Evaluating Spanish proficiency: The validity of the Profgram test and the role of grammar versus vocabulary in language assessment

Master's Thesis

Jytte Beelen (s1019285)

MA Language Variation and Multilingualism

Radboud University

Supervisor: Aurora Troncoso Ruiz

Abstract

This study evaluates the effectiveness and validity of the Profgram test in assessing CEFR-levels of Spanish proficiency, by comparing its results with the established LexTALE-Esp test and an adjusted DELE test. It also explores the relative contributions of grammar and vocabulary knowledge to second language proficiency. Forty participants, aged 18 to 54 years (M = 27.25, SD = 7.71), took part in the study, completing the Profgram, LexTALE-Esp, and DELE tests. The results showed that Profgram had strong validity as a language proficiency test. It demonstrated a moderate to strong correlation with DELE scores (r = .61), indicating that it effectively measures similar aspects of Spanish proficiency. A moderate correlation was found between Profgram and LexTALE scores (r = .48), suggesting that Profgram assesses more than just vocabulary. Regarding the role of grammar and vocabulary in second language performance, both components were equally correlated with LexTALE and DELE scores. In conclusion, the Profgram test is a valid tool for assessing Spanish proficiency, with vocabulary and grammar playing an equal role.

Introduction

Second language (L2) proficiency is a complex and multifaceted concept that refers to an individual's ability to communicate effectively in a second or foreign language. It includes different aspects of language use, such as speaking, listening, reading, and writing, and reflects the capacity to engage meaningfully in diverse communicative situations. Proficiency levels are commonly used to describe the extent to which individuals can function within these contexts, with higher levels indicating greater fluency and competence in the language (Cook et al., 2018).

The assessment of L2 proficiency plays a crucial role across various domains, including education, employment, and immigration. In education, accurately measuring L2 proficiency is essential for determining language learners' abilities and informing placement decisions. Reliable and accurate assessments ensure that students are placed in appropriate levels according to their linguistic abilities, facilitating effective learning. These assessments also provide insights for formative evaluation, allowing teachers to monitor student progress and adjust instruction as needed. Furthermore, they play an important role in summative evaluation and serve as a basis for language certification, which is often required for academic and professional advancement (Leclercq & Edmonds, 2014).

In employment, L2 proficiency assessments are crucial for determining candidates' suitability for roles that require bilingual or multilingual skills, ensuring that employees can communicate effectively within the workplace (De Jong, 2023). Similarly, in L2 immigration contexts, language proficiency is a critical factor in determining eligibility for citizenship, residency, or refugee status, as it impacts an individual's ability to integrate into the social fabric of a new country (Pont-Grau et al., 2023).

In addition to its applications in everyday settings, L2 proficiency assessment plays a vital role in research. Researchers rely on valid and reliable proficiency assessments to ensure that their findings are meaningful and accurate (Norris & Ortega, 2012). Without robust assessments, research outcomes may be misleading, undermining the credibility and applicability of the study. Thomas (1994) emphasizes the critical importance of valid L2 proficiency measures in empirical research, arguing that inadequate assessments restrict the generalizability of research outcomes and undermine the empirical claims made within the field. In this study, we designed and test a new tool for measuring L2 Spanish proficiency and test its robustness.

Proficiency assessment in research contexts

Despite its importance, proficiency assessment in research faces several ongoing challenges. As Thomas (1994) pointed out three decades ago, there remains a need for better control over the proficiency variable in second language acquisition (SLA) studies. When carrying out studies that require a proficiency assessment, researchers can choose among different possibilities, including standardized tests of language proficiency (e.g., TOEFL, IELTS), self-reported measures, and tests specifically designed for research (e.g., LexTALE). However, the choice between these tests comes with advantages and disadvantages.

Standardized tests, such as the Test of English as a Foreign Language (TOEFL), International English Language Testing System (IELTS), and Diplôme d'Études en Langue Française (DELF), are widely recognized for providing reliable measures of L2 proficiency. These tests are well-established, with questions that have been validated to assess multiple skills, including listening and reading. However, despite their reliability, a significant challenge lies in their practical implementation within research. These tests are often not suitable for research purposes due to being time-consuming and resource-intensive (Olson, 2024). For instance, the DELE exam typically requires 2 to 3 hours to complete, and administering such tests often comes with substantial financial costs, which can pose barriers for researchers, especially in contexts with limited resources.

An alternative to standardized testing is self-reported proficiency measures, where participants rate their own language skills. These measures are often used in large-scale studies because they are quick to administer and cost-effective. Self-reports, therefore, help resolve the logistical issues associated with traditional tests, as they do not require lengthy testing periods or significant financial investment. However, self-reported measures can be subjective, with participants overestimating or underestimating their own proficiency levels, which makes them less accurate than objective testing. The lack of standardization and the influence of individual biases are significant drawbacks, leading researchers to approach self-reports with caution despite their logistical advantages.

In response to these logistical challenges and reliability concerns, other attempts have been made to develop quicker proficiency assessments, such as the LexTALE tests. These tests focus primarily on vocabulary knowledge, providing a fast and simple way to measure a learner's familiarity with words in the target language. While they address the need for efficiency and are often seen as a practical solution for research contexts, they do have their limitations. Specifically, they do not assess grammar or other essential language skills, which may be critical for a comprehensive evaluation of L2 proficiency. Therefore, while tests like

LexTALE offer a convenient tool for certain research settings, they do not provide a full picture of a learner's language abilities.

The logistical challenges of standardized tests, the reliability issues of self-reported measures, and the limitations of current research tests underscore the need for alternative proficiency assessment methods that are cost-effective, easier to implement in diverse contexts and reliable. This need forms the core motivation for this research, which aims to explore more reliable and efficient measures of L2 proficiency that can be used in both educational and research settings without the limitations of traditional standardized testing.

This study presents and tests Profgram, a new Spanish language test that uses both grammar and vocabulary questions to assess proficiency. As an adaptive test, Profgram adjusts question difficulty based on participant performance, offering a more personalized assessment experience. While grammar and vocabulary are both fundamental components of language proficiency, it is unclear how each individually contributes to overall proficiency measurements. By investigating the roles of grammar and vocabulary in Profgram, this thesis aims to provide a clearer understanding of how each area is associated with language assessment and ensures the test's effectiveness in measuring proficiency. In the following sections, we discuss the CEFR and its relevance to Spanish language testing, the importance of grammar and vocabulary in L2 learning, and the role of Spanish proficiency tests in evaluating language skills.

Theoretical framework

The Common European Framework of Reference for Languages (CEFR) and its relevance to Spanish language testing

The Common European Framework of Reference for Languages (CEFR) provides a standardized framework for assessing language proficiency across European languages, including Spanish. Developed by the Council of Europe, it defines proficiency levels from A1 (beginner) to C2 (native/near native), offering a common reference point for language assessment practices (Council of Europe, 2024). In the context of Spanish language testing, the CEFR serves as a valuable tool for evaluating learners' language skills and aligning assessment objectives with established standards (North, 2020).

The significance of the CEFR lies in its provision of standardized reference points and descriptors for each proficiency level, enabling educators, test developers, and learners to align curricula, assess language proficiency, and reflect on teaching practices (North, 2020).

Spanish language assessments based on the CEFR offer thorough evaluations of listening, speaking, reading, and writing skills within a globally recognized framework. However, despite its widespread adoption, the CEFR is not always straightforward to implement in practice. Many of its descriptors are broad or open to interpretation, making them difficult to translate directly into measurable test items. While the CEFR encourages the creation of standardized, transparent materials (North, 2020), applying its descriptors meaningfully requires careful adaptation and expert judgment to ensure validity, reliability, and fairness in testing.

To address these challenges and build on the CEFR's advantages, the Instituto Cervantes (2006) created guidelines for teaching Spanish in the *Plan curricular del Instituto Cervantes: Niveles de referencia para el español*. This book presents an updated version of the objectives and content for teaching Spanish, based on the six levels of language learning outlined by the CEFR. While the CEFR descriptors focus solely on communicative functions, the *Plan curricular* connects these descriptors to grammatical and vocabulary content. Considering both grammar and vocabulary is important for fully understanding of language proficiency, as they reflect the different skills required for effective communication. This approach helps create a more complete way to assess language ability, which brings us to the next section where we explore the role of grammar and vocabulary in second language acquisition and its importance in proficiency testing.

Importance of grammar and vocabulary in second language learning

The integration of grammar and vocabulary is essential for effective language acquisition, as supported by two key theoretical frameworks: the Construction-Based Theory (Goldberg, 2003) and the Usage-Based Approach (Lieven et al., 2003). Both theories emphasize the interdependence of grammatical structures and lexical items in language learning, but they conceptualize the process differently.

The Construction-Based Theory suggests that language learners acquire linguistic forms through exposure to constructions – specific pairings of form and meaning (Goldberg, 2003: 219). These constructions are not limited to sentence structures but also include more specific, formulaic expressions. According to this theory, both grammatical rules and vocabulary are learned together through repeated usage in meaningful contexts. In this view, learners do not simply internalize abstract grammatical rules in isolation. Instead, they learn patterns of usage that combine both grammar and vocabulary. As such, grammatical

competence is not viewed separately from lexical knowledge, but as part of a unified system of linguistic constructions.

Developments in linguistic theory have increasingly highlighted the relevance of Construction Grammar for understanding language acquisition. Now widely recognized as both a descriptive and processing model, Construction Grammar is grounded in a substantial body of research (e.g., Herbst, Schmid & Faulhaber, 2014; Hilpert, 2019; Hoffman, 2022; Stefanowitsch, 2011, as cited in Pakzadian, 2023). While originally developed to account for L1 acquisition, scholars have begun exploring its implications for L2 and foreign language learning (e.g., Boas, 2022; De Knop & Gilquin, 2016, as cited in Pakzadian, 2023). Within this theory, constructions are viewed as the fundamental units of linguistic knowledge. If constructions are central to L1 development, they are equally vital in L2 acquisition.

In the context of Spanish language acquisition, a study by Mellado Blanco (2020) used Construction Grammar to analyse the expression "(no) me importa un comino" and its regional variations. Using a corpus-based approach, she examined how this construction changes across Spanish-speaking areas while keeping its core meaning. She found that even highly fixed idiomatic expressions show flexible, systematic variation. For research on language learning, this supports the idea that learners need exposure to both the form and functional use of constructions, not just isolated grammar or vocabulary.

The Construction-Based Theory is highly relevant to the current study, as it emphasizes the interconnectedness of grammar and vocabulary in L2 acquisition. Given that the study focuses on assessing Spanish language proficiency, the Construction-Based Theory aligns with the approach of examining how grammatical and lexical knowledge are integrated in the learning process. Specifically, the test components of Profgram reflect the idea that language assessment is not about isolated skills but rather about acquiring form-meaning pairings that become increasingly internalized and utilized by learners. By investigating how these two elements (grammar and vocabulary) interact in the context of Spanish L2 acquisition, the study aligns with the constructionist view, which posits that language competence is rooted in the usage of constructions. From this theoretical standpoint, we predict that participants' vocabulary and grammar scores on the Profgram test will show strong correlations, as they would likely be drawing upon integrated constructions in their use of the language.

In contrast, the Usage-Based Approach (Lieven et al., 2003) also emphasizes the connection between grammar and vocabulary, but it focuses more directly on the role of input frequency and the patterns of language use in real-world contexts. According to this

approach, language acquisition is driven by the frequency and distribution of linguistic forms encountered in everyday communication. Language is seen as a product of cognitive processes that are shaped by social interaction and exposure to input. From this perspective, both grammar and vocabulary emerge from the frequency and patterns of use in real-world contexts. The theory suggests that learners develop mental representations of language based on the co-occurrence of grammatical structures and lexical items, making them mutually reinforcing. For instance, when a learner encounters a verb in various contexts with different arguments or complements, they do not only learn the word itself but also its grammatical properties.

Research has supported these ideas, with studies showing that learners tend to acquire vocabulary and grammar simultaneously through exposure to language use in context, particularly in communicative settings (Ellis, 2002). Ellis (2002) explored the significant role of frequency and context in language acquisition, highlighting that frequent exposure to language forms, whether words, grammatical structures, or collocations, leads to their effective internalization. He highlights that grammar and vocabulary are not learned separately but as interconnected elements, with learners acquiring language patterns through their interactions. Ellis emphasizes that this simultaneous acquisition enhances fluency, as learners build automatic connections between lexical items and grammatical structures. Thus, exposure to language use in context allows learners to integrate vocabulary and grammar as part of a unified system, reinforcing each other in real-world communication.

In essence, while both models agree that grammar and vocabulary are learned together, the Construction-Based theory focuses on the storage and use of stable constructions, whereas the Usage-Based Approach emphasizes the emergent nature of language based on usage patterns. The literature has largely supported the hypotheses of both theories. Studies have consistently shown that vocabulary and grammar are not independent of each other in the acquisition process (Goodman, 1997). Therefore, these theories not only support integrating grammar and vocabulary in language learning but also suggest that acquiring them together is a natural and potentially essential aspect of L2 proficiency.

While grammar and vocabulary are interrelated in language acquisition, evaluating one component alone may not fully capture a learner's proficiency. While it is true that in some cases, measuring one aspect (e.g., vocabulary) might offer some insights into a learner's overall proficiency, this does not provide a complete picture. Grammar and vocabulary are both essential components of linguistic competence, and each plays a different role in communication. Testing both grammar and vocabulary together in an integrated

proficiency test ensures a more comprehensive assessment, capturing both the structural and lexical aspects of language. This approach aligns with theories of grammar-vocabulary interaction, accounting for the complex ways in which learners use and combine these components in real-world communication. An integrated assessment better reflects the nuances of language proficiency, allowing for a fuller understanding of a learner's abilities.

However, ongoing debates continue in the field regarding the relative importance of grammar and vocabulary in L2 proficiency evaluations and testing. While some researchers advocate for a balanced approach that integrates grammar and vocabulary instruction (Richards, 2002), others emphasize the primacy of communicative competence and argue for a greater focus on vocabulary acquisition in language education (Laufer, 1997). However, other researchers underscore the importance of grammar, arguing that it continues to play a crucial role in the development of learners' language systems (Ly, 2020). Grammar provides the structural foundation for organizing thoughts and forming clear, accurate sentences. Without strong grammar knowledge, learners may struggle with sentence structure and meaning, limiting their ability to communicate effectively, especially in more complex contexts. This ongoing dialogue underscores the complexity of language learning and highlights the challenges in L2 proficiency assessment. The debate over the roles of grammar and vocabulary in language acquisition directly impacts how proficiency is measured, as it raises questions about which components should be prioritized in assessments.

The use of Spanish proficiency tests

Studies have shown a growing interest in creating assessments that accurately measure learners' proficiency according to the CEFR guidelines (Smyk et al., 2013; Izura et al., 2014). These assessments utilize multiple-choice questions, cloze tests, speaking tests, and writing prompts to evaluate language abilities across different CEFR levels. Multiple-choice questions assess knowledge by asking participants to select the correct answer from a set of options. Cloze tests, where participants fill in missing words in a passage, evaluate context understanding and grammar application. Speaking tests assess fluency, pronunciation, and grammar through spoken responses, while writing prompts test the ability to produce coherent, grammatically correct texts. Each format provides unique insights into different aspects of language proficiency.

However, challenges persist in the development and implementation of such assessments. Proficiency assessments are often time-consuming and resource intensive.

Developing and administering these tests requires significant financial investment, personnel, and logistical effort, particularly for comprehensive assessments that cover multiple skills.

In the field of Spanish language proficiency assessment, various tests are widely recognized and used, including SIELE, DELE, CELU, and LexTALE. In the context of Spanish proficiency within education, the Diploma of Spanish as a Foreign Language, or DELE (In Spanish: *Diploma de Español como Lengua Extranjera*) is the most widely recognized and prestigious certification. DELE is an official exam administered by the Instituto Cervantes on behalf of the Spanish Ministry of Education, Culture, and Sport, providing permanent proof of Spanish proficiency once obtained. It evaluates the language proficiency of non-native Spanish speakers across the levels of the CEFR, from A1 to C2 (Instituto Cervantes, 2024).

The DELE exam assesses four key language skills: reading comprehension, listening comprehension, written expression, and oral expression (Instituto Cervantes, 2024). The reading and listening section consist of multiple-choice and cloze-type questions, testing the ability to understand written and spoken Spanish in various contexts. The writing section requires candidates to produce texts based on prompts, such as composing an essay or responding to a formal letter, which evaluates grammar, vocabulary, and the ability to organize ideas clearly. The speaking test involves a one-on-one interview with an examiner, where candidates must demonstrate their ability to engage in a conversation, respond to questions, and express their thoughts coherently.

This comprehensive structure makes DELE a widely recognized certification that assesses both receptive and productive Spanish skills. For academic institutions, it provides a standardized measure for admissions and placements. However, its fixed format may not always reflect the diverse learning styles or specific needs of students, especially those who excel in informal communication but struggle with test settings. For employers, DELE offers a reliable proficiency indicator, though it may emphasize formal language over the conversational skills needed in some jobs. Lastly, for learners, DELE can be a motivating goal, but its preparation can be stressful and may not align with individual needs. Additionally, the costs and focus on formal Spanish may limit its accessibility and relevance for some groups.

The *Servicio Internacional de Evaluación de la Lengua Española* (SIELE) is another important Spanish proficiency exam, recognized globally and administered by the Instituto Cervantes. Unlike DELE, which provides a separate certificate for each specific CEFR level (A1 to C2), SIELE does not assign a certificate to a particular level. Instead of certifying

proficiency at one specific level, SIELE evaluates language skills across multiple levels (A1 to C1) and provides a single overall score. This score reflects the individual's proficiency across the entire range of levels tested, without specifying a particular level of achievement, unlike DELE which clearly identifies the level attained. Additionally, while DELE certifications are valid for life, SIELE certificates are only valid for five years (Instituto Cervantes, 2024). Therefore, SIELE is considered a more time-sensitive measure of proficiency, reflecting current language skills, whereas DELE certifications remain valid indefinitely, serving as a permanent record of one's language proficiency at the time of testing. SIELE also tests four key areas: reading comprehension, listening comprehension, written expression, and oral expression. The entire exam is conducted online and includes multiple-choice questions for reading and listening, as well as tasks requiring written and spoken responses. In the writing section, candidates produce text based on prompts, while the speaking test is also conducted online, assessing the ability to engage in conversation and express ideas clearly.

While both the SIELE and the DELE are aligned with the CEFR framework and endorsed by the Instituto Cervantes, they differ in format, administration, and delivery. SIELE is a digital exam that can be taken year-round at authorized testing centres, offering greater scheduling flexibility. In contrast, DELE is paper-based and offered only on specific dates throughout the year. SIELE typically delivers results within three to five days, whereas DELE results may take several weeks. Additionally, SIELE provides a breakdown of performance across the different skills tested, which can be useful for learners seeking targeted feedback, while DELE follows a pass/fail format with less detailed post-test reporting. The choice between the two often depends on practical considerations, such as scheduling needs, preferred test format, or the specific requirements of institutions or employers.

The CELU (In Spanish: *Certificado de Español: Lengua y Uso*) is an important Spanish language proficiency test, primarily used in Argentina and other Spanish-speaking countries in Latin America. It is recognized by the Argentine Ministry of Education and is aimed at assessing the language proficiency of non-native Spanish speakers (Schlatter, 2008). CELU evaluates proficiency across four key skills: reading comprehension, listening comprehension, written expression, and oral expression, aligning with the CEFR levels. The test is recognized by various institutions, universities, and employers, particularly in Argentina, as a valid measure of Spanish language competence (Celu, N.D.). CELU provides a certification that can be used for academic, professional, and personal purposes, with results

typically valid for an indefinite period. It is a key test for individuals seeking to work or study in Argentina or other Spanish-speaking regions in Latin America.

However, in research, LexTALE is frequently used as a tool for assessing vocabulary proficiency in L2 learners, primarily due to its efficiency and ease of administration (Lemhöfer & Broersma, 2012). Unlike comprehensive exams like DELE and SIELE, which assess multiple language skills across various CEFR levels, LexTALE focuses specifically on vocabulary knowledge. LexTALE is easier to administer as it is an online, self-paced test, requiring no in-person components or complex scheduling. It is a quick and reliable test and often employed in studies on L2 acquisition (e.g., Silva et al., 2024) and has been validated by correlating its scores with other well-established tests, such as the Quick Placement Test (Lemhöfer & Broersma, 2012).

The LexTALE test consists of a word recognition task, where participants are presented with a list of words, some of which are real words in the target language, and others are non-words. Participants must identify which words are genuine, providing a measure of their vocabulary size. This simple format allows researchers to quickly gauge vocabulary proficiency without the extensive time commitment required by more comprehensive tests. As such, LexTALE has become a popular choice in research settings where assessing vocabulary knowledge is critical, and time constraints are a concern (Izura et al., 2014).

This study will focus on DELE and LexTALE. DELE is chosen for its prestigious status as the most official test of Spanish proficiency, recognized globally by institutions such as the Instituto Cervantes. LexTALE, on the other hand, is selected for its frequent use in research, especially in evaluating vocabulary knowledge in L2 learners.

While assessments like DELE and LexTALE are widely used, they each present limitations in certain contexts. DELE, though a highly respected measure of Spanish proficiency, is time-consuming and requires significant preparation, making it less practical for situations where quick results are needed. On the other hand, LexTALE, while efficient and useful for measuring vocabulary, is limited in its scope, as it only assesses lexical knowledge and does not provide a comprehensive evaluation of other language skills. This narrow focus leaves a gap in measuring higher levels of proficiency, where complex linguistic abilities must be assessed across various dimensions. Therefore, while both tools serve important functions, there is a need for a more comprehensive approach that can measure proficiency in a broader context and address the challenges of accurately evaluating advanced learners.

This study addresses the gap in existing proficiency assessments by creating a Spanish proficiency test, Profgram, that is both easy and fast to administer and covers both grammar and vocabulary. Relying solely on one aspect could give an incomplete or misleading picture of overall language ability. For example, a learner may have a large vocabulary but struggle with grammar, hindering their ability to construct grammatically correct sentences. Alternatively, a learner with strong grammar may struggle to express themselves fluently if their vocabulary is limited. Recognizing the importance of evaluating both aspects, the Profgram test aims to provide a more balanced assessment of language proficiency. By integrating both elements into a single, efficient tool, this test may offer a more accurate reflection of an individual's language abilities, catering to the need for quick but thorough proficiency evaluations.

The present study

This thesis has two goals. First, it aims to introduce a new more comprehensive L2 Spanish proficiency test, the Profgram Spanish test, and examine its validity. Second, it aims to investigate how grammar and vocabulary are associated to L2 proficiency.

Thus, our first research question (RQ1) is to what extent is the Profgram Spanish test valid in accurately assessing individuals' CEFR levels within a short timeframe? We hypothesize that Profgram will demonstrate strong validity in measuring Spanish proficiency compared to standardized and research-specific tests (H1). This hypothesis is grounded in the expectation that Profgram will align well with established proficiency measures like DELE and LexTALE. These tests were selected for their strong track record as reliable, well-established measures of language proficiency, making them ideal benchmarks for assessing the validity of Profgram. Furthermore, the selection of these tests is grounded in their accessibility and the availability of test content, which is essential for effectively administering and evaluating participants. Both tests were chosen because their materials are publicly available online, which facilitates their use in this research context. Utilizing publicly available test content also enhances the transparency and replicability of the study, as other researchers can verify the materials and procedures employed.

Proficiency tests that assess similar linguistic constructs, such as vocabulary and grammar, are expected to show strong correlations, as they measure overlapping skills essential to language competence. In the case of Profgram, its assessment of both grammar and vocabulary is grounded in the CEFR-level guidelines of the Instituto Cervantes, which are widely accepted for defining and measuring language proficiency in Spanish. Since these

guidelines provide a well-established framework for evaluating language competence, it is reasonable to expect that Profgram will demonstrate a positive correlation with other recognized proficiency measures, such as DELE and LexTALE.

The second research question (RQ2) investigates how specific grammar proficiency and vocabulary proficiency are associated to L2 proficiency evaluation and performance in testing contexts? The second hypothesis (H2) posits that grammar proficiency will have a stronger association with L2 test performance than vocabulary proficiency. While grammar and vocabulary are interconnected and jointly contribute to communicative competence, this hypothesis is based on the assumption that grammar provides the structural framework necessary for the effective and contextually appropriate use of vocabulary (Ly, 2020). Given that tests like LexTALE focus solely on vocabulary, demonstrating a stronger predictive role for grammar would highlight the added value of including grammar in L2 assessments and support the need for integrated measures, like Profgram, that assess both components for a more comprehensive evaluation of L2 proficiency.

To explore these research questions, participants completed a questionnaire made up of three tests: Profgram, a shortened and adapted version of the DELE, and LexTALE-Esp. Their results were used to examine the relationship between Profgram test scores, LexTALE-Esp test scores, DELE test scores, and the relative contributions of vocabulary and grammar proficiency to the Profgram, LexTALE-Esp and DELE test performance.

Method

Participants

To recruit participants for the study, several strategies were used, like spreading the test online. However, the most effective method involved directly engaging individuals in person. Specifically, the questionnaire was set up on a laptop, and potential participants were approached and invited to complete the test. This approach provided immediate access to the questionnaire, allowing participants to take part easily and efficiently.

The participants in this study were 40 individuals, ranging in age from 18 to 54 years (M = 27.25, SD = 7.71). 42 participants took part in the experiment but two of them were excluded from the analyses because they did not finish all the tests. The sample was predominantly female (67.5%, n = 27), with 32.5% male participants (n = 13). In terms of education, 72.5% of participants was either in or had completed University (n = 29), 17.5% was in or had completed University of Applied Sciences (n = 7), 7.5% was in or had

completed vocational education (*n* = 3), and 2.5% was in or had completed secondary education (*n* = 1). Regarding language background, 95% of participants were native Dutch speakers (*n* = 38), 2.5% were native English speakers (*n* = 1), and 2.5% were bilingual Dutch-German speakers (*n* = 1). The participants all spoke languages from the Germanic language family: Dutch, English and German, which share significant linguistic similarities in vocabulary, grammar, and sentence structure, and their common Germanic roots likely minimized potential L1-related differences in performance, allowing the study to focus more on their Spanish language proficiency rather than L1 interference. No participants reported any hearing impairments. In terms of vision, 75% of participants had no vision impairments (*n* = 30), while 25% had vision impairments, but these were corrected with glasses (*n* = 10).

In this study, participants self-rated their proficiency across four language skills: listening, speaking, reading, and writing, using the Common European Framework of Reference for Languages (CEFR) levels. The levels included were A0 (No Spanish), A1 (Very basic Spanish), A2 (Basic Spanish), B1 (Intermediate Spanish), B2 (Intermediate-advanced Spanish), C1 (Advanced Spanish), and C2 (Near-native/native Spanish). There were no participants with a self-rated proficiency higher than B2. The distribution of self-ratings for each skill is shown in Table 1. The self-rated proficiency levels indicate that participants feel most confident in listening and reading skills, particularly at the lower levels (A0 and A1). However, as proficiency increases, both speaking and writing skills show a significant decline, especially at higher levels (B1 and B2). This suggests that while participants are comfortable with passive skills like listening and reading, they face more challenges in active language production. The self-assessed proficiency of the participants was not part of the analysis in this thesis because it was not relevant to the main research questions.

Table 1

Distribution of Self-Ratings for Language Skills According to CEFR Levels

Skills	A0	A1	A2	B1	B2
Listening	25%	22.5%	27.5%	17.5%	7.5%
	(*n* = 10)	(*n* = 9)	(*n* = 11)	(*n* = 7)	(*n* = 3)
Speaking	40%	20%	20%	17.5%	2.5%
	(*n* = 16)	(*n* = 8)	(*n* = 8)	(*n* = 7)	(*n* = 1)

Reading	17.5% (n = 7)	27.5% (n = 11)	30% (n = 12)	15% (n = 6)	10% (n = 4)
Writing	40% (n = 16)	20% (n = 8)	22.5% (n = 9)	15% (n = 6)	2.5% (n = 1)

Participants were recruited via word-of-mouth to do the online survey, which was accessible to individuals aged 18 and above. Participation was on a voluntarily basis and they did not receive any compensation.

Materials

Profgram

For the Profgram test, multiple-choice questions were created for each CEFR level (179 for A1, 144 for A2, 118 for B1, 93 for B2, and 91 for C1), covering the grammar and vocabulary topics outlined by the Instituto Cervantes (2006), as expected by the CEFR for Spanish. The initial selection of questions was made based on the *Socios* Spanish learning textbooks. Specifically, the A1 and A2 questions were derived from *Socios 1* by Corpas & Martínez (2007), the B1 questions were based on *Socios 2* by Corpas et al. (2007), and the B2 questions were taken from *Expertos* by Tano (2012). These books served as key sources for the item selection.

The items were then categorized based on whether they focused on grammar or vocabulary. Each item was assigned to a specific CEFR level and further labelled according to the type of grammar or vocabulary it assessed. Once the items were categorized, they were reviewed by the supervisors of this thesis to ensure they aligned with the appropriate proficiency levels and labels. To further enhance the accuracy of the categorization and ensure the quality of the items, two additional Spanish teachers were consulted, both of whom were native Spanish speakers. Following these reviews, several revisions were made, including correcting typographical errors, adjusting the labels assigned to items, and reassigning certain items to more appropriate proficiency levels. This thorough process of validation ensured that the final selection of items accurately represented the intended CEFR levels.

All questions were fill-in-the-gap questions, and had four answer options, with only one being correct. In Table 2, examples of both grammar and vocabulary questions of each level are shown. The full set of questions used in this study is available in Appendix 1.

Participants were presented with a total of 50 questions, which were pseudo-randomly selected from the full dataset of questions. The questions were randomly selected from the different proficiency levels labels, and from both grammar and vocabulary categories. These questions were grouped into blocks of 10, which each block containing 5 grammar questions and 5 vocabulary questions. The order of the questions within each block was also randomized, ensuring that both the types of questions and their sequence varied for each participant. The test was designed to be adaptive: in the first block all questions were selected from the A1 category and, depending on the performance of the participant, the subsequent blocks could be made of questions increasing difficulty or from the same block. If a participant answered more than 6 questions correctly, the questions for the following block were selected from the next level of proficiency (e.g., from A1 to A2). If they answered fewer than seven questions correctly, the questions in the next block were selected from the same proficiency level. This adaptive mechanism went on for every block until participants completed five blocks of questions, i.e., 50 questions in total. The highest achievable level was C1 (if participants scored correctly to at least 7 questions in all 5 blocks and, therefore, moved up to a block with C1 questions) and the lowest level was A1 (if they failed to answer at least 7 questions correctly in all blocks and therefore only saw A1-level questions).

Table 2

Multiple-Choice CEFR Questions of Profgram Test

CEFR level	Grammar	Vocabulary
A1	... María y Sofía. **a. Estas son** b. Esta son c. Esto es d. Estos son	Mi abuela habla inglés porque es ... **a. americana** b. francesa c. alemana d. holandesa
A2	Mi casa es ... moderna ... la de Elena. **a. más/que** b. mejor/que c. un poco/que d. la más/que	No quiero otro café, gracias. ... me ha tomado dos. **a. Ya** b. Todavía c. Nunca d. Siempre

B1	Yo, en tu lugar, ... en coche. **a. iría** b. iré c. irás d. irías	Le ... que nos envíen sus condiciones de compra. **a. agradecemos** b. rogamos c. complacemos d. esparamos
B2	En estos momentos, no me parece que ... urgente sacar otra foto. **a. sea** b. es c. era d. son	... sea un excelente negocio. **a. No estoy muy seguro de que** b. Personalmente c. Cree que d. Yo mismo
C1	... mucho tiempo esperando mi turno. **a. Llevo** b. Me llevo c. Estoy llevando d. Me estoy llevando	Creo que no está todo, ... los folletos. **a. faltan** b. falta c. sobran d. sobra

Note. Correct answers are in bold.

The test results for Profgram were calculated as follows: for each correct answer, participants received a weighted score based on the level of difficulty. Specifically, for level A1, each correct answer was multiplied by 0.1, for level A2 by 0.2, for level B1 by 0.3, for level B2 by 0.4, with level C1 answers receiving a multiplier of 0.5. This scoring system allowed participants to earn more points for more difficult questions and therefore created a more drastic distinction between the different levels. The maximum score was 15. By assigning higher multipliers to more advanced levels, the system ensures that participants cannot achieve the maximum score by only answering lower-level questions correctly. This method guarantees that participants must demonstrate proficiency at higher levels to reach the maximum score. For example, answering a correct question at C1 level, which requires advanced language skills, will contribute significantly more to the total score than answering a question at the A1 level. The weighted score allows to make a difference between two participants, as can be seen in Table 3. Although both participants achieved the same raw score of 29, their weighted scores differ substantially due to the difficulty levels of the questions they answered correctly.

Table 3

Example showing how weighted scores differentiate between participants

	Block 1 – Correct responses	Block 2 – Correct responses	Block 3 – Correct responses	Block 4 – Correct responses	Block 5 – Correct responses	Score	Weighted score
Participant A	6 (A1)	6 (A1)	6 (A1)	6 (A1)	5 (A1)	29	2.9
Participant B	7 (A1)	7 (A2)	7 (B1)	7 (B2)	1 (C1)	29	7.5

To investigate the relative role of grammar proficiency versus vocabulary proficiency in L2 proficiency assessment, two extra scores were calculated based on participants' answers to the Profgram questions: one score was based on the grammar questions and the other on the vocabulary questions. The scores were again weighted according to proficiency level of the question (following the same equation as with the general score previously described). Specifically, the number of correct grammar answers at the A1 level was multiplied by 0.1, at the A2 level by 0.2, at the B1 level by 0.3, at the B2 level by 0.4, and at the C1 level by 0.5. The same procedure was applied to the vocabulary section.

Profgram_Grammar = A1GrammarScore *0.1 + A2GrammarScore *0.2…

Profgram_Vocabulary = A1VocabularyScore *0.1 + A2VocabularyScore *0.2...

LexTALE

For this study, the LexTALE-Esp (Spanish version) test, developed by Izura et al. (2014), was used in Qualtrics. The test consisted of 90 words, including 60 real words and 30 nonwords. The words were extracted from the materials of Izura et al. (2014). For the complete test, please refer to Appendix 1. Participants were presented with these words in isolation (e.g., *antar*) and asked to identify whether the word was a real Spanish word (Yes) or a non-existent word (No). The score for each participant was calculated using the following formula provided by Izura et al. (2014):

LexTALE_Score = Nyes to words – 2 * Nyes to nonwords.

This formula calculates a score taking into account the number of correct answers and subtracting from this number the number of incorrect responses multiplied by two, to penalize errors on nonwords. It accurately penalizes for guessing behaviour, as a participant who responds randomly (i.e., saying yes to half of the words and half of the nonwords) is expected to have a score around 0. The maximum score of this test is 60.

DELE

A shortened and adjusted version of the DELE exam was administered, consisting of 50 multiple-choice questions: 25 for reading comprehension and 25 for listening comprehension. Unlike the standard DELE, which is typically designed for a specific proficiency level, this version was adapted to encompass a broader range of levels, from A1 to C1. A task from each level (A1 to C1) was included in both the listening and reading sections of the adapted DELE version. Each task contained 5 questions, ensuring that a range of language proficiency levels was assessed across both components. These tasks included in the official exams, representing the different levels of DELE, are available on the Instituto Cervantes website (Instituto Cervantes, 2025).

Due to time constraints, no speaking or writing exercise were included. The participants started with the reading comprehension section, answering five questions at each CEFR level (A1 to C1). Then, they proceeded to the listening comprehension section, again answering five questions at each CEFR level. Each question had three answer options, with only one correct answer. The results were scored on a scale of 1 to 50, based on the number of correct answers. A chance correction was not applied, as the DELE neither penalizes errors nor uses chance correction (Instituto Cervantes, 2014).

DELE_Score = Number of correct answers.

The full set of questions used in this study is available in the appendix *Survey questions for Profgram study* and were obtained from the Instituto Cervantes (2025).

Procedure

The experiment took place in different digital platforms: Wix (a site-building tool), where the informed consent, demographic questions and Profgram were presented, and Qualtrics, including the DELE and LexTALE-Esp. Participants were given a link that took them to the Wix website. Participants were encouraged to take the test on a computer for

optimal usability. The survey was anonymous, and all responses were submitted confidentially.

Upon accessing the survey, participants were first presented with a consent form and information about the study. After consenting, they completed demographic questions, including age, gender, education level, and self-rated Spanish proficiency (ranging from A1 to C2). Following this, participants completed the Profgram test. Once they were done with the Profgram, the site re-directed them with a link to the Qualtrics. Participants were assigned a number when entering the Wix and this number was carried to the Qualtrics survey as an attachment to the link. This number made it possible to match the data from the Qualtrics questionnaire (DELE and LexTALE tests) with the data from the Profgram test on Wix, ensuring that all participant data could be accurately merged and analysed together. For some participants, the number assignment did not work properly, and they were assigned the same number. This issue was solved by comparing the timestamps of the Qualtrics output and the Wix output.

Participants were able to complete the survey at their convenience. The participants took an average of 31.68 minutes (*SD* = 7.17) to complete both the LexTALE and DELE tests.

Ethics

This research was conducted following the ethical guidelines of Radboud University. All participants provided informed consent to take part in the study, and their data was treated confidentially. A data management plan was developed, and this research study has been approved by the Ethics Assessment Committee Humanities of Radboud University (EACH file number 2024-2683).

Data analysis

Participants' Profgram scores (general, vocabulary and grammar scores), as well as their LexTALE and DELE scores, were analysed. Data processing involved several steps to ensure the accuracy and integrity of the analysis. The initial phase of the study involved a pilot with 6 participants, who were later determined to be valid for inclusion in the final sample. During the study, the data collected from these 6 pilot participants, which was originally stored in separate raw data files from two different platforms (Qualtrics for the DELE and LexTALE tests, and Wix for the Profgram test), was merged with the final sample. Specifically, the data from the 6 participants was first combined into a single file,

integrating the responses from both the DELE and LexTALE tests (on Qualtrics) and the Profgram test (on Wix). Once this pilot data was successfully integrated, it was then merged with the processed data file from the remaining 34 participants, resulting in a complete dataset comprising a total of 40 participants. This process ensured that all data from the pilot and the subsequent participants were consistently formatted and could be analysed together, maintaining the integrity and continuity of the study's data collection. Two participants who did not complete all tests were excluded from the final analysis, while the remaining data were included.

To answer RQ1, evaluating the validity of the Profgram test in measuring the CEFR-level Spanish proficiency, Pearson correlation analyses were conducted between the Profgram general scores and the DELE scores, and the Profgram general scores and the LexTALE scores. The correlations help to assess the degree to which Profgram aligns with these established tests, thereby providing insights into its validity.

Initially, a regression analysis was planned; however, due to high correlations between the variables, the assumption of multicollinearity for regression was violated, making this analysis not possible.

To answer RQ2, investigating how vocabulary and grammar are associated to general proficiency, additional correlation analyses were performed. These analyses explored the association between the DELE score and the Profgram grammar score and the DELE score and the Profgram vocabulary score. We did not include self-reported proficiency scores in these analyses, as the aim was to compare the reliability of our test to that of established, standardized proficiency tests.

It is important to note that the analyses in this study focus solely on external validity, specifically how well Profgram scores align with existing standardized tests. Internal validity, which refers to how consistently the test measures language proficiency within itself, was not assessed.

Results

Descriptive statistics

First, we present descriptive statistics for the sample data, followed by a brief interpretation of each test's results.

The Profgram test had a mean score of 5.57 (*SD* = 3.61), with scores ranging from 1.2 to 12.2, out of a maximum score of 15. When broken down into components, the grammar

section had a mean score of 2.71 (*SD* = 1.86), with scores ranging from 0.5 to 6.3, while the vocabulary section had a mean score of 2.86 (*SD* = 1.81), with scores ranging from 0.7 to 6.4. These results suggest that the test was moderately challenging for participants, as no one achieved the maximum score. Additionally, the similar mean scores and standard deviations for the grammar and vocabulary sections indicate that both components were of comparable difficulty.

The LexTALE-Esp test showed a mean score of 10.70 (*SD* = 16.70), with scores ranging from -9 to 45 out of a possible 60. Since LexTALE penalizes participants for incorrectly identifying nonwords as words, it is possible for participants to receive a negative score. The relatively low mean score and high deviation suggests that participants found this test more challenging, and that there was considerable variability in vocabulary knowledge among them.

The DELE test had a mean score of 26.35 (*SD* = 9.03), with scores ranging from 10 to 42, out of a maximum of 50. This suggests that participants performed at an intermediate level on average and the test was neither too easy nor too difficult.

A summary of the correlation results is presented in Table 4. The table includes correlation coefficients (*r*) for each pair of tests: LexTALE-DELE, DELE-Profgram, and LexTALE-Profgram. Also, the table includes the correlation results between Profgram grammar- and vocabulary scores, and the LexTALE and DELE scores.

Table 4

Correlations (r) between Profgram Grammar, Vocabulary, and total scores, and External Test Scores (DELE, LexTALE)

Variable	Profgram Grammar	Profgram Vocabulary	Profgram	LexTALE	DELE
Profgram Grammar	1	.930*	.983*	.474**	.602*
Profgram Vocabulary	.930*	1	.982*	.473**	.587*
Profgram	.983*	.982*	1	.482**	.605*
LexTALE	.474**	.473**	.482**	1	.377**
DELE	.602*	.587*	.605*	.377**	1

Note. * = $p < .001$, ** = $p < .05$

The correlation between LexTALE and DELE was checked to ensure the validity of LexTALE. This correlation was moderate, $r(38) = .38$, $p = .017$, indicating a statistically significant but not exceptionally strong relationship. Therefore, LexTALE demonstrates partial validity as a measure of language proficiency.

In contrast, the correlations between the individual Profgram component scores and the overall Profgram score were extremely strong. The correlation between Profgram grammar scores and the overall Profgram score was $r(38) = .98$, $p < .001$, and the correlation between Profgram vocabulary scores and the overall Profgram score was also $r(38) = .98$, $p < .001$. This very high correlation is expected, as the overall Profgram score is the sum of the grammar and vocabulary scores.

Research Question 1: To what extent is the Profgram Spanish test valid in accurately assessing individuals' CEFR levels within a short timeframe?

The correlation between DELE and Profgram general scores was strong, $r(38) = .61$, $p < .001$, indicating a high association between the two tests. This suggests that Profgram is highly effective in measuring similar aspects of Spanish proficiency as the shortened and adjusted DELE focused on reading and listening skills.

A moderate to strong positive correlation was found between LexTALE and Profgram general scores, $r(38) = .48$, $p = .002$. This significant correlation suggests that Profgram shares some common ground with LexTALE.

As an exploratory analysis, we also compared the correlations between the DELE and LexTALE and the DELE and the Profgram general scores. This comparison will shed some light on which of the two research-designed tests (LexTALE or Profgram) better correlates to an adapted version of the standardized test DELE. To compare the strength of the correlations between DELE and Profgram and DELE and LexTALE, a statistical test using Fisher's r-to-z transformation was conducted. The correlation between DELE and Profgram ($r = .61$) was not significantly stronger than the correlation between DELE and LexTALE ($r = .38$), $z = 1.30$.

Research Question 2: What is the relative association of grammar proficiency versus vocabulary proficiency on second language proficiency evaluation and performance in testing contexts?

A Pearson correlation showed a strong association between Profgram grammar scores and DELE scores, $r(38) = .60$, $p < .001$, suggesting that participants' grammar proficiency as measured by Profgram was closely aligned with their DELE performance, a recognized standard for CEFR-level Spanish proficiency. Additionally, a strong positive correlation was found between Profgram vocabulary scores and DELE scores, $r(38) = .59$, $p < .001$, suggesting that vocabulary proficiency as measured by Profgram is strongly aligned with DELE performance. Although both correlations are strong and highly significant, the grammar component showed a slightly stronger association with DELE scores than the vocabulary component, suggesting that grammar proficiency may be a marginally better predictor of overall Spanish proficiency in this sample.

In addition, we studied the correlations between the LexTALE score, which is supposed to be vocabulary-based measure, and the Profgram vocabulary and grammar scores. The correlation between Profgram grammar scores and LexTALE scores was moderate, $r(38) = .47$, $p = .002$, indicating a significant positive relation between grammar proficiency and LexTALE. Furthermore, a moderate correlation was found between Profgram vocabulary scores and LexTALE scores, $r(38) = .47$, $p = .002$. Since both correlations are identical in strength, the results suggest that LexTALE is equally related to both vocabulary and grammar performance as measured by Profgram. This may imply that, despite its intended focus on vocabulary, LexTALE performance also reflects broader linguistic knowledge.

Fisher's r-to-z transformations were conducted to compare the correlation scores for grammar-LexTALE ($r = .47$) and vocabulary-LexTALE ($r = .47$), and for grammar-DELE ($r = .60$) and vocabulary-DELE ($r = .59$). Neither difference was statistically significant, indicating no substantial differences between the correlations.

Discussion

Present study

The present study sought to present a new test for Spanish proficiency testing, Profgram, and assess its validity in measuring CEFR-level Spanish proficiency (Research Question 1). In addition, this study also aims to explore the relative association of grammar and vocabulary knowledge on L2 proficiency evaluation and performance assessment in testing contexts (Research Question 2). Participants completed three online tests: Profgram, the new adaptative assessment of Spanish grammar and vocabulary; LexTALE-Esp, a

vocabulary proficiency test; and a shortened and adjusted DELE, a widely recognized measure of Spanish language proficiency.

The results showed a strong correlation between Profgram and DELE scores ($r = .61$, $p < .001$), indicating a high association between the two tests. A moderate to strong correlation was found between Profgram and LexTALE scores ($r = .48$, $p = .002$). Further analysis revealed strong correlations between Profgram grammar and vocabulary scores with DELE ($r = .60$ and $r = .59$, respectively). Fisher's r-to-z transformation showed no significant differences between the correlations involving DELE, Profgram and LexTALE, suggesting similar relationships for both grammar and vocabulary.

To what extent is the Profgram Spanish test valid in accurately assessing individuals' CEFR levels within a short timeframe?

Regarding Research Question 1, the hypothesis that Profgram would demonstrate strong validity in measuring Spanish proficiency (H1) was supported by the results. The results showed that Profgram had strong correlations with both DELE and LexTALE, supporting its validity as an effective tool for assessing Spanish proficiency.

Given the high correlation between Profgram and DELE scores ($r = .61$), it is clear that both tests assess related aspects of language proficiency, despite their different formats. Profgram focuses on grammar and vocabulary, while DELE emphasizes reading and listening comprehension. However, both tests ultimately evaluate similar underlying language skills, such as vocabulary and grammar, which may explain their strong correlation. This correlation highlights that proficiency in grammar and vocabulary can have a significant impact on reading and listening comprehension. However, the high correlation between Profgram and DELE scores might also stem from other factors beyond shared language constructs. One possibility is that both tests may draw on overlapping areas of language exposure or experience, such as academic or formal language settings, which could lead to similar performance outcomes. Furthermore, it is possible that individuals with stronger general language proficiency or better test-taking strategies might perform well on both tests, regardless of their different formats. Therefore, the strong correlation could reflect not just the shared components of language assessed by the tests, but also broader factors such as general language aptitude or familiarity with test structures.

The moderate correlation between Profgram and LexTALE ($r = .48$) provides valuable insights into how these two tests measure different dimensions of language proficiency. While the moderate correlation suggests a relationship between the two tests, it

also highlights that Profgram is not simply a vocabulary test, despite its inclusion of vocabulary components. If the correlation were very high, it would suggest that both tests are measuring almost the same thing (likely just vocabulary). However, because the correlation is moderate, it suggests that Profgram assesses a wider range of language proficiency, integrating not only vocabulary but also grammar. This distinction is further underscored by the different approaches used by LexTALE-Esp and Profgram to measure vocabulary proficiency. LexTALE-Esp measures vocabulary through a receptive test, assessing word recognition and lexical knowledge. In contrast, Profgram includes both grammar and vocabulary components. This approach allows Profgram to capture additional linguistic dimensions, such as sentence structure, syntactic accuracy, and grammatical rules, which are not emphasized in LexTALE-Esp. Therefore, the moderate correlation between Profgram and LexTALE could emphasize that Profgram is a more comprehensive tool.

However, despite the moderate to strong correlations with DELE and LexTALE, the results suggest that Profgram is not significantly more effective or reliable in measuring Spanish proficiency compared to LexTALE: the correlation between DELE and Profgram was not found to be significantly higher than the correlation between DELE and LexTALE. Therefore, it implies that the Profgram test is not necessarily better than LexTALE in terms of its correlation with DELE results. Nevertheless, Profgram has distinct advantages that could still make it a more appealing option in certain contexts. For instance, it is shorter in duration compared to other tests, making it a more time-efficient tool.

In conclusion, in terms of validity, Profgram shares similarities with other validated language tests, such as LexTALE, which has been shown to be a reliable tool for measuring language proficiency through its correlation with established tests like the Quick Placement Test (Lemhöfer & Broersma, 2012). Like LexTALE, Profgram's correlation with established proficiency tests, such as DELE, support its validity as a tool for assessing individuals' CEFR levels.

What is the relative association of grammar proficiency versus vocabulary proficiency on second language proficiency evaluation and performance in testing contexts?

In response to Research Question 2, the study examined the relative association of grammar proficiency versus vocabulary proficiency on second language proficiency evaluation and performance in testing contexts. The hypothesis that grammar proficiency will have a greater influence on performance in language proficiency tests than vocabulary proficiency (H2) was not supported. The correlation results indicated that both grammar and

vocabulary proficiency were positively related to scores on both the LexTALE and DELE tests. Specifically, grammar proficiency was moderately correlated with both LexTALE (r = .47) and DELE (r = .61). Vocabulary proficiency showed similar correlations with both tests (r = .47 for LexTALE and r = .59 for DELE). The equal moderate correlation (r = .47) between both Profgram grammar and Profgram vocabulary scores with LexTALE scores is surprising, as one would expect the vocabulary scores to correlate more strongly with LexTALE, given that it is a vocabulary test. However, this suggests that grammar proficiency may also influence performance on LexTALE, as recognizing and processing vocabulary often requires grammatical knowledge. These findings also point to the interconnectedness of grammar and vocabulary in language proficiency, indicating that proficiency in one area may support proficiency in the other.

The results of the second research question support the Construction-Based Theory (Goldberg, 2003) and the Usage-Based Approach (Lieven et al., 2003), which emphasize the interdependence of grammar and vocabulary in language acquisition. The results confirm the idea that grammar and vocabulary are learned concurrently through exposure, as both theories suggest.

Limitations

Despite the valuable insights provided by this study, there are several limitations to consider. First, while the sample size of 40 participants is adequate for preliminary analysis, it is relatively small. A sample size of at least 70 participants would be more suitable for enhancing the reliability and generalizability of the findings. A larger sample would account for more variability in language proficiency and learning backgrounds, improving statistical power and confidence in the results. This would also allow for subgroup analyses, such as examining differences between beginner, intermediate, and advanced learners. Ultimately, a larger sample size would make the findings more robust and applicable to a broader range of language learners.

Second, the participants were predominantly students, which may limit the diversity of language learning backgrounds and contexts (Deygers & Vanbuel, 2022; Vanbuel & Deygers, 2024). Future research could explore the applicability of the Profgram test in different populations, such as adults in professional settings or language learners with varying learning strategies. Furthermore, while this study provides a comparison between Profgram, LexTALE, and DELE, it is limited to these three tests, and additional comparisons with other

widely used language proficiency tests, such as the SIELE or CELU, could further substantiate the validity of Profgram.

Another important limitation lies in the differences in scoring methodologies across the tests, which may influence performance comparisons. Specifically, LexTALE penalizes errors, and Profgram employs chance correction, whereas the DELE does neither. Although we adhered to each test's official scoring guidelines, this inconsistency in correction methods constitutes a limitation. Furthermore, the study did not account for the potential influence of factors such as motivation, learning environment, and test-taking strategies, which may also play a role in participants' performance across the tests (Setiyadi et al., 2016; Taguchi, 2008; Bumbálková, 2021).

Finally, the study focused solely on external validity. While this offers valuable insights into the test's alignment with recognized proficiency measures, internal validity was not examined. Future research could therefore incorporate methods such as factor analysis to evaluate internal validity and provide a more comprehensive assessment of the test's effectiveness.

Implications

The Profgram test has significant implications for both research and education. For researchers, it offers a quick and reliable tool for assessing participants' Spanish proficiency levels, which is essential for studies requiring a clear understanding of language competence. Unlike more formal tests like DELE and SIELE, which require extensive testing procedures, Profgram delivers an efficient and immediate evaluation through just 50 multiple-choice questions, providing a fast estimate of proficiency levels in both grammar and vocabulary. Its adaptive design, which adjusts question difficulty based on participant performance, ensures a more individualized and precise assessment.

For educators, Profgram serves as a practical tool for determining the initial language level of students, allowing for informed decisions about instructional strategies. Its speed and adaptability make it especially useful in settings where time is limited or when a quick proficiency assessment is needed. While tests like LexTALE focus solely on vocabulary, Profgram's inclusion of both grammar and vocabulary makes it a more comprehensive option for initial assessments.

Moreover, the very high correlation between grammar and vocabulary components and the overall score suggests a high degree of interconnectedness between these two domains. This raises the possibility that, depending on the specific goals of a study or

educational context, assessing only one component, either grammar or vocabulary, might be sufficient. This proposition could be explored in future research to determine whether one component alone can reliably estimate overall proficiency, especially in time-constrained or domain-specific scenarios.

As a result, Profgram can complement existing proficiency tests, providing a rapid and targeted measure of language competence in a more accessible format. Further research can explore its use across diverse educational contexts and populations, enhancing its applicability and supporting its potential as a valuable resource for both research and practical teaching.

Conclusion

In conclusion, this study found that the Profgram test is a valid and effective tool for assessing CEFR-level Spanish proficiency, with strong correlations to the DELE test and moderate correlations to the LexTALE vocabulary test. Both grammar and vocabulary proficiency play key roles in language performance, highlighting their interdependence. These results emphasize the interconnectedness of grammar and vocabulary in language learning and suggest that Profgram can serve as a reliable measure of Spanish proficiency. Further research is needed to explore additional factors influencing language performance in testing contexts.

References

Bumbálková, E. (2021). Test-taking strategies in second language receptive skills tests: A literature review. *International Journal of Instruction*, *14*(2), 647-664. https://doi.org/10.29333/iji.2021.14236a

Celu. (n.d.). Validity. Centro de Exámenes de Lengua y Uso. Retrieved February 10, 2025, from https://www.celu.edu.ar/en/content/validity

Cook, S. V., Sweet, S. J., Lancaster, A., Pandza, N. B., Jackson, S. R., Pelzl, E., Gor, K., & Doughty, C. J. (2018). Linguistic correlates of proficiency (LCP): At the intersection of testing and teaching. In J. McE. Davis, J. M. Norris, M. E. Malone, T. H. McKay, & Y.-A. Son (Eds.), *Useful assessment and evaluation in language education* (pp. 95-114). Georgetown University Press. https://doi.org/10.2307/j.ctvvngrq.9

Corpas, J., & Martínez, L. (2007). *Socios 1: Curso de español orientado al mundo del trabajo. Cuaderno de ejercicios. A1 A2*. Difusión.

Corpas, J., Martínez, L., & Sabater, M. L. (2007). *Socios 2: Curso de español orientado al mundo del trabajo. Cuaderno de ejercicios. B1.* Difusión.

Council of Europe. (2024). *The CEFR levels - Common European Framework of Reference for Languages (CEFR)*. https://www.coe.int/en/web/common-european-framework-reference-languages/cefr-descriptors

De Jong, N. H. (2023). Assessing second language speaking proficiency. *Annual Review of Linguistics*, *9*(1), 541-560. https://doi.org/10.1146/annurev-linguistics-030521-052114

Deygers, B., & Vanbuel, M. (2022). Gauging the impact of literacy and educational background on receptive vocabulary test scores. *Language Testing*, *39*(2), 191-211. https://doi.org/10.1177/02655322211049097

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143-188. https://doi.org/10.1017/S0272263102002024

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in cognitive sciences*, *7*(5), 219-224. https://doi.org/10.1016/S1364-6613(03)00080-9

Goodman, E. B. J. C. (1997). On the Inseparability of Grammar and the Lexicon: Evidence from Acquisition, Aphasia and Real-time Processing. *Language and Cognitive Processes*, *12*(5–6), 507-584. https://doi.org/10.1080/016909697386628

Instituto Cervantes. (2006). *Plan curricular del Instituto Cervantes: Niveles de referencia para el español*. Instituto Cervantes; Biblioteca Nueva.

Instituto Cervantes. (2014). *Guía del examen DELE B1*. https://examenes.cervantes.es/sites/default/files/guia_examen_dele_b1.pdf

Instituto Cervantes. (2024). Informatie SIELE Spaans. https://utrecht.cervantes.es/nl/certificaat_spaans_siele/info_siele_spaans.htm

Instituto Cervantes. (2024). Qué son los DELE. https://examenes.cervantes.es/es/dele/que-es

Instituto Cervantes. (2025). Exámenes DELE A1. https://examenes.cervantes.es/es/dele/examenes/a1

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-ESP: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicologica*, *35*(1), 49-66.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. *Second language vocabulary acquisition: A rationale for pedagogy*, *1*, 20-34. https://doi.org/10.1017/CBO9781139524643.004

Leclercq, P. & Edmonds, A. (2014). How to assess L2 proficiency? An overview of proficiency assessment research. In P. Leclercq, A. Edmonds & H. Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 3-23). Multilingual Matters. https://doi.org/10.21832/9781783092291-004

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behaviour Research Methods*, *44*(2), 325-343. https://doi.org/10.3758/s13428-011-0146-0

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of child language*, *30*(2), 333-370. https://doi.org/10.1017/S0305000903005592

Ly, C. K. (2020). The importance of grammar in language teaching and learning. In *Workshop on innovation in language teaching and learning* (pp. 186-195). Zenodo. https://doi.org/10.5281/zenodo.3947215

Mellado Blanco, C. (2020). (No) me importa un comino y sus variants diatópicas: Estudio de corpus desde la Gramática de Construcciones. In P. Huerta & A. Cuadrado Rey (Eds.), ELUA: Fraseología y variaciones (socio)lingüísticas y diatópicas (Anexo VII, pp. 89-111). Universidad de Alicante. https://doi.org/10.14198/ELUA2020.ANEXO7.06

Norris, J. M., & Ortega, L. (2012). Assessing learner knowledge. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 573–589). Routledge.

North, B. (2020). The CEFR renewed: Inspiring the future of Language education. *Italiano LinguaDue*, *12*(1), 549-560. https://doi.org/10.13130/2037-3597/13945

Olson, D. J. (2024). A systematic review of proficiency assessment methods in bilingualism research. *International Journal of Bilingualism, 28*(2), 163-187. https://doi.org/10.1177/13670069231153720

Pakzadian, M. (2023). Pedagogical construction grammar: The case of collocations and collostructions in foreign language instruction. *AILA Review*, *36*(2), 135-162. https://doi.org/10.1075/aila.22017.pak

Pont-Grau, A., Lei, Y., Lim, J. Z. E., & Xia, X. (2023). The effect of language training on immigrants' integration: Does the duration of training matter? *Journal of Economic Behavior & Organization*, *212*, 160-198. https://doi.org/10.1016/j.jebo.2023.05.027

Richards, J. C. (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667190

Schlatter, M., Scaramucci, M., Prati, S., & Acuña, L. (2008). *Celpe-Bras and CELU proficiency exams: Implementation, characteristics and impacts of the exams on Brazil-Argentina relations.* Paper presented at the Third International ALTE Congress, Cambridge, Inglaterra. http://www.celu.edu.ar/es/node/20

Setiyadi, A. B., Sukirlan, M., & Rahman, B. (2016). Language motivation, metacognitive strategies and language performance: A cause and effect correlation. *International*

Journal of Applied Linguistics and English Literature, *5*(7), 40-47. http://dx.doi.org/10.7575/aiac.ijalel.v.5n.7p.40

Silva, B. B., Kutyłowska, K., & Otwinowska, A. (2024). Learning academic words through writing sentences and compositions: Any signs of an increase in cognitive load?. *Language Teaching Research*, *28*(3), 1143-1175. https://doi.org/10.1177/13621688211020421

Smyk, E., Restrepo, M., Gorin, J., & Gray, S. (2013). Development and validation of the Spanish-English language proficiency scale (SELPS). *Language, Speech, and Hearing Services in Schools, 44*, 252-265. https://doi.org/10.1044/0161-1461(2013/12-0074)

Taguchi, N. (2008). The role of learning environment in the development of pragmatic comprehension: A comparison of gains between EFL and ESL learners. *Studies in Second Language Acquisition*, *30*(4), 423-452. https://doi.org/10.1017/S0272263108080716

Tano, M. (2012). *Expertos: Curso avanzado de español orientado al mundo del trabajo. Cuaderno de trabajo. B2.* Difusión.

Thomas, M. (1994). Assessment of L2 proficiency in second language acquisition research. *Language Learning, 44*, 307-307. https://doi.org/10.1111/j.1467-1770.1994.tb01104.x

Vanbuel, M., & Deygers, B. (2024). Assessing Receptive Skills in an Adult L2 Learner Population with Diverse Educational Backgrounds. *Language Assessment Quarterly*, *21*(4–5), 379-403. https://doi.org/10.1080/15434303.2024.2438151