

Underrepresentation of the construct in Spanish standardized foreign language exams: A computer assisted exam proposal

MIGUEL FERNÁNDEZ ÁLVAREZ
Universidad Politécnica de Madrid

JESÚS GARCÍA LABORDA
Universidad de Alcalá-Instituto Franklin

TERESA MAGAL-ROYO
Universitat Politècnica de Valencia

RECEIVED: 30 MAY 2021 / ACCEPTED: 6 JANUARY 2022

DOI: 10.30827/PORTALIN.VI.21393

ISSN PAPER EDITION: 1697-7467, ISSN DIGITAL EDITION: 2695-8244

ABSTRACT: Standardized tests are a reality in a globalized world and, consequently, in the Spanish system. Language test developers have already claimed that the current model of assessment has little to do with real language use, as factors such as the shift from paper to computer, information search and other uses of language or multimodal communication are not being considered (García Laborda & Fernández Álvarez, 2021). This theoretical paper proposes (1) a novel construct for testing that incorporates new functionalities which have been shown able to test appropriately language performance in academic settings; and (2) the proposal and design of an online multimodal exam that adapts to the new needs for language assessment in the University Entrance Examination. It begins with an analysis of current research to show the need for a change in the construct, and concludes by proposing a model for this high-stakes exam, based on a new construct of language proficiency and multimodal interaction for computer-based exams. The paper concludes that this model can change the way we understand technology-based exams and could be applicable to most current exams. **Key words:** language competence, language use, multimodality, standardized tests, University Entrance exams.

Subrepresentación del constructo en exámenes estandarizados de lengua extranjera en España: propuesta de examen asistido por ordenador

RESUMEN: Las pruebas estandarizadas son una realidad en un mundo globalizado y, en consecuencia, en el sistema español. Los desarrolladores de pruebas de lenguaje ya han afirmado que el modelo actual de evaluación tiene poco que ver con el uso real del lenguaje, ya que no se están teniendo en cuenta factores tales como el cambio del papel al ordenador, la búsqueda de información y otros usos del lenguaje o la comunicación multimodal (García Laborda & Fernández Álvarez, 2021). Este artículo propone (1) un constructo novedoso para las pruebas que incorpora nuevas funcionalidades que han demostrado ser capaces de evaluar adecuadamente el rendimiento lingüístico en entornos académicos; y (2) la propuesta y diseño de un examen multimodal online que se adapte a las nuevas necesidades de evaluación de idiomas en el Examen de Acceso a la Universidad. Comienza con un análisis

de la investigación actual para mostrar la necesidad de un cambio en el constructo y concluye proponiendo un modelo para este examen, basado en un nuevo constructo de competencia lingüística e interacción multimodal para pruebas computerizadas. El documento deduce que este modelo puede cambiar la forma en que entendemos los exámenes basados en tecnología y podría aplicarse a la mayoría de los exámenes actuales.

Palabras clave: competencia en el lenguaje, exámenes estandarizados multimodalidad, pruebas de acceso a la universidad, uso de la lengua.

1. INTRODUCTION

The difference between language competence and linguistic performance (Chomsky, 1965) is a concept that is little discussed today. However, in the case of language testing, the two concepts can be said to be related in terms of the knowledge of a foreign or second language and the time taken by a candidate to answer a particular test. While language teaching has for many years aimed at developing action skills that enable communication, language tests have attempted to fairly assess candidates' innate abilities, sometimes at the expense of communication. The successive revisions of language test design discussed in this paper have attempted to answer a simple question: "How good is a student on a test?" or in other words, "How good is a particular test?". Test design features should generally serve to make a test fair and equitable for all test takers.

Standardized testing is a reality in the current Spanish educational system. Despite the relatively low recognition that the field of language testing receives, in the context of globalization, new approaches to the creation and development of assessment tasks in the field of foreign languages are being sought (Boals et al., 2015). Researchers such as García Laborda and Fernández Álvarez (2021) consider that the current assessment model is ignoring many of the changes which second language (L2) learning has experienced due to some crucial factors that have influenced learning methods, such as the introduction of computer-based learning, the search for information in digital environments, the use of collaborative communication and learning environments, and the promotion of multilingualism in traditionally monolingual areas.

Other areas that also have had an impact are social and gender factors related to teaching content areas that require support and mastery of a foreign language. Research has focused only on specific problematic aspects such as the age and maturity of candidates taking a rigid and predetermined exam based on *Content and Language Integrated Learning* (CLIL) programs (Winke et al., 2018). However, there are no assessment studies that address the relation between language and content and its processing. While there is a clear distinction between 1) CLIL and 2) content generated through the use of language in open digital environments, such as through the use of instant messaging, virtual libraries, or social media-oriented applications, to name a few, exams do not reflect advances in language learning and do not account for the many new ways of learning.

There is no methodological vision in creating assessment tasks that are aligned with the type of content that is based on existing digital environments. This assertion by García Laborda and Fernández Álvarez (2021) is based on their study of the construct of eight foreign language certification exams used in higher education to assess foreign language

competence. For them, the problem is that, although the process of developing language tests has evolved, language use has not been adapted to the most common digital environments based on social media found nowadays in students' social and academic environments.

The present work, which questions the current state of test development, raises the issue of whether current tests reveal how well students will perform in real life. By questioning the relationship between the types of tasks in current tests and current language use, it proposes multisensory tasks to improve the field of language testing. The first part of the paper provides an overview of the conceptualization of language proficiency, focusing on some of the major theories that have been proposed in recent decades for both language teaching and language testing. These concepts form the basis for the proposal presented in this paper. Then, the concept of dynamic assessment is briefly discussed as it relates to the notion of multimodality. This section ends with a description of online interaction and linguistic mediation as presented in the CEFR Companion Volume (Council of Europe, 2020), two of the elements that are nowadays present in language teaching and that should also be part of language testing.

The second part of the paper reviews recent technological developments in the field of language testing, focusing mainly on multimodality, which is one of the foundations of the proposal presented here. This section presents the advantages of including multimodality in language tests and draws a comparison between more traditional tests and the new approach that takes into account other factors such as visual, auditory and sensory aspects in the assessment of language skills. This section is followed by the proposal of a new proficiency construct for multimedia language tests based on performance-related principles. Finally, this conceptualization is illustrated by an example of a possible task for the English component of the University Entrance exam.

2. LITERATURE REVIEW

For many years, language tests viewed a candidate's mastery of that language as an internal construct of knowledge that also required numerical or graded scoring that focused on the candidate's competence rather than on his actual performance. Underlying this account is the idea that competence is directly related to performance. In fact, this is not very different from Unitary Competence, as presented by Oller in his Integrative Model (1979, 1983). Since then, much work has been done in the field of language assessment, especially in terms of test validation, task integration, and implementation. However, practicality has prevailed over other features of tests. Prompts have been enriched by videos and audio recordings, and integrated competence tasks, such as those included in the iBT TOEFL test, aim to somehow illustrate language actions and linguistic scenarios. In this context, a number of models have served to identify the main linguistic features that will later model the test items.

Oller (1979) also called for test items to resemble those in real life by linking skills and language knowledge. However, as too much attention was paid to vocabulary and grammar, progress was very limited. In the early 1980s, Canale and Swain's Communicative Competence Model presented competence in terms of components by focusing on interaction and

sociolinguistics in addition to grammatical and strategic competence. Thus, communication was viewed as a dynamic process. However, in terms of language tests, it was difficult to reproduce acts of communication in a room or in a paper-and-pencil test. This probably prompted Canale (1983) to focus on conscious and unconscious processes, with a parallel situation in language teaching, especially in the Natural Approach. Discrete point tests were prevalent then (and still today).

Bachman's Language Competence Model (Bachman, 1990; Bachman & Palmer, 1996, 2010) considered strategic competence, metacognitive strategies, and higher-order thought processes as cornerstones of fluency, with affective skills ascribed a significant role that may not be significantly different from the Communicative Approach - although they were never specifically described or developed. This view is complemented by the work of Kramsch (1986), who examined the importance of interactional competence based on speakers' shared knowledge of the world and language. The problem with this view is that measuring only interaction can reflect a test taker's individual competence only at a superficial level. Celce-Murcia, et al. (1995), in the same vein as Canale (1983), emphasized the dynamics of linguistic communication from a sociocultural perspective of discourse co-construction. They also believed that language must be contextualized in its adequate scenario.

As it has been illustrated, the concept of language competence has evolved over the years to include other aspects of language that were not as prevalent or did not exist when the concept was first formulated. In the second decade of the 21st century, linguistic competence has become even more complex, as language users are exposed to new forms of interaction (such as online communication or video conferencing) and situations in which speakers from different linguistic and social backgrounds converse. As a result, language users are expected to adopt different and new strategies that can be achieved through aspects such as linguistic mediation. Assessment can therefore no longer be conceptualized as a linear process, as test takers must not only interact but also negotiate meaning to make the process as authentic as possible. Dynamic assessment encompasses all of these new aspects of language. The following sections address these concepts (fluency, dynamic assessment, language negotiation, and online interaction) as an introduction to the use of multimodality in assessment, which is the core aspect of this paper.

2.1. The Construct of Language Competence

One of the most important aspects of language test construction is the definition of the language construct to be assessed. Davies et al. (1999) define the term construct as

The trait or traits that a test is intended to measure. A construct can be defined as an ability or set of abilities that will be reflected in test performance, and about which inferences can be made on the basis of test scores. A construct is generally defined in terms of a theory; in the case of language, a theory of language. A test, then, represents an investigation of what a test actually measures and attempts to explain the construct (p. 31).

Brown (1994, p. 256) stated that "proficiency is a construct. Communicative competence is a construct. Self-esteem is a construct. Virtually every theoretical category (...) is a

theoretical construct.” Therefore, it is very important to first define the construct of a test. In 1965, Chomsky established the concept of linguistic competence, which refers to any speaker’s unconscious knowledge of the functional rules of a language, based on a hypothetical innate Universal Grammar. In the 1970s, this concept was redefined when Hymes (1971) described what he called communicative competence, which focuses on the use that people make of language rules based on the context in which they find themselves. According to this conceptualization, non-linguistic factors are also taken into account in language production, as the social dimension of language also plays an important role.

Canale and Swain (1980) extended the concept of communicative competence. They developed a construct of communicative competence consisting of four components: grammatical, sociolinguistic, discursive and strategic competence. The first refers to the degree of mastery of the linguistic code, including vocabulary, pronunciation, grammar and syntax. The second has to do with the ability to produce statements that are appropriate to the communicative situation, both in form and meaning. In other words, sociolinguistic competence is the knowledge of the sociocultural rules of language. Discursive competence is the ability to use different types of discourse and organize them according to the parameters of the communicative situation in which they are produced and interpreted. Finally, strategic competence reflects the potential of interlocutors to progressively define and qualify the meanings they convey and to make adjustments or clarifications. This competence allows people to use all the linguistic and extra-linguistic resources at their disposal to prevent communication from breaking down or taking undesirable paths.

The conceptualization presented by Canale and Swain (1980) has undergone some modifications. Bachman (1990) reorganized it and broke down the components. He included three main components in his model: language competence, strategic competence and psychophysiological mechanisms. The first component, which is our main interest in this paper, includes different types of knowledge we use in communication. According to Bachman (1990), language competence is the knowledge about language, as shown in Figure 1.

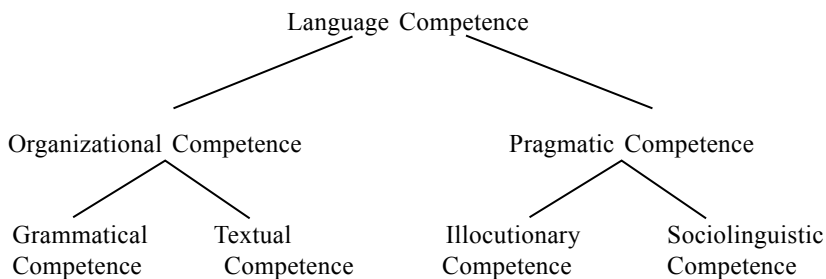


Figure 1. Bachman's Language Competence

Source: Bachman, 1990

For Bachman (1990) and Bachman & Palmer (1996, 2010), language competence includes two types of capabilities: organizational and pragmatic competence. The former refers to both the formal structure of language (grammatical competence) and knowledge of how

a discourse is constructed (textual competence). Grammatical competence includes mastery of vocabulary, morphology, syntax, and phonemic and graphemic elements. Finally, textual competence comprises aspects related to cohesion and rhetorical organization.

The second type of ability is called pragmatic competence, and refers to the functional use of language, i.e., illocutionary competence and sociolinguistic competence, the knowledge of the appropriate use of language according to the context in which it is used. Illocutionary competence involves the mastery of functional features of language such as the ability to express ideas and feelings (ideational functions), to achieve something (manipulative functions), to use language to teach, learn and solve problems (heuristic functions) and to be creative (imaginative functions). Finally, sociolinguistic competence takes into account aspects such as sensitivity to dialect and register types, proximity to the characteristic features of the language, and understanding of cultural references and idiomatic figures.

However, there are other aspects to be considered within language competence that go beyond the conceptualization proposed by Bachman (1990). To the competences described above, we need to add performance competence, within which we need to speak of collaboration between speakers in meaning-making, which can vary depending on the number of speakers involved in a communicative act and is defined as interpersonal competence (Schmidt-Fajlik, 2011). This is closely related to the concept of negotiation of meaning, which is very much in vogue today with language mediation and the CEFR Companion Volume (Council of Europe, 2020). Performance competence is also influenced by personal innate characteristics of the speaker and the context in which the communication takes place. Depending on these factors, speakers have to use situational strategies to get the message across to their interlocutor. These are complemented by negotiation strategies, defined by Long (1991, 1996) as adjustments or modifications in encoding and decoding that the learner and his or her interlocutor make in the communicative process.

2.2. Dynamic Assessment

Within the framework of sociocultural theory, dynamic assessment is a pedagogical intervention based on an integrated view of instruction and evaluation as parts of the same process to understand and promote the development of students' skills. Although it starts from a socio-constructive paradigm, dynamic assessment is fully linked to formative assessment. The aim of this approach is to develop mechanisms to help learners overcome the difficulties they face in performing a task. The objective is not so much to ensure that the learners can do the task correctly, but rather to provide them with the appropriate mediation to develop their skills so that they can apply them in similar situations or transfer them to more complex tasks.

Dynamic assessment is a concept that comes from the sociocultural psychological tradition (Aljaafreh & Lantolf, 1994). Unlike static assessment, it focuses on the progress and incorporation of all the necessary tools a learner needs to solve the learning tasks at the right time in the best possible way. In this way, the learner is guided in the learning process through corrective feedback that regulates the activity performed. In dynamic assessment, effective feedback has the following characteristics: it is contingent, that is, it reflects the actual need and stops when the learner can already solve the task independently; it is graded, that is, it provides the minimum amount of help that the learner needs to be able to solve

the task; and it is dialogical, since it is given to the learner's zone of proximal development.

Only using dynamic assessment, where cooperation between speakers is essential, has a true sociocultural assessment been made. However, McNamara (2000) criticizes that language tests would lead to a test of identity rather than of proficiency if all sociocultural features were taken into account. This brings us back to a point where there is a conflict between the purpose of communication and personal competence. The final goal is still up for debate today.

An approach combining both communicative interaction and personal competence was taken by Poehner (2007, 2008) and Poehner and Lantolf (2005), who assumed that tests at the time limited candidates' communicative abilities. Their sociocultural model placed interaction at the heart of communicative events, with reciprocal influence and affective factors playing an important role in the emergence of communication. Although the basics are not very different from Kramsch's Model of Interactional Competence (1986), it is more obvious that the commonalities and simplicity of interaction lead to evaluating the potential candidate only in terms of other elements such as the interlocutor, context, and the support from other elements (e.g., the media), among others.

Fatemipour and Jafari (2015) also see this danger and criticize that tests usually only diagnose the candidate's competence at a certain point, but hardly ever their potential. The main problem is that dynamic assessment seems to be more suitable for formative assessment than for summative language tests, despite later work by Poehner et al. (2015) and Poehner (2018), since a major task of the examiner is to promote the candidate's performance. In a very interesting review of dynamic assessment in foreign language teaching, Vergara Cabarcas et al. (2019) do not mention any significant studies where dynamic assessment is used after 2012. However, despite the lack of use of dynamic assessment in high stakes tests, we must acknowledge that its use is a more authentic way of assessing candidates' language proficiency, especially as the focus is on mediation, which is an aspect that has not been considered in the past.

2.3. Online interaction and linguistic mediation

As mentioned earlier, the concept of language competence has expanded to include new forms of communication such as online interaction or the use of telecommunications. Language users nowadays have to make use of skills and competencies that were not part of their everyday lives decades ago. Thus, our vision of language proficiency also has to focus on those aspects which have recently been addressed by the recent work of the Council of Europe (2020). Apart from the notion of online interaction, the concept of linguistic mediation, which was already present in the CEFR, has now received much more attention with the Companion Volume. The new conceptualization of linguistic mediation now goes beyond translation and paraphrasing (Fernández Álvarez & García Hernández, 2021), as there are many tasks that fall under the three types of activities envisaged in the work: textual, communicative and conceptual. In the case of Spain, the Official School of Languages have already integrated linguistic mediation into their assessment practices. At other levels of education, it is beginning to be integrated into the foreign language curriculum, as in the case of the Comunidad de Madrid, where students must demonstrate their competence in the use of mediation strategies in the program for Advanced English in the Baccalaureate.

According to the Council of Europe (2020), the role of mediation has received more attention “because of its relevance in increasingly diverse classrooms, in relation to the spread of CLIL (Content and Language Integrated Learning), and because mediation is increasingly seen as a part of all learning, but especially of all language learning” (p. 36). For this reason, and taking into account the idea of the authenticity of tasks simulating real situations as suggested by Oller (1979), the question of whether mediation should be part of the construct needs to be explored. The processing of linguistic mediation tasks involves several components, both linguistic and non-linguistic, which are interrelated. It is a multimodal process that facilitates communication between language users (Choi & Yi, 2016).

On the other hand, online interaction is now considered a third branch of interaction alongside oral and written interaction, as evidenced by the Companion Volume. While the use of technology and communication tools through apps, social media and other platforms has been well accepted in the classroom, the field of language assessment, as mentioned earlier, has not developed much in this regard. Recently, some interesting studies have been conducted on the integration of online interaction in the classroom, such as the Italian project by Cinganotto (2019, p. 150), who claims that “online interaction becomes a way to help our students become European and global citizens, developing the global and transversal competences needed to face the challenges of the knowledge society.” In view of this, and due to globalization, the field of language testing should adapt to the needs of 21st century language users.

All of these concepts represent new areas of research in language testing. Dynamic assessment is based to some extent on the concept of mediation in that there is an interaction between a test administrator, who acts as a mediator, and a test taker (Campione & Brown, 1987). However, conceptualization must go beyond this, and the field must begin to focus on the interrelation between speakers (or in this case, test takers) and their language use in order to construct meaning. Moreover, not much has been published about online interaction and its applicability to an exam situation, at least in the context of standardized testing. Co-construction of meaning, as claimed by Celce-Murcia et al. (1995), is present in the language classroom today (Council of Europe, 2020). However, there is still a huge gap between 21st century classroom practice and language testing, which has not kept pace with the development, globalization and needs of today’s language users and learners.

3. RECENT EVOLUTIONS OF TECHNOLOGY IN LANGUAGE TESTING: MULTIMODALITY

So far, little significant progress towards integrating technology in language testing seems to have been made in recent years, despite calls for better assessments that meet the current requirements for aligning language tests with new teaching methods and the use of technology (García Laborda et al., 2011). Certainly, the scenarios have improved, but the nature of the tasks remains very similar to those from the early 20th century. This is especially true for many tests that are still administered in paper form today. García Laborda and Fernández Álvarez (2021) state that versatile multilingual tests follow old-fashioned models in which grammar, vocabulary, reading, writing and listening comprehension tasks are based on traditional forms, and candidates are asked to deliver monologues as oral performance responses. This means that these tests use a monolithic approach and do not appear to be

very different from their paper-and-pencil predecessors.

Due to practicality, much effort has been put into the validation of technology-based tests although the question remains as to whether technology has served to implement “better” and more reliable tests. For example, the introduction of the TOEFL® iBT exam implied that higher-order thinking was introduced in the form of integrated skills (Chapelle et al, 2008), but the exam was not very different from the TOEFL 2000 version, which reflected the theories of language testing in the late 1990s (Chapelle et al., 1997). Consequently, it seems technology has been used primarily to implement the shift from paper-and-pencil testing to web-based testing, albeit following similar test models despite its potential. Technology could help focus on online interaction, as different tasks where test takers interact in an online platform could address this form of communication.

Furthermore, new technology-based scenarios have brought the possibility of integrating new combined tasks into language assessment. The development of a task implies the simultaneous use of different perceptual channels by humans. Similarly, multiple perceptual channels working simultaneously allow perceiving, checking and answering a task. The main problem seems to be the integration of techniques of multimodal (also multisensory) human-computer interaction in order to achieve an increasing normalization of interaction in language tests, as well as the integration of new models of communication use, mostly based on the advancing theories of competence and communication.

The use of pattern recognition interfaces such as facial recognition, eye-tracking, cerebral sensors that allow for the assessment of non-verbal language such as glances, gestures and other communication inputs can be used to enrich and approximate other types of visual communication between a student and the examiner. For example, the question «Do you like apples and bananas?» can be answered in different ways such as «Yes, I do, and you?», «Yes, I do», «Yes» or simply nodding, and in real life, none is better than the other. However, the last two are rarely considered correct or appropriate in language tests. So, when scoring tests, it is important to remember that prompts are not necessarily just verbal input, but a combination of speech, figurative language, and body language - involving not just the face or head, but also the hands and the rest of the body. In this sense, attentive user interfaces with some pattern recognition will be useful, taking into account the user's interest and the importance of time for data collection (linguistic and communicative cues) by incorporating visual recognition technologies such as eye-tracking, facial expression and gesture recognition.

Multimodal interfaces, thus, enable the organization and transmission of the user's enactive knowledge in front of a real or fictional environment that serves to evaluate their knowledge (Pacheco & Souza-Concilio, 2013). The purpose of education is not to impart knowledge, but to promote the learner's problem-solving skills, which can then be transferred to different situations (Bruner, 1961; McLeod, 2019). In other words, computer-based assessment environments, whose interfaces enable the expression and transfer of enactive knowledge through the integration of different sensory aspects, can be created (Ho et al. 2009). For example, the student's natural gestures during an assessment task (efferent component of the system) and the use of different communication channels (afferent component) can be used to assess linguistic competences.

Additionally, enactive interfaces can be considered as they are used to add hands and bodies to perform a task. In our case, only the hands can suffice. In a study of an early

version of the APTIS test, García Laborda et al. (2017) observed different reaction in students when taking an oral test according to their proficiency level. Advanced students tended to move the body to enrich their responses, even when their body language was not recorded, while lower-level students tended to become static. It would also be interesting to see if supporting such hand and head movements has an effect not only on the way we communicate in real life, but also in speaking tests. Of course, body language has been but only marginally associated with language tests, despite the various culturally determined different patterns that non-native speakers use when speaking an L2. As Table 1 shows, all the features described here imply there are important differences between traditional Computer Assisted Language Testing (CALT) and multimodal enriched CALT.

Table 1. Differences between traditional and multimodal enriched CALT

TRADITIONAL ENRICHED CALT	CHARACTERISTICS
–Skill based input (usually one at the time)	–Multimodal input
–Item based	–Continuous
–Sequential processing	–Multi-process
–Umbrella oriented (candidates can take any section at any one time)	–Linear but also with certain freedom (especially in speech)
–Unimodal	–Multimodal

Source: Universidad de Alcalá de Henares, 2021

Other benefits of using multimodal channels simultaneously include more creative tasks (as we show later in this paper), the use of higher order thinking skills such as understanding, evaluation and creation, which are very limited in many language tests, an increasing need for problem solving, and the inclusion of individual differences. For example, if we ask a candidate to negotiate the solution to a problem with another candidate using a virtual library and creating a new object.

As mentioned earlier, these performances are not to be regarded as incidental, but as an essential part of human communication in an L2 and must be evaluated accordingly. This can be done through the use of Graphical User Interface (GUI), which uses a camera and a keyboard mouse to recognize voice and gesture input and performs an evaluation. In addition, like current testing platforms, they require a human to interpret nonverbal language. Even the most sophisticated oral exams today use a dual scoring system that includes an automated system for recognizing writing and speech, and a human. Many believe that soon all scoring will be done automatically by scoring machines, but if there is no human scoring, important evidence is likely to either be neglected or simply lost. Therefore, even with guesses, there should be either patterns for automatic image recognition or human intervention. The problem with automatic body recognizers would be that body language includes

a number of different types of gestures (such as symbolic, deictic, iconic, and pantomimic) and also that there may be no correspondence between the L1 and L2 at initial language levels. Overall, the test criteria should include speech output, consideration of candidate behavior and nonverbal language, and the use of various skills such as collaboration and interaction between candidate and computer. We will discuss types of interaction in further sections of this paper.

As we have seen so far, it seems that language tests do not take into account or take advantage of recent advances in technology, especially multimodality, originally called multisensory or multimodal communication. There are three kinds of multimodal interaction elements developed or used by the candidate, which could be evaluated:

- Visual: facial and eye movements, use of facial expressions, head movement, gestures (any body parts).
- Auditory: speech and nonverbal sounds (such as “*neh neh*”, “*uhum*” “*eh*” and the like).
- Sensory: emotions and others, such as the use of time.

Surprisingly, some of these features, especially body movements, eye-tracking, or gaze movements, are used in remote tests. However, the question of why a combination of two or more input modes is not used in speech tests remains unanswered.

4. DESIGNING A NEW COMPETENCE CONSTRUCT FOR MULTIMEDIA-BASED LANGUAGE ASSESSMENTS

All of what has been said so far proves the need to reconceptualize the construct of competence for new types of assessments, because so far this work has shown that measuring competence without including different kinds of performance only maps part of what communicative competence includes. Therefore, competence models based only on personal capability cannot meet the new demands of dynamic communication. Therefore, personal communication would be only one part of a much larger construct in which the way of communicating with virtual media has a significant meaning to describe not only what the students “know” but also what they “can” do. To this end, Figure 2 proposes a new model of language competence that takes interaction into account, rather than a mental, built-in model that seems to be the current paradigm in language testing. According to Figure 2, there are several aspects to consider, the most important of which are the relationship between linguistic competence and (1) performance competence, (2) organizational competence, (3) pragmatic competence, (4) ideational competence, (5) sociolinguistic competence, and (6) interpersonal competence. In addition, linguistic competence is influenced by personal characteristics (i.e., memory, age, sociocultural level, studies, etc.) and context. Of all these interrelationships, the one with pragmatic competence is beyond the scope of this article, as we will focus only

on those aspects that have traditionally been neglected in the development of language tests.

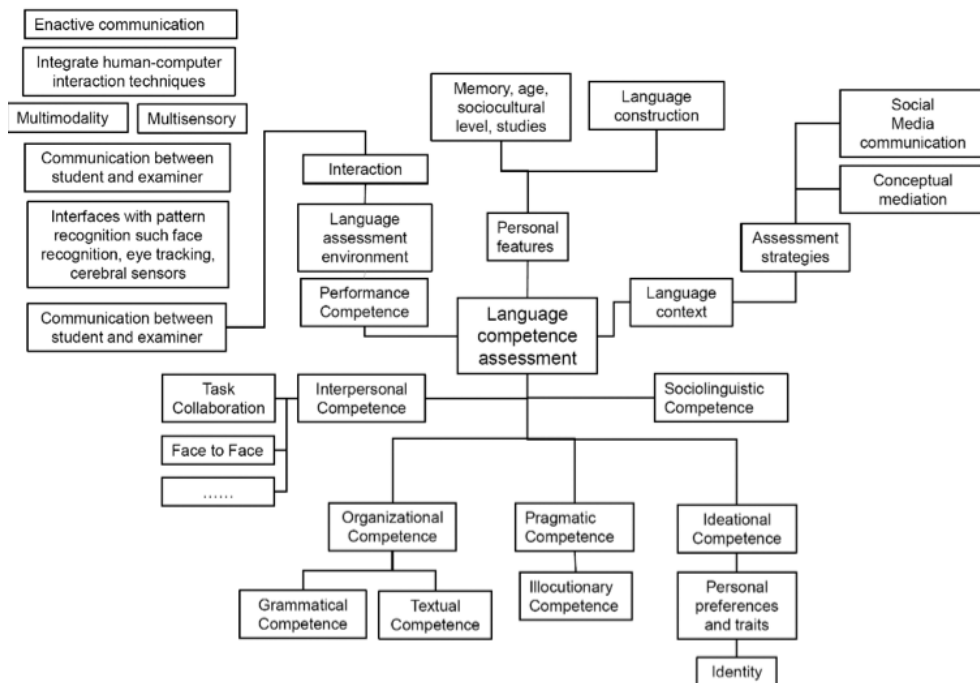


Figure 2. New model for language competence assessment

With respect to the Ideational Competence, it is related first to personal forms of communication, and second to the choices and characteristics of the person that interacts with the external (and even internal) speech and define the speaker’s personality. As will be shown, this has a significant impact on Performance Competence in terms of interaction and engagement in individual and group work. For example, a test taker working with another may make choices during a test to make the speech and ideas accessible or simply retain the ideas to show better performance than the other person, as might be the case in normal problem/task solving group work. As for the influence of context, this is evident in the use of socio-situational strategies such as use of time or knowledge of the communication channel (such as media, virtual conferencing and others).

This also has implications for language management and interaction due to external factors (such as taking an online language test with or without an examiner), which can also have a significant impact on the negotiation of meaning (Long, 1994), an idea that is clearly explained in Bachman’s (1990) and Bachman and Palmer’s work (1996) when defining and describing language competence. However, performance competence seems to be what needs to be revised in contemporary language use. According to this proposal, what needs to be measured is the ability to interact with context and external audiences through language.

In this sense, it is necessary to measure in language tests the interactionist ability of the individual, which cannot be observed individually (as mentioned when candidates “show off” in front of an examiner) in the form of interaction and collaboration, but as dual or group engagement. This, in turn, leads to the development of interpersonal-interactional competence, which is contrary to the traditional perspective of individual competence.

Following this language construct, a framework for language testing has been developed in which, in contrast to the traditional view, the focus is on performance rather than on individual competence (Figure 3). Individual competence includes the four traditional skills (speaking, listening, reading, and writing), and the use of higher order thinking skills, as this paradigm introduces problem-based tasks, and motivation and personality.

The mental processes shown by the test taker are used to link the required tasks and the candidate’s competence. In terms of tasks, test designers and administrators need to ensure quality by validating the entire test (Weir, 2004) and its items, as well as accessibility, especially in computer/technology-based tests. They need to keep in mind the test objectives and the construct definition to achieve these objectives, which are also shaped by the test characteristics. The use of specific hardware/software tools has become one of the most important aspects of item construction as it allows for new types of items, as will be seen later. Performance is also influenced by test timing, the personal preferences of the participants, and finally by the state of mind (mood) of the test taker.

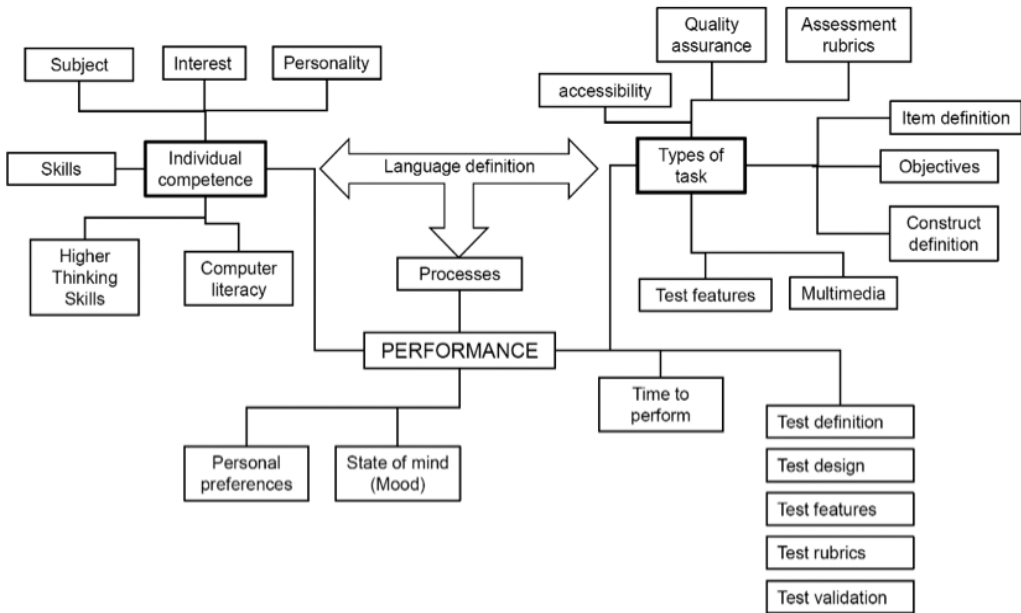


Figure 3. Framework for language testing

According to this framework, the use of technology is most relevant because it allows for the creation of new types of items that require test takers to show their performance competence by performing, as will be discussed in the following section.

5. THE APPLICATION OF THE NEW MODELS TO THE UNIVERSITY ENTRANCE EXAMINATION

In a very recent work, Ruíz-Lázaro et al. (2021) found that not all students in the Spanish education system take a similar English exam for university entrance, as each region adapts the legislation differently. This means that there is an unequal, perhaps unfair, situation between all candidates wishing to attend university. There is not a single region in Spain that specifies the definition of the construct of language competence used in their University Entrance Examination (EVAU/EBAU), and this is what we are proposing here with this alternative model of language assessment, based on multimodality and emphasizing the use of real collaborative tasks.

As with other university entrance exams, the aim of this test is to check whether students are able to use English as a means to work in academic contexts, so all four traditional skills should be included, however, the exams currently used in Spain usually neglect speaking and listening skills (except in Galicia and Catalonia). The problem is that the current English test has not been changed for many years, despite the numerous studies showing its lack of validity and reliability (Amengual Pizarro, 2006; Díez Bedmar, 2011; Sanz Sainz & Fernández Álvarez, 2005), and the various proposals that have been made to modify and expand the construct of the test (García Laborda, 2006, 2012), to include new tasks (García Laborda & Martín-Monje, 2013), oral assessments (Amengual-Pizarro & Méndez García, 2012; Bueno-Alastuey & Luque Agulló, 2012), or to incorporate the use of computers in its administration (García Laborda & Gimeno Sanz, 2008; García Laborda & Magal Royo, 2009; García Laborda, 2010).

This exam exemplifies how the evolution of language teaching and language assessment has not been parallel. While language teaching has focused on preparing students to be part of a multilingual and multicultural society that incorporates the latest technologies, and the CEFR has also included online interaction, the use of telecommunications and linguistic mediation in the Companion Volume (Council of Europe, 2020), language assessment is carried out as in the 20th century. Table 2 shows the comparison between our multimodal, problem-based proposal, which incorporates some of those elements already taken into consideration for teaching, and the current construct of the English component of most University Entrance exams in Spain.

Table 2. Comparison between the current English component of most University Entrance exams in Spain and our multimodal proposal

SKILL	CURRENT IN MOST REGIONS IN SPAIN	MULTIMODAL PROPOSAL
Reading	Open questions, True/False, multiple choice	Online libraries, online dictionaries, other media texts (blogs, websites) requiring reading, summarizing, and analyzing, not just “knowing”

Writing	Essay writing, short answers	Incorporating images (from libraries) and other objects. Creation of a multimodal (visual, auditory) product like an (approximate) scientific mini-paper
Speaking	Not included	Collaboration between test takers, negotiation with a goal, interventions, different types of language, and interactions from negotiation to presentation of results, not from an external input, but from a self-created product.
Listening	Included in some regions as a multiple-choice task	Digital objects from libraries, but can also be just interaction with the contributor(s).
Language use	Grammar questions	Wide variety of ways

In order to illustrate how to include all the main points we have analysed and as the main goal of this paper is to design an alternative model of language testing based on multimodality, an example will be provided based on an initial problem-solving situation:

“The candidate will work with a peer to design and implement a small 300-word essay on current considerations about the solar system. To do this, the student will have access to the libraries provided, which will include reading and listening data. Each student will have access to certain individual information that will not be shared by both candidates.”

This test will be recorded and sent to the test raters who will need to do both holistic and analytical evaluation. Candidates will have to read and understand texts at the required level (in this exam it is usually B1 in the CEFR), negotiate in an online environment and share their individually obtained information to create their own individual written output. In this sense, students will have to interact collaboratively while using all four traditional skills with input which will be provided both in writing and orally. The testing platform will record both linguistic and non-linguistic interactions. The time period will not have to be longer than one hour, which makes it very realistic as well as interactive and provides reliability about real language use, as in the previously addressed language competence described in section 4. Care will also be taken to ensure that what is assessed are the three main concepts of Figure 3. In this way, it is assumed that students will be working closer to the actual language use that candidates encounter in real life.

This alternative model proposed presents new tasks that are richer, as well as more creative and authentic, and even though ‘practicality’ issues (human and material resources, equipment, time constraints, safety issues, etc.) might arise, it is important to make sure that all the tasks can be implemented in a high-stakes testing situation without any problem to ensure that it follows the principles of validity, reliability and practicality.

6. CONCLUSIONS

Multimodal interactivity has been an important part of language teaching for years, and even more so now that we are in a technological age where communication is becoming more virtual every day. The communicative process, and, therefore, language competence, has evolved rapidly in recent decades. However, even though multimodality has brought many new opportunities for language testing, there is still much to be done. One of the first tasks is to revise the language competence model to include new forms of communication. The proposal presented in this paper addresses the importance of non-conventional, enriched assessments that can be used in summative assessments and standardized tests with limited resources and that are not very different from the way people and students use technology-based resources. There are also significant areas that require change in the nature of items, including the need for human assessment (as opposed to current machine-based assessment), platform implementation, cultural differences between different nationalities, and natural access to training materials, among others.

Shifting from one paradigm to another is always a challenge that must be overcome. Although the field of language testing is evolving and trying to adapt to the times, it is doing so at a slow pace, and in the case of Spain, that pace is even a little slower. As has been mentioned, the language component of the University Entrance Examination has changed little (or almost not at all) in recent decades, and this is indicative of the resistance to change that exists in this field. Nonetheless, it is our professional task to propose improvements to the testing system in order to assess a real and authentic representation of the language use that students will need in the future. Implementing this type of multimodal assessment is possible once it is piloted and would bring an increase in test reliability, but possibly at the expense of practicality. So, while it is a great way to assess interpersonal and individual competencies, it could be challenging to implement, as it happened in the early 2000s when most exams went online. Consequently, we need to look at multimodality not as a limitation, but as a way to enrich language testing in the future.

One of the limitations of this work, as with any other theoretical work, is the need for further research and applicability. Test developers might assume that a person is fluent with language, which is more convenient to diagnose than actual ability to interact with and use language in real life. However, this would not be representative enough of other language constructs that are widely used in the 21st century and were not relevant a few years ago. Our further research relates to creating an appropriate context in which this can be implemented. A second concern is how to move from criterion-referenced tests to multilevel tests. This might be possible within certain limits, e.g., by constructing tests whose input can mean one level above or below the competence being measured (similar to what many Cambridge Assessments examinations do when certificates of a lower level are issued).

As a final remark, it is worth noting that there is still a long way to go and, despite the resistance to change and to the integration of other competences into language tests, it is our hope that this change will soon occur and that this proposal can serve as a starting point for an assessment that is not only more valid and reliable, but also much fairer and with a much more positive impact. There is no doubt that the 21st century student has nothing to do with the student of the last century. Therefore, both teaching and assessment must adapt to the new realities.

7. REFERENCES

- Aljaafreh, A. & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal*, 78, 465-483.
- Amengual Pizarro, M. (2006). Análisis de la prueba de inglés de selectividad de la Universitat de les Illes Balears. *Ibérica*, 11, 29-59.
- Amengual-Pizarro, M. & Méndez García, M. C. (2012). Implementing the oral English task in the Spanish university admission examination: An international perspective of the language. *Revista de Educación*, 357, 105-127.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (2010). *Language assessment in practice*. Oxford University Press.
- Boals, T., Kenyon, D. M., Blair, A., Cranley, M. E., Wilmes, C., & Wright, L. J. (2015). Transformation in K-12 English language proficiency assessment: Changing contexts, changing constructs. *Review of Research in Education*, 39(1), 122-164. <http://dx.doi.org/10.3102/0091732X14556072>.
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, 31, 21-32.
- Bueno-Alastuey, M. C., & Luque Agulló, G. (2012). Foreign language competences required in the university admission examination: A proposal for the evaluation of oral aspects. *Revista de Educación*, 357, 81-104.
- Campione, J. C., & Brown, A. L. (1987) Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.) *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82-115). Guilford Press.
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 2-27). Longman.
- Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Celce-Murcia, M., Dornyei, Z. & Thurrell, S. (1995). Communicative competence: A pedagogically model with content specifications. *Issues in Applied Linguistics*, 6(2), 5-35.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language*. Routledge.
- Chapelle, C., Grabe, W. & Berns, M. (1997). *Communicative language proficiency: Definition and implications for TOEFL 2000*. Educational Testing Service.
- Choi, J., & Yi, Y. (2016). Teachers' integration of multimodality into classroom practices for English language learners. *TESOL Journal*, 7(2), 304-327.
- Cinganotto, L. (2019). Online interaction in teaching and learning a foreign language: an Italian pilot project on the companion volume to the CEFR. *Journal of e-Learning and Knowledge Society*, 15(1), 135-151.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.
- Council of Europe (2020). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing.
- Diez Bedmar, M. B. (2011). The English exam in the university entrance examination: An overview of studies. *Revista Canaria de Estudios Ingleses*, 63, 101-112.
- Fatemipour, H., & Jafari, F. (2015). The effect of dynamic-assessment on the development of passive vocabulary of intermediate EFL learners. *Journal of Educational and Management Studies*, 5(1), 41-51.

- Fernández Álvarez, M. & García-Hernández, S. (2021). Teachers' perceptions of linguistic mediation in the curriculum for advanced English in Madrid secondary schools. *Language Teaching Research*. <https://doi.org/10.1177/13621688211005602>.
- García Laborda, J. (2006). Analizando críticamente la selectividad ¿Todos los estudiantes tienen las mismas posibilidades? *TESOL-Spain Newsletter*, 30 (2), 9-12.
- García Laborda, J. (2010). ¿Necesitan las universidades españolas una prueba de acceso informatizada? El caso de la definición del constructo y la previsión del efecto en la enseñanza para idiomas extranjeros. *Revista de Orientación y Psicopedagogía*, 21(1),71-80.
- García Laborda, J. (2012). Presentación. De la selectividad a la prueba de acceso a la universidad: Pasado, presente y un futuro no muy lejano. *Revista de Educación*, 357, 17-27.
- García Laborda, J., Amengual Pizarro, M., Litzler, M. F., García Esteban, S. & Otero de Juan, N. (2017). *Student perceptions of the CEFR levels and the impact of guided practice on APTIS oral test performance*. British Council.
- García Laborda, J. & Fernández Álvarez, M. (2021) Multilevel language tests: Walking into the land of the unexplored. *Language Learning & Technology*, 25(2), 1-25.
- García Laborda, J. Giménez López, J. L., & Magal Royo, M. T. (2011). Validating mobile devices in the Spanish university entrance exam English paper. *New Educational Review*, 25(3), 160-171.
- García Laborda, J. & Gimeno Sanz, A. (2008). Adaptación del examen de inglés de las pruebas de acceso a la universidad a un entorno informático: Estudio sobre la tipología de preguntas. *Proceedings of the XXV Congreso Nacional de Lingüística Aplicada* (pp. 723-730). Servicio de publicaciones de la Universidad de Murcia.
- García Laborda, J., & Magal Royo, T. (2009) Training senior teachers in compulsory computer based language tests. *Procedia - Social and Behavioral Sciences*, 1(1), 141-144.
- García Laborda, J., & Martín-Monje, E. (2013). Item and test construct definition for the new Spanish Baccalaureate final evaluation: A proposal. *International Journal of English Studies*, 13(2), 69-88.
- Ho, T., Yin, P., Hwang, G., Shyu, S. J., & Yean, Y. (2009). Multi-objective parallel test-sheet composition using enhanced particle swarm optimization. *Educational Technology & Society*, 12(4), 193-206.
- Kramsch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal*, 70(4), 366-372.
- McNamara, T. (2000). *Language testing*. Oxford University Press.
- McLeod, S. A. (2019). *Bruner - learning theory in education*. Simply Psychology. Retrieved from: <https://www.simplypsychology.org/bruner.html>
- Oller, J. W. (1979). *Language tests at school*. Longman.
- Oller, J. W. (1983). *Issues in language testing research*. Newbury House.
- Pacheco, B. A. & Souza-Concilio, I. A. (2013). Multimodal interfaces: An enactive approach, *IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom)*. (pp. 51-58), Budapest, Hungary, doi: 10.1109/CogInfoCom.2013.6719302.
- Poehner, M. E. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *Modern Language Journal*, 91(3), 323-340.
- Poehner, M. E. (2008). *Dynamic assessment. A Vygotskian approach to understanding and promoting L2 development*. Springer.
- Poehner, M. E. (2018). Probing and provoking L2 development: The object of mediation in dynamic assessment and mediated development. In J. P. Lantolf, M. E. Poehner & M. Swain (Eds.), *The Routledge handbook of sociocultural theory and second language acquisition* (pp. 249-265). Routledge.

- Poehner, M. E., Zhang, J. & Lu, X. (2015). Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing*, 32, 337–357.
- Poehner, M. E. & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research*, 9(3), 233-265.
- Ruiz-Lázaro, J., González Barbera, C. & Gaviria Soto, J.L. (2021). Las pruebas de inglés para acceder a la universidad. Una comparación entre comunidades autónomas. *Educación XXI*, 24(1), 233-270, <http://doi.org/10.5944/educXX1.26746>
- Sanz Sainz, I. & Fernández Álvarez, M. (2005). La validez del examen de inglés en selectividad. In H. Herrera Soler & J. García Laborda (Eds.). *Estudios y criterios para una selectividad de calidad en el examen de inglés* (pp. 149-164). Editorial de la Universidad Politécnica de Valencia.
- Schmidt-Fajlik, R. (2011). Improving interpersonal competence in the communicative classroom. *The Language Teacher*, 35(3) 25-30.
- Vergara Cabarcas, L. K., López Caraballo, J. L., Castellón Barrios, D. J., Vásquez Rossi, C. A. & Becker Arroyo, E. A. (2019). Dynamic assessment approach in language teaching: A review. *Zona Próxima*, 30, 82-99.
- Winke, P., Lee, S., Ahn, J. I., Choi, I., Cui, Y., & Yoon, H. (2018). The cognitive validity of child English language tests: What young language learners and their native-speaking peers can reveal. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect*, 52(2), 274-303. <http://dx.doi.org/10.1002/tesq.396>.