TUMOLO, C. H. S.. **Assessing reading comprehension: the fair focus of test items**. In: Tumolo, C. H. S.; Figueiredo, D. C.; Moritz, M. E.; D'ely, R.. (Org.). Advanced Research in English Series: Contextualized practices in EFL teaching and assessment. Florianópolis: PGI - UFSC, 2011, v. , p. 17-42.

ABSTRACT

Many tests have been developed and used to assess reading ability in English as a foreign language. The use of a test has consequences, and it can be considered more defensible if it allows for interpretations and actions based on the relevant factors being assessed. Discrimination based on other factors will violate the principles of fairness. One factor pointed out in literature as a potential source of unfairness in testing is background knowledge. In this article, I aim at discussing the aspect of background knowledge within fairness and unfairness in test development and use, and at pointing to test items and item characteristics that may reduce unfairness in terms of test bias. In so doing, I contribute with discussions and suggestions for the development of tests with features of fairness.

RESUMO

Vários testes têm sido desenvolvidos e usados para medir a habilidade de leitura em inglês como língua estrangeira. O uso de um teste tem conseqüências e pode ser considerado mais defensável se ele permite interpretações e ações baseadas nos fatores relevantes sendo medidos. Discriminação baseada em outros fatores violará princípios de justiça. Um fator apontado na literatura como fonte potencial de injustiça em testes é conhecimento anterior, de mundo e/ou de assunto. Neste artigo, discuto o aspecto do conhecimento anterior dentro das questões de justiça e injustiça no desenvolvimento e uso de testes, e aponto para itens de testes e características de itens que podem reduzir injustiça. Desta forma, contribuo com discussões e sugestões para o desenvolvimento de testes que incorporam características de justiça.

1. INTRODUCTION

The ability of reading in English as a foreign language has probably been the most required to function in our modern society for all purposes, especially today with so much information in English available on the internet. It has particularly been a requirement for university studies, since many academic articles are published in this international language.

Many instruments have been developed and used to assess reading ability, suitable for both regular education and distance education. As I have shown in Tumolo (2005), the use of a test has consequences, and a test is considered more defensible if the interpretations based on it can be argued to be appropriate, i.e., based on the intended consequence of discriminating those with the various levels of the relevant factor(s). Any discrimination based on other factors will be considered unfair. That is, performance on tests should be related to the relevant factor(s), not related to features of unfairness.

One of the features of unfairness widely debated is bias, particularly if it is predictive. Shohamy (2000), for example, raises several questions in talking about fairness in testing, being one of them "do they [the tests] create biases?" (p. 17). For that reason, test use must be examined closely, since the actual use may be different from the intended purposes, with the unintended consequence of biased against some test takers. Test use must, thus, be appropriate in terms of the intended consequences, and responsible in terms of the potential unintended consequences. As Bachman (1990) stresses, the benefits of test use must outweigh the negative consequences caused for individuals and for the society. Otherwise, it is not defensible.

In Tumolo (2005), I investigated university entrance examinations, analyzed test items and concluded that some of the items were not defensible based on an investigation of validity. In this article, I aim at extending the discussion by focusing specifically on the

aspect of fairness and by pointing to items and item characteristics that may reduce the unintended consequences of test bias. Ultimately, this article aims at contributing with discussions and suggestions for the development of tests with features of fairness, which are required in most international codes of testing practice currently available.

Next, I present the motivation for this article. In section 2, I discuss background knowledge as a potential source of unfairness. In section 3, I review the literature to show how background knowledge affects reading comprehension, in terms of multiple readings and of unified representations and processing gains, being, thus, a source of bias in testing. In section 4, I present the processes underlying reading comprehension and ways to minimize background effects for testing purposes. In section 5, I present reading tasks to be used in a language test which complies with features of fairness. Lastly, in section 6, my final remarks are made.

1.1. Motivation for the article

The motivation for this article is given by the university entrance examinations used for the admission to UNICAMP. The concept of reading used for the development of the examinations is in line with what scholars believe of reading today. It is: reading is not a passive decoding of meaning, but an active task of negotiating meaning based on global comprehension, resulting in a new text by the reader[1]. The skills tested include explicit and implicit or inferential aspects, global and local comprehension, and micro and macro level skills (Scaramucci, 2002), that is, high-level skills are expected for the test.

[1] Taken from the session concerning English as a foreign language on the *manual do candidato*.

The concept of the test is also in line with the purpose of the whole examination. The 2003 candidate's manual mentions that the entrance examinations have had, since 1987, the aims of selecting students who can think, draw correlations, develop hypotheses. All the skills tested are in accordance with what Clapham (1996) found to be essential for university studies.

The emphasis on these high-level skills for the entrance examination may be accounted for as one way to overcome what Norton and Stein (1998) consider to be a validity paradox: while universities want students who can think, draw correlations, be critical and independent learners, testing instruments used as entrance examinations do not always allow for the demonstration of such skills.

But there is a note in the candidate's manual: *not any reading is allowed*. This suggests the test developers are concerned with what is the fundamental characteristic of comprehension as an active process, which, as mentioned in the manual, will lead to the creation of a new text by the reader. Comprehension is active, and a "reader cannot help but interpret and alter what he reads in accordance with prior knowledge about the topic under discussion" (Pearson & Johnson, 1978, p. 24).

Ultimately, the test developers seem to be concerned with the extent of the alteration, particularly with the possibility of textual intrusion and/or script intrusion. These two types of intrusion are discussed by Pearson and Johnson (1978), who characterize the former as a reading coming from the text, but being the result of random selection of segments, with no argument for them to be considered plausible, and the latter as coming from the reader's head for which there is no plausible line of reasoning. My first question was, then, how is it possible to know what is plausible or intrusion if comprehension is active?

However, the issue seems to be more complex. Allowing some readings and not allowing some others, that is, considering some readings plausible or intrusion, correct or

incorrect, may be based on specific knowledge and on value judgment. This, thus, raises ethical questions: What is the line between correct and incorrect reading? Who is to decide what reading is to be allowed and not allowed? Who has the right answer? Does that decision result in any form of discrimination against some groups based on irrelevant factors? Can that be a source of bias, hence of unfairness?

2. SOURCES OF UNFAIRNESS

One aspect presented in the specialized literature as a potential source of bias, of unfairness, is background knowledge[2]. Kunnan (2000) summarizes the most important concerns associated with fairness in language testing, including content and format bias, and stresses that bias causing unfairness include "topical knowledge and technical terminology, specific cultural content and dialect variation" (p. 3).

Background knowledge may be a source of bias when favoring groups or individuals with more familiarity with the topic. However, sources of bias can only be claimed if the factor affecting performance is not included in the construct[3] to be used for the development of the test. The scope of a construct must be delimited to define what factors to include or not, which has implications for the development and use of tests and for the interpretation of their results.

[2] Background knowledge in this article refers to any prior knowledge such as content, topical, technical, and cultural knowledge.

[3] According to the Dictionary of Language Testing, published by Davies et al (1999), construct can be defined as "an ability or set of abilities that will be reflected in test performance, and about which inferences can be made on the basis of test scores. A construct is generally defined in terms of a theory; in the case of language, a theory of language" (p. 31).

Bachman and Palmer (1996) offer three options for defining the scope of the construct in relation to topical knowledge[4], considering the characteristics of the examinations and of the test takers' expected knowledge. The options are: 1) construct definition including only language ability, not including topical knowledge; 2) construct definition including topical knowledge; and 3) separate constructs defined for language ability and topical knowledge.

Considering language testing situations in English as a foreign language (EFL) in which test takers are heterogeneous in relation to topical knowledge, the scope of the construct should follow the first category, where language ability is the only relevant factor for the interpretations and decisions based on the test. Thus, for the purpose of assessing language ability with heterogeneous test takers in relation to knowledge, background knowledge must be a factor to be controlled and its effects minimized, or else bias may result.

3. EFFECTS OF BACKGROUND KNOWLEDGE

In this section, I discuss some effects of background knowledge for reading comprehension. For that, I a) consider background knowledge as promoting multiple readings; and b) bring to bear scholars and researchers showing where background knowledge has its most effect, contributing to reading comprehension as a ready-made structure and as a resource-saving structure.

[4] The definition of topical knowledge by the authors is that it refers to knowledge schemata or real-world knowledge (Bachman & Palmer, 1996, p. 65).

3.1. Background knowledge promoting multiple readings

The concept of reading used for the entrance examination aforementioned is in line with what most recent accounts of the reading comprehension processes claim. Reading is considered today an interactive process. Rumelhart (1977), in proposing his interactive model for reading, claims that the reading involves the application of all knowledge sources, that is, a reader will draw on sources of information such as visual, orthographic, lexical, semantic, syntactic and schematic (world knowledge).

The sources of information contributing to the reading are, thus, top-down, coming from the reader's background knowledge stored in the long-term memory as schemas, and bottom-up, coming from the conveyed meaning in the text. Since readers will have different background knowledge, the mental representation resulting from the reading process may be somewhat different for each reader.

Studies carried out by renowned scholars such as Gernsbacher (1997), van Dijk and Kintsch (1983), Kintsch (1988), Lorch and O'Brien (1995), Kintsch (1998) on the mental representation resulting from the reading process point to the construction of two networks of comprehension, one called text model and the other situation model. Grabe (1999, 2000) review these studies as well as others relevant in the area and summarize the conclusions concerning the two networks. Text model reflects the information in the text, being a closer representation of comprehension up to the level of propositional integration. Situation model represents an interpretation of the text information, involving reader's background knowledge to a further extent, based on the reader's goal for reading, motivation, attitudes and evaluations of the information in the text.

In conclusion, Grabe (1999) stresses that the reader will be able to "both recognize and understand the information in the text, and also to create an interpretation that is unique"

(p. 19). Readers will, thus, provide similar and also distinct summaries, and the mental representation will be the result of the similarities and the distinctions, with features of interpretation.

Similarly and based on many other studies on reading comprehension, Urquhart and Weir (1998) claim that there will be different readings of the same text. The authors accept variations caused by the different background knowledge of readers, and call them interpretations. Thus, it is possible to conclude, based on the accounts and claims presented, that the outcome of reading will be somewhat different from reader to reader.

This difference can, however, be taken further. Norton and Stein (2000) reported on an experience that challenged their views of testing in reading particularly, in their own words, "disrupted our assumptions about tests, texts, and teaching" (p. 232). They were carrying out the pilot testing phase of a reading comprehension test designed to select some students for university admission. It was a post apartheid situation in South Africa. The passage was based on an article published on a local newspaper and was about a factual account of monkeys disturbing one family in a town. This is what both researchers thought was the 'universal understanding'.

What the researchers found relevant was the possible alternative interpretation of the passage. In a retrospect interview with the participants, the researchers found that some of them thought that the text reflected the interests of a class and a race over the interests of another less powerful, resulting in what can be considered a divergent or insurgent reading: "it is about Black people, who are the 'monkeys' 'on the rampage' in the White people's home", "it is about who owns the land", "it is about violence in our society" (Norton & Stein, 2000, p. 244).

The researchers, surprised by the alternative interpretation based more on a metaphor than on the factual information, concluded that their intended meaning was a dominant

reading, with meanings that hegemonically frame text interpretation, based on assumptions of knowledge shared by the writer and the intended audience. They also concluded that divergent or insurgent readings may occur, produced at specific time and space, which challenges meanings that hegemonically frame text interpretation.

Considering the alternative interpretation legitimate, they explain that there is an inherent instability of textual meaning in that it is the product of its social occasion, a

> complex tapestry in which the status of the participants, their use of body language, their race (among other characteristics), the time and place of interaction, and the purpose of the interaction have a direct bearing on the social meaning of the texts apprehended within the occasion (Norton & Stein, 2000, p. 244).

In fact, many studies confirm that meaning is social, contextualized, situated. Gee (2000) briefly presents many areas which are converging to the idea that mind is social, with the consequence that thinking and communicating are also defined as socially shaped. In sociohistorical psychology, thinking is mediated through tools historically and culturally shaped for certain functions and meanings. Cognitive linguistics is based on the idea that the way language is organized shapes the way we interpret the world. Modern sociology stresses that thinking is structured by institutional forces leading to human social practices. And in poststructuralist and postmodern work, thinking and acting are thought around discourses, socially and culturally formed.

Based on all these studies, and furthering the discussion, Gee (2000) concludes that meaning is situated in specific sociocultural practices and experiences, and that thinking and using language is assembling situated meanings, defined based on one's sociocultural experiences and specific shared practices, routinized and normed by the groups and their members.

Nonetheless, the instability of the textual meaning may be under some control of the writer, who makes all the efforts to limit the use of inferences on the part of the reader. Ashcraft (1994) refers to this process as implication, used by the writer in the expectation

that readers will draw certain conclusions through inference making[5] based on the information in the text. He recognizes, however, that inferences made during a communication event can be authorized and unauthorized.

Successful communication, according to the author, relies on authorized inference, whereas unauthorized inference will lead to some miscommunication. It is possible to argue, however, along with Rumelhart (1980), that, although miscommunication will result when readers do not make the expected inferences, hence misunderstanding the author, there will be some understanding(s) of the text as the result of the reading activity.

In sum, multiple readings may result from the interaction of readers' characteristics with texts, being mostly influenced by the different background knowledge, including sociocultural schemata. How is it possible to accommodate this within a test? Or as Urquhart and Weir (1998) put it "the theory insists that the good reader makes sense of the text by supplying knowledge based on his or her own unique experience. The testers, on the other hand, are obliged to look for 'correct answers'" (p. 113).

The question for fairness is: if multiple readings or interpretations are an integral part of comprehension, if there can be different legitimate readings and consequently different legitimate answers, particularly considering the discussion on the divergent or insurgent readings presented above, who has the correct answer? Or rather, is there a correct answer?

3.2. Background knowledge: essential for higher-level processes

[5] The process of drawing connections and conclusions not mentioned in the text, based mostly on background knowledge.

The effects of background knowledge to reading comprehension have been extensively studied. The studies are concerned with whether or not it contributes decisively and in what levels or operations it contributes the most.

Many scholars have shown positive contribution of background knowledge to reading comprehension. Gagné, Yekovich and Yekovich (1993), for example, in their model of reading, claim that background knowledge has its most effect on the inferential comprehension, involving integration and summarization. During integration, inferences necessary to understand the information only implicitly present in the text are made to connect two or more propositions occurring within sentences, across sentences and across paragraphs, based on the rules of pronominal reference or on the activation of some schemas.

Consider the example, similar to the one provided by the authors: *The dog walked towards John. He ran for shelter*. Integrating these sentences requires making inference based both on the rule of pronominal reference and on the existing schema that dogs represent danger. Integration organizes the new information present in the text by building a coherent meaning representation with the help of existing knowledge.

The summarization processes are implied in the production of a macrostructure that expresses the main ideas of the text. Conceptual understanding of the topic of the passage, essential for this process to occur, is involved in the inference of the missing relevant information to connect large amounts of information. The production of the macrostructure involves a system to find relations among concepts, through the identification of the topic and the search for the relations among concepts, based on hierarchically structured propositions, on the activation of associated concepts, and on the convergence of the activated concepts into a summary statement. Summarization, like integration, organizes the new information by building a coherent meaning representation with the help of existing prior knowledge.

Research has also shown the positive contribution of background knowledge for reading comprehension, which may be direct as well as indirect. The direct contribution refers to the provision of an organized structured knowledge for the resolution of what is only implicitly given in the text. During comprehension, knowledgeable readers will: a) map the incoming relevant information onto already existing knowledge structure, allowing for the encoding and retrieval of domain-related information more readily and the interpretation of events in a unified manner (Chiese, Spilich & Voss, 1979); and b) know in advance how facts are interrelated, being able to draw the appropriate inference when the relations among facts are left implicit (Just & Carpenter, 1987), relating different propositions in the text, resulting in a more interrelated representation (Smith, 1992).

This is also true for specific cultural knowledge. Knowledgeable readers with specific type of background knowledge in terms of cultural schemata will provide a better integrated understanding and the development of a unified meaning of the text. Pritchard (1990) carried out a research to investigate the role of specific cultural knowledge comparing the readings produced by members of two different cultures for the same texts and concluded that the content of culturally familiar materials made possible the integration of local understandings and the development of a unified meaning of the text.

Knowledge, thus, allows for a more accurate representation of the text contents, since it facilitates the integration of local comprehension and the development of unified meaning of the text, and provides a coherent understanding of the contents at between-sentence, between-paragraph, and between larger units of text (Fincher-Kiefer, Post, Greene & Voss, 1988).

There is also the indirect effect of background knowledge, related to the fact that the provision of an organized knowledge structure will make it possible for the cognitive operations to consume less resource from the limited working memory capacity, i.e.,

knowledgeable readers will demand less resource from working memory (WM) in terms of information computation. The two functions of WM – processing and storage – have a limited pool of resources to share, having to trade off against each other, meaning that "a computationally demanding task may leave less capacity for storing information and vice versa" (Just & Carpenter, 1987, p. 472). In this sense, a knowledgeable reader will have more resources available for storing information longer in an active state, allowing for more relations to be established. Hence, the reader will have more concepts and relations from the previous parts of the text still active in the WM (Daneman & Carpenter, 1980), with the positive consequence that the chunking produced will be richer, more coherent and with qualitatively different information.

Main idea construction will benefit from the use of cognitive resources made available. It gains in processing efficiency for prediction, comprehension monitoring, derivation of word meanings, and assignment of importance (Afflerbach, 1990), underlying processes essential for main idea construction. Other underlying operations benefited are finding referent, processing complex syntactic structures and ambiguities in garden-path sentences, relating predictive pairs (Tomitch, 2000), and achieving balanced inference strategies in terms of trade-off between information needed for local and global coherence (Whitney, Ritchie & Clark, 1991).

In sum, there is enough evidence that background knowledge has a positive effect on comprehension. The more the background knowledge, the more efficient the underlying processes of reading, particularly integration and summarization. Conversely, the poorer the background knowledge, the less integrated the information, the weaker the meaning constructed.

All in all, in the case of language tests, considering the effects of background knowledge in allowing for multiple readings and even legitimate divergent readings, and in

providing gains in processing efficiency for a more integrated and unified representation, as presented above, there seems to be reasons enough to minimize the effects of background knowledge to find evidence of the relevant factor of language ability. In so doing, features of fairness are enhanced and biased discrimination is reduced.

4. MINIMIZING THE EFFECTS OF BACKGROUND KNOWLEDGE

As aforementioned, reading has been characterized as an activity resulting from both lower- and higher- level processes. The higher the processes, the more background knowledge is involved. Once test takers will have different background knowledge, having its effects on a language test can be considered unfair, biased in favor of those who share it and against those who do not share it. Fairness in testing can, thus, be best achieved if its effects are minimized.

Two possible ways may be considered to minimize the effects of background knowledge for the purpose of enhancing fairness in language testing used for selection, that is, to comply with the features of fairness: by topic and by item level. By topic implies choosing one specific topic for the texts used as part of the assessment instruments, and make it public in advance so as to allow for similar opportunities for all test-takers to become knowledgeable on the topic to be able to really understand the test texts and perform at their best in a language test.

By item level implies choosing items demanding specific levels of cognitive operations involved in the reading process. Gagné, Yekovich and Yekovich's (1993) account of the reading processes can be used. In their account, the underlying lower-level processes

are decoding[6], lexical access, and parsing whereas the higher-level processes are the inferential processes of integration, summarization, and elaboration.

Both summarization and elaboration, the highest levels, involve background knowledge to an extent further than the advocated in this article as possible for fairness. Summarization is the process leading to main idea construction[7], and involves background knowledge for the selection of relevant information and for the organization of the propositions in a hierarchical outline. Elaboration is when the reader brings his prior knowledge to make some sense out of the new information presented in the text aiming at acquiring declarative knowledge[8]. Test items requiring summarization and elaboration will not comply with the features of fairness.

Lexical access and parsing, on the other hand, are at the level advocated here. Lexical access is responsible for the identification and selection of the appropriate meaning of the words. Parsing uses the syntactic and linguistic rules of the language to derive meaning from larger units of meaning, such as the meaning of a phrase, a clause or a sentence. Any item requiring lexical access and parsing will comply with the features of fairness.

The decision for choosing items requiring integration is more complex. Integration is the process of connecting the propositions together, which results in a coherent representation of the ideas in the text. Readers will have to relate some elements given as separate in the text, with no explicit mention of the relation. Integration requires, thus, referencing[9], the process of connecting elements in the text, of establishing intersentential relations. Integration also requires inferences for the construction of a coherent mental

[6] Decoding is mostly perceptual, thus not a process to be measured in testing for the purposes discussed here.
[7] Pearson and Johnson (1978) consider the construction of the main idea as the highest level in their taxonomy.
[8] See Ballstaedt and Mandl (1984) for a study on the role of elaborations.
[9] Examples of referencing are given as appendix 3.

representation. This is the highest level possible advocated here to comply with features of fairness.

However, inferences required for integration may be at two levels: at the propositional level, inferences based on the information recoverable from the text, called propositional inferences, and at the pragmatic level, based on information not recoverable from the text, called pragmatic inferences.

To understand this difference, consider the examples given by Hughes (2003). Propositional inference refers to the inference made to understand that Harry was working at her studies, based on the information *Harry worked as hard as she had ever done in her life. When the exam results came out, nobody was surprised that she came top of the class*. Making the required inference is mostly a matter of relating the propositions: Harry got a good grade in the exam because she worked hard, where grade and exam are associated with the schema of school which defines the meaning of work as the activity of studying.

Pragmatic inference[10], on the other hand, refers to the inference made to understand that some drivers drove fast or slowly with the information *it took them twenty minutes by road to get from Reading to Heathrow airport*. The inference is based on the knowledge of the distance between the two places, Reading and Heathrow airport. It can be inferred that they drove fast if it is known that the places are very distant and that they drove slowly if it is known that the places are close to each other. If the specific information about the distance is not part of the reader's background knowledge, such inferences are not possible.

When the pragmatic inference is based on factual knowledge as the example of the airport above, readers will be able to understand if they have the required knowledge. Conversely, if they do not have it, inference will not be made and comprehension will be

[10] Pragmatic inference is based on pragmatic knowledge. Bachman and Palmer (1996) define pragmatic knowledge as referring to the characteristics of the language use context, being divided in functional (intentions) and sociolinguistic knowledge (appropriate use and cultural reference).

impaired. On the other hand, when the pragmatic inference is based on conceptual knowledge, various plausible inferences may occur.

In their explanation of the nature of comprehension, Pearson and Johnson (1978) mention plausible inference, determined by the default value dictated by the schema with the slots constraining the inferences. To explain the process of inference drawing, they provide the following short story of a situation in a restaurant:

> John went to Vescio's, his favorite Italian restaurant. He ordered lasagna. When the waiter brought it, John was so enraged that he left without leaving a tip. He even forgot his umbrella.

The authors propose some questions whose answers will require inferential comprehension. Considering the question *why did John go to Vescio's?*, the authors point out that possible answers are *because it was his favorite restaurant*, requiring only literal comprehension, or *because it was convenient for him, since it is close to his workplace*, requiring plausible inference. The plausible inference is, in fact, pragmatic inference, since inference making relies on information not recoverable from the text about the distance separating John from the restaurant. The inference is considered only plausible, since more than one answer is possible for the question.

This plausibility criterion is, according to the authors, what distinguishes between plausible inferences and wrong inferences, called intrusions in that no argument can be given for them to be considered plausible answers. In their account, intrusions seem to be easily identifiable because of what is plausible to expect as default to fill the slots, and also because deviations from the default are usually mentioned by the writer.

However, the target reader with specific knowledge presupposed by the text writer may be different from the real readers who may not share the presupposed knowledge. This may be particularly true when the reader is a test taker and does not choose the texts to read.

Since the reasoning underlying the answers to inferential questions involve shared knowledge and generalizations about the world, not necessarily shared by the test takers, different reasonings may occur, resulting in different answers.

Pragmatic inference relies on factual knowledge and on conceptual knowledge. In both cases, it relies on the idea of shared knowledge. The question for fairness now is: Is it possible to presuppose shared factual or conceptual knowledge for readers/test takers? In a language test whose construct does not include background knowledge, the logical answer is no.

As we move from lower-level to higher-level processes, we are increasing the contribution from the reader's background knowledge. Since this knowledge is always different and unknown, we are moving from stability to volatility (Pearson & Johnson, 1978). This must be considered within test development to choose items which focus on the propositional inferences as the highest level to comply with the features of fairness.

5. FAIR FOCUS IN LANGUAGE TESTS: READING TASKS

Nuttall (1996) describes reading task and divides them into reading for plain sense and reading into discourse. Reading for plain sense involves understanding syntax, recognizing and interpreting cohesive devices, and interpreting discourse markers whereas reading into discourse is concerned with what the writers mean by what they say, that is, what is either presupposed or implied by the writer.

Since the aim of the article is to pursue items with features of fairness, tasks under reading into discourse are not presented, in that they are mostly top-down oriented, involving pragmatic inferences. Rather, it will present the mostly lower-level tasks of reading for plain

sense. Although test items may focus on any of the tasks, some of them may deserve the focus, namely, tasks for a) establishment of reference; and b) for understanding syntax in terms of establishing the boundaries of each clause, of recognizing the constituents of a noun phrase, and of identifying participles and infinitives in non-finite clauses.

The process of establishing reference, called by the author understanding cohesive devices, is divided in the following subtasks: interpreting proforms and ellipsis, and establishing lexical cohesion. Proforms are words such as *it*, *our*, *this*, *one*, *so/not* (as in I think so/not), and comparatives (smaller, same, other). They can be anaphoric and cataphoric[11]. Ellipsis is the omission of certain words already mentioned, used by the writer to avoid unnecessary repetition.

The establishment of lexical cohesion may be in terms of synonyms, hyponyms, text-structuring words and pin-down words. Both synonyms and hyponyms are used to avoid repetition. Text-structuring words are lexicalized within their context, that is, the reader must refer to some information previously stated. Examples are *issue, methods*, *events*, *views*, *explanations* and *phenomena*. Pin-down words refer to propositions, thus carrying their underlying propositional meaning, for example the word *approach.*

Test items may assess readers' ability to establish reference, i.e, to relate some words to some information previously given in the text. The ability will involve background knowledge to the level that a reader is trained (Pearson & Johnson, 1978), that is, a reader knows that certain words are related to some referent backward or forward, and that some words are omitted to avoid repetition.

Tasks for understanding syntax involve a) the establishment of the boundaries of each clause, b) the recognition of the constituents of a noun phrase, and c) understanding participles and infinitives in non-finite clauses. The establishment of the boundaries of each

[11] Anaphoric is a reference to a previously stated word. Cataphoric is a reference to a word used later in the text.

clause requires the identification of main verbs and other finite verbs, as well as the subject and object. Readers must be able to recognize all these elements to parse the sentence and know how the words relate to one another. In Nuttall's words, get the signification of the clauses and sentences. Test items may focus on the identification of anyone of the sentence elements and will provide evidence about test takers' ability to relate information under the constraints of the language system, not based on random selection of segments of the text.

Adding emphasis to some of Nuttall´s tasks for understanding syntax, items may assess readers' ability to identify the various functions of the suffix *ing*, which makes the grammar system of English rather complex. Test item may focus, for example, on the understanding of participles in clauses as in the sentence *smoking is bad for your health*, as well as on its uses after a preposition or as an adjective/modifier. Emphasis is justified once this is a source of difficulty and misunderstanding for Brazilian readers.

Likewise, items may focus on readers' ability to recognize and understand the constituents of a noun phrase, particularly to identify their heads, which actually express the substance of the phrase. The emphasis in this case is justified once noun phrases have different formation in English and in Portuguese, thus being a source of difficulty and misunderstandings for Brazilian readers.

Test items must provide evidence for the discrimination of test takers according to their levels of the relevant ability. They are only considered good items if they have high discrimination indices. Items focusing on propositional inferences, and on the tasks of establishing reference and understanding syntax are likely to allow for high discrimination indices, providing essential information for the interpretation of relevant factor of reading ability in English as a foreign language, when dealing with heterogeneous groups, particularly for the purpose of selection. And most importantly, complying with features of fairness.

6. FINAL REMARKS

In considerations of test use, today, justifications must be provided with evidence and/or argumentation as to relevance and to consequences. This claims for judgmental questions such as why should something be measured? The judgmental question discussed in the article is why to assess high-level processes once there may be the unintended consequence of test bias, that is, discrimination of test takers on the basis of the irrelevant factor of background knowledge rather than language ability?

As one moves from lower-level processes to higher-level processes in reading, one moves from stability to volatility (Pearson & Johnson, 1978). As one moves from assessing lower-level processes to assessing higher-level processes, one also moves from stability to volatility. And when this volatile arena is being assessed, uncertainty arises, bringing with it power relations[12], that is, someone must have the correct readings to provide the correct answers. And when there is power relation due to different social statuses, test takers may become submissive and may seek to provide answers they think will be considered correct understandings, affecting directly their answers in the test (Norton & Stein, 1998).

As aforementioned, background knowledge will always affect reading comprehension. Readers will have different background knowledge and construct somewhat different interpretations of the same text, resulting in multiple mental representations. Traditional testing cannot accommodate multiple readings. It is mostly based on the positivist/psychometric paradigm, which allows for value-neutral interpretations of reality, a reality "governed by immutable laws and mechanisms that are essentially independent of

[12] See Shohamy (2001) for an extensive discussion on power relations and testing situations.

who, when, and how it is being examined" (Hamp-Lyons & Lynch, 1998, p. 259). It does not incorporate the idea that meanings cannot be determined in advance without considering the socio-cultural aspects of the test takers, the particular assessment contexts, and that multiple readings or even insurgent readings may result as the outcome of a reading activity.

Based on what has been presented, it seems logical to claim that the effects of background knowledge must be minimized, hence reducing the possibility of muddied measurement (Urquhart & Weir, 1998), with reading-irrelevant factors associated with the test items. Since the role background knowledge plays on reading comprehension has been shown to be essential, one possible way of achieving fairness seems to be by choosing test items which have low demands on this factor.

For the purpose of assuring fairness in language assessment, remaining on the stability seems to be defensible, that is, assessing text model up to the level of propositional integration involving propositional inferences, or plain sense of the language. Reading in a foreign language has been accounted for as an interactive process (Rumelhart, 1977, 1980; Eskey, 1988), that is, it also involves lower-level processes or bottom-up processing. A good reader makes inferences, construct main idea(s), analyzes passage organization, recognizes author's tone and style, but also constructs literal meaning, being sensitive to semantic and syntactic cues. A good reader has linguistic knowledge for efficient syntactic and lexical processing necessary for the integration of information across sentences and, hence, necessary for the construction of coherence in a text (Zwaan & Brown, 1996).

Human communication is an interplay involving people. Due to its principle of efficient communication, information that can be presupposed or inferred is usually omitted (Nuttall, 1996). A writer will not include in the text what may be presupposed for the reader, which is usually part of shared knowledge and experience, shared opinions, attitudes and emotions. Also, the writer will not include in the text, but will imply, the information

expected for the reader to be able to infer. For communication to take place, the reader must be able to reconstruct the presuppositions and draw the expected unstated conclusions.

In this interplay, background knowledge resulting from membership in specific socio-cultural groups will play a role, leading to interpretations. Language tests used for selection do not seem to be able to accommodate that, and cannot have the unintended consequence of discriminating based on the irrelevant factor of socio-cultural membership.

7. REFERENCES

Afflerbach, P. (1990) The influence of prior knowledge on expert readers' main idea construction strategies. *Reading Research Quarterly*, p. 31-46.

Ashcraft, M. (1994) *Human memory and cognition*. Harper Collins College Publishers.

Bachman, L. (1990) *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.

Bachman, L. & Palmer, A. (1996) *Language testing in practice*. Oxford, UK: Oxford University Press.

Ballstaedt, S. & Mandl, H. (1984) Elaborations: Assessment and analysis. In: H. Mandl, N.L. Stein & T. Trabasso (eds.) *Learning and comprehension of text*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Chaui, M. (2003) *Convite à filosofia* (13ª edição). Editora Ática.

Chiese, H. Spilich, G. & Voss, J. (1979) Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, v. 18, p. 257-273.

Clapham, C. (1996) *The development of IELTS: A study of the effect of background knowledge on reading comprehension*. Cambridge, UK: Cambridge University Press.

Cumming, A. (1996) Introduction: the concept of validation in language testing. In A. Cumming & R. Berwick (eds.) *Validation in language testing*. UK: Cromwell Press.

Daneman, M. & Carpenter, P. (1980) Individual differences in working memory and reading. *Journal of verbal Memory and Verbal Behavior*, v. 19, p. 450-466.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. & McNamara, T. (1999) *Dictionary of language testing*. Cambridge, UK: Cambridge University Press.

Eskey D. E. (1988) Holding in the bottom: an interactive approach to the language problems of second language readers. In: P. L. Carrell, J. Devine & D. E. Eskey (eds.) *Interactive approaches to second language reading*. Cambridge: Cambridge University Press.

Fincher-Kiefer, R., Post, A., Greene, T., & Voss, J. (1988) On the role of prior knowledge and task demands in the processing of text. *Journal of Memory and Language*, v. 27, p. 416-428.

Gagné, E., Yekovich, C. & Yekovich, F. (1993) *The cognitive psychology of school learning*. New York: Harper Collins College Publishers.

Gee, J. P. (2000) Discourse and sociocultural studies in reading. In: M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (eds.) *Handbook of reading research*, v. 3, p. 195-207, New Jersey, USA: LEA.

Gernsbacher, M. A. (1997) Two decades of structure building. *Discourse Processes*, n. 25, p. 265-304.

Grabe, W. (1999) Develpoments in reading research and their implications for computer-adaptive reading assessment. In: M. Chalhoub-Deville (ed.) *Issues in computer-adaptive testing of reading proficiency*. Cambridge, UK: Cambridge University Press.

Grabe, W. (2000). Reading research and its implications for reading assessment. In A. J. Kunnan (Ed.) *Fairness and validation in language assessment*. Cambridge, UK: Cambridge University Press.

Hamp-Lyons, L. & Lynch, B. K. (1998) Perspectives on validity: a historical analysis of language testing conferences abstracts. In: A. J. Kunnan (Ed.) *Validation in language assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Heberle, V. (2000) Critical reading: integrating principles of critical discourse analysis and gender studies. *Ilha do Desterro* n. 38, p. 115-138. Florianópolis, SC: Editora da UFSC.

Hughes, A. (2003) *Testing for language teachers* (2nd ed). Cambridge, UK: Cambridge University Press.

Just, M. & Carpenter, P. (1984) Reading skills and skilled reading in the comprehension of text. In: H. Mandl, N.L. Stein & T. Trabasso (Eds.) *Learning and comprehension of text*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Just, M. & Carpenter, P. (1987) *The psychology of reading and language comprehension*. Boston, MA: Allyn and Bacon.

Kintsch (1988) The role of knowledge in discourse comprehension: a construction-integration model. *Psychology Review*, n. 95, p. 163-82.

Kintsch, W. (1998) *Comprehension: A paradigm for cognition*. New York: CUP.

Kintsch, W. & van Dijk, T. (1978) Toward a model of text comprehension and production. *Psychological Review*, v. 85, n. 5, p. 363-394.

Kunnan, A. J. (1998) Approaches to validation in language assessment. In: A. J. Kunnan (Ed.), *Validation in language assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Kunnan, A. J. (2000) Fairness and justice for all. In: A. J. Kunnan (ed.) *Fairness and validation in language assessment*. Cambridge, UK: CUP.

Lorch, R. & O' Brien, E. (1995) *Sources of coherence in reading*. New Jersey: LEA

Lorch, R., Klusewitz, M. & Lorch, E. (1995) Distinctions among reading situations. In R. Lorch & E. O' Brien (eds.) *Sources of coherence in reading*. New Jersey: LEA.

Norton, B. & Stein, P. (1998) Why the 'Monkeys Passage' bombed: Tests, genres, and teaching. In A. J. Kunnan (ed.) *Validation in language assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Nuttall, C. (1996) *Teaching reading skills in a foreign language* (new edition). Oxford, UK: Macmillan Education.

Pearson, P. D. & Johnson, D. D. (1978) *Teaching reading comprehension*. New York: Holt, Rinehart and Winston.

Pritchard, R. (1990) The effects of cultural schemata on reading processing strategies. *Reading Research Quarterly*, v. 25, n. 4, p. 273-295.

Rumelhart, D. E. (1977) Toward an interactive model of reading. In: Dornic, S. (ed.) *Attention and performance* VI. New York: Academic Press.

Rumelhart, D. E. (1980) Schemata: the building blocks of cognition. In: R. J. Spiro, B. C. Bruce, & W. F. Brewer (eds.) *Theoretical issues in reading comprehension*. Hilldale, N J: Erlbaum.

Scaramucci, M. (2002) Entrance examinations and TEFL in Brazil: A case study. *Revista Brasileira de Lingüística Aplicada*, v. 2, n. 1, p. 61-81. Belo Horizonte, MG: Faculdade de Letras da UFMG.

Smith, E. & Swinne, D. (1992) The role of schemas in reading text: a real-time examination. *Discourse Processes*, v. 15, p. 303-316.

Shohamy, E. (2000) Fairness in language testing. In: A. J. Kunnan (ed.) *Fairness and validation in language assessment*. Cambridge, UK: CUP.

Shohamy, E. (2001) *The power of tests: A critical perspective on the uses of language tests*. Pearson Education Limited.

Tomitch, L. (1991) Schema activation and text comprehension. *Fragmentos*, v. 3, n. 2, p. 9-43.

Tomitch, L. (2000) Designing reading tasks to foster critical thinking. *Ilha do Desterro* n. 38, p. 83-90. Florianópolis, SC: Editora da UFSC.

Tumolo, C. H. S. (2005) *Assessment of reading in English as a foreign language: investigating the defensibility of test items*. Unpublished dissertation. Universidade Federal de Santa Catarina, Florianopolis, SC, Brazil.

Urquhart, S. & Weir, C. (1998) *Reading in a second language: process, product and practice*. Longman.

van Dijk, T. A. & Kintsch, W. (1983) *Strategies of discourse comprehension*. New York: Academic Press.

Whitney, P. Ritchie, B. & Clark, M. (1991) Working-memory capacity and the use of elaborative inferences in text comprehension. *Discourse Processes*, n. 14, p. 133-145.

Zwaan, R. & Brown, C. (1996) The influence of language proficiency and comprehension skills on situation-model construction. *Discourse Processes*, n. 21, p. 289-332.