting from the process of elicitation are deemed to be incomplete reports, and poor reflections of cognitive processing (NISBET and WILSON, 1977; ERICSON, 1983). To counter this possibility, it is advisable to have post TAM interviews to elicit as much data as possible.

An additional concern which is of particular significance in bi- or multilingual contexts has to do with the capacity of participants to express themselves and the differences in the linguistic and speaking competences of individuals (SMITH and KING, 2013). Those with well- developed language skills in both the target and language of instruction will provide different, if not more intelligible, accounts of the task than others with language shortcomings and thus their perception of a task and the way they perform it may result in faulty or inaccurate reporting. However, faulty or incomplete reporting can also be the result of frequent interruptions – in cases where the researcher prompts at inappropriate moments or may ask *wh*-questions, and in dissonance with the instructions of TAM use, and a consequent heavy cognitive load (SELIGER, 1983; STRATMAN and HAMP-LYONS, 1994) as mentioned above.

To conclude, as Pressley and Afflerbach (1995) commented in the 1990s, despite the controversies surrounding it, 'think aloud methodology is still maturing with much interesting work already accomplished and considerable work to be done' (PRESSLEY and AFFLERBACH, 1995, p. 1). In fact since then TAM has been increasingly deemed a valid research tool by cognitive researchers and practitioners despite issues

raised around veridicality (SMITH and KING, 2013). In the next section I look at issues concerning the reactivity and veridicality of TAM.

Reactivity and veridicality of verbalized reports from Think Aloud Methods

Despite criticisms of Think Aloud Methods in the 1980s and 1990s, such as those mentioned above, TAM's popularity continued to grow during that period (COHEN, 1996). Nevertheless, there have been and continue to be ongoing concerns with the veridicality and reactivity of verbal protocols.

In my search to understand the terms *reactivity* and *veridicality*, I was baffled by the absence of straightforward definition in the literature. The few I found revolve around research studies in L1 (ERICSSON and SIMON, 1979; 1980; 1981; 1993), hence the need to see what happens in L2 research.

Reactivity had been long proposed by Ericsson and Simon (1993) and also appears in earlier work on TAM in a subtle manner. I found, in the course of my quest a possible definition of reactivity, namely 'the impact verbalizations may have on the way participants handle tasks, the time it takes them to carry out tasks, and their eventual success in task completion' (MAAIKE, MEMMO and SCHELLENS, 2003, p. 339) and the effect this has on second language reading. The impreciseness of the consequences and/or real impact does contrast with what most studies and reviews of TAM have yielded.

For instance, Ericsson and Simon (1981, 1983, 1993) had argued that the TAM methodology is a valid one for research and that the data yielded does not interfere with participants' cognitive and reading processes while they are engaged in resolving problems/tasks, as I have mentioned above. Although, as already mentioned, frequent interruptions or a heavy cognitive load have also been claimed as sources of possible changes and disruption of learners' cognitive and reading processes, resulting in incomplete reporting (SELINGER, 1983; STRATMAN and HAMP-LYONS, 1994, apud YOSHIDA, 2008); these factors are not seen by researchers such as Ericsson and Simon (1981) and by more recent cognitive researchers as being of major concern (ERICSSON and SIMON, 2006; YOSHIDA, 2008; SMITH and KING, 2013). Ericsson and Simon (1981) claimed that, 'in a review of studies, mostly in L1, comparing subjects thinking aloud with subjects performing the same tasks silently, we found no differences in such measures of cognitive processes as success rate, methods employed, or speed of performance, […]' (ERICSSON and SIMON, 1981, p. 3).

There are, however, some exceptions to the findings mentioned above. For instance, 'in tasks where subjects used non-verbal codes in their thinking', that is, in tasks with a large visual perceptual component, performance 'was slowed down somewhat in the verbalizing conditions' (ERICSSON and SIMON, 1981, p. 3). Despite this, reactivity in terms of speed, time and successful task completion does not seem to be a major problem in TAM. Overall, Ericsson and Simon (1979, 1980, 1981, 1993) found no reactive effects for Think Alouds in L1 research.

Reactivity can then be explained better, perhaps, if one uses what Leow and Morgan-Short (2004) and Bowles and Leow (2005) have defined as the act of thinking aloud potentially triggering changes in learners' cognitive processes while performing the task. The term '*potentially'* used by Bowles and Leow contrasts with terms and or expressions like *'may have'* and *'their eventual success in task completion'* used by Maaike et al (2003, p. 339). As can mentioned above, the definitions revolved around L1 and as such Leow and Morgan-Short (2004) and Bowles and Leow (2005) call for a greater

understanding of reactivity effects of thinking-aloud on L2 reading comprehension and processes, especially those related to the type of task readers engage in while reading –the aim of the Part II of this paper. The need to fully investigate these impacts is crucial for L2 and or EFL studies, so that, perhaps, we may be geared towards minimizing the adverse effects over the results of research actions using TAM in L2.

As for defining veridicality, one can say that this appears to be associated with validity and the 'probability that processes underlying behaviour may be unconscious and thus not accessible for verbal reporting …' and to the 'possibility that verbalizations, when present, may not be closely related to underlying thought processes' (ERICSSON and SIMON, 1993, p. 109). Thus, in terms of veridicality in a cognitive process study, certain factors may be said to come into play to the disadvantage of TAM. One of these factors is the automation of processes which do not often allow for the provision of a full picture of the cognitive process especially when the text being read is of a low degree of complexity thus resulting in the inaccessibility of verbalizations for the readings are so automatic and subjects just *don't really make an effort* when resolving them and as such don't see the need to remember to verbalize the though process need for TAM (YOSHIDA, 2008).

Another veridicality issue is linked to the limited nature of verbalizations observed using TAMs; verbalizations are limited to the conscious processing of tasks that the participants can verbalize (LYONS, 1986; LEOW and MORGAN-SHORT, 2004; YOSHIDA, 2008). This automaticity is coupled with another factor, that of a hidden automated process that participants do not tend to and/or do not report, resulting in incomplete data and having the potential to reflect poor cognitive processing of tasks (LYONS, 1986).

As mentioned above, the relative ease with which participants are able to verbalize their processes, or not, due to their level of language skills development, is another factor to be borne in mind. Because individuals develop oratory skills differently, depending on various individual factors and histories of language socialisation, they tend to provide different levels of reporting on thoughts and cognitive processes, and this may be exacerbated by their individual capacities to perceive the task accurately and to perform it successfully. Aspects of gender, personality, social milieu, and previous experience also play a role in the ways in which individuals perceive a task and their reporting on their cognitive processes. The provision of verbalizations is thus not immaculate or infallible, and as such can produce degrees of faulty or distorted data.

In the following section I discuss key issues with regard to reactivity and veridicality in second or foreign language research.

Reactivity and Veridicality in SLA and FL Research

Yoshida (2008) sees SLA research as having benefited from TAM over the past few decades. He reports that TAMs have been used in SLA to 'observe the cognitive processes involved in the use and acquisition of language' (p. 199) and that the major SLA areas in which think-alouds have been extensively utilized are Reading, Writing, and Testing, Language acquisition, Discourse research, and research on attention and awareness. Turning to the issue of reactivity in L2 acquisition, and the act of thinking aloud potentially triggering changes in learners' cognitive processes while performing a task, results of recent studies (LEOW and MORGAN-SHORT, 2004; BOWLES and LEOW, 2005) seem to be in line with the initial quest posed by Ericsson and Simon (1981): whether the assumption of non-reactivity is applicable to tasks in SLA.

In order to answer this question, I resort to Leow and Morgan-Short's (2004) study, where there is clear mention of TAM being used in SLA research to observe the cognitive processes taking place in the use and acquisition of language. There is also clear mention of other major variables in SLA that have been studied using TAM, for example, linguistic competence, reading strategies, background knowledge, etc. Think-aloud methods have provided language acquisition researchers with information about the types of strategies employed by learners when interacting with L2 tasks, the types of input that induce most noticing strategies and skills in reading comprehension on the part of participants, and the types of processes that can be predicted by a particular type of verbal report. However, the question that needs to be asked is whether reactivity has been clearly addressed in such studies. The answer to this is inconclusive at this stage.

The reactivity effects of thinking-aloud on L2 reading comprehension and processes have not as yet been fully investigated in terms of the specific types of tasks or cognitive processes in which second or additional language readers engage while reading. One relatively recent study by Bowles and Leow (2005) that addressed this issue investigated the differential effects of types of verbalization (non-metalinguistic and metalinguistic) with more advanced language learners. The results showed that there was no significant reactivity given that none of the think-aloud protocols caused reactivity in general, but that metalinguistic verbalizations, concurrently, appeared to cause a decrease in text comprehension (BOWLES and LEOW, 2005). While these findings gave rise to speculations by these scholars pointing to the idea that reactivity varies according to task type, text variables, and individual differences, Yoshida (2008) calls for further research to clearly determine the veracity of these findings and speculations.

Overall, verbalizations from thinking aloud have not been conclusively found to lead to a reliable change in the cognitive process, specifically with regards to the accuracy of response to any given task. Consequently, there is 'no empirical evidence that the kinds of reports [above mentioned] will fail to reflect what the subject is actually heeding or has just heeded' (ERICSSON and SIMON, 1981, p. 5).

Another issue in the context of reactivity relates to individual linguistic competence and the ease with which individuals are able to verbalize their thoughts. Participants' perceptions about the task may also differ and, as has been mentioned, gender, personality, and previous experience are other variables with which researchers need to engage to better and more fully comprehend what contributions these make to the verbalization process, and to use this knowledge to improve SLA learning and teaching processes. Cabinda (2013, 2014) found that a possible misinterpretation of instructions (language competence) might have influenced the outcome of text comprehension. Moreover, female participants performed rather better than their counterparts. Some of these aspects are explored further in Part II.

For EFL academic learning contexts therefore, TAM seems to offer the potential to illuminate reading processes, the usage of reading skills and strategies and the hidden actions that occur while reading, task resolution in construing meaning from text, and even the hidden reactions of learners to task taking. However, care needs to be exercised in the use of TAM in this context, for as Ericsson (2006, p. 228) cautions, when participants explain why they are selecting actions or have to describe carefully the structure and detailed content of their thoughts, they 'are not able to merely verbalize each thought as it emerges, they […] engage in additional cognitive processes' that result in the generation of thoughts that match the 'required explanations and descriptions', but at the

same time can result in changes to their thought sequence. In my experience this blurs most results of the initial propose of using TAM in EFL given the language problems linked to threshold and oratory competencies. Most of the actual thought disclosure fades away and participants tend to invent so they do not lose face.

With regard to the issue of veridicality, SLA research has been grappling with issues of whether veridicality of retrospective data is trustworthy or not. Data retrospectively collected has been questioned as representing a true reflection of the cognitive processes applied by a participant verbalizing his or her thoughts at the time of taking the task (LEOW and MORGAN-SHORT, 2004, p. 49). These authors have battled to provide evidence and convincing arguments for the reliability or veracity of retrospective data and the question continues to linger; retrospective protocols cannot as yet be seen as being accurate reflections of cognitive processes employed by participants while interacting with L2 data. Leow and Morgan-Short call for more combined research on veridicality, validity, reactivity and reliability involving L2 participants, where stimulated recall procedures can be used in an effort to document learners' cognitive processes while engaged in a previous L2 interaction. Their plea is echoed by Johnstone *et al*. (2004), Gass and Mackey (2000), and Leow (2002) who have pointed to the existence of memory decay or double-input exposure as variables in need of exploration, and thus the need for more empirical research on veridicality to confirm assumptions and claims made for it in the SLA field. For instance, a concern associated with the adequacy of the language skills of the participants to verbalize thought processes may result in inaccurate verbalizations in both concurrent and retrospective processes and this issue is particularly important where TAMs are used in second or foreign language reading; it is difficult to decide with any certainty whether a problem with verbalization is a reading task-related or a processing problem, or even a language needed for verbalization of a problem. The issue of text and/or task familiarity or simplicity can also be problematic in the sense that participants can guess solutions (to easily solvable issues) and appear to accurately verbalize their reasoning behind the resolution. However, care should be taken not to confuse their reasoning behind the resolution, as this may not be a true reflection of the entire, or part, of the cognitive process that may have occurred.

Recently, Smith and King (2013) expanded the discussion concerning veridicality and present an analysis in a review in which they highlight in summarized form the issues which cognitive researchers need to consider carefully when using TAM as a research tool. For instance, they call to the attention of researchers the recommendations related to the veridicality of verbal protocols made in the 1980s and 1990s by researchers such as Ericsson and Simon (1984, 1993). In their review Smith and King (2013) show evidence indicating the failure of researchers to slow down processing, to consider variations in participants' verbal abilities within interpretations of the data, and to predict the probable contents of participants' self-reports (SMITH and KING, 2013, p. 715). They alert researchers to the fact that the failure of researchers and theorists to take into account and observe the above within a cognitive framework may result in 'protocols with embedded erroneous data' (SMITH and KING, 2013, p.716) resulting in problems of veridicality.

Fundamentally in their discussion, Smith and King (2013) show that TAM is still a valid data collection tool but alert researchers using it to three fundamental arguments related to the presence of non-veridicality in verbal protocols of language learners (LLs), in which the assumption is that veridicality of verbal reports is present when verbal output matches mental operations, and the contrary, when this is

not the case. In both these contexts non-veridicality is theorized to stem from two major types of errors involved in the data elicitation process, errors of omission and commission (RUSSO, JOHNSON and STEPHENS, 1989), i.e. errors related to the absence of matches between verbal output and mental operations disclosed by the learner and an error labelled by Smith and King (2013, p. 715-6) as failing to consider 'the presence of language(s) as an inherent variable'. Given that language(s) are an inherent part of my study and that the participants are multilingual and are communicating within and dealing with a multiplicity of contexts and language competences issues, this aspect is of utmost importance to the study. The issue I need to address is whether I have omitted, or left aspects of language(s) *undealt* with in the process of collecting data. The validity of the issues raised by Smith and King (2013) is crucial for research focusing on validating the effective use of cognitive and metacognitive reading strategies, including support strategies, by FL learners.

Using my study as an example, I have followed some of the recommendations and used TAM and primarily concurrent (introspective) verbalizations, only using retrospective verbalizations when needed for clarification (with one participant) as per Ericsson and Simon (1984, 1993). This aided in ensuring rigor and veridicality in my use of TAM. However, the issues related to veridicality and non-veridicality, as touched on in the studies mentioned above, need to be elucidated for them to be of value in any search, via empirical studies. As these processes are interlinked I have thus borne in mind the recommendations for maintaining rigor and veridicality in the use of TAM data collection made by Ericsson and Simon (1984, 1993).

In what follows I draw on critiques of these recommendations by Smith and King (2013, pp. 711-715) and highlight issues of fundamental importance for EFL.

a) ***Increase Representativeness of Thought through Concurrent Protocols.*** Ericsson and Simon (1993) recommend the use of concurrent protocols and reports based on verbal cognitions to augment the possibility of deriving protocols with reflections of thought process verbalized by participants. I have, in accordance with several studies mentioned by Smith and King (2013), paid primary attention and given weight to concurrent verbalizations. This is so for multilingual EFL students are more prompt to disclosure thought process even when dealing with more than one language, i.e. three, when using concurrent method and several aspects of the target languages can be examined, for example, vocabulary-knowledge as participants derive meaning for unknown words, language-oriented strategies/content-oriented (based on orientation of processing), regulatory/cognitive/cognitive-iterative strategies (based on type of processing), and above-clause/clause/below-clause (based on domain of processing) strategies. There is a wide room for learners to compensate for the absence of linguistic knowledge or processing ability with concurrent data collection process and this can be used as a means of increasing representativeness of verbal protocols, (ERICSSON and SIMON, 1984; 1993).

b) ***Slow Down Processing.*** Ericsson and Simon (1993) clearly underline the importance of slowing down automatized processes. 'End-of-paragraph prompting' for verbal protocols is essential in order to sufficiently interrupt otherwise automatized processes. Smith and King (2013) do not see this as

interfering with the concurrent verbalization process, although Ericsson and Simon (1984, 1993) recognize that fully automatic processes such as reading are hard to self-report and thus recommend the use of retrospective protocols without seeing this as representing a contradiction of their initial recommendation (use of concurrent verbalizations), where participants have to specify their thoughts in response to the specific signal which had previously interrupted the automatic process (i.e., reading). However, a researcher following a concurrent verbalization process needs to be aware of deliberately not encouraging participants to 'provide descriptions or explanations of their processing' (ERICSSON and SIMON, 1993, p. 109). In accordance with other SLA studies mentioned by Smith and King (2013), in terms of this particular recommendation I recognized the importance of participants' slowing down the automated process of reading, as well as their intent to preserve comprehension through the use of complementary protocol formats, i.e. there is a need to aid he already complex and hard reading process of EFL learners but the slowing down process means not to interrupt the LL is the middle of the paragraph but at the conclusion of the paragraph, not at sentence or word level. A researcher would be more likely to tap comprehension as a completed product and less likely to intercept comprehension as a process in this manner; the use of immediate retrospection with at least with one participant, as recommended by Nassaji (2003), Upton and Lee-Thompson (2001), Wesche and Paribakht (2000), and other research methods (SMITH and KING, 2013, p. 713) is crucial with EFL LLs to verify if comprehension process was effective or not.

c) ***Emphasize Process over Product.*** There exists a potential for researchers using verbalizations to collect and process data to give prominence to the products of cognitive or thought processing rather than to the awareness on the part of participants about their own thought processes (ERICSSON and SIMON, 1984; 1993). In this context, as Smith and King (2013) demonstrate, many studies, even those which are fairly recent, have been product-oriented (DRESSLER *et al*, 2011; GASCOIGNE, 2002; NASSAJI, 2003; PARIBAKHT, 2005; LEE-THOMPSON, 2008) and involved products and/or tasks that were inclusive of drawing inferences, answering questions, and retelling. Recent studies more consistent with a process-oriented approach, and which would seem to have heeded Ericsson and Simon's (1984; 1993) cautioning regarding prominence of product over process, have investigated reading difficulties and cognitive and metacognitive strategies deployed by bilingual students while reading. These studies include Alsheikh (2011), Geladari *et al*. (2010), Stevenson, Schoonen and de Glopper (2007), Upton and Lee-Thompson (2001), Wesche and Paribakht (2000), Yang (2006) and Zhang, Gu and Hu (2007) and Smith and King, 2013, p. 714). These studies focus on how studied SLA participants

understood the meanings of words and employed reading strategies for their understanding of text. According to Smith and King (2013), the danger inherent in giving prominence to product rather than process is that, for the participants there is a 'greater likelihood that the verbal protocols would reflect the anticipated task rather than be a representation of their awareness of the ongoing reading process' (SMITH and KING, 2013, p. 174). However, as asserted by the authors, Ericsson and Simon (1984; 1993) have not explicitly stated that process-oriented tasks would place a greater burden on the participant to report the process. And as such, Smith and King (2013) hypothesize that this should have been the case.

In conclusion, they posit that 'the research tasks should be geared towards maximizing the probability that the verbal protocols obtained during the reading process would be most representative of that participant's processing, and, therefore, process-oriented studies would more than likely be the norm than would those with product-influenced protocols' (Smith and King, 2013, p. 714). This was the key aspect informing my study, and the core of the aim of the study: to establish whether reading is to be understood as a process or a product. My study has focused on the former but without discarding the latter.

d) ***Tap Current Processing***. Attention is drawn here to the need for researchers not to consciously solicit participants to provide a generalized description of their processing across trials. In this context Smith and King (2013, p. 715) warn of the 'possibility that conscious attention would be placed only on operations involved in earlier trials of the verbal reporting process' and observe a 'general adherence to this recommendation' by all but one study, Wesche and Paribakht (2000) out of the 20 studies they reviewed. On the basis of these studies they considered that non-observation of this cautionary rubric would render the data of dubious veridicality given that the results would likely be affected by participants' predisposition to report information on operations involved in earlier trials of the verbal reporting process of the reading exercise.

e) ***Direct Participants to Provide Non-explanations***. This directive is related to the nature of the directions provided by researchers to participants of a given study that uses TAM, and these directions, as Smith and King (2013, p. 715) put it, 'should be such as to discourage participants from providing descriptions or explanations of their processing as reports of intermediate and final products of processing….'. They support the suggestion that such directions can be open-ended or can be framed to encourage participants to 'report on a specific type of information in their working memory' (Ericsson and Simon, 1984; 1993, pp. 10-11). As descriptions or explanations of cognitive processes constitute certain introspective protocols, as noted by Smith and King (2013, p. 715), researchers should give prominence to concurrent verbalizations as recommended above (see ***A***.), given that these result in data collected as close to real time as possible, and during task completion, and are closest to

actual thought processes. Thus Smith and King (2013) warn of the danger of having non-veridical data if the above concurrent verbalization protocols are not observed. The danger here, with EFL leaners, could be a end product that is not realistic given issues related to insufficient linguistic knowledge, unfamiliar reading topics, misunderstandings of the very same instructions/directives, provision of made up images of the reading and task completion tasks, fantasizing at own pleasure to not lose face, etc. Smith and King (2013, 2013, p. 715) emphasize that researchers should recognize that *directions impact the nature of reports* and that they should be willing to acknowledge this impact on the presentation of their findings.

f) ***Consider Participants' Verbal Abilities to Generate Verbal Protocols***. This sixth recommendation made by Ericsson and Simon (1984; 1993) relates to differences in individuals' abilities to produce think-aloud protocols, and that an aspect to be borne in mind by researchers is the possibility that an increased general verbal ability could provide individuals with an advantage when reporting verbal protocols. The ways participants vary in their linguistic competence, their background knowledge relative to a target text, and their specific individual experiences in the interpretation of texts is of paramount importance in TAM studies (SMITH and KING, 2013) and, as these scholars put it, this applies 'not only with regard to their ability to verbalize, but in relation to their background experiences as individual language learners (LLs)' (SMITH and KING, 2013, p. 715). As has been noted, this issue is a problematic one and Smith and King (2013, p. 715) argue that researchers do not clearly address the language competence of participants as a factor in TAM, and seem to be 'oblivious to the nuances between individual participants as they undertake a myriad of reading tasks'. They argue that researchers using TAM tend to refer to the linguistic status of a participant as either a Spanish, English, Portuguese or French student, for example, or mention his/her current level of linguistic competence according to a rigid and/or narrow classification system. They draw attention to the fact that crucial information such as an individual participant's first exposure to the L1, L1 learning period or experience, languages spoken at home and/or in other countries; language of instruction is not mentioned and these variables in fact significantly affect participants' abilities to verbalize thoughts in conjunction with reading tasks (Smith and King, 2013). Both Bernhardt (2011) and Smith and King (2013) emphasize the importance of taking into consideration in any TAM research the fact that LLs may vary in their origin, come from diverse and multiple language backgrounds, and that their experiences with the language(s) may bear little resemblance to one another, and thus attention should be paid to these variables so that veridicality issues, i.e. unreal or made up accounts of the thought process or task resolution account resulting in inadequate and invalid verbal accounts, are minimized and 'interpretation of the protocol data might be allowed to reflect these

differentiated abilities' (SMITH and KING, 2013, p. 716).

g) ***Predict Study Participants' Self-reports:*** one final recommendation by Ericsson and Simon (1984; 1993) is focused on researchers being able to predict the ability of the participants to self-report while they are completing (or attempting to) a task. In this context great importance is placed upon the researcher's ability to foresee what set of prior knowledge the participant might possess and thus 'anticipate the procedures in which a study participant might engage to arrive at a particular solution to the task parameters' (Smith and King, 2013, p. 716). To do this, the researcher engages in task analysis to define the probable sequential elements of a task that may result in a probable set of possible thought sequences for its successful performance. As Smith and King (2013) observe, there is mention in some studies of expected responses (strategies, inferences) from study participants (examples being the studies by Chun, 2001; Bengeleil and/and Paribakht, 2004; Lee-Thompson, 2008), but none of the studies provided task analysis as 'an indication of the probable and possible sequences to be expected for alternative procedures in a task or a given series of tasks' (Smith and King, 2013, p. 716). Citing Ericsson (2003) with reference to mathematical tasks used for illustrating task resolution sequencing, Smith and King (2013, p. 716) conclude by suggesting the probability of a similar procedure being followed to 'appropriate a method for determining predictability of verbal protocols of reading, in an effort to enhance veridicality'. I would suggest this not to be the case in an EFL context like mine, given the multiplicity of contexts, backgrounds, dialects, age and possibly gender and not to mention the unequal and inequalities in competence levels in various schools providing similar courses. Thus using a mathematical model to determine predictability of verbal protocol verbalizations in reading (in EFL contexts), and enhance veridicality as concluded above could hinder the probability of adequate and true results from TAM. In fact, this is a major concern in this context according to Smith and King (2013, pp. 716-718).

Ultimately, because language as an inherent variable has been neglected in most L2 studies and the dearth of attention to this may be a source of veridicality issues (Smith and King, 2013), there is a need to concentrate on how this variable, language, plays its part in the production of thought disclosure and or verbal protocols of EFL participants compare with L1 learners. Smith and King focus their argument on the value of verbal reports with second language learners and consider them to be the "elephant in the room" issue. They make reference to lack of control of language as a variable in several studies, which seem to use mostly monolingual LLs (examples are reviews by Ericsson and Simon, 1984 and 1993, and Pressley and Afflerbach, 1995). What needs to be borne in mind are issues of the credibility or reliability of verbal reports due to the second language learners' linguistic abilities that may further 'confound representation of memory processes' (SMITH and KING, 2013, p. 716). The complexity of engaging in this process is illustrated by Ericsson and Simon (1984; 1993) who argue that individuals who are fluent in a second language will usually verbalize in that L2 but will be thinking internally in the oral code of their native language or in non-oral code, and as

such there will be (almost) a one-to-one mapping between structures in the oral code of the first language and the code of the second language that is used for vocalization (as cited in Smith and King, 2013, p. 717). Given the multilingual complex context of my study with EFL learners (a range of L1 and L2 with differing levels of competence) I permitted the participants to verbalize in any language, i.e. the target (English) or native (one of the Bantu languages for some) or the lingua franca (Portuguese for all) and certainly I am aware of the constraints that come with this, i.e. the difficult to have a one-to-one mapping of the used oral codes. For instance, as Smith and King (2013, p. 717) point out, 'the challenges inherent in reading and performing a task in a second language (usually English), subsequently conducting interpretation through the native language, and deciding whether to revert back to English or to relay the contents of memory in the native language are significant and do influence the composition of protocols'.

Thus since language is an additional inherent variable in SLA research, and this seems to have been neglected, there is a need to do further research in the field and to propose TAM verbalization procedures and/or trends that may render verbalized data validity levels less questionable given that language dictates the linguistic product of such learners, therefore any attempt to verbalize reports not only undergoes transformation during verbalization, but also experiences alteration due to linguistic interference and, as Smith and King (2013, p. 717) argue, 'the language task required, and the demand to verbalize that task, find themselves competing for the linguistic capacity, ultimately affecting completeness (omission) *and* accuracy (commission) of the verbal protocols' and this is very true with EFL learners in a context such as mine.

CONCLUDING REMARKS

At the beginning of the present paper I set out to review a number of studies with the intention of providing a sound basis to validate TAM; also I had proposed this paper as a means to provide an additional ground to the understanding of Think Aloud Methodologies (TAM), its use in L1 and SLA research for a variety of purposes and explore the history of the concept, by looking at the major hurdles, and issues related to reliability and veridicality.

Upon reviewing a multiplicity of papers and studies that have used Think Aloud Methodologies (TAM) this paper has provided an exploratory journey onto the history, the concept and methodological concerns about TAM and have expanded the definitions for *reliability* and *veridicality*, and underlined some of the major hurdles researchers face when using the tool.

Furthermore, the review has made it possible to present and emphasise several keys issues (resulting from several pioneering studies and the use of the ground-breaking work by Ericsson and Simon, 1981;1984; 1993) and other followers like Yoshida (2008), Bernhardt (2005, 2011), Smith and King (2013) and show that for the EFL context such as mine the following aspects need an added attention: there is a clear need for comprehensible and clear definition of concepts such as reactivity and veridicality so that research in SLA can provide proper and adequate results when TAM is used with EFL learners; the ground-breaking work by Ericsson and Simon (1981, 1984, 1993) on TAM has been crucial in L1 and within SLA research studies and has been proven to be used as an evolutionary tool to the identification of and effective use of reading comprehension strategies, both cognitive and metacognitive, which are more commonly investigated through questionnaires, taxonomies and surveys, especially in EFL contexts where reading comprehension strategies, types of strategies learners apply in L2 tasks,

discourse analysis and problem-solving tasks are yet to be fully comprehended and researched upon to their wide spectrum; theoretical background on TAM in the present paper is a trampoline to a practical use with in my context with Mozambican tertiary learners, given the lack of studies discussing reading strategies of adult learners in tertiary FL multilingual contexts in general and Portuguese speaking countries in particular where learners speak an array of L1 languages, i.e. a diverse repertoire of Bantu languages, and differing linguistic competences, knowledge backgrounds of the target text and language, specific individual experiences in the interpretation of texts and so forth.

The paper has brought to the attention of researchers the emphasis that is placed upon the observation of most of the cautionary rubrics offered by Ericsson and Simon (1984; 1993) and unpacked and discussed by Smith and King (2013) in their review. This paper is one that unpacks empirical studies to consubstantiate most of the theoretical aspects discussed above; one that uses and looks onto data collected via TAM and which upon analysis has provided solid basis for validating TAM, its veridicality and reliability in the field of SLA research. Hence, I invite the reader to delve into the second part of this paper, entitled "Dissecting think aloud methods: towards the understanding of procedural issues, data collection to identify and study reading comprehension strategies in FL multilingual contexts through TAM (Part II)."

REFERENCES

AFFLERBACH, P. The influence of prior knowledge and text genre on readers" prediction strategies. **Journal of Reading Behavior** 22, 131-148,1990.

AFFLERBACH, P. The influence of prior knowledge and text genre on readers: prediction strategies. **Journal of Reading Behavior,** v.22, p. 131-148, 1990.

AFFLERBACH, P. Verbal reports and protocol analysis. In: KAMIL, M. L.; MONSETHAL, P. B.; PEASON, P. D. and BARR, R. (Eds.). **Handbook of Reading Research**. Mahwah: Erbaum, 2000.

AFFLERBACH, P.; JOHNSTON, P. On the use of verbal reports in reading research. **Journal of Reading Behaviour**, v.16, n. 4, p. 307-322, 1984.

AFFLERBACK, PETER. Verbal reports and protocol analysis. In Kamil, Michael L., et al. (Eds.) Handbook of Reading Research. Mahwah, NJ: Erbaum, 2000.

ALANEN, R. Input enhacement and rule presentation in second language acquisition. In R. Schmidt (Ed.), Attention and awareness in foreign language learning and teaching, pp.259-302. Honolulu: **University of Hawai'I Press**, 1995.

ALANEN, R. Input enhacement and rule presentation in second language acquisition. In R. Schmidt (Ed.), **Attention and awareness in foreign language learning and teaching**. Honolulu: **University of Hawai'I Press**, 1995.

ALDERSON, J. CHARLES. Testing reading comprehension skills (Part one). **Reading in a Foreign Language**, v.6, n.2, p.425-438, 1990.

ALSHEIKH, N.O. Three readers, three languages, three texts: The strategic reading of multilingual and multiliterate readers. **The Reading Matrix, v** 11, n.1, p. 34-53, 2011.

AUSTIN, J.; DELANEY, P. F. Protocol analysis as a tool for behaviour analysis. **Analysis, Verb and Behaviour**, v. 15, p.41-56, 1998.

BENGELEIL, N.F. & T.S. PARIBAKHT. L2 reading proficiency and lexical inferencing by university EFL learners. **The Canadian Modern Language Review**, v. 61. n.2, pp. 225-249, 2004.

BEREITER, C.; BIRD, M. Use of thinking aloud in identification and teaching of reading comprehension strategies.

Cognition and Instruction, v. 2, n. 2, p. 131-156, 1985.

BERNHARDT, E. B. Progress and procrastination in Second Language Reading. **Annual Review of Applied Linguistics,** v.25, pp.133–150. 2005.

BERNHARDT, E.B. **Understanding advanced second-language reading**. New York: Routledge, 2011.

BLOCK, C.C. and ISRAEL, S.E. The ABCs of performing highly effective think-alouds. **The Reading Teacher,** v. 58, n. 2, pp. 154-167, 2004.

BLOCK, ELLEN. See how they read: Comprehension monitoring of L1 and L2 readers. **TESOL Quarterly**, v.26, n.2, p.319-343.,1992.

BOWLES, M. A. and LEOW, R. P. Reactivity and type of verbal report in SLA research methodology. **Studies in Second Language Acquisition**, v. 27, p. 415-440, 2005.

BRANCH, J. L. Investigating the information-seeking processes of adolescents: The value of using think-alouds and think afters. **Library and Information Science Research**, v. 22, n. 4, pp. 371–392, 2000.

BRANCH, J. L. Investigating the information-seeking processes of adolescents: The value of using think-alouds and think afters. **Library and Information Science Research**, v. 22, n. 4, p. 371–392, 2000.

CABINDA, M. The need for a needs analysis at UEM: Aspects of and attitudes towards change. **Linguistics and Education**, v.24, p. 415– 427, 2013.

CABINDA, M. Purported use and Self-awareness of Cognitive and Metacognitive Foreign Language Reading Strategies in Tertiary Education in Mozambique. **Afrika Focus**, v. 29, n.1, 2016 – pp25-47, 2016.

CASTEK, J. ET AL. Developing new literacies among multilingual learners in the elementary grades. In L. Parker (ed.). **Mediated learning environments for young English learners**: Connections in and out of school. Mahwah, NJ: Lawrence Erlbaum Associates, p.111-153, 2007.

CHUN, D. M. L2 reading on the web: Strategies for accessing information in hypermedia. **Computer Assisted Language Learning,** v.14, n.5, pp. 367-403, 2001.

COAKSEY, R. Judgment Analysis, Think-Aloud Protocols, Cause Mapping, Image Theory and Neural Network Simulation. **The Brunswik Society Newsletters**. v. 26, p. 79-122, 2000.

COHEN, A. D. Verbal Protocols as a Source of Insights into Second Language Learner Strategies. **Applied Language Learning**, v. 7, n. 1 &2, p. 5-24, 1996.

CRUTCHER, R. J. Telling what we know: The use of verbal report methodologies in psychological research. **Psychology Science**, v. 5, p. 241-244, 1994.

DAVEY, B. Think aloud: Modelling the cognitive processes of reading comprehension. **Journal of Reading**, v. 27, n.1, p. 44-47, 1983.

DOLAN, R. P. *et al*. Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. **Journal of Technology, Learning, and Assessment**, v. 3, n. 7, 2005.

DOMINOSWSKI, R. L. **Verbalization and Problem Solving**. In: HACKER, D. J.; DUNLOSKY, J.; GRAESSER, A. (Eds). Metacognition in Educational Theory and Practice. Mahwak: LE Ass.,1998. P. 25-45.

DRESSLER, C.; *et al.* Spanish-speaking students' use of cognate knowledge to infer the meaning of English words. **Bilingualism**: **Language and Cognition**, v.14, n.2, pp. 243-255, 2011.

DUNCKER, K. On problem-solving. In: DASHIELL, J. F. (Ed.). **Psychological**

monographs. Washington: American Psychological Association,. P. 1–114, 1945.

ERICSSON, K. A. & HERBERT, A. S. **Protocol analysis**: verbal reports as data. Cambridge, Massachusetts: MIT Press, 1993.

ERICSSON, K. A. & HERBERT, A. S. **Protocol Analysis**: verbal reports as data. Cambridge: MIT Press, 1993.

ERICSSON, K. A. & HERBERT, A. S. **Protocol analysis**. Cape Town: UWC, 2010. (28 December 1981).

ERICSSON, K. A. & HERBERT, A. S. **Protocol analysis**: verbal reports as data. Cambridge: MIT Press, 1984.

ERICSSON, K. A. & HERBERT, A. S. **Protocol analysis**: Verbal reports as data. Rev. Edition. Cambridge: MIT Press, 1993.

ERICSSON, K. A. & HERBERT, A. S. Verbal reports as data. **Psycholinguistical Review**, v. 87, n. 3, p. 215-251, 1980.

ERICSSON, K. A. & SIMON, H. A. **Protocol analysis: Verbal reports as data.** Cambridge, MA: MIT Press, 1984.

ERICSSON, K. A. Concurrent verbal protocols on text comprehension: a review. **Text**, v. 8, n. 4, p. 295-325, 1988.

ERICSSON, K. A. **Concurrent verbal protocols on text comprehension**: A Review Text, vol. 8, no. 4, pp. 295-325, 1988.

ERICSSON, K. A. **Protocol Analysis and Expert Thought**: concurrent verbalizations of Thinking during experts' performance on representative tasks. The Cambridge Handbook of Expertise and Expert Performance, 2006. p. 223-242.

ERICSSON, K. A. **Protocol Analysis and Verbal Reports on Thinking**: an update and extracted version. 2002b. Available at: http.//www.psy.fsu.edu/faculty/ericsson/ericsson.proto.thnk.html. Accessed on: Set. 2016.

ERICSSON, K. A. **Protocol Analysis and Verbal Reports on Thinking: An update and extracted version.** 2002b. Available in: http.//www.psy.fsu.edu/faculty/ericsson/ericsson.proto.thnk.html. Accessed in 11/11/2019

ERICSSON, K. A. Towards a procedure for eliciting verbal expression of non verbal experience without reactivity: Interpreting the verbal overshadowing effect within the theoretical framework for protocol analysis. **Applied Cognitive Psychology**, v. 16, p. 981-987, 2002a.

ERICSSON, K. A. Towards a procedure for eliciting verbal expression of non verbal experience without reactivity: Interpreting the verbal overshadowing effect within the theoretical framework for protocol analysis. **Applied Cognitive Psychology**, v. 16, pp. 981-987, 2002a.

ERICSSON, K. A.; & SIMON, H. A. Verbal reports as data. **Psycholinguistical Review,** v.87, n.3, p. 215-251,1980.

ERICSSON, K. A.; & H. A. SIMON. **Protocol analysis**: Verbal reports as data. Cambridge, MA: MIT Press, 1984/ 1993.

ERICSSON, K. A.; SIMON, H. A. Thinking-aloud protocols as data: effects of verbalization. Unpublished Manuscript. Carneige-Mellon University, 1979. (CIP working paper No. 317).

ERICSSON, K. A.; & H. A. SIMON. Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press, 1984/1993. (Original work published 1984).

GASCOIGNE, C. Documenting the initial second language reading experience: The readers speak. **Foreign Language Annals**, v.35, n.5, pp.554-560, 2002.

GASS, S. and MACKEY, A. **Stimulated recall methodology in second language research**. Mahwah: Erlbaum, 2000.

GELADARI, A. E.; GRIVA; K. MASTROTHANASIS: A record of bilingual elementary students' reading

strategies in Greek as a second language. **Procedia-Social and Behavioral Sciences**, v.2, n.2, pp.3764-3769, 2010.

GRABE, W. & STOLLER, F. L. **Teaching and Researching: Reading.** 2nd Ed. Pearson Education ESL, 2011.

HENDERSON, R. D.; SMITH, M. C.; PODD, J.; VARELA-ALVAREZ, H. A comparison of the four prominent user-based methods for evaluating the usability of computer software. **Ergonomic***s,* v. 39, p. 2030-2044, 1995.

HUEY, E. **The Psychology and Pedagogy of Reading.** New York: Macmillan, 1908.

ISRAEL, S. E. **Understanding strategy utilization during reading comprehension: relations between text type and reading levels using verbal protocols**, 2002. Thesis (PhD, Teachers College, Ball State University, Muncie, 2002).

JOHNSTONE, C. J.; BOTTSFORD-MILLER, N. A.; THOMPSON, S. J. **Using the think aloud method (cognitive labs) to evaluate test design for students with disabilities and English language learners. Minneapolis**: University of Minnesota, 2006.

KARBALAEE KAMRAN, S. Does reading strategy use predict and correlate with reading achievement of EFL learners? **International Journal of Research Studies in Language Learning***,* v. 2, n. 2, p. 29-38, 2013.

KERN, R. G. The role of mental translation in L2 reading. **Studies in Second Language Acquisition***,* v. 16, n. 4, p. 441-461, 1994.

KIBBY, M. W. Thinking Aloud and Reading Comprehension. Center for Literacy and Reading Instruction, 1997. **Univ. Buffalo**. Acessed 11/11/2019. Available at http://wwwreadingcenter.buffalo.edu/center/research/think.html.

KOPRIVA, R. ELL validity research designs for state academic assessments**:** an outline of five research designs evaluating the validity of large-scale assessments for English language learners and other test takers. In: COUNCIL OF CHIEF STATE SCHOOL OFFICERS MEETING, Houston, 2001. **Proceedings**… June 22–23, 2001.

KUCAN, L. and BECK, I. L. Thinking aloud and reading comprehension research: Inquiry, instruction and social interaction. **Review of Educational Research**, v. 67, n.3, p. 271-299, 1997.

LEE-THOMPSON, L. An investigation of reading strategies applied by American learners of Chinese as a foreign language. **Foreign Language Annals**, v.41, n.4, pp.702-721, 2008.

LEIGHTON, J. P. **Avoiding misconception, misuse, and missed opportunities:** the collection of verbal reports in educational achievement testing. **Educational Measurement: Issues and Practice**, v. 23, p. 6–15, 2004.

LEOW, R. P. A study of the role of awareness in foreign language behavior: aware versus unaware learners. **Studies in Second Language Acquisition**, v. 22, p. 557–584, 2000.

LEOW, R. P.; MORGAN-SHORT, K. To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. **Studies in Second Language Acquisition**, v. 26, p. 35-57, 2004.

LEOW, RONALD P. The effects of amount and type of exposure on adult learners' L2 development. **Modern Language Journal**, v.82, p.49–68, 1998a.

LEOW, RONALD P. Do learners notice-enhanced forms while interacting with the L2? An online and offline study of the role of written input enhancement in L2 reading. **Hispania,** v.84, p.496–509, 2001b.

LEOW, RONALD P. Attention, awareness, and foreign language behavior. **Language Learning**, v.51, p.113–155, 2001a.

LEOW, RONALD P**.** Toward operationalizing the process of attention in SLA: Evidence for Tomlin and Villa's 1994 fine-grained analysis of attention. **Applied Psycholinguistics**, v.19, p.133–159, 1998b.

LEOW, RONALD. P. A study of the role of awareness in foreign language behaviour: Aware versus unaware learners. **Studies in Second Language Acquisition**, v. 22, p. 557–584, 2000.

LYONS, W. **The disappearance of introspection**. Cambridge: MIT Press, 1986.

MAAIKE, J. H.; MENNO, D. T.; SCHELLENS, P. J. Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. **Behaviour and information technology**, v. 22, n. 5, p. 339–351, 2003.

MOHKTARI, K. and SHEOREY, R. Measuring ESL students' awareness reading strategies. **Journal of Development Education**, v. 25, n. 3, p. 2-10, 2002.

MOHKTARI, K.; REICHARD, C. Assessing students' meta-cognitive awareness of reading strategies. **Journal of Educational Psychology**, v. 94, n. 2, p. 249-259, 2002.

NASSAJI, H. Issues in Second-Language Reading: Implications for Acquisition and Instruction. **Reading Research Quarterly**, v. 46, n.2, p.173-184, 2011.

NISBET, R. E.; WILSON, T. D. Telling more than we know: Verbal reports on mental processes. **Psychological Review**, v. 84, p. 231-259, 1977.

of Cognitive Processes to Survey Methodology. Jossey-Bass, San Francisco, 1986.

PARIBAKHT, T. S. The influence of first language lexicalization on second language lexical inferencing: A study of Farsi-speaking learners of English as a foreign language. **Language Learning**, v. 55, n.4, pp.701-748 2005.

POMERANTZ, J. Factors Influencing Digital Reference Triage: A Think-Aloud Study. **The Library Quarterly**, v. 74, n. 3, p. 235-264, 2004.

PRESSLEY, M.; AFFLERBACH, P. Verbal protocols of reading: the nature of Constructively Responsive Reading. Hillsdale: Erlbaum, 1995.

RENKL, A. Natural Language Mediation. **Cognitive Psychology**, v. 2, p. 1-56, 1997.

ROSA, E. M.; & O'NEILL, M. Explicitness, intake, and the issue of awareness: Another piece to the puzzle. **Studies in Second Language Acquisition**, v.21, p. 511–556, 1999.

ROTT, S. Relationships between the process of reading, word inferencing, and incidental vocabulary Acquisition. In J. F. Lee & A. Valdman, (Eds), **Form and meaning:** Multiple perspectives. pp. 255–282, 1999. Boston: Heinle & Heinle.

RUSSO, J.E.; JOHNSON, E. J.; & STEPHENS, D.L. The validity of verbal protocols. **Memory & Cognition,** v.17, n.6, p.759-769, 1989.

SELIGER, H. W. The language learner as linguist: Of metaphors and realities. **Applied Linguistics**, v.4, p.179–191, 1983.

SHEOREY, R.; & MOKHTARI, K. Differences in the meta-cognitive awareness of reading strategies among native and non-native readers. **System,** v.29, p. 431-449, 2001.

SIMON, H. A.; KAPLAN, C.A. Foundations of cognitive science. In: OSNER, M. I. (ed.). **Foundations of Cognitive Science**. Cambridge, MA, MIT Press, 1989.

SMITH, P. and KING, J. R. An Examination of Veridicality in Verbal Protocols of Language Learners. **Theory**

and Practice in Language Studies, v. 3, n. 5, p. 709-720, May 2013.

STRATMAN, J. F.; HAMP-LYONS, L. Reactivity in concurrent think-aloud protocols. In P. Smagorinsky (Ed), **Speaking about writing: Reflections on research methodology**, p. 89–112. London: Sage, 1994

SUDMAN, S.; BRADBURN, N. M.; SCHWARZ, N. (Eds.). **Thinking about Answers**: the Application SHEOREY, R., & MOKHTARI, K. Differences in the meta-cognitive awareness of reading strategies among native and non-native readers. **System,** n.29, p.431-449, 2001.

THOMPSON, S. J.; JOHNSTONE, C.J.; and THURLOW, M.L. **Universal design applied to large scale assessments**. Minneapolis: National Center on Educational Outcomes, 2002.

UPTON, T. A.; & L. LEE-THOMPSON. The role of the first language in second language reading. **Studies in Second Language Acquisition,** v.23, p.469-495, 2001.

VAN SOMEREN, M. W.; BARNARD, Y. F.; SANDBERG, J. A. C. **The think-aloud method**: A practical guide to modeling cognitive processes. San Diego: Academic Press, 1994.

WESCHE, M. B.; & PARIBAKHT, T. S. Reading-based exercises in second language vocabulary learning: An introspective study. **The Modern Language Journal,** v**.** 84. n. 2, p.196-213, 2000.

WHITNEY, P.; BUDD, D. Think-aloud protocols and the study of comprehension. **Discourse Process**, v. 21, n. 3, pp. 341-35, 1996.

YOSHIDA., M. Think-Aloud Protocols and Type of Reading Task: The Issue of Reactivity in L2 Reading Research. Kyoto University of Foreign Studies 2008. In: BOWLES, MELISSA; *et al.,* (ed.). **Selected Proceedings of the 2007 Second Language Research Forum**. Somerville: Cascadilla Proceedings Project, 2008. p. 199-209.