

# Language Testing for Minority Students in Portuguese Schools: Teacher's Decision Making Based in Common European Framework

SANDRA FIGUEIREDO

*Universidade Autónoma de Lisboa, Portugal*

MARGARIDA ALVES MARTINS

*ISPA Instituto Universitário, Portugal*

CARLOS FERNANDES DA SILVA

*University of Aveiro, Portugal*

Received: 8 July 2015 / Accepted: 31 July 2016

ISSN: 1697-7467

**ABSTRACT:** Recent research on cognitive and academic performance assessment of immigrant students evidences inconsistencies in teachers' practice regarding performance assessment of L2 students. We aim to verify whether students classified by schools as having different proficiency levels in accordance to the Common European Framework of Reference for Languages (CEFR, 2001) perform differently in tests on vocabulary and verbal reasoning. The study included 23 learners of Portuguese as L2 aged 9-18, divided into three groups according to their proficiency level (A1, A2, B1). Assessment tools included four tasks - verbal analogies, semantic associations, picture naming and morphological extraction. Our expectations are that A1 students perform more poorly in all tasks than A2 students and that these perform more poorly than students classified as having a B1 level in Portuguese language. Results show that, on the one hand, there are performance differences among proficiency groups and, on the other hand, there is inconsistency in the expected order of performance differences in the tasks (from the beginner to the most advanced level). Results suggest inconsistencies in teachers' practice regarding testing of the performance of minority groups with implications in the academic success of the identified students.

**Keywords:** Second language Assessment; Linguistic Minorities; Proficiency Level; Verbal Reasoning; Vocabulary; Common European Framework.

## **Pruebas de lengua para estudiantes portugueses en escuelas minoritarias: Decisiones de profesores basadas en el Marco de Referencia Europea**

**RESUMEN:** Investigaciones recientes sobre la evaluación del rendimiento cognitivo y académico de los estudiantes extranjeros ponen en evidencia las incoherencias de la práctica de los docentes en relación a la evaluación del desempeño de los estudiantes de L2. Nuestro objetivo es verificar si los estudiantes clasificados por las escuelas, que tienen diferentes niveles de competencia de acuerdo con el Marco Común Europeo de Referencia para las Lenguas (MCER, 2001), actúan de forma diferente en las pruebas de vocabulario y razonamiento verbal. En el estudio participaron 23 estudiantes de portugués como L2 9 -18 años, dividido en tres grupos según su nivel de competencia (A1, A2, B1). Las herramientas de evaluación incluyen cuatro tareas - analogías verbales, asociaciones semánticas, reconocimiento de imágenes y extracción morfológica. Nuestras expectativas son que los estudiantes A1 tienen un

rendimiento peor que los estudiantes A2 en todas las tareas y que éstos realizan aún peor que los estudiantes clasificados con un nivel B1 en lengua portuguesa. Los resultados muestran que, por una parte, hay diferentes capacidades de rendimiento entre los grupos de competencia y, por otro lado, hay inconsistencia en las capacidades esperadas en cada uno de los diferentes grupos (desde el principiante hasta el nivel más avanzado). Los resultados sugieren inconsistencias en la práctica de los docentes respecto a las pruebas de la actuación de los grupos minoritarios con implicaciones en el éxito académico de los estudiantes identificados.

**Palabras clave:** Evaluación Segunda lengua; Minorías Lingüísticas; Nivel de competencia; Razonamiento verbal; Vocabulario; Marco Común Europeo.

## 1. INTRODUCTION

Recent studies in Second Language (L2), specifically on academic success of immigrant populations, have focused on analysis of age effects, of being exposed to the language and on differences in the type of speakers (specificities in mother-tongues). Variability in performance has been thoroughly and frequently tested in international school areas, especially in the American context, since the 1960s (American Council on the Teaching of Foreign Languages, 1996; Bachman, 2000; Bailey & Huang, 2011; Bygate, Swan & Skehan, 2013; Coombe, 2013; Ramirez, Gen, Geva et al., 2012). In the specific field of assessment, international researched has focused on the development and validation of tools for immigrant school population (Bachman, 2000; Coombe, 2013; Bailey & Huang, 2011), though each national system has developed its own guidelines, as is the case of the American Council on the Teaching of Foreign Languages (ACTFL) in the USA, the Canadian repository of tools and tests - ALBERTA (2012) and the Association of Language Testers in Europe (ALTE), in Europe. The European and American resources are very different in terms of scientific research and of school practices, which have an impact on enhancing evidence and validity of new materials and new data in this scientific field (Bachman, 2000; Bailey & Huang, 2011). Which is the best reference? The type of linguistic minorities will determine the necessary tools for assessment, placing and learning for each country. However, the materials available are mostly restricted to English as a Second Language and as a Foreign Language, there being vast experience in PhD theses and exploratory studies as assessment tools since 1960 (Coombe, 2013).

In the European context, besides assessment and certification systems, the reference available to education professionals in the field of designing materials for teaching and learning may be found in the document Common European Framework of Reference for Languages (CEFR - Council of Europe, 2001). In Portugal, the CEFR is applied in schools to determine the proficiency of immigrant learners of different ages and nationalities and from different school levels. Yet, the impact of this application in the performance levels of the immigrant population is not known (Pérez-Sabater, 2012). Moreover, despite the fact that the guidelines and rules of reference such as the CEFR (Council of Europe, 2001) and the Framework for Portuguese Foreign Language Teaching (Grosso et al. 2001) are considered as having been developed on scientific research (Grosso et al. 2011:21), there has been no psychometric validation or longitudinal study on the scores that define the profiles and cut-off points (Bachman, 2000; Bailey & Huang, 2011; Jones, 2013). The existing frame-

works of reference are only a small part of the intervention and identification of linguistic minorities (Antón, 2011; Figueiredo & Silva, 2012; 2013). The descriptors of these guiding frameworks do not specify the assessment methods (Grosso et al.) to assess proficiency, and the certification procedures in the first language context do not complete assessment in a formal education environment. These are separate methods and do not meet the need to analyze linguistic and cognitive profiles of immigrant population in European schools.

In the case of Portugal and of learners of Portuguese as L2, CEFR has helped teachers to determine linguistic proficiency of the different minority groups in Portuguese schools. Identifying proficiency is linked to general academic proficiency and with students' cognitive language proficiency (Cummins, 1979; Hulstijn & Schoonen, 2012), as also with students' attitudes on the autonomous language learning (Pérez-Sabater, 2012), which makes it crucial that proficiency measures be designed. The cognitive and academic dimensions are not always reflected on the idea of proficiency or skills "level" these tests indicate (Jones, 2013). Hulstijn and Schoonen (2012) analyze how the proficiency level does not provide information on cognitive functioning, which explains why CEFR is not effective in assessing proficiency in L2 and development in L2 in the European school context. In CEFR, proficiency levels are determined from A1 (basic proficiency level: beginner) to C2 (advanced proficiency level; 'proficient user'). From the six levels defined in CEFR, only three are analyzed in this study - A1, A2 and B1 - which may indicate inconsistencies from a scientific and assessment point of view (Nunes & Lorke, 2011). Reference to skills aimed at defining proficiency levels does not complete the idea of assessment of objectives attained described in the framework (CEFT, Council of Europe, 2001). As Nunes and Lorke (2011) indicate, contexts are an important variable which CEFR does not contemplate in its guidelines.

It is rather common that research aimed at analyzing differences between average of academic performances of L2 learners and scores obtained in assessment tests, conducted by the scientific and the non-educational community, as is the case of the Test of English as a Foreign Language (TOEFL), in non-European context and on English as the main language (Cho, Yeonsuk & Bridgeman, 2012). The authors analyzed how the academic performance of university students was proportionally similar to the results those very same students attained in TOEFL and realized the correlation was low though it did not compromise the validity and benefit TOEFL had for speakers of English as a Second or Foreign Language. On the other hand, Kokhan (2013) analyzes TOEFL's limitation, with implications on the placement of students in specific classes according to the results they obtained in the Test and the standardized approach of tests such as TOEFL in schools (Llosa, 2011). This type of analysis is needed within the context of Romance languages, especially associated to 'European' languages as second languages, so as to assess the validity of tests based on the scales proposed by CEFR and the implications that the inconsistencies between the scales and actual proficiency (in actual contexts) may have in terms of negative impact in the ambiguous assessment of immigrant students with different levels of exposure to the host language (Coombe, 2013).

This study is part of a wider research project whose aim is validating specific tasks in the assessment of proficiency and cognitive profile (for example, vocabulary and verbal reasoning) of different immigrant groups so as to understand the varied measures teachers must adopt to assess new immigrant students. According to Halle et al. (2011), assessment of linguistic proficiency must focus on linguistic construct though not independently from

academic performance (Hulstijn & Schoonen, 2012). So as to be considered in the area of L2 education, assessment must encompass the maximum of dimensions possible. The descriptors of the framework of reference are guidelines but are not enough as (validated) tool to assess competences and design specific measures for different language levels.

This paper aims to verify if there are inconsistencies between the practice of teachers in Portuguese schools regarding assessment (by using the only existing framework: CEFR) of L2 students, divided according to their level of linguistic proficiency (levels A1, A2 and B1), and assessment of those students in specific tasks on vocabulary and verbal reasoning - verbal analogies, semantic associations, picture naming and morphological extraction. Differences observed among the proficiency levels and the profile of subjects' answers suggests inconsistencies in the practice of teachers regarding assessment of minority groups, with implications in the academic success of students considered as having difficulties and placed in inadequate support programs.

## 2. METHOD

### 2.1. Participants

This sample has intentionally diverse characteristics concerning proficiency levels and precedence (country of origin and languages spoken) that are crucial to understand the reality of the students' pluralism and mainly regarding the proficiency differences. After a control on sampling criteria we found a small group (23 individuals) that constitutes an important sample of specificities of inconsistencies concerning the evaluation and achievement comparison and inconsistency examination. Additionally, we found in previous recent studies the similar criteria for small samples regarding the European framework in examination for minority groups other than English second language learners. The study includes 23 students with different proficiency levels: in the A1 group (beginner level) subjects are 9 to 18 years old ( $M=12,6$ ;  $SD=2,7$ ), 5 (62,5%) are male and 3 (37,5%), are female, in the A2 level (elementary level) subjects are 12 to 18 years old ( $M=15,6$ ;  $SD=2,3$ ), 2 (18,2%) are male and 9 (81,8%) are female, in the B1 group (threshold level) they are 11 to 14 years old ( $M=12$ ;  $SD=1,4$ ) and 1 (25%) is male, 3 (75%) are female. All participants attend state primary and secondary schools in the district of Lisbon. Of the 23, 7 (31,8%) are from China and speakers of Mandarin, 3 (7,1%) are speakers of Romance languages and come from European countries and Latin America, 1 (4,5%) speaks a Slavic language, from Eastern Europe, 5 (22,7%) speak Creole and are from Portuguese Speaking Countries (PALOPs), and 6 (27,3%) speak an Indo-Aryan (e.g., Urdu) language and come from Asia.

### 2.2. Instrument

#### *Task 1 - Verbal analogies*

The 6-item verbal analogy test was adapted from *Verbal Analogies* (Test n.º 2) by "Woodcock-Munoz Language Survey-Revised (WMLS-R, 2005) and the Portuguese adapted version presents Cronbach Alpha of .60. The test aims to, within verbal reasoning, assess how the student, whose mother-tongue is not Portuguese, completes six sentences based on

vocabulary and the proposed analogies. The verbal analogy task aims to measure understanding of logical association within random phrase contexts. The test scores as follows: 1 point for each correct answer (Total score: 6 points). Example 1: Fill in the missing word by logic association: ‘Star is to sky as fish is to \_\_\_\_.’

#### *Task 2 - Semantic associations*

The 6-item semantic association test was adapted from “Woodcock-Munoz Language Survey-Revised” (WMLS-R, 2005). The Portuguese adapted task evidenced high reliability with internal correlation consistency of .86. This test, within the area of vocabulary, aims to assess students’ ability to identify semantic relations among randomly placed words; students are asked to complete the six items by providing synonyms and antonyms of each word in Portuguese. The task was thus assessed: 2 points for each correct answer (Total score: 12 points). Example 1: Fill in the missing word by identifying the correct synonymous or antonymous of the presented word: ‘Rich \_\_\_\_ (synonymous) \_\_\_\_ (antonymous)’.

#### *Task 3 - Picture naming*

The picture naming test of *Teste de Diagnóstico do Português Língua Não Materna* (Mateus, 2009) - Diagnostic Test of Portuguese as Second or Foreign Language - includes 36 pictures and presents high internal consistency (.94). Students are asked to name, in writing rather than orally, as in the original test, the pictures as they are sequentially shown in 5 pages. The test aims to identify the vocabulary skills of immigrant students in the context of second or foreign language and in a non-complex frequency level (according to CORLEX, a Portuguese index of vocabulary frequency and difficulty). Total score for this task is 12 points.

#### *Task 4 - Morphological Extraction*

The morphological extraction test includes 4 items and was adapted from the *Morphological Extraction Test* by August, Kenyon, Malabonga et al. (2001). This study presents a Cronbach Alpha of .53. The objective of the test is to assess students’ ability to make changes and extractions from a derived word in a specific sentence. In the case of L2 students, we aim to verify, in terms of vocabulary and verbal reasoning, interference from first language to host language in the process of morphological change. Change is analyzed regarding conversion of morphemes in new words in the main language. The task was thus assessed: 2 points for each correct extraction, 1 point for each partially correct answer (total score: 8 points). Example 1: Word: Friendship / Sentence: “My schoolmates are my \_\_\_\_\_.”

In terms of psychometric properties of the above mentioned tests, only that on morphological extraction evidenced limited internal consistency (.53). The original test in English presents a high Cronbach coefficient (.93).

### **3.3. Procedure**

Data collection took place in 2013 and 2016 in primary and secondary state schools in the district of Lisbon. We contacted teaching facilities in the district of Lisbon to propose the study and disseminate the research objectives. Communicating with the schools allowed us to identify 108 immigrant students, 23 of whom were informed of their proficiency by

the school. Upon informed consent and authorization and selection of demographics of school population, the four tasks were applied and assessed (scored) in accordance with the information of the original tasks. Students completed the battery of tests (15 tasks) in 60 minutes in a classroom. Data collection was made in one group session by specific groups of students in the schools involved. Participants were divided into groups so that test application would be more effective and they were carried out under the supervision of the researcher.

Tests were applied both in paper and in computer formats (sound files for phoneme and text decoding). Information regarding the classification of proficiency levels were provided by the schools after informed consent and once the school year had started (when assessment is carried out). It was perceived that assessment is different among schools and that not all measure proficiency of linguistic minorities in order to present level categorization. Data was treated using SPSS, version 21.

### 3. RESULTS

We aim to verify if there is inconsistency between the practice of teachers in Portuguese schools regarding assessment of L2 learners, separated by levels A1, A2 and B1, and assessment of those students in the tasks included in this study. Differences are expected among groups classified according to their linguistic proficiency (according to the Common European Framework of Reference for Languages) and that these differences will be made visible in terms of increasing order of performance (A1, beginner level, A2, elementary level, B1, threshold level). Differences are also expected among groups so as to identify lower levels of proficiency related to less positive performance in the four tasks on vocabulary and verbal reasoning.

Subjects were classified by schools according to proficiency levels defined by the CEFR (Council of Europe, 2001) and our sample includes the first three groups: A1 (beginner level), A2 (elementary level) and B1 (threshold level) These groups correspond to the first three of the CEFR and includes two types of users: the basic user (A1 and A2) and the independent user (B1).

#### 3.1. One-Way Analysis of Variance

We produced a one-way analysis of variance to examine descriptive statistics for the children's performance in the four tasks, from the factor I (factorial analysis), according to their levels of proficiency (evaluated by teachers and based on CEFR). ANOVA analyses were made considering the proficiency groups a variable and their performance in the four tasks the dependents. Through the univariate ANOVAs we also examined the effect sizes for the groups mean differences, using the L2 proficiency levels. The groups differed significantly in the results obtained for all the four tasks but the effect sizes showed to be substantially higher for the morphological extraction task ( $\eta^2 = .440$ ) and for the naming task ( $\eta^2 = .461$ ), considering the benchmarks of Cohen for statistical value of  $\eta^2$  (Cohen, 1988).

Table 1 shows the means, standard deviations and effect sizes for the groups determined by their proficiency levels. The three groups presented different performance averages and differ significantly across the proficiency groups, according to the four tasks administered.

Table 1. *Performance Average in the Three Tasks in Portuguese as L2, in Accordance to the Three Proficiency Level Given by the Schools*

Proficiency Groups (CEFR)	Verbal analogies			Semantic associations			Picture naming			Morphological Extraction		
	M	S.D.	$\eta^2$	M	S.D.	$\eta^2$	M	S.D.	$\eta^2$	M	S.D.	$\eta^2$
A1	19.17	14.08	.225	2.50	2.59	.163	2.00	1.26	.461	0.83	0.98	.440
A2	32.40	4.88		5.56	3.45		4.27	1.565		3.91	2.07	
B1	34.75	4.03		4.50	3.70		3.25	2.37		4.00	0.82	
Total	28.90	10.47		4.48	3.39		3.43	1.86		3.05	2.13	

In the task on verbal analogies, group A2 has more correct answers, followed by groups B1 and A1. ANOVA results evidenced significant differences among the groups:  $F(2,18)=3,72$ ;  $p<.05$ ;  $\eta^2=.225$ . A post hoc (Tukey) test revealed significant differences among groups A1 and A2 ( $p<.05$ ) and no significant differences between group B1 and the other groups (A1 and A2).

In the task on semantic associations, group A2 (second group) presents higher results, followed by group 3 (B1) and group 1 (A1). However, ANOVA results did not evidence significant differences among the groups.

In the task on picture naming, group A1 (first level of proficiency) evidences the lowest scores and groups A2 (second level of proficiency) and B1 (third level) have similar averages. ANOVA results evidenced significant differences among the groups:  $F(2,17)=5,61$ ;  $p<.05$ ;  $\eta^2=.461$ . Post-hoc analysis revealed significant differences between group A1 and the other two groups (A2, B1) ( $p<.05$ ), there existing no significant differences between A2 and B1.

In the task on morphological extraction, the group who performed the best was B1 (the third group) whose average is close to that of group A2 (second group); the group that performs more poorly is A1 (first group), similarly to what occurs in the other tasks. ANOVA results evidenced significant differences among the groups:  $F(2,18)=7,46$ ;  $p<.05$ ;  $\eta^2=.440$ . Post-hoc analysis revealed significant differences between group A1 and the other two groups (A2, B1) ( $p<.05$ ), there existing no significant differences between A2 and B1.

#### 4. DISCUSSION

This paper aims to verify the consistencies between the practice of teachers in Portuguese schools regarding assessment (based in CEFR) of L2 learners, divided according to their level of linguistic proficiency (levels A1, A2 and B1) and assessment of those students in

specific tasks on verbal analogies, semantic associations, picture naming and morphological extraction. Though differences exist among the proficiency groups (previously classified as such by the schools), there exists also inconsistency between the groups and their performance in the tasks applied in this study in terms of the expected order of results (we expected a link between lower levels of proficiency and poorer performance in the four tasks and in verbal reasoning). We expected that a higher level of proficiency would lead to higher level of correction in the tasks, which did not always occur. Moreover, considering the difference in performance in the four tasks and the three proficiency levels, we aimed to understand why the differences were more significant in some tests.

In the task on verbal analogies, the order of results is unexpected, since group A2 performs significantly better than group A1 but group B1, though its performance is better, the differences between A2 and B1 are not significant. These results contradict the assumption that students classified as having a higher level of linguistic proficiency (B1) perform better in verbal reasoning and cognitive processing within interpretation of more rhetorical language (Johnson & Rosano, 2009). On the one hand, the specificity of the topic in this type of tests may hinder good performance because L2 learners tend to have poor results in specific assessment situations, in which themes or topics are imposed without a wider context being given for the answer (Lee & Sawaki, 2009). On the other hand, in tasks such as identifying verbal analogies, we realize that a wider difference is expected in terms of results between the least proficient group (A1) and the other groups, considering these tasks require that students master more vocabulary and higher proficiency in the dominant language, Portuguese (Grigornko, Sternberg & Ehrman, 2000; McBride-Chang et al., 2005). Identifying verbal analogies in the several sentences proposed and the manipulation of morphemes (the task on morphological extraction) from derived words are assessment contexts within the complex task field for L2 and sequential bilingual learners (Woodcock-Munoz, 2005). By sequential bilinguals, we mean those bilinguals (learning L2) who did not learn the two languages simultaneously (when they are exposed to two languages, usually in the first infancy). In the case of the verbal analogies tests, an average to high level of competence is required regarding understand word and meaning association with words having different levels of frequency, the so-called “comprehension-knowledge and fluid reasoning” (Roomaney & Koch, 2013). Mastering vocabulary is the main influence in performance differences in this type of task. Performance in specific tasks as these should be more researched in the context of linguistic minorities, especially from the point of view of Indo-Aryan, Mandarin and alphabetic speakers, so as to further knowledge on their cognitive profiles and processing.

Additionally, order is not the expected in the context of semantic associations, though there is no significant differences in results by A2 and B1 groups. However, we must emphasize that the A2 group performs slightly better than the B1 group, which is unexpected. Identifying synonyms and antonyms are often complex tasks for learners with poor linguistic skills, who may take longer in decoding vocabulary, especially unfamiliar one, and who cannot resort to transference from the mother-tongue (Crossley, 2013). The amount of vocabulary learned determines the ability to identify synonyms and antonyms in L2 (Schmitt, 2008). Age is also frequently a predictor because those less skilled and younger tend to identify a small number of semantic relations (Sheng et al. 2013). Mastering semantic relations allows the student access to inter-cultural and multilingual environment, which are added skills to knowing a second or foreign language. L2 learners who do not learn multiple meanings (and



a grade of synonymy and of autonomy) will not perform thoroughly and identify semantic relations (Fernandes, 2009). The skills involved in completing such a task are not only lexical but also semantic, involving a conceptual system under analysis in a semantic associations task, as was applied in this and in previous studies which determined the need of a context (placing the words in a text) for understanding and cognitive activation of the semantic dimension of the words in the exercise (Fernandes). On the other hand, the results of this study, though not significant, indicate an inconsistency in the expected order of the groups, since expectation would be that a group classified by the school as A2 would perform worse than a group classified as B1.

In the picture naming task, and as expected, differences arose between the groups and in the order of language proficiency - A1, A2, B1. However, the differences are noticeable only regarding A1 and the other groups. This is not a structurally demanding task at cognitive level, as remarked in a similar previous study (Wang & Lam, 2009). Group A1 is the one that names less pictures (19), when compared to the other groups (which correctly named 32 and 34 pictures) Probably, group A1, as associated to more limited vocabulary in the early stages of L2 learning, will face higher challenges in naming a series of pictures that have no immediate lexical correspondence in L2, considering their starting point is the mother-tongue vocabulary (Jared, Poh & Paivio, 2013). There were mostly omissions and naming in the English language, which is the mediation language. Naming in a different language was rather common, which evidenced no transference but rather lack of knowledge of the Portuguese lexicon. In the study by Hulstijn and Schoonen (2012), the task on vocabulary in Dutch as L2, and in more complex vocabulary tasks, B1 and B2 groups (within the “independent user” range) evidence only a slight differences, whereas in less complex tasks (low and medium-level frequency), the differences were higher. The conclusion is, thus, that L2 learners show differences in terms of degree rather than category. Hulstijn and Schoonen (2012) therefore question the CEFR and the validation of its tests.

In the morphological extraction task, the results are those expected, there being significant differences between A1 and the other groups; no significant differences were seen between group A2 and group B1. In this task, on morphological extraction or change (root extraction, August et al.) another verbal reasoning is at stake, one which implies average to high mastering of vocabulary (August et al., 2001; McBride-Chang et al., 2005), which is linked to the spelling and phonological effect. The subjects’ mother-tongue is here a variable (Ramirez et al., 2010). In the context of English as L2 and of speakers of Spanish and Portuguese as mother-tongues, recent studies have proved the effect of mother-tongue in the variation of results in assessment tasks (Abreu, 2012; Geva & Massey-Garrison, 2013; Ramirez et al., 2010). The morphological extraction task analyzes the subjects’ ability for morphological manipulation (based on phonological and spelling clues in the mother-tongue) as well as detects language transference (Abreu, 2012; Geva & Massey-Garrison, 2013; Ramirez et al. 2010). This way, phonological or spelling mistakes can occur and are quite frequent (in the A1 group, there were more frequent phonological and spelling mistakes. Examples of mistakes for the answers, in Portuguese: “espiritualmente”, “cuidadosa”, “amigos” and “felizmente”: ‘espiritamente’, ‘espiritual’, ‘cuidamente’, ‘cuidados’, ‘cuidadinhos’, ‘amizade’, ‘felizemente’, ‘feliozo’, ‘espírito’, ‘espiritual’). However, the phonological mistakes do not explain the spelling mistakes and vice versa. High correctness in this task evidences high phonological awareness (Wang & Lam, 2009), as well as high cognitive

awareness, attentiveness to L2 and to clues (noticing hypothesis, Mota & Zimmer, 2005) in the transference process (between L1 and L2), which will determine commitment in mental processes that connect the structures of the new language and produce knowledge in L2 - production hypothesis (Mota & Zimmer, 2005). However, all groups in this study have low scores (a maximum of 4 in 10). Cut points, even in very low scores, should be considered to differentiate poor performance and established ordered levels which are not considered by the CEFR and even in American tests, which have been studied by research in terms of how students are distributed among the cut points in the tests and analyzes the criteria for class distribution according to TOEFL results (Kokhan, 2013).

Results usually reject the congruence hypothesis among teaching practice in Portuguese schools in regards to assessment of L2 learners and assessment of those students in this study's tasks. Studies have confirmed that performance differences in the study's tasks organize groups A1, A2 and B1 in an increasing order in terms of performance, as expected and as previous research in Europe. The latter has concluded that differences among proficiency groups in vocabulary and grammar tasks are in line with the expected skills for the groups classified according to the CEFR levels (Hulstijn & Schoonen, 2012). Our results are also similar to those of previous studies carried out in the US (Cho & Bridgeman, 2012), which concluded that students who performed well in the TOEFL tasks, their performance did not correlate as expected in terms of performance in school assessment tasks. Group A2 performed better than the supposedly more skilled group - B1 - in the tasks on verbal analogy and semantic association. Nevertheless, students in the least skilled group - A1 - did perform the poorest. Yet, students in groups A2 and B1 have very similar performance levels, as in the morphological extraction tasks, in which there was no significant difference between the two. Considering all results and the relation between the tasks selected for this study and the importance of vocabulary skills for performance in the tasks, we conclude that it is extremely relevant for teachers to have access to an organized and valid information corpus and tools on vocabulary acquisition and development in L2, especially in regards to vocabulary selection and evaluation (McCarthy, 2008; Pulido, 2006). This will enable them to understand the L2 proficiency levels and cognitive aspects of students from different minorities and with different proficiencies in the dominant language (Coombe, 2013; Tidball & Treffers-Daller, 2008). Vocabulary measurements are only reliable for L2 assessment if vocabulary task results show differences that identify different proficiencies and high internal consistency in regards to the scales used (Laufer & Nation, 1999). On the other hand, the results of this study suggest two issues: limitations in the classification teachers award students and limitations in CEFR assessment guidelines for teachers' classifications. These results confirm previous studies that analyze the same limitations and inconsistencies in the proficiency levels assigned (Alderson et al., 2002; Hulstijn & Schoonen, 2012; Nunes & Lorke, 2011; Weir, 2005).

In regards to the limitations of this study, besides the low internal consistency of one of the tasks - morphological extraction (.53) - we must state that the number of variables related with the features of the sample (mother-tongues, nationalities, mother-tongue instruction, and exposure to the language) may skew the interpretation of some results. In conclusion, performance differences among the groups are confirmed; yet, these differences do not show similarity between the proficiency levels in increasing order (A1, A2, B1) and performance level of those groups in the referred tasks. Moreover, the study puts forth new

data in L2 research on performance and cognitive and linguistic profiles in Portuguese as L2 through the development of evidence mostly on verbal reasoning.

## 5. REFERENCES

- Abreu, L. (2012). "Subject Pronoun Expression and Priming Effects among Bilingual Speakers of Puerto Rican Spanish", in K. Geeslin, & M. Díaz-Campos (eds.), *Selected Proceedings of the 14th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Proceedings Project, 1- 8.
- Alderson, J.C. (Ed.). (2002). *Case studies in the use of the Common European Framework*. Strasbourg. Council of Europe. ISBN 92-871-4983-6.
- Alberta Education. (2012). "Language Proficiency Assessment." Available from: <http://www.learnalberta.ca/content/eslapb/languageproficiencyassessment.html>, accessed 12 February 2013.
- American Council on the Teaching of Foreign Languages. (1996). "American Council on the Teaching of Foreign Languages". Available from: <http://www.actfl.org/>, accessed 20 June 2014.
- Antón, M. (2011). "A review of recent research (2000-2008) on applied linguistics and language teaching with the reference to L2 Spanish", in *Language Teaching*, 44,1: 78-112.
- August, D., Kenyon, D., Malabonga, V., Louguit, M., Caglarcan, S., & Carlo, M. (2001). *Extract the base test—English*. Washington, DC: Center for Applied Linguistics.
- Figueiredo, S., & Silva, C. (2012). "Programa de intervenção em contexto educativo: dispositivo electrónico de avaliação da descodificação fonética de aprendentes de Português Língua Segunda.", in *Revista Portuguesa de Educação*, 25,1: 113-1356.
- Figueiredo, S., & Silva, C. (2013). "Learning analytics and Second-Language Context: a digital instrument for measurement of real-time data regarding second language learners achievement", in *International Journal of Technology Enhanced Learning (IJTEL)*, 5,2: 178-194.
- Bailey, B., & Huang, N. (2011). "Do current English language development/proficiency standards reflect the English needed for success in school?", in *Language Testing*, 28,3: 343-365.
- Bachman, L. (2000). "Modern language testing at the turn of the century: assuring that what we count counts", in *Language Testing*, 17,1: 1-42.
- Bygate, M., Swain, M., & Skehan, P. (2013). *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing*. London and New York: Routledge.
- Cho, Yeonsuk, & Bridgeman, B. (2012). "Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities.", in *Language Testing*, 29,3:421-442.
- Conselho da Europa. (2001). *Quadro Europeu Comum de Referência para as línguas*. Porto: Edições ASA.
- Coombe, C. (2013). "Book review: Assessing Academic English: Testing English Proficiency 1950-89.", in *Language Testing*, 30,1:143-145. DOI: 10.1177/0265532212458893.
- Council of Europe. (2001). *Quadro Europeu Comum de Referência para as línguas [European Common of Reference for Languages]*. Porto: Edições ASA.
- Cummins, J. (1979). "Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters.", in *Working Papers on Bilingualism*, 19: 121-129.
- Crossley, S. (2013). "Assessment automatic processing of hypernymic relations in first language speakers and advanced second language learners: a semantic priming approach.", in *The Mental lexicon*, 8,1: 96-116.

- Fernandes, N.(2009). *Relações semânticas de sinonímia e antonímia: contributo para o desenvolvimento da competência lexical na aula de Português Língua estrangeira*. Master Program Porto: Faculdade de Letras da Universidade do Porto.
- Geva, E., & Massey-Garison, A. (2013). “A Comparison of the Language Skills of ELLs and Monolinguals Who Are Poor Decoders, Poor Comprehenders, or Normal Readers. *In Journal of Learning Disabilities, 46, 5:387-401*.
- Grigornko, E., Sternberg, R., & Ehrman, M. (2000). “A Theory-Based Approach to the Measurement of Foreign Language Learning Ability: The Canal-F Theory and Test.”, in *The Modern Language Journal, 84:390-405*. Doi: 10.1111/0026-7902.00076
- Grosso, M<sup>a</sup>, J., Soares, A., Sousa, F., & Pascoal, J. (2011). *Quadro de Referência para o Ensino do Português no Estrangeiro. Documento Orientador*. Direção-Geral da Educação e do Desenvolvimento Curricular.
- Halle, T., Hair, E., Wandner, L., McNamara, M., Chien, N. (2012). “Predictors and outcomes of early versus later English language proficiency among English language learners.”, in *Early Childhood Research Quarterly 27: 1- 20*.
- Hulstijn, J., & Schoonen, R. (2012). “Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of Reference for Languages (CEFR).”, in *Language Testing, 29,2: 203-221*.
- Jared, D., Poh, R., & Paivio, Al. (2013). “L1 and L2 picture naming in Mandarin-English bilinguals: A test of Bilingual Dual Coding Theory.”, in *Bilingualism: Language and Cognition, 16,2: 383-396*.
- Johnson, J., & Rosano, T. (2009). “Relation of cognitive style to metaphor interpretation and second language proficiency.”, in *Applied Psycholinguistics, 14,2: 159-175*.
- Jones, N. (2013). “Design an inclusive framework for languages.”, in E. Galaczi & C. Weir (Eds.), *Studies in Language Testing-Exploring Language Frameworks* (pp. 105-118). Cambridge.
- Kokhan, K. (2013). “An argument against using standardized test scores for placement of international undergraduate students in English as a Second Language (ESL) courses.”, in *Language Testing, 30,4: 467-489*.
- Laufer, B., & Nation, P. (1999). “A vocabulary size test of controlled productive ability.”, in *Language Testing, 16,1: 33-51*.
- Lee, Y., & Sawaki, Y. (2009). “Application of three cognitive diagnosis models to ESL reading and listening assessments.”, in *Language Testing,23,2:131-166*.
- Llosa, L. (2011). “Standards-based classroom assessments of English proficiency: A review of issues, current developments, and future directions for research.”, in *Language Testing, 28,3: 367-382*.
- Mateus, H. (2009). *Teste de Diagnóstico do Português Língua Não Materna*. Direção-Geral de Educação.
- McBride- Chang, C., Wagner, R. K., Muse, A., Chow, B. W, & Shu, H. (2005). “The role of morphological awareness in children’s vocabulary acquisition in English.”, in *Applied Psycholinguistics, 26: 415- 435*.
- McCarthy, M. (2008). “Assessing and interpreting corpus information in the teacher education context.”, in *Language Teaching, 41, 4: 563-574*.
- Mota, M., & Zimmer, M. (2005). “Cognição e aprendizagem de L2: o que nos diz a pesquisa nos paradigmas simbólico e conexionista.”, in *Revista Brasileira de Linguística Aplicada, 5,2: 155-187*.
- Nunes, E., & Lorke, F. (2011).” O problema da adequação dos parâmetros do Quadro Europeu Comum de Referência e a “necessidade de emergir como os outros de nós mesmos.”, in *Revista X, 2: 40-60*.

- Papageorgiou, S. (2014). "Book review: the common european framework of reference: the globalization of language education policy.", in *Language Teaching*, 18: 543-544.
- Pérez-Sabater, C. (2012). "A pioneer study on Online Learning Environments Following the Common European Framework of Reference for Languages.", in *Procedia-Social and Behavioral Sciences*, 46: 1948-1955.
- Pulido, D. (2006). "Vocabulary in a second language: selection, acquisition, and testing.", in *Studies in Second Language Acquisition*, 28, 1: 143-144.
- Ramirez, G., Chen, X., Geva, E., & Kiefer, H. (2010). "Morphological awareness in Spanish-speaking English language learners: within and cross-language effects on word reading.", in *Reading and Writing*, 23: 337-358. DOI 10.1007/s11145-009-9203-9.
- Roomaney, R., & Koch, E. (2013). "An item and construct bias analysis of two language versions of a verbal analogies scale.", in *South African Journal of Psychology*, 43,3: 314-326.
- Sheng, L., Bedore, L., Pena, E., & Fiestas, C. (2013). "Semantic development in spanish-english bilingual children: effects of age and language experience.", in *Child Development*, 84,3: 1034-1045.
- Schmitt, N. (2008). "Review article: instructed second language vocabulary learning.", in *Language Teaching Research*, 12,3: 329-363.
- Tidball, F., & Treffers-Daller, J. (2008). "Analysing lexical richness in French learner language: what frequency lists and teacher judgements can tell us about basic and advanced words.", in *Journal of French Language Studies*, 18,3: 299-313.
- TOEFL (2012). *TOEFL® Institutional Testing Program (ITP) & TOEFL Family Product*, available from: <http://www.ets.org/pt/toefl/>, accessed 13 September 2013.
- Wang, W., & Lam, A. (2009). "The English Language Curriculum for Senior Secondary School in China Its Evolution from 1949.", in *RELC Journal*, 40,1: 65-82.
- Weir, C. J. (2005). "Limitations of the Common European Framework of Reference for Languages (CEFR) for developing comparable examinations and tests.", in *Language Testing*, 22,3: 281-300.
- Woodcock, R. W., Munoz-Sandoval, A. F., Rued, M.L. & Alvarado, C. G. (2005). *Woodcock-Munoz Language Survey – Revised, English*. Itasca, IL: Riverside Publishing.
- Zhao, M., Meng, H., Du, X., Liu, T., Li, Y., & Chen, F. (2011). "The neuromechanism underlying verbal analogical reasoning of metaphorical relations: An event-related potentials study.", in *Brain Research*: 62-74. <http://www.learnalberta.ca/content/eslapb/languageproficiencyassessment.html>, accessed 12 February, 2013. <http://www.actfl.org/>, accessed 20 June, 2014.