Can the Reading Construct Be Examined outside the Reading Context? An Investigation of the Construct Validity of an English as a Foreign Language Reading Test

Godefroid B. Katalayi

Institut Supérieur Pédagogique de Kananga, DR Congo

Abstract—This paper examines the construct validity of a multiple-choice reading test from the epistemological stance that supports a close relationship between test construct and test context. It aims to evaluate the extent to which the test tasks not only target the reading construct, but also come to grip with the specific test context (Katalayi & Sivasubramaniam, 2013). Fifty (50) multiple-choice items taken from the 2017 English state examination (ESE) were administered to 496 Grade 12 secondary school students and a concurrent strategies questionnaire was used to elicit information on participants' use of strategies during test writing (Katalayi & Sivasubramaniam, 2014). The findings indicate that, although the fifty item questions cover the whole range of reading construct by targeting all the processing levels; there is, however, a predominance of items that require text processing at higher level than those items that require text processing at lower level. This is to suggest that since the ESE includes higher level test items it does not reflect the context of the ESE; and therefore, it is untenable in the educational practices of language teaching and testing.

Index Terms—construct validity, reading construct, reading context, processing levels, social context, tasks complexity

I. PREAMBLE

In conceptualizing reading as a process, many reading models consider text comprehension as a multi-componential process where a reader must attempt and activate various cognitive processes in order to successfully comprehend a text (Katalayi & Sivasubramaniam, 2014; Embretson & Wetzel, 1987; Gao & Rodgers, 2007; Field, 2004; Wagenmakers, Steyvers, Raaijmakers, Shiffrin, van Rijn, H, & Zeelenberg, 2000; Perfetti, 1997; van Gompel, 2007). These processing models have generally investigated the construct validity of a reading test in terms of the different levels of structure by associating the specific cognitive processes with each level (Katalayi & Sivasubramaniam, 2014). Although research indicates some variations in the number of levels (see Gao & Rodgers, 2007; Grabe, 2009), six processing levels are generally recognized. The first level, called lexical encoding, relates to the reader's capacity to encode words (Field, 2004; Wagenmakers *et al.* 2000). The second level, termed syntactic parsing, relates to the reader's capacity to analyze a sentence or phrase in terms of the grammatical constituents by identifying the different syntactic relations (van Gompel, 2007). The third level, termed propositional meaning, deals with how a reader interprets clauses semantically, how he/she integrates the information in the sentence as a whole, and how he/she abstracts the sentence gist (Haberlandt & Graesser, 1985). The fourth level, called inferencing, relates to a reader's capacity to integrate the information provided in the text with the information that he/she already knows (Kintsch & van Dijk, 1978). The fifth level, called building a mental model, relates to the reader's capacity to identify main ideas, to relate them to previous ideas, to distinguish between macro-propositions and micro-propositions and to impose some hierarchical structure of the information in the text (Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983). The last and sixth level, called 'creating text level macrostructure', relates to reader's capacity to recognize the hierarchical structure of the whole text and to determine which information is central to the meaning of the text (Weir & Khalifa, 2008a).

Katalayi and Sivasubramaniam (2013) argue that studies that have adopted this hierarchical-level taxonomy to validate the reading construct (see van Steensel, Oostdam & van Gelderen, 2012; Weir & Khalifa, 2008b) have generally suggested that test items that require examinees to process the text at a higher level (text level) are more cognitively demanding than those that require examinees to process the text at a lower level (sentence/paragraph level). More specifically, these studies have reported that lower level test questions are generally easy as they require examinees to read at sentence and/or paragraph level by quickly searching for text information at the level of text micro-propositions. On the contrary, higher level test questions are generally difficult as they require examinees to read at text level by focusing on text macro-propositions and by attempting to generate more connections between their knowledge and text information.

This way of viewing the process of reading and its assessment from a purely cognitive perspective appears to emphasize a reductionist view of reading. Yet, research suggests that reading is not only a cognitive process; but it is also a social process where a reader interacts with a text to construct meaning (Bloome, 1985; Green & Meyer, 1991; Katalayi, 2014). As a social process, reading is a situated activity that takes place in a specific context with specific task demands (Katalayi & Sivasubramaniam, 2014; Bloome, 1985; Knoester, 2010). Such a social perspective to reading premises that the social context within which students read influences their reading behaviours. This is to suggest that the social context determines how students engage with reading tasks and the cognitive processes they internalize from and for reading (Bloome, 1985; Katalayi, 2014).

In the context of reading in a second/foreign language, one route taken by the studies that have adopted a social activity stance to reading has been the understanding of the socio-contextual variables that influence reading activity; and therefore impact on reading performance (Katalayi & Sivasubramaniam, 2014). Some of these studies have specifically investigated the impact of the students' characteristics on reading performance by arguing that students' capacity to deploy appropriate strategies to read the text and complete the test tasks also depends on the extent to which the test tasks are appropriate to their individual characteristics; including their skills to comprehend the text (Neiles, 2012; Katalayi & Sivasubramaniam, 2014), their motivation for reading (Mucherah & Yoder, 2008; Taboada, Tonks, Wiogfield, & Guthrie, 2009; Katalayi, 2014; Katalayi & Sivasubramaniam, 2014), their attitudes towards the test (Katalayi & Sivasubramaniam, 2014; Murray, Riazi & Cross, 2012; Rasti, 2009) and their experience with the test (Reeve, Heggestad, & Lievens, 2009; Katalayi, 2014). Other studies have addressed the issue of the impact of the socio-contextual factors on reading performance in terms of the impact of the organizational characteristics of the school under which the students read; including the policies adopted by the school in terms of discipline (Lee & Shute, 2010), the collaboration between teachers (Lee & Shute, 2010), and the demographics of the class size (Lee & Shute, 2010; Katalayi, 2014). Further to this, other studies have addressed the issue of the availability, quality and relevance of reading resources and their impact on reading performance (Katalayi, 2014; Bukenya, 2005; Kalibala, 1999). Still, some other studies have addressed the issue of teacher's qualification and experience (Lee & Shute, 2010), his/her motivation and commitment for teaching (Bishay, 1996), the kind of support system provided to him/her (Jahangir, Saheen & Kazmi, 2012; Katalayi, 2014) and these studies have attempted to correlate these characteristics to student's reading performance. In sum and spirit, the socio-contextual perspective on reading activity premises that the whole range of aforementioned factors can function as predisposing conditions affecting what one reads, why he/she reads, and how well he/she reads. Therefore, this perspective on reading can be useful to the reading assessment researcher by helping him/her understand the many specific reading behaviors that can further explicate the nature of reading activity and reading performance (Katalayi & Sivasubramaniam, 2014).

Is the social dimension of reading reflected in current assessment processing models to examining the reading construct?

Many current processing models to examining the reading construct appear to be strongly influenced by the union of cognitive psychology and assessment (Embretson & Wetzel, 1987; Cohen & Upton, 2006, 2007; Gao & Rodgers, 2007). These models generally support that reading is a complex and multidimensional process and the best way to assess the reading process is to describe the subskills that underlie it (Song, 2008). Therefore, this can be made possible by gathering information regarding examinees' knowledge structure, skills and strategies while completing test tasks (Cohen & Upton, 2006, 2007; Gao & Rodgers, 2007). This way of examining the reading construct suggests that it is possible to link particular reading items to specific sub-skills that they are said to tap into (Alderson, 2000; Katalayi & Sivasubramaniam, 2014; Song, 2008). Therefore, one way to achieve this goal has usually been to model item statistic properties in terms of the cognitive demands required by test tasks (Embretson & Wetzel, 1987; Gao & Rodgers, 2007).

Katalayi and Sivasubramaniam (2014) argue that although validity studies that have adopted these models have incorporated current cognitive theories in examining the reading construct; thus providing useful information on examinees' performance on the reading test, they nevertheless appear to have ignored to link the critical features of these cognitive theories to the specific testing context so as to reveal the meaningful cognitive processes in a particular test context. Methodologically, these models have used some current methods that help to understand how examinees complete test tasks such as protocol analysis and task analysis. However, these models appear to rely heavily on factorial analyses and linear regression methodologies. Results from factor analysis methodologies tell whether or not the different reading items included in the test load on the same factor (Guthrie & Kirsch, 1987; Young, 2008). On the other hand, results from linear regression methodologies indicate a number of text and item variables and the effect of these variables on item statistics (Freedle & Kostin, 1999) and/or the particular variables that significantly account for the variance in item difficulty (Kasai, 1997). As Katalayi and Sivasubramaniam (2014) argue, since these two methodologies are psychometrically driven, results from the studies that use these methodologies hardly reveal the interaction between the different processes that are actually involved in answering the individual test items and the actual context of reading that is expected to clarify the particular test domain. Therefore, I believe that in examining the reading construct of an ESL/EFL test, there is an inevitable need to use a model that signposts and signifies the actual context of reading activity.

II. AIM AND SIGNIFICANCE

This paper examines the construct validity of the DR Congo English state examination, a multiple-choice reading test administered to final year secondary school students as a subtest of the national test for certification (Katalayi & Sivasubramaniam (2014). The paper aims to evaluate the construct validity of this test within the epistemological stance that supports a strong relationship between test construct and test context. More specifically, the paper aims to evaluate the extent to which the test tasks not only target the reading construct as it appears to be actually conceptualized in literature but also come to grip with the specific test context.

In order to achieve this aim, I hypothesize that since cognitive processing always occurs within and is significantly affected by the reading context, examinees' use of strategies to comprehend the text and complete test tasks depends not only on their individual processing skills, but also on the knowledge domain and the test tasks. In light of this hypothesis, I suggest that the test tasks be evaluated in terms of two dimensions: first, the cognitive processes examinees deploy while completing the test tasks, and then the extent to which the cognitive demands required to complete the test tasks are proportional to the examinees' characteristics and actual conditions under which reading instruction occurs.

By examining the reading construct within this suggested larger context of reading activity, I hope, together with Katalayi and Sivasubramaniam (2014) that the insights generated by this study can help reading researchers and ESL/EFL teachers view reading construct as one concept that is difficult to understand if it is explained and/or investigated outside the specific context of reading activity. Furthermore, these insights can be gainfully deployed in the educational practice of reading assessment, and they can be used by reading assessment researchers to inform their validation studies. The study is particularly important in the Congolese context where literature points to inadequate and deficient descriptions of reading assessment practices (Katalayi & Sivasubramaniam, 2014; Katalayi, 2012; 2014). Therefore, the findings are meant to serve English state examination (ESE) developers to design a test that includes tasks that are adjusted to the actual context of reading activity. This actual context is characterized by a complex quadriglossic linguistic structure that positions English as a fourth language that is learnt as a school subject from grade 9 to grade 12 (Katalayi & Sivasubramaniam, 2014; Katalayi, 2012, 2014) while French is the official language that is used in administration and education; whereas the four national languages (Ciluba, Kikongo, Kiswahili, and Lingala) serve inter-group communication in their areas of influence; and an estimated more than 200 'vernaculars' serve intra-group communication (Katalayi & Sivasubramaniam, 2013; Kasanga, 2010; Kasanga 2012). In such a French-dominated linguistic landscape with an infinitesimal presence of English in intra-national transactions and in public display (Kasanga, 2012), a relevant way to investigate the construct validity of the ESE would be to adopt an approach that helps to understand students' reading behaviours by considering the specific situations in which reading occurs as well as the reading tasks that readers confront. Therefore, while validating the ESE, the test tasks need to be described in terms of cognitive processes and socio-contextual parameters.

III. RESEARCH QUESTIONS

To examine the construct validity of the DR Congo ESE from a conceptual stance that supports a relationship between reading construct and reading context, I propose and address the following two questions:
• What are the processing levels targeted by the ESE items?
• To what extent the ESE tasks are appropriate to the actual ESE context?
The answer to the first question is meant to determine the processing levels that are targeted by the ESE items and to determine the extent to which the ESE includes tasks that target the whole range of the reading construct; one main criterion of its construct-validity. On the other hand, the answer to the second question is expected to determine the extent to which the degree of complexity of ESE tasks is proportional to the actual test context; a condition that aligns with the epistemological stance that considers test context as relevant to test construct.

IV. METHOD

A process-oriented approach to examining the reading construct needs a method that encourages informal comments on the actual reading process (Katalayi & Sivasubramaniam, 2013; Katalayi & Sivasubramaniam, 2014). Hence, protocol analysis can provide an understanding of the processes examinees engage to comprehend the text and complete test tasks. Protocol analysis is a methodology that uncovers the cognitive processes a person goes through while performing a task (Katalayi & Sivasubramaniam, 2013; Ericsson & Simon, 1993). It premises that, since people have "privileged access to their experiences" (Katalayi & Sivasubramaniam, 2013; Ericsson & Simon, 1993: xii), and that the information in their verbal reports is trustworthy (Park, 2009), it is possible to verbalize their thoughts in a way that does not alter the sequence of thought that mediate the completion of the task (Ericsson & Simon, 1993). Protocol analysis methodology has gained a strong ground in validating reading tests (see Rupp, Ferne, & Choi, 2006; Yamashita, 2003; Katalayi & Sivasubramaniam, 2013; Katalayi & Sivasubramaniam, 2014).

In using protocol analysis in this study, the underlying assumption is that the way participants search for text information, evaluate item alternatives, and choose the best option can be registered through their verbalizations and this can be analyzed to determine their response decision processes (Katalayi & Sivasubramaniam, 2013). This is why

in this study the data were collected during the task completion process in order to ensure a close connection between thinking and verbal reports (Ericsson & Simon, 1993).

Furthermore, in order to address the social context of reading practices in the DR Congo, the paper builds upon the results from a mixed-methods study that has comprehensively described the context of the English state examination (Katalayi & Sivasubramaniam, 2014). This study was conducted in the same setting (Kananga, DR Congo) with participants sharing the same characteristics to those included in this study. Katalayi and Sivasubramaniam (2014) used a contextual questionnaire that he respectively administered to 793 student-participants, and a contextual questionnaire that he administered to twenty-seven (27) Grade 12 English language teachers. The contextual questionnaire administered to student-participants included questions relating to their proficiency level in English language, their experience with the ESE, their motivation for reading, their attitudes towards the ESE, their reading frequency at school and at home, the reading resources used in classroom and at home, their frequency of attendance in reading classes, their preferences for the number of hours to learn English, their preferences for the number of hours to write the ESE, their preferences for the number of questions to be included in the ESE, their preferences for the number of alternatives to be included in the ESE, and their preferences for the number of parallel forms to be used in the ESE. On the other hand, the contextual questionnaire administered to the 27 teacher-participants included questions that related to their qualification and experience, their motivation for teaching, their involvement in in-service training programs, their use of assessment methods, the reading resources they use and the frequency of use of these resources. The two questionnaires were both close and open in that each statement was followed by a number of suggestions the participants had to choose from and an empty box where they had to provide an explanation/reason of their choices (Katalayi & Sivasubramaniam, 2014). The data from these two questionnaires were analyzed both quantitatively and qualitatively. The quantitative analysis aimed to provide a picture of the frequency of occurrence of different observations while the qualitative analysis aimed to provide deeper insights necessary for grasping the reasons of the participants' choices. The findings relating to the description of the DR English state examination context as reported by Katalayi and Sivasubramaniam (2014) can be summarized as follow:

(1) Students' abilities in English language in general, and in reading in English in particular, are poor as they hardly cope with basic language skills (listening, speaking, reading and writing); they have problems to match the English spelling and pronunciation; and they believe that English is a difficult language one cannot easily learn;

(2) Students have low motivation for reading and this is reflected in a low reading amount at school and at home and low attendance to English classes;

(3) Some students have negative attitudes towards the ESE as they believe that they must not write the ESE since they perceive this test to be a way of failing them on the national test. They also believe that the ESE is not a fair test since it is based only on text passage comprehension and neglects other language skills. Furthermore, students seem to prefer an extended time to write the ESE, a reduced number of test questions, a reduced number of item alternatives and a reduced number of alternate forms;

(4) Students have less experience with the ESE because the multiple-choice technique is almost not used by the teachers as the classroom assessment method on the ground that it is difficult to construct, and it cannot properly evaluate the students' language abilities;

(5) Reading instruction is severely hampered by an abject paucity of reading materials for reading at school and at home as well as by the use reading materials of poor quality;

(6) The English curriculum is old-fashioned; and some teachers do not have this document as part of their teaching documents. Therefore, they do not use this document in the design and planning of reading course;

(7) There is a total absence of in-service training as well as continuous training for English teachers. Consequently, teachers are not able to keep abreast of new developments in teaching. Besides, there is a substantial degree of confusion in their use of teaching methods.

Sample, instrument and materials

The participants were 496 Grade 12 secondary school students who were in the final year of secondary school and who were preparing for the national test in order to get a national certificate (Katalayi & Sivasubramaniam, 2013).

As instrument, I used a concurrent strategies questionnaire adapted from Weir and Khalifa (2008a) and Katalayi and Sivasubramaniam, 2013) in order to elicit information on participants' use of strategies during test writing. In this questionnaire (see Appendix), I instructed the participants to select from a list of strategies, individual strategies they had used to answer each test question immediately after they had answered that individual test question.

The test material consisted of 50 multiple-choice reading test items taken from the 2017 edition test and that was administered to the participants under normal testing conditions (Katalayi & Sivasubramaniam, 2013).

Procedure

Following the procedure suggested by Katalayi and Sivasubramaniam (2013), the participants were told that the test was accompanied with a questionnaire that was to be completed while taking the test. They were instructed to fill out the questionnaire each time immediately after answering an individual test question. They were reminded that they had to carefully go through the list of suggested strategies in the questionnaire and check all the strategies they remembered they had actually used to find the answer to each individual test question.

Analysis

The processing levels targeted by individual test items were determined through an examination of strategies the participants used to answer individual test questions (see Appendix for strategies coding). Therefore, questions that requested lower level processing were classified in levels 1, 2 and 3. Level 1 (lexical encoding) included all test questions that required participants' skill to understand vocabulary, level 2 (syntactic parsing) included the questions that required participants' skill to understand grammar, while all test questions that required participants to add any external knowledge to the text in order to understand the actual meaning of individual propositions were classified in level 3 (propositional meaning). On the other hand, questions that requested higher level processing were classified in levels 4, 5 and 6. Level 4 (inferencing) included those questions that required participants to link information necessary to answer the test question to information not explicitly stated in the text. Level 5 (building a mental model) encompassed all test questions that required participants' ability to establish accurate comprehension of explicitly stated main ideas and supporting details across sentences; while level 6 (creating text level structure) included all test questions that required participants to read the whole text in order to construct its macro-structure.

In order to examine the occurrence of strategies the participants had used, the frequency of strategies was computed and percentages were calculated. Percentages closer to 100.0% indicated that the majority of participants used the strategy in answering individual test questions while percentages closer to 0.0% indicated that few participants used the strategy to answer individual test questions.

In order to establish the complexity of test tasks, the mean item difficulty (mean ID) of test items included in each processing level was computed. Item difficulty is a straightforward statistics that consists in taking the number of candidates who got an item right and divide this number by the total number of candidates who answered that item. From the results of item difficulty per individual item, the mean item difficulty was computed for different test items included in each of the six processing levels. Values closer to 1 suggest that the item is easy; values closer to .0 suggest that the item is difficult whereas values that cluster around .5 suggest that the item is of moderate difficulty.

V. RESULTS

The processing levels targeted by the test items

Table 1 on the following page presents information on the different processing levels targeted by the 50 test items. The data suggest that the 50 test items included in the sample targeted all the six levels in different proportions. In descending order, we have (1) inferencing level (14 items; 28%), (2) syntactic parsing (11 items; 22%), (3) creating text level structure (10 items; 20%), (4) lexical encoding (8 items; 16%), (5) building a mental model (5 items; 10%) and (6) propositional meaning (2 items; 4%).

Level 1: Lexical encoding items

The information presented in Table 1 indicates that the participants used five (5) reading strategies and four (4) test taking strategies in order to answer the eight (8) test questions that required their skills to identify or/ and understand some words used in the text.

As regards the use of reading strategies, the data indicate that all the participants had to read the specific sentence where the vocabulary item was located (S9=100%) first, and the majority of them had to use of their knowledge of vocabulary (S6=73%). However, nearly a third of participants had to also carefully read a portion of the paragraph (S3=29%) in order to grasp the context in which individual words were used so as to use context clues to figure out the word meanings (S7=27%). On the other hand, only very few participants had also to resort to their knowledge of grammar (S5=10%) in order to answer vocabulary items.

As regards the use of test taking strategies, the majority of participants considered the five options and immediately focused on the "correct" answer (S11=70%). However, few participants could show uncertainty by first considering the five options and then delaying to decide which option was the correct answer (S10=23%) before they could finally selected their answers either through elimination of other options (S14=15.0%) or through guessing (S15=8.0%). The mean item difficulty index (.71) suggests that these test questions were generally easy.

TABLE 1:
PROCESSING LEVELS TARGETED BY THE 50 TEST ITEMS

Processing levels		N=50	%	Mean ID	Strategies used	
Level 1	Lexical encoding	8	16	.71	Reading	S9 [100%]; S6 [73%]; S3 [29%]; S7 [27%]; S5 [10%].
					Test taking	S11 [70%]; S10 [23%]; S14 [15%]; S15 [8%].
Level 2	Syntactic parsing	11	22	.73	Reading	S9 [100%]; S5 [79%]; S3 [21]; S6 [6%].
					Test taking	S11 [61%]; S12 [28%]; S10 [11%]; S14 [8%]; S15 [3%].
Level 3	Propositional meaning	2	4	.56	Reading	S5 [73%]; S6 [48%]; S3 [24%]; S7 [14%]
					Test taking	S10 [80%]; S11 [55%]; S14[45%]
Level 4	Inferencing	14	28	.41	Reading	S8 [80%]; S1 [73%]; S2 [59%]; S3 [49%].
					Test taking	S10 [100%]; S13 [50%]; S14 [32%]; S15 [18%].
Level 5	Building a mental model	5	10	.43	Reading	S3 [100%]; S2 [40%]; S8 [10]
					Test taking	S10 [90%]; S14[76%]; S15 [14%]; S11 [10%]
Level 6	Creating a text level representation	10	20	.38	Reading	S1 [100%]; S2 [78%]; S3 [56%]; S4 [42%].
					Test taking	S10 [82%]; S14 [50%]; S15 [32%]; S11 [18%].

(ID=Item difficulty; S1=strategy #1, S2=Strategy #2, S3=Strategy #3, S4=Strategy #4; etc.)

Level 2: Syntactic parsing items

The data presented in Table 1 indicate that participants used four (4) reading strategies and five (5) test taking strategies in order to answer the fifteen (11) test questions that required their skills to understand syntax.

As regards the use of reading strategies, all the participants had to read the specific sentence where the grammar structure was located (S9=100%) before they could use their knowledge of grammar (S5=79) in order to correctly parse the sentence. However, for some questions, the participants had also to carefully read the entire portion of the paragraph where the grammar structure was located (S3=21%) in order to understand the actual use of the structure. In rare cases, participants had also to resort to their knowledge of vocabulary in order to answer grammar items (S6=6%).

Regarding the use of test taking strategies, the majority of participants, on the basis of their knowledge of grammar, considered the five options and immediately focused on the option they perceived was the "correct" answer (S11=61%) although some participants produced their own answers immediately after they had read the item question, and then could finally look at the five options to confirm their answers (S12=28%). However, few participants could ponder the five options but postponed their selection of the 'correct' answer until further consideration (S10=11%) before they could make their final choice either through elimination of other options (S14=8%) or simply through blind guessing (S15=3%). The mean item difficulty index (.73) suggests that these test questions were easy.

Level 3: Propositional meaning items

The findings presented in Table 1 suggest that the participants used four (4) reading strategies and three (3) test taking strategies in order to answer the two (2) test questions that required their skill to establish propositional meaning.

Considering the use of reading strategies, the majority of participants used their knowledge of grammar (S5=73%) and/or their knowledge of vocabulary (S6=48%). However, few participants had to read a portion of the text carefully (S3=24%) before they used their knowledge of grammar and/or vocabulary. Furthermore, a minority of participants used context clues to figure out the meaning of difficult words in the proposition (S7=14%).

As regards test taking strategies, the majority of participants considered the five options, but postponed consideration of the 'correct' option (S10=80%) before they could either focus on the familiar option (S11=55%) or select the 'correct' option through elimination of other options (S14=45%). The mean item difficulty (.56) suggests that these test questions were of moderate difficulty.

Level 4: Inferencing items

The data presented in Table 1 suggest that participants used four (4) reading strategies and four (4) test taking strategies in order to provide answers to the fourteen (14) inferencing test questions.

As regards reading strategies, the main strategy used was a careful reading of the whole text (S1=73%). Some examinees either preceded or followed this careful reading with quick reading/rereading (S2=59%). On the other hand, for some other items, the participants had to carefully read a specific paragraph (S3=49%) before they could draw some conclusions based on what the text/paragraph implied (S8=80%).

As regards test taking strategies, all the participants considered the five options and then postponed to choose the "correct" option (S10=100%)) until they could finally make their choice either through text/paragraph global comprehension (S13=50%) or through elimination of other options (S14=32%) or again simply through blind guessing (S15=18%). The item difficulty index (.41) suggests that these test questions were difficult.

Level 5: Building a mental model items

The participants used three (3) reading strategies and four (4) test taking strategies in order to answers the five (5) items that requested their skills to establish accurate comprehension of explicitly stated main ideas and supporting details across sentences.

With regard to the reading strategies, all the participants read the target paragraph carefully (S3=100%) with the aim to comprehend all supporting details as elaborated across different sentences. However, some participants had to read first the whole text rapidly (S2=40%) in order to construct its representation before they focused on the target paragraph and read it carefully. Few participants augmented their careful reading of the paragraph by trying to draw conclusions on what the paragraph implied (S8=10%).

Concerning the use of test taking strategies, most participants first considered the five options and then postponed consideration of the "correct" option (S10=90%) before they finally made their choice either through elimination of other options (S14=76%) or through guessing (S15=14%). However, few participants considered the five options and then immediately focused on the option they considered to be the "correct" one (S11=10%). The mean item difficulty index (.43) suggests that these questions were difficult.

Level 6: Creating text level structure items

The participants used four (4) reading strategies and four (4) test taking strategies in order to answer the ten (10) test questions included in this level. As to reading strategies, all the participants had to carefully read the whole text (S1=100%) in order to construct its macro-structure. In order to locate the pieces of information scattered throughout the text, the majority of participants read the whole text expeditiously (S2=78%). However, since in many cases the pieces of information necessary for answering individual test questions were located in specific paragraphs, participants had to read those specific paragraphs first carefully (S3=56%) in order to build the paragraph gist, and then sometimes expeditiously (S4=42%) in order to search for specific information.

Concerning the use of the test taking strategies, most participants considered the five options and then postponed consideration of what could appear to them as the best option (S10=82%) before selecting their answers either through elimination of other alternatives (S14=50%) or through guessing (S15=32%). However, only few participants could consider the five options and immediately select the option that appeared familiar to them (S11=18%). The mean item difficulty index (.38) suggests that these test questions were difficult.

ESE focus: Lower level processing or higher level processing?

The findings provided in Table 2 suggest that of the 50 test items included in the sample, 21 items (42%) requested participants to process the text at lower level while 29 items (58%) requested them to process the text at higher level.

As far as task complexity is concerned, the data in the same table indicate that the mean item difficulty for the 21 items that required lower level processing was .67, suggesting that the participants perceived these items to be of moderate difficulty. On the other hand, the mean item difficulty for the 29 items that required higher level processing was .41, suggesting that the participants perceived these items to be difficult.

TABLE 2:
LOWER LEVEL PROCESSING VERSUS HIGHER LEVEL PROCESSING

Processing levels				N=50	%	Mean ID
1	Lower level	1	Lexical encoding	8	16	.71
		2	Syntactic parsing	11	22	.73
		3	Establishing propositional meaning	2	4	.56
	Total lower level processing item questions			**21**	**42**	**.67**
2	Higher level	1	Inferencing	14	28	.41
		2	Building a text mental model	5	10	.43
		3	Creating a text level representation	10	20	.38
	Total higher level processing items			**29**	**58**	**.41**

VI. DISCUSSION, IMPLICATIONS, CONCLUSION

The ESE tasks appear to cover the whole range of reading construct by targeting all the six processing levels. However, there is a predominance of test questions that require higher level processing than those that require lower level processing. As my theoretical stance supports a strong relationship between test construct and test context, this predominance of higher level test items does not appear to reflect the context of the ESE and therefore, it is untenable in the educational practices of language teaching and testing. This finding therefore raises the issue of the ESE construct validity on the ground that the ESE construct validity depends, not only on the extent to which the test includes tasks that target the whole range of reading construct, but also on the extent to which the test tasks are appropriate to the test context (Katalayi, 2014).

The study results have revealed that higher level questions were more difficult for examinees than lower level questions; a result that appears to concord with findings reported by Weir and Khalifa (2008b), van Steensel, Oostdam and van Gelderen (2012) and Rouet *et al*., (2001)**.** While lower level questions mainly required examinees to read at sentence and/or paragraph level and recourse to their knowledge of grammar and/or vocabulary, higher level questions mainly required them to read at text level by attempting to generate more connections between their knowledge and text information. Furthermore, in order to answer lower level questions, examinees generally considered the five options and focused on the familiar option, or they produced their own answers after reading the questions and then looked at the five options to confirm their choices. On the other hand, in order to answer higher level questions, examinees usually had to recourse to elimination of options and guessing strategies. Yet, evidence suggests that elimination of options and guessing are two test wiseness strategies examinees use as a short-cut to arriving at answers (Cohen & Upton, 2006, 2007; Rupp, Ferne & Choi, 2006). Therefore, the use of these two strategies in the context of this study appears to suggest that the process of answering higher level questions may have been a problem-solving process instead of being a reading process where examinees actually engage with the reading test tasks by actively working to understand the text, to understand the expectations of the questions, to understand the meanings and implications of the different item alternatives and to select or discard alternatives based on the way they understand the text.

Since the focus on higher level questions does not appear to reflect the actual ESE context described previously, we are inclined to conclude that the ESE fails to take account of the actual characteristics of the students and the actual conditions under which reading instruction occurs. Consistent with this conclusion, the ESE appears to be construct-irrelevant as it includes tasks that are less appropriate to the test context. This conclusion appears to concur with the conceptualisation of validity as the relevance of inferences made on the basis of test scores (Messick, 1989, Kane, 2011). Therefore, in order to be construct-valid, the ESE test scores need to closely reflect the actual reading context by including more tasks that require examinees to read at sentence and/or paragraph level as such tasks appear to be less cognitively demanding than those that require them to read at text level. Since examinees' performance on test tasks appears to be influenced by the context of reading practice, it can be suggested that by including more tasks that require examinees to read at sentence level and paragraph level, the ESE would closely reflect the actual context of English/reading instruction characterized by students' low skills in English, in general and in reading, in particular; the low motivation for reading, some negative attitudes towards the ESE, less experience with the ESE, limited reading at

school and at home due to an abject paucity of reading materials, poor teacher quality and lack of commitment of their English teachers.

This insight is worth considering in conducting validation studies. It suggests that a test item can be valid when the information required answering this items taps into the target construct; but still it can be inappropriate to the specific context of the test. In order to further our knowledge of the construct validity of reading tests, we need to agree that the development of valid reading tests can be possible if we first examine as extensively as possible the nature of reading activities and different response processes examinees engage during test taking. This entails eliciting the cognitive processing involved in not only the text and test tasks, but also in contexts beyond the test itself, that is, in those contexts involved in performing reading tasks in classroom and real life (Katalayi, 2014). By ignoring this route, reading appears to be an activity that occurs in a vacuum; therefore, the results generated by psychometrically-driven studies that ignore the actual reading context tend to miss the salient aspects of the reading construct as we hardly understand the actual reading behaviours. Further to this, examinees' completion of test tasks can become more of a problem-solving activity rather than a reading activity. Therefore, reading construct cannot be examined outside the appropriate context of reading activity.

APPENDIX. STRATEGIES QUESTIONNAIRE

Instructions:
- You will have 2 hours to do the test and fill out the questionnaire.
- Read the text and provide answers to questions based on the text.
- After answering each test question, please fill out the questionnaire for that question. To fill out the questionnaire, you will have to identify and list in order ALL the strategies you remembered you have used to answer that individual question. The list of strategies and their code are present in the table below.
- For example, if to question 1 you first read the text carefully, and then reread it rapidly and finally look at the question options and produce your own answer after reading the question, you must write: S1; S4; S11.

Code	Strategy
S1	read the whole text carefully
S2	read the whole text rapidly
S3	read a portion of the text carefully
S4	read a portion of the text rapidly
S5	use my knowledge of grammar
S6	use my knowledge of vocabulary
S7	Use context clues to figure out the meaning of difficult words
S8	draw conclusion based on what the text implies
S9	search for the sentence where the vocabulary item or grammar structure is located
S10	consider the options and postpone consideration of the 'correct' option
S11	consider the options and focus on a familiar option
S12	produce my own answer after reading the question, then look at the options
S13	Select option through paragraph/passage overall comprehension
S14	select option through elimination of other options
S15	select option through guessing

REFERENCES

[1]   Alderson, C.J. (2000). Assessing reading. Cambridge: Cambridge University Press.
[2]   Bishay, A. (1996). Teacher motivation and job satisfaction: A study employing the experience sampling method. *J. Undergrad. Sci. 3*, 147-154.
[3]   Bloome, D. (1994). Reading as a social process IN A MIDDLE school classroom, in D. Graddol & J. Maybin (Eds.), *Researching language and literacy in social context* (pp.100-130). Clevedon: Multilingual Matters Ltd.
[4]   Bukenya, M. (2005). Findings and recommendations of the NAPE Report: Levels of competency in literacy and numeracy of primary school Pupils in Uganda. *Journal of Career and Technical Education, 18* (1), 69–80.
[5]   Cohen, A. & Upton, T. (2006). Strategies in responding to the new TOEFL reading tasks. (TOEFL Research Reports), Princeton, N.J.: Educational Testing Services.
[6]   Cohen, A. & Upton, T. (2007). I want to go back to the text: Response strategies on the reading subtest of the new TOEFL. *Language Testing, 24*(2), 209-250.
[7]   Embretson, S. & Wetzel, D. (1987). Component latent trait models for paragraph comprehension tests. *Applied Psychological Measurement 11*, 175-93.
[8]   Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.
[9]   Field, J. (2004). Psycholinguistics: the key concepts, London: Routledge.
[10]   Freedle, R. & Kostin, I. (1993). The prediction of TOEFL reading item difficulty: implications for construct validity. *Language Testing 10*(2), 133-170.
[11]   Gao, L., & Rodgers, T. (2007). Cognitive-psychometric modeling of the MELAB reading items. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.
[12]   Grabe, W. (2009). Reading in a second language: Moving from theory to practice. New York: Cambridge University Press.

[13] Green, J.L., & Meyer, L.A. (1991). The embeddedness of reading in classroom life; Reading as a situated process, in C.B. Baker & A. Luke, (Eds.), *Towards a critical sociology of reading pedagogy: Papers of the XII World Congress on reading on reading* (pp.141-159). Queensland: John Benjamins.

[14] Guthrie, J.T. & Kirsch, I.S. (1987). Distinctions between reading comprehension and locating information in the text. *Journal of Educational Psychology*, *79*, 220-227.

[15] Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General 114*(3), 357-374.

[16] Jahangir, S. F., Saheen, N., & Kazmi, S. F. (2012). In-service training: A contributory factor influencing teachers' performance. *International Journal of Academic Research in Progressive Education and Development, 1*(1), 31- 38.

[17] Kalibbala, G. (1999). Sustainable textbook provision and utilization in Uganda PETDP for the Government of Uganda. *Journal of Career and Technical Education, 17* (2), 29–40.

[18] Kane, M. T. (2011). Validating score interpretations and use: Messick lecture, Language Testing Research Colloquium, Cambridge: SAGE.

[19] Kasai, M. (1997). Application of the rule space model to the reading comprehension section of the Test of English as a Foreign Language (TOEFL). Unpublished Doctoral Dissertation, University of Illinois, Urbana.

[20] Kasanga, L. A. (2012). English in the Democratic Republic of the Congo. *World Englishes, 31*(1), 48-69.

[21] Kasanga, L, A. (2010). Streetwise English and French advertising in multilingual DR Congo: symbolism, modernity, and cosmopolitan identity. *International Journal of the Sociology of Language*, *7*(4), 33-42.

[22] Katalayi, G. B. & Sivasubramaniam, S. (2014). A context-based perspective to investigating the construct validity of reading assessments. *Mediterranean Journal of Social Sciences, 5(*27*)* 847-860.

[23] Katalayi, G. B. & Sivasubramaniam, S. (2013). Careful reading versus expeditious reading: Investigating the construct validity of a multiple-choice reading test. *Theory and Practice in Language Studies, 3*(6), 877-884.

[24] Katalayi, G.B. (2012). The DR Congo English state examination: Some fundamental validity issues (Context validity evidence). Unpublished Master Thesis, University of Western Cape, South Africa.

[25] Katalayi, G. B. (2014). Fundamental validity issues of an English as a foreign language test: A process-oriented approach to examining the reading construct as measured by the DR Congo English state examination. Unpublished Doctoral Dissertation, University of the Western Cape, South Africa.

[26] Kintsch, W.M., & van Dijk, T. A. (1978). Towards a model of text comprehension and production. *Psychological Review, 85*(5), 363-394.

[27] Knoester, M. (2010). Independent reading and the 'Social Turn': How adolescent reading habits and motivation relate to cultivating social relationships. *Networks, 12*(1), 1-13.

[28] Lee, J., & Shute, V. J. (2010). Personal and social-contextual factors in k-12 academic performance: An integrative perspective on student learning. *Educational Psychologist, 45*(3), 185-202.

[29] Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York: American Council on Education.

[30] Mucherah, W., & Yoder, A. (2008). Motivation for reading and middle school students'performance on standardized testing in reading. *Reading Psychology, 29*(1), 214-235.

[31] Murray, J. C., Riazi, A. M., & Cross, J. L. (2012). Test candidates' attitudes and their relationship to demographic and experiential variables: The case of oversea trained teachers in NSW, Australia. *Language Testing 29*(4), 577-595.

[32] Neiles, K. Y. (2012). An Investigation of the effects of reader characteristics on reading comprehension of a general chemistry text. Unpublished Doctoral Dissertation, The Catholic University of America, USA.

[33] Park, S. (2009). Verbal report in language testing. *The Journal of Kanda University of International Studies, 2*, 287-307.

[34] Perfetti, C.A. (1997). Sentences, individual differences, and multiple texts: three issues in text comprehension, *Discourse Processes 23*, 337-355.

[35] Rasti, I. (2009). Iranian candidates' attitudes towards IELTS. *Asian EFL Journal, 11*(3), 75-92.

[36] Reeve, C. L., Heggestad, E. D., & Lievens, F. (2009). Modeling the impact of test anxiety and test familiarity on the criterion-related validity of cognitive ability tests. *Intelligence 37,* 34–41.

[37] Rupp, A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, *23*(4), 441-474.

[38] Song, M. Y. (2008). Do divisible subskills exist in second language (L2) comprehension? A structural equation modeling approach. *Language Testing, 25*(4), 435-464.

[39] Taboada, A. Tonks, S.M., Wiogfield, A. & Guthrie, J.T. (2009). Effects of motivational and cognitive variables on reading comprehension. *Reading and Writing, 22*, 185-209.

[40] van Gompel, R. P. G., & Pickering, M. (2007). Syntactic parsing. Gaskell, G. (Ed.), In: *Oxford Handbook of Psycholinguistics.* (1st ed.) (pp. 289-307). Oxford: Oxford University Press.

[41] van Steensel, R., Oostdam, R., & van Gelderen, A. (2013). Assessing reading comprehension in adolescent low achievers: Subskills identification and task specificity. *Language Testing, 30*(1), 3-21.

[42] Wagenmakers, E. J., Steyvers, M., Raaijmakers, J.G., Shiffrin, R.M., van Rijn, H, & Zeelenberg, R. (2000). A Model for evidence accumulation in the lexical decision task. Paper presented at the 41st Annual Meeting of the Psychonomic Society, New Orleans.

[43] Weir, C.J., & Khalifa, H. (2008a). A cognitive processing approach towards defining reading comprehension, *Cambridge ESOL: Research Notes*, *31*, 2-10.

[44] Weir, C.J., & Khalifa, H. (2008b). Applying a cognitive processing model to Main Suite reading papers, *Cambridge ESOL: Research Notes 31*, 11-16.

[45] Yamashita, J. (2003). Processes of taking a gap-filling test: comparison of skilled and less skilled EFL readers. *Language Testing, 20*(3), 267-293.

[46] Young, J.W. (2008). Ensuring valid content tests for English language learners. *R&D Connections. 8*, ETS Research and Development, Princeton, NJ.

Godefroid B. Katalayi holds a doctorate degree in language education from the University of the Western Cape, South Africa. He currently lectures English teaching methodology and language testing modules at the Department of English at Institut Supérieur Pédagogique de Kananga, DR Congo. He has also lectured Academic literacy and Methods of English modules in the Faculty of Education, University of the Western Cape, South Africa. His research interest includes language assessment, the validity of reading tests, the assessment of narratives, and the teaching of grammar and reading to EFL learners.