

NEW METHODOLOGIES FOR SECOND LANGUAGE ASSESSMENT: MEASURING AND IDENTIFYING PROFILES IN MIGRANT SCHOOL CONTEXTS

Sandra Figueiredo¹, Margarida Martins², Carlos Fernandes³

¹*Higher Institute for Applied Psychology (PORTUGAL)*

²*Higher Institute for Applied Psychology (PORTUGAL)*

³*Department of Education, University of Aveiro (PORTUGAL)*

Post-doctoral research sponsored by Foundation for Science and Technology (grant with reference SFRH/BPD/86618/2012).

Abstract

In Education, the policy related to immigrant school population is of utmost importance considering the migratory history and the current scenarios of Multilingual classrooms. The research evidence in second language area is a very recent period in science, mainly regarding the assessment concerns and responsiveness. To be inclusive, school professionals need differentiated tools to assess and place students. The diagnostic evaluation is the first step in educational scientific research, in this field. On one hand, the language assessment in second language context is still a gap in Portuguese schools and at research settings; on the other hand, the international scientific research still presents contradictory frameworks regarding standardization of norms and benchmarks to develop tests and scoring. In this work new methodologies are addressed as instrument for school improvement, specifically regarding second language learners. Our research is focused in the development of a diagnostic test to assess proficiency and skills of students in Portuguese language, creating validated measuring and reliable training in this area. Literacy practitioners should be aware of profiles and needs of immigrant students, so far this was not achieved in our national system using oversampling in research studies. Particularly here we will provide characteristics of examination tools to identify type of responses that Portuguese language learners present in such second language diagnostic tests, integrating several tasks to ensure enough observations and reliable scores, to determine profiles that inform how to score constructed-response items. This research will be able to describe language behaviour profiles and specific language errors that would be useful as training clues (as rubrics for each content area) to define correct language and grammar to be expected to receive positive scores in the test. Teachers and students are still unfamiliar with these tests and the orientation about fulfilment and scores should be supported by research results and validated material to help to apply specific evaluation to specific responses, in real school contexts.

Key-words: second language, assessment, scores, immigrant students, validation.

1 Evaluation of second language proficiency in school context: gap in methodologies and procedures.

Using data from specific and diverse international repositories and tests, this study developed a diagnostic test for immigrant school students. The empirical research is on course and all the tests are applied to Portuguese students with the main goal to examine language profiles and to determine levels, regarding not solely the language proficiency but mainly speakers profiles according to age, arrival time and home languages. The type of first language and age could be the crucial influence in determining cognitive academic proficiency in the second language (Figueiredo, 2009; Figueiredo & Silva, 2013). It is expected that second language learners perform with more difficulty than native students in reading, writing, listening and comprehension skills, which affects generally the performance in academic contents (science, literature...). Cognitive predisposition could be masked by limitations of language proficiency (Abedi, 2002; Stowe, 2006) which leads to the emergent research about cognitive and linguistic predispositions of different second language learners, with the additional factor of migratory experience. According to the language assessment, in second language area, is of utmost importance to identify students at risk and different language speakers

circumstances (Howard, 2009). One of the concerns in the assessment of cognitive abilities and language proficiency of foreign speakers is the fact that the tests are usually administered in the main language of tuition (second language for immigrant students) which could compromise the performance in those tests,

For all test takers, any test that employs language is, in part, a measure of their language skills. This is of particular concern for test takers whose first language is not the language of the test. Test use with individuals who have not sufficiently acquired the language of the test may introduce construct-irrelevant components to the testing process. In such instances, test results may not reflect accurately the qualities and competencies intended to be measured....Therefore it is important to consider language background in developing, selecting, and administering tests and in interpreting test performance. (p. 91) (cited by Abedi, 2002, p. 233).

Although the probability of construct-irrelevancy constraints, the main focus placed into cognitive and linguistic abilities are assured by simple statements and debriefing administered by the researcher, in classroom, during tests application. According to Direção-Geral de Inovação e Desenvolvimento Curricular (2005), in Portugal, there is only a certification system of Portuguese as foreign language that operates with several tests, organized by proficiency levels from Common European Framework (2001). This system has a gap regarding school population, once there is evaluation of non-school individuals only, over 14 years old. The construction and selection of tests are conducted with no supervision or previous scientific validity. One implication detected is the fact that the evaluation addresses foreign language proficiency which is different from second language proficiency. In terms of learning and acquisition, the difference has other implications that need to be observed. Additionally, such exams neglects certain components involved in evaluation of cognitive and linguistic academic proficiency. Also, individuals show no knowledge about proficiency levels and standards in second language communication. Validated tests, submitted to empirical research, could inform low and advanced language learners. This research context is centered in school evaluation and placement in support programs.

There are several variables involved in language learning, particularly in second language acquisition, and that awareness has restricted investigation to specific factors and its impact in language coding and decoding. Some validation work should be enhanced in order observe: teacher's knowledge and methodologies (created, adopted) as also student's knowledge about instruments and rhythms; effective application immersed in classroom; and results documented and disseminated (coherence of data). Still in national context (Portugal) the Instituto de Linguística Computacional presents some advance in this area, with recent information about instruments or corpora of results that are mainly descriptive (Direção-Geral da Educação, 2009).

1.1. Literature review update

A literature review update conducted by Howard (2009) shows important issues that are investigated in last years of international scientific research, regarding English second language learners, since kindergarten until grade 12. Evidence reveals that foreign students after several years of immersion did not reach to the same syntactic skills as the native English-speakers (Howard). This fact leads to other issue: the amount of instruction in first language could influence the performance in second language. In the other hand, specific other factors were underestimated such as the amount of first language instruction and its impact in second language proficiency. Other variables are confirmed by recent research as determinant for second language achievement such as arrival date, gender, parental involvement, age, socioeconomic status, newcomer programs, sheltered instruction, previous schooling. Related to previous schooling, teacher's experience and teaching methods would be an important factor to be analyzed more depth. Teachers credentials and knowledge could be a disadvantage for immigrant student's progress. Although several factors could explain second language proficiency, only specific variables are strong predictors,

The strongest predictors of academic success of ESL students include proficiency in first language, proficiency in second language, past academic achievement in L2 and the amount of formal schooling prior to ESL instruction. Apart from being strong predictors, the evidence supporting these relationships is also relatively strong. More modest evidence supports that the

amount of ESL instruction time, parental involvement, age at time of arrival or length of residence, individual differences (including motivational factors) and socioeconomic status are also important predictors of success of ESL students. In addition, there is also some evidence suggesting that teacher credentials and resource investment may also be related to academic achievement of ESL students. (Howard, p. 12).

In fact, in our current study we have, in parallel, a study applied to language teachers from several Portuguese schools to evaluate the representations of professionals about methods and materials of second language learning. Scaffolding strategies, adopted in some cases by teachers in classroom context, would be probably important to valorise the first language and heritage culture of immigrant students, enhancing motivation and predisposition of young children placed in dual language programs. This meta-analysis found that this factor has been underestimated by scientific research.

Considering the study of processes and the study of instruments for assessment, international research has provided great evidence and materials (Alderson & Banerjee, 2002, Alberta Education, www.learnalberta.ca) for the intervention in second language learning. English is the main language verified for different published tests and offers some advantages for adaptation purposes,

This discussion of language proficiency assessment instruments concentrates on English proficiency assessment because that is the one constant in all ESL and bilingual education programs. Some of the instruments have non-English forms, and some can easily be adapted to other languages. Implications and applications for other languages are discussed in the context of each instrument. "Assessment of second language acquisition", p. 23).

However in portuguese context, and considering that Portuguese is a significant spoken language worldwide, there were no materials for school support. Additionally, digital educational resources, considering European context, are underestimated in this area. There are several platforms and packages of educational tools but little reference to second language learning. The instruments available in our schools have a limitation: psychometric validation. The great concern is to develop valid measures, with previous exploratory research. One of the causes of school dropout is the situation of migrant students who leave school once they do not have support and encouragement to pursue academic achievement (Eurostat, 2011) which posits Portugal in the lower levels of academic success indicators. Since 2005 several recommendations were made to schools in order to develop measures and procedures that contribute to the integration of minorities and for their nourishment in academic plan (OI, 2011; 2012). The students with "origin in immigration" became the priority of the II Plan for the Integration of Immigrants for the period 2010-2013, through valid knowledge of the school population (demographic information on immigrant students, specificities of the minorities) and information about training of the teachers assigned for the instruction of Portuguese as second language.

The diversity of instruments and worldwide research is positive but could present some constraints when professionals need to select or adapt methodologies for assessment. According to Gascoigne & Parnell (2013),

Given the tremendous amount of professional literature devoted to tasks, either as a vehicle for focus on form or as a subject in its own right, it was surprising to find so little mention or use of tasks in the 18 commercially-available, post-secondary French materials reviewed. Indeed, only three out of the 18 post-secondary French textbooks reviewed and published in the U.S. between 1996 and 2012 (...). The tremendous variability between what we say about tasks and what we do in terms of the actual success of these applications—at least as represented in many post-secondary French textbook programs—is disappointing. One recommendation stemming from this admittedly anecdotal investigation is that, instead of taking the introductory descriptions of commercially-available materials at face value, language educators in the U.S. (in French or any other language) might be well served by borrowing the Common European Framework of Reference Guide for User's six-step auto-reflection tasks and four materials selection questions presented above. (...). By sharing materials, techniques, and guidance as evidenced in the Standards (in the U.S.) and the Frameworks (in Europe), we can move one step closer to our goals of multitasks, multiskills, and multiconnections." (p. 98).

One of the specificities of the second language investigation should focus on the awareness that instruments and research evidence related to foreign language acquisition should not be misunderstood as transferable tools for the second language acquisition field (for evaluation of skills both in foreign and second languages. They are different domains). Additionally, there is some

ambiguity generated by the transference of measures and procedures applied in mother tongue that are perceived, mostly by teachers, as accountable for second language learners. This current project will follow mainly the structure of international tests which are available for English second language (ESL) context with valid psychometric properties such as Woodcock-Munoz Language Survey Revised (Woodcock, Munoz-Sandoval, Ruef, & Alvarado, 2005) and LEXTUTOR (Nation, 1990). The selected tests (after the adaptation process) will focus on orality and writing skills (comprehension and production) and will provide evidence regarding “intelligibility” and “comprehensibility” in second language. Our focus will be on the evaluation of the cognitive academic language proficiency to provide indicators of different skills. For the adaptation work we based our tests on the resources and benchmarks provided by: International Language Testing Association (ILTA), Repository Alaska Department of Education, Minnesota Language Proficiency Assessments, Digital Repository for Students with Special Education Needs, Teachers of English as Foreign Language (TOEFL), among others. After an extensive analysis we selected these sources as the most adequate to our research purpose, focusing ESL models in American contexts (and not models applied to romance languages, in European domain, considering that no complete approaches were found specifically for these research goals).

2. New methodology in validation process: Portuguese migrant students and cognitive profiles

The main interest of this Portuguese research is to understand the sources of variability in two ways: variability of performance (Myford & Wolfe, 2000) considering the different language speakers that integrate the sample of this specific study; and the variability of achievement observed in other types of assessment applied to the similar groups of speakers. It is recommended that literacy practitioners should be aware of the necessity of validated and accurate instruments to assess and to place students at risk. Immigrant or Bridging students are identified as being school population at risk due to the limited proficiency. These instruments should be developed in order to respect and to understand that different speakers (different mother tongues) lead to different minds and decoding abilities. In this study we have critical issues that became the main goals of this project:

What are we testing?

The cognitive academic language proficiency is the principal concern considering that proficiency is emergent and dependent on the formal instruction and school support.

Which type of testing?

Comprehensive diagnostic.

Will the test be normed with individuals of the same cultural and linguistic groups?

The main goal is to test different speakers and different nationalities.

Are the test items comprehensible according to the fact that individuals are second language learners?

A pilot study was conducted in 2013 before the ongoing application in schools. The test will be analysed regarding if it is feasible and adequate, including to provide school staff with the training they need to apply the instrument and measure the results.

Which levels of proficiency are involved?

The individuals were selected regarding their proficiency between A1 – B1.

Score of the test?

The score will be based on scoring method from Woodcock-Munoz Language Survey – Revised Normative Update (2005). The quantity of variability of performances will be examined in what extent that defines the different profiles of speakers. Second language learner is the target, excluding the second language user.

And, does the test have a strong theoretical basis?

All the tasks and theoretical framework are based on the main sources and evidence consulted for the tests construction and adaptation: Alberta Education, Educational Testing Service, Test of English as a Foreign Language, Diagnostic English Language Needs. The framework of this research is greatly related to practical evidence, considering that there is few theoretical strong data concerning this area. There are other criteria to be aware,

According to Kane [2001, cited by Uysal, 2009] however, insistence on the necessity of a theory base for all types of assessments was meaningless especially in the areas where there is little theory, and the uniform approach to validation that is too theoretical and ambiguous caused confusion understanding what construct validity and validation study meant. (...) Kane (2001), on the other hand, suggested an alternative approach that was unified, but flexible in which, different kinds of validity arguments to support different kinds of inferences could be made according to the context. While the details of the validity argument for each interpretive argument would be unique, the general approach to specify and evaluate the inferences would be consistent or unified. Kane's validity definition did not require a theory; yet, it was still reflecting on the general principles in the construct model. Kane suggested adopting an argument-based approach, an interpretive argument rather than validation research. (...)

According to this, measuring performance and giving scores will be important to determine profiles and patterns for the generalization of scores to the population represented in similar contexts of evaluation. Despite of the american evidence that underlies the components of this test, the principles of assessment and contribute for placement and supervision of school programmes are consistent with the european educational policies and immigrant emergent needs.

The main organization criteria are the inclusion of subtests regarding different skills and levels, including several independent items and resulting in several observations. Concerning the score, rubrics will be determined to inform raters about specificities to be assessed according to each skill to be measured (for example, fluency and coherence in writing tasks). Then, some 'calibration' will be considered to meet the patterns of answers from second language learners. Conventional and particular errors gathered from overall performances will be evaluated and registered to determine profiles of Portuguese second language learners, in school context. Specific tasks of the diagnostic tool are related to cognates and pseudowords, testing the knowledge of learners about spelling conventions. Positive and negative transfers are also observed with those tasks, analyzing inappropriate transfers between languages. Recall after reading will be tested as also comprehensibility and general phonological awareness. Such verbal behaviours, resulting from the tests battery, could inform about specific homopheneity (Binnie, Montgomery & Jackson, 1974; Figueiredo & Silva, 2013) across languages, taking into account the type of languages spoken at home by individuals.

Besides de validation process, our research intend to provide a complete instrument to be recognized as educative policy applied as academic support for the large population of immigrant pupils. This Diagnostic Test has specific features that differentiate it from other instruments available in Portuguese language. It is not an assessment for certification or admission, it is not simply a research instrument, it is a diagnostic evaluation that will enhance management of placement into adequate programs that should be implemented at schools, with specific training for language teachers. These are the main principles of Diagnostic English Language Needs Assessment (DELNA), an important language screening and assessment procedure of New Zealand (Read, 2008).

2.1. Test configuration: new insights into second language education

Reading, writing and comprehension skills are the main target of this diagnostic test. Oral production is not assessed at this project. This study presents a battery of 9 tasks constructed on the following selected levels:

- verbal analogy,
- phonetics perception,
- foreign accent,
- story recall,
- cognates,
- morphological manipulation,
- transference.

Based on previous tests, in English language, were adapted several tasks:

- lexicon identification: a words list with different cognitive levels of correspondence between words (adapted from Lextutor, Nation, 1990). The Portuguese test considers the index of Portuguese (European variant) words frequency (the criterion was stated in 1000 occurrences), even considering that this index is available for adults (based on lexical production and comprehension) only.

Furthermore in the lexicon task, was added an exploratory task to observe which words frequency should be provided for the (second language) speaker to accomplish their acquisition. Three texts present different three semantic cores (with repeated words along each text) and those texts were examined with LEXILE measure (measure of difficulty on texts for grades and readers, by calculating in a scale; available in English and Spanish only). All the texts should be placed between 610 L e 1000L, according to Lexile indicators.

- Morphological transformation: the individuals should make changes on the morphological segments of words in different contexts (adapted from "Extract the Base Test", August & Kenyon , 2001).
- Transference test: individuals should recognize correspondances between words, to be observed transference between first and second language (adapted from "Cognate awareness test", August, Kenyon, Malabonga et al., 2008).
- Identification of words and pseudowords based in dichotic hearing task (adapted from "Phonological Transfer Test", Rolla & August ,2002).
- Verbal analogy and verbal naming: images and analogies to identify (adapted from "Woodcock-Munoz Language Survey-Revised (WMLS-R)", 2005).
- Comprehensibility and intelligibility in the L2 speech: listening of brief narrative and judgement about difficulty and foreign accent.
- Writing test (with story recall task) based on images sequence and descriptors of http://www.learnalberta.ca/content/eslapb/writingsamples/docs/grade1_2.pdf; for the story recall, a text will be listened in an electronic device and should be write as complete as possible. It is expected appropriated vocabulary, events narrated in appropriated order, grammatical sentences with correctness in syntax.
- Oral comprehension test based on "Test Of English as a Foreign Language" (TOEFL).

This diagnostic test uses authentic language samples (in Portuguese) and work sampling in order to report language performance in real-time and 'in function', excluding interviews, analytic teaching, criterion-referenced tests, curriculum-based assessment, classroom observations, review of records ("The assessment of second language acquisition"). Again, all these techniques involve different tasks which enhance the capacity to inform teachers and schools about a more complete picture of proficiency,

(...) do not rely on one instrument to meet all needs. Select a variety of instruments or strategies for the most complete picture of a given student or a group of students. Instruments that work well for initial identification and program exit may not be useful in helping a teacher plan for a given class of students or for individualizing instruction. ("The assessment of second language acquisition", p. 22).

The construction of our test will be focusing on students from basic and high school (8 years old, minimum) and we determine the following demands:

- Linguistic lexicon (knowledge of 1500 words) related to familiar objects and based on images (level 1),
- Syntax (knowledge of simples sentences and related with familiar routines, based on images),
- Questions of strategic type (knowledge of directives, yes/no questions, based on images),
- Clarification (using gesture, mimic and simple words),
- Sociolinguistic awareness (comprehension of social and familiar expressions),
- Speech (comprehension of few words in familiar sentences, based on images),
- Auditory discrimination (decoding of several phonemes, not necessarily explaining readiness for writing comprehension).

All the tasks are being applied since May 2013, in schools of Lisbon district, in paper and electronic format (sounds files). Several criteria were considered for the sample to observe how different independent variables would be influent to determine performance and profiles to be defined. Is expected that nationalities, home languages, age, parent's home languages, parent's professional

activities, lateralization (hand dominance), amount of instruction in L1 will be important to establish differences between individuals and, mainly, to identify specific profiles of language learners in Portuguese context. However should be noticed that "According to Haynes (2007), it is a myth that children learn second languages either "faster or easier" than adults, although children may master English accents better than adults." (Howard, p. 9). The mastery on intelligibility of discourse, mainly related to the native accent, is not an exclusive factor that distinguishes young from older learners in second language achievement (Figueiredo & Silva, 2013). In fact we understand, based in previous research with Portuguese immigrant learners, that adults could perform better in several levels of language. These conditions will be analyzed to replicate or reject recent data. Those profiles will be new insight and contribute for scientific research in Portuguese second language learning field. Further, all the tasks, after the empirical testing, will be available and informed in a repository of instruments for second language learning and testing that constitutes the main second milestone of this scientific post-doctoral project. With this repository, available in electronic platform, data could be easily stored and provide full access to scientific community and to school's professionals. In the other hand, this repository should be placed in 'interactive' manner to receive and document incoming data related to new methodologies, provided by schools as also by researchers, worldwide.

REFERENCES

- Abedi, J. (2002) Standardized Achievement Tests and English Language Learners: Psychometrics Issues. *Educational Assessment*, 8(3), 231-257.
- Alderson, J., & Banerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching*, 35, 79-113.
- "Assessment of second language acquisition", <http://www.wce.wvu.edu/Resources/CIRCLE/Articles/The%20Assessment%20of%20Second%20Language%20Acquisition.pdf>
- Bijl, R., & Verweig, A. (2012). *Measuring and monitoring immigrant integration in Europe - Integration policies and monitoring efforts in 17 European countries*. The Netherlands Institute for Social Research.
- Binnie, C.A., Montgomery, A.A., & Jackson, P.L.(1974). Auditory and visual contributions to the perception of consonants. *Journal of Speech and Hearing Research*, 17, 619-630.
- Direção-Geral da Educação (2012). *Teste de Diagnóstico de Português Língua Não Materna* (1.º ano – 12.º anos). <http://www.dgidec.min-edu.pt/outrosprojetos/index.php?s=directorio&pid=195>. DGE.
- Direção-Geral da Educação. (2009). *Processo de validação dos testes de diagnóstico de PLNM*. DGE.
- Educational Testing Service. (2005). *Guidelines for Constructed-Response and Other Performance Assessments*. Educational Testing Service.
- European Commission (2001). *Quadro Europeu Comum de Referência para as Línguas: Aprendizagem, ensino, avaliação*. Porto: Edições Asa.
- EUROSTAT (2011). *Education trends 2000-2009. From 10 to 20 pupils per teacher in primary education across Member States in 2009*. Eurostat.
- Figueiredo, S., & Silva, C. (2013). *Ritmos e competências em avaliação: minorias linguísticas e contextos pioneiros de avaliação*. Novas Edições Académicas.
- Figueiredo, S., Silva, C. (2009, Dezembro). Comportamento textual e discriminação de identidades articulatórias: desempenho diferenciado em aprendentes de Português Língua Segunda. *Revista de Avaliação Psicológica*, 8 (3), 289-301. ISSN: 1677-0471.
- Gascoigne, C., & Parnell, J. (2013). Tasks and Focus on Form: Connections in the US and Europe. In S. Dhonau (Ed.), *Multitasks, Multiskills, Multiconnections*, pp. 89-101. University of Arkansas of Little Rock.
- Howard Research & Management Consulting Inc. (2009). *Kindergarten to Grade 12 English as a Second Language Literature Review Update*. Howard Research & Management Consulting Inc.
- Myford, C., & Wolfe, E. (2000). *Monitoring Sources of Variability Within the Test of Spoken English Assessment System*. Educational Testing Service.

Nation, P. (1990). "Lextutor": <http://www.lex tutor.ca/tests/> (1990).

Read, J. (2008). Identifying academic language needs through diagnostic assessment. *Journal of English for Academic Purposes*, 7, 180-190.

Rolla, A., & August, D. (2002). "Phonological Transfer Test". http://www.cal.org/acquiringliteracy/assessments/phono_transfer.html.

Seabra, T., Mateus, S., Rodrigues, E., & Nico, M. (2011). *Trajetos e projetos de jovens descendentes de imigrantes à saída da escolaridade básica*. Observatório de Imigração.

Stowe, L. (2006). When does the neurological basis of first and second language processing differ? Commentary on Indefrey. In Gullberg & Indefrey (Eds.), *The Cognitive Neuroscience of Second Language Acquisition* (pp. 305-311). Oxford: Blackwell Publishing, Ltd.

TOEFL (2012). *TOEFL® Institutional Testing Program (ITP) & TOEFL Family Products*. Brazil.

<http://www.ets.org/pt/toefl/>

Uysal, H. (2009). *Developments in validity research in second language performance testing*. <http://www.ubplj.org/index.php/bjll/article/view/15>

Woodcock, R. W., Munoz-Sandoval, A. F., Rued, M.L. & Alvarado, C. G. (2005). *Woodcock-Munoz Language Survey – Revised, English*. Itasca, IL: Riverside Publishing.