EXPLORING THE MEASUREMENT OF VOCABULARY SIZE TO DIFFERENTIATE BRAZILIAN PORTUGUESE-ENGLISH BILINGUALS' ACCESS TO GRAMMATICAL KNOWLEDGE IN THE L2[1]

por Ricardo Augusto de Souza (UFMG)** e Jesiel Soares-Silva (UFMG)*

ABSTRACT

This study explored the validity of a measure of vocabulary size – the Vocabulary Levels Test (VLT) – as a predictor of Brazilian Portuguese-English college level bilinguals capacity to access grammatical representations when using their L2 under temporal pressures. We compared performances in the VLT and in a global test of ability in English L2: the Oxford Placement Test (OPT). Afterwards, participants performed a speeded acceptability judgment task in English, in which a ceiling of 8 seconds was set for each judgment call. Results show that only those participants who classified as high proficiency in both the VLT and the OPT were capable of detecting grammatical violations. We interpret the results as indicating that a measure of vocabulary size is a predictor of both fluency in lexical access and fluency in grammatical knowledge access of L2 speakers.
KEYWORDS: Vocabulary size; acceptability judgments; automaticity; L2 proficiency.

RESUMO

Este estudo explorou a validade de uma medida de tamanho de vocabulário – o Vocabulary Levels Test (VLT) – na previsão da capacidade de acesso a representações gramaticais de bilíngues do português do Brasil e do inglês durante o uso de sua L2 sob pressões temporais. Os desempenhos no VLT e em um teste global de habilidade no inglês como L2 – o Oxford Placement Test (OPT) foram comparados. Em seguida, os participantes desempenharam uma tarefa de julgamento de aceitabilidade em inglês temporalizada, na qual um teto de 8 segundos foi estabelecido para a emissão dos julgamentos. Os resultados demonstraram que somente aqueles participantes que foram classificados como de alta proficiência tanto no VLT quanto no OPT foram capazes de detectar as violações gramaticais. Os resultados são interpretados como indicativos que uma medida de tamanho do vocabulário prevê tanto a fluência no acesso lexical quanto a fluência no acesso ao conhecimento gramatical de falantes de L2.
PALAVRAS-CHAVE: Tamanho de vocabulário; julgamentos de aceitabilidade; automaticidade; proficiência em L2.

* ricsouza.ufmg@gmail.com

**Universidade Federal de Minas Gerais, Brasil

1. INTRODUCTION

It is broadly known that L2 proficiency pairs up with language dominance as fundamental constructs in bilingualism studies. The former construct entails specifications of observable L2 ability and fluency, whereas the latter entails patterns and preferences of use of languages in everyday life. In studies with L2 speakers, participants' L2 proficiency level is very often taken as an independent variable, and researchers often take language ability profiles as a screening factor for between-subject comparisons. Although employment of some sort of measurement of L2 proficiency and/or language dominance has been a common practice in the scientific investigation of bilingualism, such practice has been for some time criticized for being at many times inconsistent and lacking sufficiently powerful generalizability (Grosjean, 1998; Hulstijn, 2012). This concern has led several researchers to consider the study of measures of language dominance (e.g. Dunn & Fox Tree, 2009; Bedore et al., 2012) and measures of specific dimensions of L2 proficiency (e.g. Alderson, 2005; Souza, Duarte e Berg, 2015) as a primary goal in the L2 research agenda.

A measurable aspect of L2 proficiency that has been amply studied is L2 lexical knowledge (Nation, 1990; Meara, 1996; Laufer & Nation, 1999; Read, 2000; De Groot, 2011). Tests explicitly measuring L2 lexical knowledge may focus on vocabulary size or on the amount of representational detail and connectedness in the mental lexicon (or vocabulary depth). According to Meara & Alcoy (2010), one of the most widely accepted instruments to measure vocabulary size in Nation's (1990) *Vocabulary Levels Test* – the VLT.

Souza, Duarte & Berg (2015) have conducted a study to establish the validity of cut-points of a timed version of the VLT among Brazilian college level speakers of English L2. The primary goal of the present study is to further validate the timed version of the VLT by investigating its ability to discriminate L2 proficiency profiles in relation to performance in a psycholinguistic task: a timed acceptability judgment task in which both grammatical and ungrammatical sentences in English were displayed as stimuli. Acceptability judgment tasks may vary in stimuli presentation mode and task requirements. But there is evidence that more than one of its varieties can yield data that reflects differences in the state of L2 representations (R. Ellis, 2005; Souza & Oliveira, 2014). Therefore, the acceptability judgment is a behavioral off-line task that is likely to be sensitive to differential profiles in L2 proficiency.

There were also two specific objectives in the present study. We sought to compare the discriminatory behavior of VLT scores vis-à-vis an objective written test measuring global proficiency in English L2. The scores of this test – the Oxford Placement Test (OPT) – roughly meet a broadly accepted framework of L2 proficiency: the Common European Reference Framework (CERF) (Council of Europe, 2001). Furthermore, we sought to estimate the minimum time window within which bilinguals (Portuguese/English) could make accurate judgment calls about a sentence in English. To achieve this specific goal, we mostly replicated the design of the study reported by Souza et al. (2015), in which the authors have established the mean least it took monolinguals of both English and Portuguese with a college level education to accurately judge the grammaticality of sentences in their L1.

In the next section, we discuss the construct of L2 proficiency as a cognitive trait and the significance of objective and practical measures of L2 proficiency in the context of experimental L2 research. Afterwards, we briefly revise models that tap into the role of L2 vocabulary size as a component of L2 proficiency. We then pass over to describing the methods and the analyses of data we compiled in the present study. We finish this paper with a presentation of our interpretation of our findings, and a brief discussion of what we believe to be the very next steps to be taken in our research agenda.

2. THE CONSTRUCT OF L2 PROFICIENCY AND ITS ROLE IN SECOND LANGUAGE RESEARCH

Proficiency in a second language (L2) may be generally understood as the ability of fluent use of that language. This definition frames L2 proficiency as associated with global achievement in communicative situations involving L2 use. By this definition the construct of L2 proficiency is not typically conceived of as a single unidimensional cognitive trait, but rather as resulting from interactions among multiple components (Bachman, 1990; Hulstijn, Anderson & Schoonen, 2010). This global construct may be operationalized as correlations among measures that tap into linguistic knowledge, and which are dependent on systematic protocols for the observation of language-related behavior. Such measures are intended to reflect an individual's latent capacity for efficient implementation of L2 knowledge in linguistic performance that is both formally accurate and contextually appropriate.

The picture of proficiency as a global effect of interactions of specific knowledge repositories and skills has a long tradition. As reviewed by Hulstijn (2015), in the early 1960s a framework for analysis and assessment of L2 proficiency emerged from work by Robert Lado and John Caroll. This was a two-dimension framework that crossed linguistic components – namely knowledge of phonology or orthography, syntax, morphology, and the lexicon – with both receptive skills (reading and listening comprehension) and productive skills (speaking and writing). As can be seen in Bachman (1990), with the rise of communicative perspectives on the conceptualizations of linguistic knowledge and of language education the breadth of levels of linguistic organization covered by frameworks of L2 proficiency was later widened by the incorporation of pragmatic, sociolinguistic and discourse-oriented features.

Hulstijn (2011) proposes a model that captures the multidimensionality of language proficiency of both L1 and L2 speakers. According to such model, the construct of overall language proficiency for either L1 or L2 can be split into "basic" and "higher language cognition". The author's notion of basic language cognition (henceforth BLC) is compatible with Ullman's (2001) declarative/procedural model for the neurocognitive base of language representation, as the notion encompasses implicit procedural computations of linguistic knowledge (phonology/phonetics, morphology, and syntax), as well lexical representations stored in declarative memory. Higher language cognition (henceforth HLC), on the other hand, encompasses relatively low frequency, more complex, and slower to process lexical items and morphosyntactic structures. Such items and structures are likely to occur in modalities, genres and registers associated with specific and specialized types of experiences and lifestyles, such as educational or workplace affiliations.

Hulstijn's (2011) proposal is also attuned with crucial differences between the ontogenesis and end-state of L1 and L2 development. The model predicts that whereas L1 speakers universally attain ceiling proficiency in BLC, high levels of HLC attainment among L1 speakers is likely to depend on interactions between individual differences in cognitive capacity and particular trajectories in learning opportunity. With respect to L2 proficiency, on the other hand, whether or not even BLC is fully attainable after critical age periods is a question open to debate. For bilingual individuals, L2 proficiency attainment is highly variable, and to varying degrees it is affected by individual differences such as age of onset of L2 learning, language learning aptitude, overall cognitive capacity, personality profiles, affective attitudes towards the L2, L2 learning circumstances, and personal histories of experience with the L2 (Harley et al., 1990; Dörnyei, 2005).

Hulstijn (2011) argues that in basic language cognition the linguistic components are added by high automaticity in processing. The author argues that this is a feature that may be related to the fact that the overt linguistic manifestations of BLC tend to be high frequency linguistic forms, which also tend to be shared across discourse modalities, genres and registers. Therefore, proficiency in BLC may be construed as deriving from the ample opportunities that mature language users will have had to process such forms and their meanings for both comprehension and production throughout their linguistic histories. We align ourselves with Hulstijn's proposal, as we understand automaticity in processing to be an essential component of high L2 proficiency.

Automatic language processing has been defined as opposed to controlled processing. The latter requires more attentional efforts than the former. Controlled processing is also more based on strategic decisions and it is more time consuming than automatic processing (Segalowitz, 1991; Segalowitz & Hulstijn, 2005). According to Segalowitz & Hulstijn (2005), automaticity can be construed as a facet of high L2 proficiency in view of an understanding of language processing as relying on limited cognitive resources. If high L2 proficiency equates with capacity to use an L2 in the performance of complex tasks, then such capacity requires aspects of the processing routines such as the ones involved in linguistic coding and decoding to be less demanding then the processing routines required for reasoning, coherence, pertinence and accuracy of situation assessment. It should be highlighted that the claim that automatic plays a role in language proficiency in no way is construed as equating proficiency with implicit learning of the L2. We understand along with N. Ellis (2005) that both implicitly and explicitly learned linguistic representations can eventually come to be accessed automatically. Ultimately, automaticity is a feature of fluent language use, and as such it is an aspect of language processing rather than a corollary of how consciousness or metalinguistic awareness is hypothesized to have specific impacts on L2 acquisition[2].

Therefore, at earlier stages in the ontogenesis of L2 BLC, in Hulstijn's (2011) terms, the L2 user should be expected to demonstrate less automatic processing than in later stages. Being at an early stage of L2 development may be equated with having had less experience with L2 processing than more developed users. Thus, if the right cognitive profiles and favorable learning situations are found, it is reasonable to expect broader linguistic representations and more automaticity in their process-

ing from more experienced L2 speakers, with experience being understood as a function of sustained practice in use of at least the linguistic manifestations of the L2 that characterize BLC. As stated by Segalowitz & Hulstijn (2005, p. 371), automaticity is "the prime psychological construct invoked for understanding frequency effects and how repetition leads to improvement in L2 skill". In the present study, we attempted at operationalization of some degree of automaticity in our tasks, a procedure we believe should be incorporated into most endeavors to analyze and measure L2 proficiency.

Ultimately, L2 proficiency is a challenging construct to conceptualize and also to measure objectively. Notwithstanding, measurement of bilingual speakers' differential proficiency profiles is a matter of absolutely critical importance for the psycholinguistics of bilingualism. Because of the eminently experimental base of research in psycholinguistics, comparability and replication of results are fundamental for the advancement of the field. Grosjean (1998) states that one of the difficulties that jeopardize consensus in bilingualism studies is the lack of standardized procedures for describing and measuring differences in profile across bilingual populations from which samples are drawn. Bilinguals' linguistic proficiency in both dominant and non-dominant languages is one of such relevant profile differences, according to the author. Also, as pointed by Hulstijn (2012), some kind of measure of linguistic skill level is quite often taken as the main – if not the single – independent variable of experimental studies in L2 acquisition and bilingual language processing. This fact alone should justify L2 scholars' careful theoretical consideration of which facets of L2 proficiency are selected for observation.

The relevance of the conceptualization of L2 proficiency is topped with the need for careful planning of how to measure it in efficient and practical manners in L2 acquisition and processing studies. Issues of practicality of proficiency assessment administration haunt designers of language tests for educational and accreditation purposes. Of course, in experimental laboratory work such issues may be even more critical, as seldom can investigators afford the time required for administration of complex proficiency test batteries. This often leads researchers to employ sections or subtests within standard test batteries, or to use scales constructed to diagnose proficiency by measuring a single dimension or but a few dimensions of the overall construct. It is our understanding that the problems with variability in psychometric instrumentation may be much worsened by lack of validation studies of the scales researchers employ. Because proficiency is a multidimensional construct, we understand that the validity of scales targeting specific dimensions to discriminate profiles in accordance to the variability in linguistic representation and processing that motivate psycholinguistic investigations is ultimately an empirical question. The present study is an attempt to address such empirical question by targeting a test of one specific dimension (vocabulary size) in relation to one specific psycholinguistic task (acceptability judgments).

3. VOCABULARY SIZE AS A MEASURE OF L2 PROFICIENCY.

One of the dimensions of L2 proficiency that has stemmed reasonably practical tests (from the standpoint of test administration) is L2 lexical knowledge. De Groot (2011) asserts that a bilingual's high proficiency level in the L2 is dependent not only on his or her accuracy in grammar, but also on a considerable level of vocabulary knowledge. Also, Hulstijn's (2011) model of split BLC and HLC

does predict variability in lexical knowledge, as discussed above. Furthermore, there are considerable empirical studies suggesting a relation between differential vocabulary knowledge and measurable differences in bilinguals' language proficiency.

When it comes to measuring L2 vocabulary knowledge, more than on dimension can be taken into consideration. In the L2 mental lexicon literature, the dichotomy between "vocabulary breadth" and "vocabulary depth" is oft cited as a descriptor of two broad dimensions in the organization and development of lexical competence (Meara, 1996; Read, 2000; Milton, 2010; Schmitt, 2014). Vocabulary breadth is understood as the quantity of words someone is capable of recognizing and connecting to a core meaning, in other words, an individual's vocabulary size. Vocabulary depth outreaches vocabulary recognition, as it entails at least access to information related to derivational morphology, collocation restrictions, subcategorization frames, membership to semantic fields and classes, and usage restrictions (Milton, 2010, Meara 2009).

According to Schmitt (2014), the research on measures of vocabulary size and depth shows that for learners with small L2 vocabularies and for high frequency words there is no distinction between the two measures, whereas for larger L2 vocabularies and low frequency words bilinguals tend to develop vocabulary depth more slowly than vocabulary size. As the dichotomy between size and depth is probably related to the distinction between receptive (recognition) and productive vocabulary (Schmitt, 2014), what the research suggests is that after a certain threshold of L2 vocabulary size bilinguals may be more able to recognize form-meaning links of L2 words than they are able to accurately use L2 words.

Notwithstanding the distinction between vocabulary breadth and depth, there is ample evidence that vocabulary size measures are consistently correlated to estimates of performance level on the four language skills of speaking, writing, listening, and reading (Alderson, 2005, Milton 2013). Milton (2010) conducted a study in order to analyze the impact of vocabulary as a dimension within the six L2 proficiency levels in the CEFR (Common European Framework of Reference for Languages). The study's focus was on the vocabulary threshold sizes necessary for bilinguals to perform according to the CEFR descriptors. The author compared the CEFR's levels with vocabulary size as measured by the XLex[3] test. The results show that L2 vocabulary range requirements steadily increase as the CEFR's levels move upwards. The correspondences found by Milton (2010) are summarized in Figure 2, where the estimates of vocabulary size refer to lemmatized items, i.e., to word families:

---

3. XLex (Meara & Milton, 2003) is a corpora frequency-based test in which participants have to affirm which words they know from a list. Then, it is calculated how many words for each list (frequency-based) each participant knows.

CERF Level	Vocabulary Descriptors	Vocabulary Size XLex (5000 max)
C2	Broad lexical repertoire including idiomatic expressions and colloquialisms.	4,500 – 5,000
C1	Little obvious searching for expressions. Good command of idiomatic expressions and colloquialisms.	3,750 – 4,500
B2	Vocabulary for matters connected to his or her field and most general topics.	3,250 – 3,750
B1	Sufficient vocabulary to express him/herself with some circumlocutions.	2,500 – 3,250
A2	Sufficient vocabulary to conduct routine, everyday transactions involving familiar situations.	1,500 – 2,500
A1	Basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.	< 1,500

Figure 1: - Vocabulary range criteria from Council of Europe (adapted from Milton, 2010)

The VLT (Nation, 1990; Schmitt, Schomitt & Clapham, 2001) is a five-level test elaborated to measure vocabulary size. It has 5 parts, each containing six items (thus 30 items in the whole test) in which test-takers must produce matching of three out of six words with three definitions. Therefore, each part of the VLT yields a ceiling of 18 correctly matched words. The VLT estimates vocabulary size levels by correspondence between level and word frequencies bands based on the Brown Corpus. Successful completion of level 1 corresponds to knowledge of the 2,000 most frequent words; completion of level 2 corresponds to the 3,000 most frequent words, level 3 corresponds to the 5,000 most frequent words, level 4 is a special section corresponding to academic and scientific vocabulary, and level 5 corresponds to knowledge of the 10,000 most frequent words. According to Nation (1990), the cut-point for successful completion of a VLT level is 12 correct matchings out of the 18 possible ones per level. It should be noted that Souza, Duarte & Berg (2015) report no discriminatory effect for level 4 (academic and scientific vocabulary), which is interpreted as a byproduct of the fact that such lexical domain is heavily made up of Latin-originated words that form cognates with Portuguese words. Figure 2 is a demonstration of the layout of VLT items.

```
1 – business
2 - clock              (  ) part of a house
3 – horse              (  ) animal with four legs
4 – pencil             (  ) something used for writing
5 – shock
6 – wall
```

Figure 2: Example of a question in the Vocabulary Level Test

According to Read (2000) the VLT is as discrete vocabulary test, as the construct underlying the test relies exclusively on vocabulary knowledge (specifically, the meaning of words). The author also analyzes the VLT as being a selective test, as the words were chosen based on corpora frequency. Finally, Read describes the VLT as a context-independent vocabulary test, since the it does not tap into knowledge of situations where words would be likely to occur.

In Nation's (1990) original proposal of the VLT, there is no specification of a time limit for test-takers. However, Laufer & Nation (2001) conducted a study of a computerized vocabulary test based on the VLT in which response latencies were analyzed. The researchers found a moderate and significant negative correlation between vocabulary size and response latency. In other words, raises in vocabulary breadth are related to a higher speed in linking word form and meaning.

In a study to explore the validity of the proposed 12-matching cut-point for Brazilian college level test-takers Souza, Duarte & Berg (2015) implemented a temporal ceiling for completion of the VLT. Such temporal ceiling suggests the administration of the VLT within a time window of 10 minutes, i.e.: 20 seconds per item or roughly 6.66 seconds per definition to be matched with a word. The authors' rationale was the integration of a component of automaticity – namely speed of task execution – to the VLT construct. Souza, Duarte & Berg (2015) suggest that 12, 13 and 14 correct matchings are equally discriminatory cut-points for the VLT. Importantly, the integration of a speed requirement seems particularly useful in light of the authors' finding that the academic vocabulary section (level 4 of the VLT) does not discriminate English L2 vocabulary knowledge of the average Brazilian college student. By introducing the speed limit, the non-discriminatory section may be functioning as a modulator of how far the test-taker will be likely to reach in face of the temporal constraint for execution of the test's task. As described below, it was the administration mode of the VLT described in Souza, Duarte & Berg (2015) that we employed in the present study.

We now pass over to the details of our present exploration of the relationship between measures of L2 vocabulary size and access to L2 grammatical knowledge.

4. MATERIALS AND METHODS

As stated above, the primary goal of the present study was to further validate the exploration of the validity of the Vocabulary Levels Test scores as a proficiency measure for Brazilian Portuguese-English bilinguals' L2 proficiency. Specifically, we sought to (1) replicate previous research results suggesting correlations between vocabulary size and overall proficiency (Alderson, 2005; Milton, 2013); (2) examine the behavior of VLT scores when discriminating test-takers' performance in a timed version of the acceptability judgment task. An ancillary goal of this study was the establishment of a minimal time window for the speeded judgment task with L2 speakers, and in order to achieve such goal we replicated the procedures described in Souza et al's (in press) study with monolinguals.

In order to achieve the goals of the present study, we administered the VLT and a general proficiency test, and we designed an acceptability judgment task covering both word order and syntax-semantics interface phenomena. We describe each of these instruments in the following sub-sections.

4.1 Vocabulary Level Test (VLT)

Following Souza, Duarte & Berg (2015), in the present study participants were allowed up to 10 minutes to complete the VLT to the best of their abilities. In order to pass from one level to another, participants should score at least 14 (78%) of the 18 possible points per section. Therefore, we chose to employ the most conservative cut-point, even though Souza, Duarte & Berg did not observe dif-

ferences between this cut-point and the 12-point cut-point originally proposed by Nation (1990). We considered high-proficient participants to be the ones who could achieve level 5 of the VLT, based on previous studies with Brazilian Portuguese-English bilinguals that employed the same test as a screening method and that identified differential behavior related to high levels of proficiency thus measured (Souza, 2012; Souza & Oliveira, 2014).

4.2 The Oxford Placement Test (OPT)

The paper-version of the OPT (quick version) published by Oxford University Press is a 60-question multiple choice placement test containing questions on grammar, reading comprehension, and language usage. Participants are allowed up to 30 minutes to complete the entire test. In a typical OPT item, participants are required to indicate the best selection to create complete sentences. Below is an example of a grammar question from the OPT:

I don't remember .... the front door when I left home this morning.			
( ) to lock	( ) locking	( ) locked	( ) to have locked

Figure 3: Example of a question of the Oxford Placement Test

The OPT scores roughly place test-takers in the CEFR levels. As previously mentioned, the CEFR levels are A1, A2, B1, B2, C1, and C2 in ascending order. The corresponding scores with the CEFR are: from 0 to 17 points (A1); from 18 to 29 points (A2); from 30 to 39 points (B1); from 40 to 47 points (B2); from 48 to 54 points (C1), and from 55 to 60 points (C2).

Therefore, an OPT score tentatively allows an interpretation of overall proficiency gauged by the "can-do" list proposed by the CEFR. Such "can-do" list specifies the communicative make-up of each of the CEFR levels: A1 level learners can understand and use familiar everyday expressions and very basic phrases; A2 level learners can understand sentences and frequently used expressions. B1 level learners can understand the main points of clear standard input on familiar matters; B2 level learners can understand the main ideas of complex text on both concrete and abstract topics. C1 level learners can understand a wide range of longer, demanding texts, and recognize implicit meaning; C2 level learners can easily understand virtually everything heard or read (Council of Europe, 2001, p. 24).

4.3 Acceptability Judgment task

The acceptability judgment (AJ) task is an offline data collection procedure in which participants are instructed to react to a series of linguistic stimuli by assessing their grammaticality. In this study, we employed the design reported Souza et al's (2015) study of speeded acceptability judgment task with monolinguals of American English and Brazilian Portuguese. Accordingly, in the present AJ task stimuli were presented on a computer screen. Participants were exposed to sentences (presented one by one) in the center of the screen. Then, they judged each sentence using a 5-point Likert scale (Figure 4). Responses were given using the numeric keys of a computer keyboard, and a time limit of 8 seconds was set for the judgment calls.

Numeric keypad	Judgment levels
1	Totally unacceptable
2	Not well-formed, almost unacceptable
3	Not well-formed, but maybe acceptable
4	Slightly ill-formed, almost perfect
5	Totally perfect

Figure 4: Levels of Likert scale to the Acceptability Judgment task, adapted from Souza et al (*in press*).

Our AJ task corpus was composed of 56 sentences, and 16 of them contained grammatical violations. There were two types of sentence violations (8 sentences each): argument structure realization violations involving unergative verbs in transitive syntax, and explicit morpho-syntactic violations involving long-distance dependencies (Wh-movement) and subject-verb agreement.

Argument structure realization violations were chosen because according to White (2003), L2 argument structure may pose a challenge to L2 learners, as "interlanguage lexical representations may not correspond to argument structures encoded in the lexicons of native speakers of the L2" (White, 2003, p. 206). Although unergative verbs do not transitivize in either Brazilian-Portuguese or English, bilinguals with Brazilian Portuguese L1 and English L2 will need to learn which argument structures are licensed in their L2 and which ones are not, as there are several cases of argument structure patterns that are productive in English, but not in Portuguese. The induced-movement alternation (Souza, 2011; 2012), the dativized bi-transitive construction (Zara, Oliveira & Souza, 2013), and the resultative construction (Souza & Oliveira, 2014) are examples.

The sample of ungrammatical sentences because of explicit morphosyntactic violations encompassed 4 sentences with violations in subject-verb agreement (henceforth VAgr), and 4 sentences with violations to WH-movement (henceforth WHm). These too are cases of ungrammaticality in both the L1 and the L2 of the bilingual population we observed. Nevertheless, it has been demonstrated that violations of this type are not necessarily perceived by L2 speakers (Clahsen & Felser, 2006; Jiang, 2007), an observation that has also been replicated with bilinguals of Brazilian Portuguese and English (Carneiro & Souza, 2012). This fact may reflect difficulties with the integration of grammatical knowledge into L2 sentence parsing, which can be caused by lack of automaticity in parsing routines. Therefore, actual detection of such violations under time pressure may reflect more automaticity in access to grammatical knowledge, hence higher fluency and proficiency in L2 use.

As control sentences, we employed a set of 8 sentences instantiating the induced-movement alternation of English (e.g.: The woman jumped her horse over the fence). As stated above, this is a case of argument structure realization that is not licensed in Brazilian Portuguese, but which is learnable by high proficiency Brazilian Portuguese-English bilinguals (Souza, 2011; 2012). By employing this type of sentence, we wanted to check whether the sample of the bilingual population of interest to this study would be capable not only to perceive ungrammaticality in their L2, but also to inhibit a restriction of their L1 that is not applicable to their L2. Such inhibition should take place under the time constraint of our AJ task.

Sentences (1) through (5) below illustrate the type of stimuli employed in our AJ task:

(1) *The man laughed the children during the party. *Transitivized unergative verb*
(2) *The girl give the cats milk twice a day. *Agreement violation:*
(3) * What did Steven read the book that Helen talked about? *WH movement violation:*
(4) The instructor ran the boys around the park. *Grammatical sentence (Induced movement alternation)*
(5) The girls melted the cheese in the bowl. *Grammatical sentence*

3.4 Participants

Thirty individuals took part in the experiment; 17 women and 13 men, with a mean age of 25.6 years (sd=6.03). All of them had completed or are completing at least a bachelor's degree. All of them were right-handed, with good (or corrected) vision, living in Belo Horizonte/MG or Goiânia/GO. All of the participants were bilinguals (Portuguese/English) and were classified into two groups of English proficiency level, based on the OPT and VLT test scores.

The proficiency classification employed both OPT and VLT scores. Following our criteria, in order to be considered high proficient, participants needed to achieve 55 points or more on the OPT (91% is C2 level according to the CEFR descriptors), and also reach and complete level 5 of the VLT (10,000 frequent words). Level 4 of the VLT was not considered. As discussed above, this level was demonstrated not to discriminate the average college level Brazilian Portuguese- English bilingual with regards to L2 vocabulary knowledge (Souza, Duarte & Berg, 2015). Furthermore, as will be shown below, there were no significant differences among participants whose VLT scores placed them in levels 1-3 with respect to their performance in either the OPT or the AJ task. On the other hand, participants whose VLT scores placed them in level 5 showed performance that were significantly different from participants at VLT levels 1-3. Descriptive information about the proficiency-based stratification of our participant sample are displayed in Table 1.

	Tests				
	VLT (Word Frequency level)			*OPT (CREF levels)*	
	2000, 3000, 5000	10000		A2, B1, B2	C2
	proficiency level				proficiency level
	Low	High		Low	High
participants (n=30)	18	12		18	12

Table 1: Participants' proficiency in Vocabulary Level Test (VLT) and Oxford Placement Test (OPT)

5. DATA ANALYSIS AND DISCUSSION

First, an exploratory data analysis was conducted to determine if reaction time (RT) means for the 30 subjects in the Acceptability Judgment task were normally distributed for each target sentence. Results from a Kolmogorov-Smirnov test for normality indicated that the distribution of the RT means

did not deviate significantly from a normal distribution in all four cases. Normality test results, means, and standard deviation are displayed in Table 2:

Sentence type	RT (msec)	Sd (msec)	Statistics
Verbal transitivity violation	4821	.746	.121*
Morphosyntactic violation	5488	.912	.104*
Induced-movement alternation	4809	.807	.143*
Grammatical	5104	1018	.130*

* p > .05 **Table 2**: Means, standard deviation and normality test of reaction time means to sentence type (n=30)

Based on the procedures of Souza et al. (2015), we estimated the maximum time one would take to make a judgment call on the sentences there were exposed to. To do so, we considered the RTs by level of proficiency. The difference of means for all sentences combined (grammatical/ungrammatical) between low proficient (M=5569, SD=.426) and high proficient (M=4509, SD= .655) was significant, t(4) 8, 268, p< .005. We suggest that a value of one RT mean of lower proficient added with a standard deviation is the maximum time necessary in an acceptability judgment under this configuration. Thus, we suggest that 6000 milliseconds is the maximum time that a bilingual (Portuguese/English) takes to judge a sentence written in English with approximately 40 characters (spaces excluded).

We then proceeded to the confirmatory investigation of the correlation of VLT scores and a general proficiency diagnostic measure (the OPT in the present study). A Pearson product-moment correlation coefficient (*r*) was computed to assess the relationship between the two tests in order to verify the degree of correlation between the scores each one produces as diagnosis of L2 proficiency. To do so, we computed the total number of scores reached by lower and high proficients in OPT and in VLT. Our hypothesis was that there would be a positive correlation between tests for each groups of proficiency. The data displayed in Table 3 confirms this hypothesis:

	Low proficient (OPT)			low proficient (VLT)		
N	mean	SD		Mean	SD	**Pearson's *r***
18	35.9	10.3		43.4	5.6	.559*
	high proficient (OPT)			**high proficient (VLT)**		
	mean	SD		Mean	SD	
12	55.7	.86		86.9	2.6	.586*

*p< .05 **Table 3**: Correlation between OPT and VLT proficiency levels

This results shows that the VLT scores indicating high proficiency (VLT level 5 in our definition) correlate significantly with differences in OPT scores. The lower correlation among low proficient subjects on both tests (*r*=.559) can be due to the higher variation among the subjects' performance, as attested by the clearly higher standard deviations observed with the low proficiency group. We interpret this result as satisfactorily validating the claim that a measure of vocabulary size correlates to measures of overall proficiency for the population of interest to the present study, namely college level Brazilian Portuguese-English bilinguals.

In order to verify the correlation among the proficiency tests (the OPT and VLT) and our timed AJ task, we first analyzed the judgments elicited by the collapsed grammatical and the ungrammatical sentences in the two proficiency level groups of the present study. As can be seen in Figure 6 below, we found differences in the behavior of the two groups. Such differences are related to both the grammaticality status of the stimuli and the proficiency level of the participants.
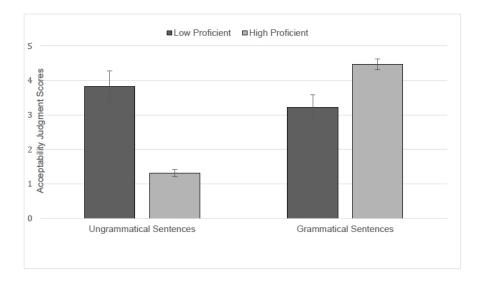

Figure 6: Means of sentence judgments by proficiency level

There are significant differences between the judgment calls for the ungrammatical sentences made by the low proficiency (M=3.8, SD=0.45) group and the high proficiency group (M=1.3, SD=0.10), t(28)=18,97, *p*< .01. This clearly indicates that the low proficiency participants were generally unable to detect the violations of the ungrammatical sentences within the 6-second window frame of our timed AJ task. This observation supports Jiang (2007) proposal that some bilinguals may experience difficulty integrating morphosyntactic information when processing L2 stimuli. We interpret this situation as an indicator of the lower level of automaticity of the low proficiency sample when compared to the higher proficiency sample of the present study. It should be noted that actually the high proficiency participants (M=4.46, SD=0.15) fared better at the identification of the grammatical sentences than the low proficiency group (M=3.21, SD=0.36), t(28)=11,14, *p*< .01.

As discussed below, we interpret this observation as a probable effect of the inclusion of the induced movement alternation sentences among our grammatical sentences. It was only the high proficiency participants' responses that yielded a significant difference between the ungrammatical (M=1.30, SD=0.10) and the grammatical sentences (M=4.46, SD=0.15), t(11)=54,10, *p*< .01. Among the low proficiency participant sample we observed, the pattern actually indicates a tendency for misjudgment, with ungrammatical sentences (M=3.83, SD=0.45) yielding higher mean judgments than grammatical sentences (M=3.21, SD=0.36), t(17)=4.25, *p*< .05. All in all, we interpret the pattern of our results as showing that only the participants whose VLT and OPT scores classify as high proficiency had sufficiently automatic access to their L2 grammatical representations to perform satisfactorily within the average 6-second ceiling of our timed AJ task.

Finally, we analyzed the specific role of each of our target sentence type in our timed AJ task for the two proficiency groups. A repeated-measures analysis of variance of AJ scores means of the low proficient group across the four groups of sentences indicated a main effect of sentence type considering subjects as a random factor, $F1(3,51)=9.45$, $p< .001$, $\eta_p^2 = .357$, and items as a random factor $F2(3,21)=4.20$, $p< .05$, $\eta_p^2 = .375$. The repeated-measures analysis of variance of AJ scores means for the high proficient group also revealed a main effect of sentence type when subjects were taken as a random factor: $F1(3,33)= 1159$, $p< .001$, $\eta_p^2 = .991$, as well as items as a random factor: $F2(3,21)=685$, $p< .001$, $\eta_p^2 = .990$.
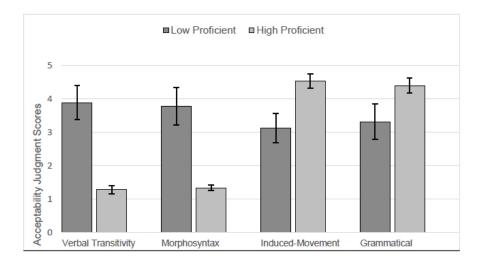

Figure 7: Means of sentence type judgments by proficiency level

However, as can be seen in Figure 7, it is only with the high proficiency group that consistent detection of ungrammaticality as opposed to grammaticality takes place. Therefore, it is clear that the sentence effect yielded among the low proficiency participants in the 6-second ceiling of our task is not driven by access to grammatical representations, some of which are shared by their L1, such as agreement and WH-movement violations. This supports our interpretation that only the participants whose VLT/OPT scores measure high proficiency demonstrate sufficient automaticity to access L2 grammatical representations under strict time constraints.

6. CONCLUSION

In this study we advanced work reported by Souza, Duarte & Berg (2015) aiming at validating a measure of vocabulary size – the Vocabulary Size Test, or VLT (Nation, 1990) – as a diagnostic tool to assess Brazilian Portuguese-English bilinguals at college level. The VLT is a practical, easy to administer test that has been used, together with other tests of vocabulary size, as a screening procedure is published studies of bilingualism and second language acquisition (Hulstjin, 2012). The VLT has also been previously employed with such a purpose in studies with Brazilian Portuguese-English bilinguals (Souza & Oliveira, 2011; Souza, 2012; Oliveira & Souza, 2014).

Specifically, we developed a comparison between VLT scores and another measure of proficiency in English L2, the Oxford Placement Test (OPT). We verified moderate and significant correlations be-

tween the scores of test-takers who achieved the highest level in the VLT (whom we refer to as "high proficiency) and the last level of the OPT, as well as moderate and significant correlations between VLT scores up to level 3 and lower bands in the OPT, which are tentatively associated with CERF labels describing levels of ability in an L2 below full communicative mastery.

Also, we investigated the performance in a timed acceptability judgment task with English sentences of a sample of bilinguals whose VLT scores indicate "high proficiency" in L2 English as compared to a sample diagnosed through the same test as having "low proficiency". This timed task was assigned with a temporal ceiling of 8 seconds, the average performance having taken place within 6 seconds. The results showed a striking contrast between the high proficiency and the low proficiency groups, with only the latter being able to make judgment calls that converge of the L2 grammar.

Three quarters of the violations instantiated in our stimuli for the acceptability judgment task were actually grammatical restrictions that do apply to he bilinguals L1. Therefore, we interpret the overall failure of the low proficiency participants to accurately detect such violations as a failure to fully access grammatical knowledge when using the L2 under strict temporal restrictions. Difficulty to integrate certain types of grammatical information when processing the L2 has been previously suggested to be a factor modulating bilingual language processing (Clahsen & Felser, 2006; Jiang, 2007). It is important to mention that the task employed in the present study does not elicit samples of online language processing. Nevertheless, in light of the requirement for speeded performance of our task, we interpret our results as suggestive that the measurement of large L2 vocabulary size does not only indicate higher fluency in lexical access, as suggested by Laufer & Nation (2001), but also fluency in access to grammatical representation repositories. As fluency comes along unplanned and subliminal performance as dimensions of automaticity (Segalowitz & Hulstijn, 2005), we believe that the measurement of L2 vocabulary can be also indirectly informative of differential profiles in L2 automaticity.

REFERENCES

Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. London/New York: Continuum.

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

Bedore, L. M.; Peña, E. D.; Summers, C. L.; Boerger, K. M.; Resendiz, M. D.; Greene, K.; Bohman, T. & Gillman, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish-English bilingual children. *Bilingualism: Language and Cognition*, vol. 15 (3), p. 616-629.

Clahsen, H.; Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, vol. 27 (1). p. 3-42.

Carneiro, M. M.; Souza, R. A. (2012). Observação do processamento online: uma direção necessário para o estudo experimental da sintaxe bilíngue. *ReVEL*, vol. 10 (18), p. 107-127.

Council of Europe. (2001). *Common European Framework of Reference for Languages*. Cambridge: Cambridge University Press.

De Groot, A. M. B. (2011). *Language and Cognition in Bilinguals and Multilinguals*. New York: Psychology Press.

Dörnyei, Z. (2005). *The Psychology of the Second Language Learner: Individual Differences in Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dunn, A. L. & Fox Tree, J. E. (2009). A quick, gradient Bilingual Dominance Scale. *Bilingualism: Language and Cognition*, vol.12 (3). p. 273-289.

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit knowledge. *Studies in Second Language Acquisition*, vol. 27 (2). p. 305– 352.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in second language acquisition*. vol 27 (2), p. 141-172.

Grosjean, F. (1998). Studying bilinguals: methodological and conceptual issues. *Bilingualism: Language and Cognition*, vol.1 (2), p. 131-149.

Harley, B.; Cummis, J.; Swain, M. &Allen, P. (1990). The nature of language proficiency. In: Harley, B.; Allen, P.; Cummis, J. & Swain, M. (Eds). *The development of second language proficiency*. Cambridge: Cambridge University Press.

Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second language assessment. *Language Assessment Quarterly*, vol. 8 (3), p. 229-249.

Hulstijn, J. H. (2015). *Language Proficiency in Native and Non-native Speakers*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Hulstijn, J. H. (2012). The construction of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition*, vol.15, p. 422-433.

Hulstijn, J.; Anderson, C.; Schoonen, R. (2010). Developmental stages in second language acquisition and second language proficiency: Are there links between them? In: Bartning, I.; Martin, M. & Vedder, I. (Eds). *Communicative proficiency and linguistic development: intersections between SLA and language testing research*. Vol. 1. European Second Language Association.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning,* n. 57 (1), p. 1-33.

Laufer, B. & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, n. 16 (1), p. 33-51.

Laufer, B. & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? *EUROSLA Yearbook*, v. 1, p. 7-28.

Levin, B. (1993). *English Verb Classes and Alternations*: A Preliminary Investigation. Chicago: University of Chicago Press.

Levin, B. & Rappaport Hovav, M. (1995). *Unaccusativity*: At the Syntax-Lexical Semantic Interface. MIT Press, Cambridge, MA.

Meara, P. (2009). *Connected Words – Word Associations and Second Language Vocabulary Acquisition*. Amsterdam/Philadelphia: John Benjamins.

Meara, P. (1996). The dimensions of lexical competence. In: Brown, G.; Malmkjaer, K.; Williams, J. (Eds). *Performance and Competence in Second Language Acquisition*. Cambridge: Cambridge University Press.

Meara, P. M.; Alcoy, J. C. O. (2010). Words as Species: An Alternative Approach to Estimating Productive Vocabulary Size."*Reading in a Foreign Language*. vol 22 (1), p. 222-236.

Meara, P. & Milton, J. (2003). *X_Lex, the Swansea Levels Test*. Newbury: Express.

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. In: Bartning, I.; Martin, M. & Vedder, I. (Eds). *Communicative proficiency and linguistic development: intersections between SLA and language testing research*. Vol. 1. European Second Language Association.

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In: Bardel, C.; Lindqvist, C. & Laufer, B. (Eds.) *L2 Vocabulary Acquisition, Knowledge and Use: New Perspectives on Assessment and Corpus Analysis*. European Second Language Association.

Nation, P. (1990). *Teaching and Learning Vocabulary*. Boston, MA: Heinle & Heinle.

Read, J. (2000). *Assessing vocabulary*, Cambridge: Cambridge University Press.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, vol. 64 (4), p. 913–951.

Schmitt, N.; Schmitt, D.; Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, vol. 18 (1), p. 55-88.

Segalowitz, N. (1991). Does advanced skill in a second language reduce automaticity in the first language? *Language Learning*, 41 (1), p. 59-83.

Segalowitz, N.; Hulstijn, J. H. (2005). Automaticity in bilingualism and second language learning. In: Kroll, J. F.; De Groot, A. M. B. (Eds.). *Handbook of Bilingualism – Psycholinguistic Approaches*. Oxford/New York: Oxford University Press.

Souza. R. A. (2011). Argument structure in L2 acquisition:language transfer re-visited in a semantics and syntax perspective. *Ilha do Desterro,* n. 60, p. 153-187.

Souza, R. A.; Duarte, D. & Berg, I. B. (2015). A avaliação de habilidades linguísticas em L2: uma questão metodológica em estudos de influências translinguísticas. In: Buchweitz, A; Mota, M. B. (eds.) *Linguagem e Cognição: Processamento, Aquisição e Cérebro*. Porto Alegre: EdiPUCRS.

Souza, R. A. (2012). Two languages in one mind and the online processing of causatives with manner-of-motion verbs. *ReVEL*, special issue, n. 6, p. 220-239.

Souza, R. A; Oliveira, C. S. F. (2014). The learnability of the resultative construction in English: a comparative study of two forms of the acceptability judgment task. *Revista da ABRALIN*, vol. 13 (2), p. 375-410.

Souza, R. A.; Oliveira, C. S. F.; Soares-Silva, J. ; Penzin, A. G. A. & Santos, A. A. (2014). Estudo sobre um Parâmetro de Tarefa e um Parâmetro Amostral para Experimentos com Julgamentos de Aceitabilidade Temporalizados. *Revista de Estudos da Linguagem*, vol. 23 (1), 2015. p. 211-244.

Ullman, M. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition*, vol.4 (2), p. 105-122.

White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

Zara, J. V.; Oliveira, F. L. P.; Souza, R. A. (2013). Selective transfer in the acquisition of the double object construction by Brazilian learners. *Alfa – Revista de Linguística,* vol. 57 (2), p. 519-544.