

## Resenha

### Measuring L2 Proficiency

LECLERCQ, Pascale; EDMONDS, Amanda; HILTON, Heather (Eds). **Measuring L2 Proficiency: Perspectives from SLA**. Bristol: Multilingual Matters, 2014.

Ana Paula Scholl\*

The concept of language proficiency has been a subject of debate among teachers and researchers for many years. The assessment of second language (L2) proficiency is the core of the language testing industry and the concern of many professionals when it comes to putting in practice the idea of proficiency they hold. *Measuring L2 Proficiency: Perspectives from SLA* aims to collect ideas on identifying and measuring L2 proficiency, brings together contributions from different branches of second language acquisition (SLA) research and narrows the gap between research and language teaching. Edited by Pascale Leclercq, Amanda Edmonds and Heather Hilton, this volume has contributions from researchers who concentrate their work on different aspects of language proficiency. Focusing on the constructs of reliability and feasibility in the assessment of L2 proficiency, the authors discuss indicators of productive or receptive proficiency in L2 that are valid and practical to implement.

The book consists of a short preface, an introduction, three sections and a brief epilogue. The first part, entitled *General Considerations for L2 Assessment* comprises three chapters. The second part, *Language Processing and L2 Proficiency*, is also composed by three chapters. The last part, *Focused Assessment Instruments*, contains four chapters.

In the introduction, entitled *How to Assess L2 Proficiency? An Overview of Proficiency Assessment Research*, Pascale Leclercq and Amanda Edmonds provide an overview of proficiency assessment research and situate the different contributions of the book. According to the authors, the book is an attempt to respond to the need for valid, reliable and practical

---

\* Mestranda em Letras (Linguística Aplicada) pela Universidade Federal do Rio Grande do Sul (UFRGS) com bolsa da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Contato: anapaulascholl@gmail.com.

methods of assessing L2 proficiency. Different conceptions and definitions of proficiency are brought together and the authors postulate that the definition which is most relevant for this volume is that of Hulstijn (2011:242): “language proficiency is the extent to which an individual possesses the linguistic cognition necessary to function in a given communicative situation, in a given modality (listening, speaking, reading, or writing)”. The term language proficiency needs to be defined when writing about proficiency assessment, so that the construct being measured is clear, and the editors successfully accomplish this in the introduction chapter of this volume. This introduction is of fundamental importance given that it clarifies many of the terms used throughout the book and the different points of view adopted by the contributing authors.

Initiating the first part of the book, Heather Hilton discusses L2 speech production and fluency by analyzing data from a corpus in the chapter *Oral Fluency and Spoken Proficiency: Considerations for Research and Testing*. After briefly discussing what proficiency entails, the author presents current L2 production models, contrasting the concepts of *production* and *fluency* and setting Kormo’s model of bilingual speech production as their theoretical framework. This contextualization is very important, so the reader knows the theory and concepts underlying the analysis. Hilton then proceeds to present the PAROLE corpus, its design and specifications, clarifying the criteria used to define different speech phenomena and how they were coded. In this study, the speech production of 45 participants with different proficiency levels (33 learners of English, and 12 of French) was collected in order to look more closely at temporal features of L2 speech in an attempt to understand more clearly what makes a speaker proficient. Based on the temporal characteristics (speech rate, hesitation rates, numbers of words/syllables between pauses, etc.) of the corpus production, three fluency levels were identified: low, mid and high FI (Fluency Index). The chapter presents both quantitative and qualitative results, which, in conjunction, provide a better understanding of the data. The quantitative part shows characteristics of the groups’ (low, mid and high FI) production regarding mean length of utterance, lexical diversity, rates of error and syntactic and informational delivery. Using qualitative results, the author establishes the link between fluency and proficiency by giving many examples of native and L2 learners’ speech productions and analyzing the phenomena found in them. The concepts throughout the chapter are specified, making clear what is meant by fluency and its components. Limitations are pointed out and suggestions for future research are given.

The following chapter, *Multiple Assessments of Oral Proficiency: Evidence from a Collaborative Platform*, written by John Osborne, raises the question of to what extent people implementing or using the Common European Framework of Reference (CEFR) have the same understanding of their scales and descriptors and apply them in similar ways when assessing learners' spoken production. In order to compare assessments and check whether the understanding of the guidelines is consensual, it is useful to have them complemented with a resource in which users can contrast different judgments. The author presents the WebCEF platform, an online tool for the collaborative assessment of oral language proficiency in which tasks are created, learners are recorded while doing those tasks, data are published to be accessed by other members, and performances are assessed using online grids. Some examples of task assessments are analyzed with quantitative measures of inter-rater agreement, considering consistency and consensus when evaluating overall spoken interaction, range, accuracy, fluency, interaction and coherence. Furthermore, the author examines the raters' annotations for each of these criteria, in order to identify which features influenced the level which the raters assigned to a given sample. By providing quantitative data as well as a qualitative analysis of some sample evaluations on the WebCEF platform, it is possible to look more closely at how examiners rate oral proficiency and how their annotations are in accordance with the CEFR descriptors. The combined analysis suggests that even though the CEFR guidelines prime for transparency, not everybody understands and uses them in the same way and there are aspects that raters find important that are not comprised in the descriptors. The chapter is insightful as to how different people rate the same samples of learners' oral production using the CEFR descriptors and showing how a collaborative platform can be an important tool to understand how proficiency descriptors are applied by different raters.

The last chapter of the first section, entitled *Using Learner Corpora for Testing and Assessing L2 Proficiency*, is written by Marcus Callies, María Belén Díez Bedmar and Ekaterina Zaytseva. The authors define the main concepts within learner corpus research (LCR) and language testing and assessment (LTA), and provide the reader with a clear picture of what these terms mean and when they are used. The chapter joins both areas to show how LCR can be used to assess L2 writing proficiency, distinguishing the different approaches to using learner corpora in LTA: *corpus-informed*, *corpus-based*, and *corpus-driven*. Later on, the authors present some challenges to the use of learner corpus research in LTA, pointing out some methodological issues and how they can affect studies in the area. The authors provide two

examples describing different approaches to the use of learner corpora for assessing L2 proficiency in writing. The first one analyzes 302 compositions of a group of L2 English learners with the same institutional status, using the CEFR to establish the proficiency level and error analysis of the written production. The second example demonstrates how a corpus (*Corpus of Academic Learner English*) can be used to assess advanced learner language in academic writing. In order to do so, the authors operationalize the concept of advancedness by identifying linguistic descriptors that are characteristic of this register. According to the authors, this kind of analysis can be useful in various stages of testing because it offers evidence of what should be measured in academic writing, influences the development of realistic tasks and provides usage-based information relevant to the evaluation process. By giving examples, the authors demonstrate how learner corpora can be used as a valuable source of information on learner language use in LTA. Also, they raise important questions about the criteria that should be met in this type of analysis and about how proficiency homogeneity is not guaranteed in a learner language corpus.

The second part of the book opens with the chapter *Listening Comprehension: Processing Demands and Assessment Issues*, written by Peter Prince. Unlike most chapters in the book, which focus on productive skills, Prince's focuses on a receptive skill: listening comprehension. Since listening cannot be observed directly, it is a difficult skill to be assessed. The author raises the question of what extra processing (e.g., memory and reading skills) are required in order for a test taker to answer either multiple choice or open-ended questions regarding what they have listened in a task. According to the author, regardless of the type of test adopted, assessing listening involves other processing demands that can interfere with listening processing. In a clear and objective manner, the nature of listening comprehension and the processes involved in it are presented, with relation to what should be considered in devising or choosing a listening test. The study described in the chapter aimed at examining the relationship between word recognition and task accomplishment - the extent to which participants accurately reported the input according to the task instruction - by having 10 students perform a news summary and a dictogloss task. The chapter ends with a discussion about what is involved in the listening tasks analyzed, comparing the results with other studies and making suggestions for future work.

The chapter *A Psycholinguistic Measurement of Second Language Proficiency: The Coefficient of Variation*, written by Carrie A. Ankerstein, focuses on lexical access - a central

feature of L2 proficiency - from a psycholinguistic perspective. The author gives an important explanation about the process of lexical access, differentiating automatic and attentional processes, followed by some examples of tasks used to measure lexical access and some considerations about the possibility of the attentional process of lexical access becoming automatic for high proficiency L2 learners. The study reported in this chapter aimed to answer three questions: (1) Is word recognition quantitatively and qualitatively different as a function of word frequency for L1 and university-level L2 speakers of English?; (2) Are there quantitative and qualitative differences in semantic priming for L1 and university-level L2 speakers?; (3) Is semantic priming stronger for low-frequency target words compared to high-frequency target words for both groups? In order to answer these questions, 30 native speakers of English and 30 L2 English speakers performed a lexical decision task in which they had to decide if the stimulus on the screen was a word or a non-word. The target words were controlled for frequency, syllable length and imageability and appeared with related, unrelated or neutral prime words. The findings are discussed by readdressing the three research questions proposed, what the author does in a well-organized and objective manner. The chapter offers some insight into how lexical access works and how native and non-native processing may be, shedding light on cognitive resources used for L1 and L2 processing.

Dominique Bairstow, Jean-Marc Lavour, Jannika Laxén and Xavier Aparicio compose the last chapter of the second section: *Evaluating the Workings of Bilingual Memory with a Translation Recognition Task*. First, the authors discuss how the links between two languages can be established in the bilingual mental lexicon. They propose the use of a translation recognition task in order to evaluate the types and strength of these links, assuming that bilingualism is not a fixed state and that the mental lexicon is flexible – additional links can be created in memory, others can weaken or even disappear. Next, they present different models of bilingual lexical memory, setting the BIA+ model as the basis for their study. Their pilot study's goal was to evaluate the ease with which participants go from a lexical to a semantic level and their capacity to establish semantic links while language switching. To this end, a translation recognition task was devised, in which two words of different languages appeared and the participants had to judge if they were translation equivalents. The words were controlled for familiarity, orthographic similarity and language order, and 25 speakers of L1 French and L2 English performed the task. The results indicate that translation ability is positively correlated with self-reported proficiency in the L2 and there is a familiarity effect. Even though

the authors are confident that the translation recognition task is a way to dissociate the words from each other based on their form and/or meaning, more evidence could have been shown in the chapter supporting this idea and to what extent their task can provide us with indications of the way links are progressively created in the course of acquiring a second language, so the reader could have a better understanding of how the task can fulfill its purposes.

Opening the third part of the book, Nicole Tracy-Ventura, Kevin McManus, John M. Norris and Lourdes Ortega give an informative description of the development of an elicited imitation task to be used to measure French L2 oral proficiency in the chapter '*Repeat as Much as You Can*': *Elicited Imitation as a Measure of Oral Proficiency in L2 French*. The authors discuss issues in defining proficiency as well as measuring it, considering the fact that the definition of proficiency can be divided into two types of language ability – basic and high language cognition – and that proficiency is a major variable that should be included in all studies related to SLA. According to the authors, a good way of measuring basic language cognition in the L2 is elicited imitation (EI), an approach that requires participants to listen to a stimulus and then repeat it as accurately as they can. The study presented in the chapter had 29 L2 French-degree students performing an EI task with 30 test sentences, with results suggesting that the French version of the EI task is valid and reliable for the measurement of oral proficiency in SLA research. With well-connected sections, the authors discuss theoretical issues and present their study with methodological precision, indicating also some ideas for future research and limitations of their work.

The subsequent chapter, *Exploring the Acquisition of the French Subjunctive: Local Syntactic Context or Oral Proficiency*, is written by Kevin McManus, Nicole Tracy-Ventura, Rosamond Mitchell, Laurence Richard and Patricia Romero de Mills. They investigate how variations in oral L2 proficiency and local syntactic context account for learners' use and knowledge of the subjunctive in French. Given that the frequency of use of the subjunctive is very low in tasks not specifically designed to elicit it, previous studies cannot confirm the difference between structures that have not been acquired yet by the learner and structures that are simply not produced. The study in this chapter set out to answer the questions: (1) Do differences in L2 proficiency and local syntactic context affect the learner's use of the French subjunctive in oral and written production? and (2) Do differences in L2 proficiency and local syntactic context affect learners' knowledge of the French subjunctive? Three tasks – an argumentative writing task, a semi-structured oral interview, and a timed grammaticality

judgment task – were completed by 23 learners of French as an L2 divided into two groups (low and high score groups) according to their scores in an elicited imitation task, as well as 10 native speakers of French. The results are presented in detail with the aid of several tables, demonstrating that only a few manifestations of the subjunctive were used, even by the native speakers of French. The authors readdress their research questions and hypotheses, confirming or refuting them in view of their findings. The goal of the study is clear throughout the chapter, making it easy for the reader to follow why the methodological decisions were made and how the questions were answered.

Chapter 10, *Testing L2 Listening Proficiency: Reviewing Standardized Tests Within a Competence-Based Framework*, written by Naouel Zoghalmi, stresses the fact that listening comprehension is a major component of language competence and that high-quality listening tests need to be theory-driven. The author investigates the listening constructs on which two tests are based: the *Oxford Placement Test* (OPT) and the *First Certificate in English* (FCE). Drawing on Buck's (2001) competence-based listening framework, Zoghalmi presents a competence L2 listening comprehension ability model and describes the basic skills or behaviors that listeners would exhibit as a result of a linguistic and strategic processing. In her study, the aim was to compare the listening segments of two commercialized English tests and investigate to which degree these tests predict overall listening proficiency in the French context. In total, 130 French learners of English as a second language performed the OPT and FCE listening sections. Quantitative findings suggested that the OPT does not reflect overall listening proficiency, given that some students who obtained above-average scores on the OPT got low scores on the FCE. A quantitative analysis indicated that the OPT listening sections does not account for higher-order listening skills and measures only recognition of prosodic and phonological patterns, operationalizing a restricted linguistic aspect of listening proficiency. On the contrary, the FCE seems to focus on higher-order abilities, fitting Zoghalmi's competence model of listening comprehension. The quantitative and qualitative analyses complement each other in an attempt to achieve the goals of the study and the author is successful in doing so.

The last chapter, entitled *Assessing Language Dominance with the Bilingual Language Profile*, is written by Libby M. Gertken, Mark Amengual and David Birdsong. In this well-written chapter, the authors conceptualize linguistic dominance, describe previous dominance assessment tools and present the Bilingual Language Profile (BLP) – an instrument for

measuring language dominance that is intended to be a reliable, valid and practical tool to describe bilingual participants in research. The authors point to the fact that language dominance and language proficiency, even though related, are distinct constructs. Dominance is thus defined as a global construct, informed by several factors related to knowledge, processing and effect, having proficiency as an essential component. By pointing out advantages and drawbacks from previous dominance assessment tools, the researchers show what the BLP owes to these previous assessments in terms of theoretical orientation and design and which shortcomings they intend to overcome. A description of the process of elaborating the instrument is given, as well as its features and how it can be used. The chapter is objective and clear, giving the reader insights into language dominance and making a case for the use of the authors' tool for assessment of language dominance.

The issues addressed in this book are very current and of crucial importance for all those working with the creation or use of tests to measure linguistic proficiency in an additional language. The first part of the book shows how corpora can be used to understand better different proficiency levels, applying the CEFR descriptors and error analysis. The critical position of the authors enables the reader to consider methodological issues when the studies involve the use of corpora in assessing oral and written L2 proficiency. The second part deals with psycholinguistic measures of L2 proficiency, focusing on how the languages are processed and what are the implications for assessing proficiency. The third part of the book contributes to the discussion about proficiency assessment by stressing the importance of clearly establishing the concepts and definitions underlying the tests used in order to know what is being measured with each test, as well as the importance of using independent measures of proficiency in SLA research.

By presenting a wide range of research tools for the measurement of L2 proficiency, this publication provides an array of options that can be used in both research and teaching contexts. Above all, this book stresses a need for having clear constructs underlying what is tested with each measure and inspires researchers and teachers to reflect on their own practices in testing L2 proficiency.

Resenha recebida em: 28.01.2015

Resenha aprovada em: 13.05.2015