NATIONAL STANDARARDIZED FOREIGN LANGUAGE TEST: ENGLISH

2024

Table of Contents

Background

The OECD Mandate

The Organization for Economic Co-operation and Development (OECD), founded in 1961, aims to stimulate economic progress and world trade. The OECD represents a forum for countries describing themselves as committed to democracy and the market economy and provides a platform to compare policy experiences, seek answers to comı problems, identify good practices and coordinate domestic and international policies of its members. Since the year 2013, during President Laura Chinchilla's time in office (2010-2014), Costa Rica openly expressed a special interest of becoming an OECD country and started the process of becoming a member of the organization. To achieve this goal, the OECD required the country to establish mechanisms to improve its citizens' quality of life in different areas, one of them being education.

The Costa Rican educational system is mainly split into General Basic Education (three cycles) and Diversified Education. The first two cycles (out of three) of the General Basic Education correspond to primary education three years each (first to third grade and fourth to sixth grade). The third cycle corresponds to lower secondary education (seventh to ninth grade). The next stage in the system is called Diversified Education. This refers to higher secondary education, from tenth to twelfth grade for the technical branch and tenth to eleventh in all other offerings. When students enter Diversified Education, they can choose a two-year track in artistic, scientific or academic high schools or opt for a three-year track in technical schools. To graduate from either branch, students must take a national exam called Bachillerato in different subjects. This includes a reading comprehension test in English.

In relation to the educational field, the OECD sponsors an international program called the *Program for International Student Assessment* (PISA). This program tests students'

competencies in three domains (Science, Math and Reading Comprehension) revealing how well-prepared students are to face the challenges and demands of adult life in the workplace. PISA is administered in the official language of each participating country from 15 to 16-year-old students enrolled in secondary education. Costa Rica has administered PISA in Spanish since 2009, when the country first participated in this international program. So far, the country has participated in four PISA cycles: 2009+, 2012, 2015, 2018 and 2022. Costa Rica has not only wanted to compare the outcomes of its educational system to those of other participating countries by means of PISA, but it also wanted to become an OECD country. To reach this goal, the OECD has requested that Costa Rica self-examine its current systems in different areas such as education, health, and economy, among others. The purpose at that time was to detect which areas to be improved, if any, and what improvements to make.

After several OECD visits (missions as the OECD calls it) to Costa Rica in 2015 and 2016, OECD representatives specialized in the education sector released a report entitled, *Reviews of National Policies for Education, Education in Costa Rica,* in which various recommendations were made. The recommendations listed are currently being addressed by responsible government entities so that the OECD consider granting Costa Rica membership into the organization. Once the recommendations are implemented, Costa Rica will become a member of the OECD, which means they will be represented in a forum where their economic and social practices will be recognized. One of the recommendations mentioned in the report is presented below.

> Costa Rica is implementing an ambitious curricular reform that emphasizes critical thinking instead of rote memorization, as well as giving increased importance to domains such as citizenship and foreign languages. This initiative has great promise as a means of engaging students as more active learners and ensuring they gain skills that are more relevant to society and the labor market. The emphasis on problem solving and critical

thinking is vital in a modern economy and reflects the kind of higher-level competences which should come to the fore in upper secondary education, at a stage in learning which goes beyond basic skills. The end-of-cycle certifying Bachillerato exam is being reformed to reflect these changed expectations [....]. Second, the Bachillerato exam itself needs reform to allow those who succeed in only some elements of the exam to receive proper recognition of their achievements. (p. 158, Reviews of National Policies for Education, Education in Costa Rica)

Regarding the Bachillerato National Exam in Costa Rica, English language proficiency was measured by means of a single modality, reading comprehension. This exam evaluated the students' knowledge of the different units and themes taught throughout Diversified Education through testlets; a stimulus (reading passage) followed by a maximum of five selected-response items. The reading passages were aligned to the topics and themes presented in the 2007 National English Curriculum. Therefore, responding to the OECD suggestions, the authorities have agreed that changes must be implemented by the year 2021 when the first generation of students taught under the new English National Curriculum, implemented in 2017, will graduate from the Diversified Education.

The mandate for a new English curriculum

Due to the needs analysis entitled, the "Diagnostic Report: English Study Programs for Preschool, Primary and Secondary Education" conducted by the national English advisors at the Curricular Division and published in 2015, it was unveiled the following areas of improvement:

First, learners needed an updated curriculum that reflects the knowledge, skills, and abilities needed to communicate in a range of language use contexts and to succeed in the information age as 21st century learners. The needs analysis that informed the curricular reform suggested that some of the target contents of pre-school, elementary

and secondary school´s curriculum had lost pertinence as manifested by anecdotal reports collected from teachers and students. Secondly, learners who receive English lessons in elementary and high schools are not reaching the expected English proficiency levels after eleven or twelve years of instruction. Among other things, these shortcomings could be attributed to the fact that the previous curriculum failed to specify the English language proficiency level that students are expected to attain at the end of each cycle and to the fact that the assessments in both classroom and standardized testing contexts were not systematically aligned with curriculum and instruction. Thirdly, for citizens to communicate effectively in the global context and to face the challenges of an interconnected world, they need to possess a number of competences.

Finally, the findings from the needs assessment administered to a sample of stakeholders including English Advisors, other language specialists (e.g., university professors, schoolteachers) and students revealed a need for:

- Better articulation of learning objectives, contents and assessments across the cycles.
- More coherence among the curricular elements.
- Improved clarity of the learning objectives.
- More detailed specification of pedagogical mediation.
- The incorporation of Information and Communication Technologies in the learning environment.
- Reduction of the number of units in the curriculum.
- More learning resources for teachers and students.
- Reconceptualization of assessment practices in both classroom and national contexts.
- Alignment of assessments with learning, instruction, and the curriculum.

- Establishing explicit criteria (e.g., language content knowledge, pedagogical content knowledge, assessment literacy, and digital literacy) for hiring and retaining English language instructors.
- The English class primarily delivered in the target language.
- Strengthening communication and coordination with universities in relation to teacher formation and teacher training.

A need to implement a new English curriculum because some of the target contents did not address the reality of the students in the classroom. Another important finding revealed that the learners were not reaching the expected proficiency level expected at the end of the Diversified Education. Among the conclusions expressed by the Curricular Division were that these findings and other anecdotal accounts expressed by English language teachers, indicated that the previous curriculum was not fulfilling the language provision required by students who are living in a globalized society and in a world that constantly changes in terms of technological resources. Another important conclusion was that both the classroom-based assessment and the large-scale assessment did not take into consideration the synergies among curriculum and instruction.

The English language curriculum is founded on a legal framework that consists of the 1949 National Constitution (clauses 76 to 89), the Basic Education Law (Article 2), the Educational Policy for the 21st Century, and a philosophical framework fostered by three philosophical trends: Humanism, Rationalism and Constructivism. It is influenced by the principles of a socio-constructivist approach to education in which learners have active participation in, interaction with, and adaptation to the context; as well as holism, where learners are seen as moving from a fragmented world vision to an integral one, where everything and everyone are interconnected and interdependent; and critical pedagogy, where learners become producers of knowledge, not just consumers.

Figure 1 Proficiency level projection for the National English Curriculum (NEC)
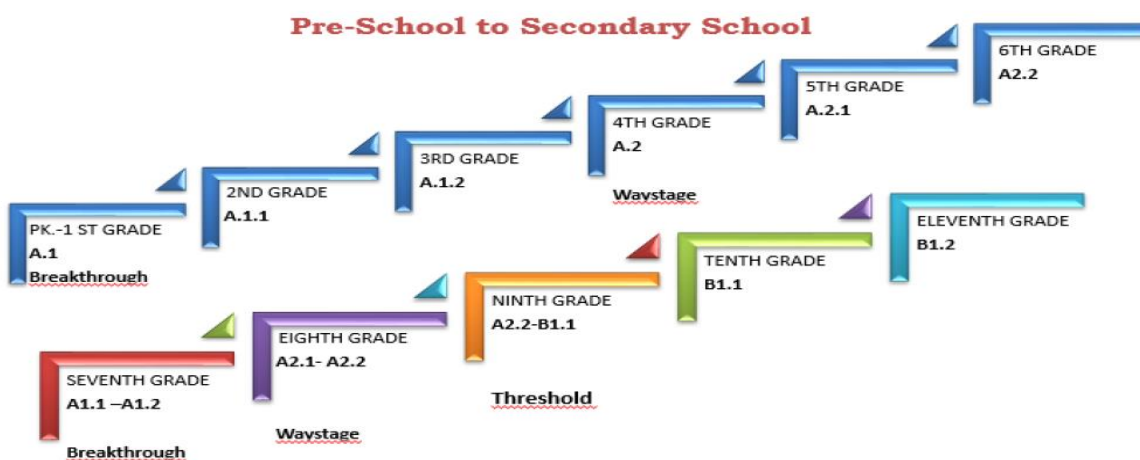
Source: National English Curriculum, 2016

The curriculum is aligned to the proficiency levels according to the Common European Framework of Reference (CEFR), these indicators describe what learners can do in the target language. The following table presents the proficiency levels starting in seventh grade with A1 and at the end of the Diversified Education learners will reach B1 level of the CEFR.

Another element to highlight is the overarching approach known as the Action-Oriented Approach (AOA) (Piccardo, 2014), which views language learners as 'social agents' who need to perform tasks in a specific environment and within a particular field of action to reach an objective, solve a problem or reach a concrete result. This approach promotes the active participation of students who plan, monitor, and evaluate simple and complex tasks individually and collaboratively using the English language. These tasks are organized around authentic situations that represent different domains, contexts, situations, or scenarios that provide authenticity to them. The National English Curriculum claims that by the end of the secondary education, learners should have the linguistic competences to communicate effectively at the B1 proficiency level, according to the CEFR, in familiar matters regularly encountered in the socio-interpersonal, socio-transactional, and academic domains in four language modalities or competencies.

To guide instruction to plan English lessons, the following teaching processes during the lessons are proposed:

Table 1 Teaching Process During Lessons

Oral Comprehension	**Planning** (pre-listening, motivating, contextualizing, explaining task goal); **listening for the first time** (general understanding); **pair/group feedback; listening for the second time** (more detailed understanding); and **self/co assessment or integration of skills.**
Written Comprehension	**Planning** (pre-reading, explaining task goal, use typographical clues, list difficulties/strategies to cope them); while-reading **reading for the first time; pair/group feedback, reading for the second time**, **post–reading** (for reacting to the content or focusing on features /language forms and self /co assessment)
Oral Production	Spoken interaction (**planning, organizing, rehearsing, and interacting**) and spoken production (**planning, organizing, rehearsing, and producing**).
Written Production	Pre-writing, drafting, revising, editing, and publishing.

Source: National English Curriculum, 2016

Assessment in this curriculum is a purposeful, continuous, contextualized, authentic, reflective, investigative, systematic, and multiphase process to serve each learner's learning and growth. For each goal or "can do performance descriptor" assessment will mainly be performance-based. Learners are required to demonstrate through integrated-skills tasks, that simulate real-life situations, what they can do with the language within a domain, scenario, and theme. Therefore, assessment must be authentic, which means that the task will be stated within domains and scenarios beyond the classroom setting, and the socio-cognitive, socio-affective, sociocultural, and linguistic demands upon the learner will be like the one of speakers in a target language setting. Therefore, standardized assessment must be coherent with classroom assessment as well. To achieve this, it is necessary for the Ministry of Public Education (MEP acronym in Spanish) to implement a test that elicits students' English language

proficiency, provides information to address learning gaps, generates evidence of learning successes, and creates more opportunities for students.

The mandate for a new national English exam

As stated in the previous section, the new English curriculum was first put into practice in the year 2017 in alignment with the Educational Policy "Educating for a New Citizenship". This policy promotes the new citizen as an individual who is an active agent of change regarding the student's own learning and, as global citizens. The curriculum conceptualizes the teaching of English through "scenarios". These are real-life situations that provide authenticity in the Costa Rican context: tasks, activities, texts (NEC, p.40). The *Prueba Nacional Estandarizada de Lenguas Extranjeras: Inglés* (Spanish acronym PNELE) will resemble the approach implemented in the English classroom by presenting authentic tasks for students to demonstrate their performance in the target language. Therefore, the new PNELE is aligned with both the demands of the OECD and the approach of the new curriculum, which introduces the Common European Framework of Reference (CEFR) levels starting in seventh grade with A1 moving up progressively to the B1 level. It is important to clarify that the bilingual high schools exit linguistic profile is B2. It is important to state that the PNELE will be framed within three different domains of language use: socio-interpersonal, socio-transactional and academic. Students may find situations in the following contexts in the PNELE: (1) The socio-interpersonal contexts where students will have to understand meanings (negotiate, establish or maintain conversations) in everyday topics. (2) The socio-transactional context refers to transactions done in everyday life (buying things at a store, for example). (3) In the academic context, the students find themselves aiming to acquire propositional information (such as understanding a lesson). The exam claims to measure the semantico-grammatical and pragmatic knowledge of the students at the end of the Diversified Education, in two linguistic modalities or competences Reading and Listening.

The challenge of the Office for the Management and Evaluation of the Quality, (DGEC acronym in Spanish), now is to come up with an exam design in which every student is given the proper cognitive test to demonstrate their general English language ability; students will reveal their capacity to do something by means of the "can do" statements. In other words, the ability they may have developed for interacting or using the English language to be successfully functional in the modern society.

PNELE will be a proficiency exam, not an achievement exam as the one that has been administered before. Furthermore, the new English exam will no longer assess only one competence (reading comprehension) as the Bachillerato test has done so for so many years. Rather, PNELE will contextualize the assessment in three language use domains and do so through two language modalities Listening and Reading. The results will indicate whether the students at the end of the Diversified Education have achieved the linguistic skills by measuring their range of achievement of the English language as established in both the CEFR and the ENC. In this way, the students can demonstrate they have used their knowledge, skills and abilities (KSA's) to demonstrate their mastery of the second language comprehension ability. This new exam will require that MEP face some challenges, such as the delivery of the test (paper-based assessment, PBA or computer-based assessment, CBA).

The paper-based assessment requires MEP to print the tests for every student--this is what has been done up for the administration of the Bachillerato exam. The other type of delivery, as the one PISA employs, would imply a more complex delivery mechanism. If this mechanism is adopted, Costa Rica will need to have as many USB devices as there are examinees as well as a software to download the students' responses, so that they can be scored automatically or else have good connection to the web in case the test being delivered online, or a combination of both depending on resources.

Costa Rica has had experience with delivering tests on computers in the context of PISA. This country has participated in five different PISA cycles whose sample has not exceeded ten thousand participating students, unlike the envisioned exam, where there are approximately sixty thousand students. This huge difference must be considered in terms of getting the necessary material resources. Based on experience administering PISA, another challenge for the PNELE is to make sure schools have enough electrical resources and enough computers. Another aspect to consider is if the platform for the test in CBA will run only in Windows or in Mac or any other operating system. If it is supposed to run in different operating system, the work is doubled since each operating system has different technical requirements. As a result, different platforms must be managed. It is of great importance that some other aspects related to the delivery be considered such as the number of proctors, computers (desktops or laptops), wires, outlet power strips, information technology (IT) specialists to give support before and during the administration. Another important detail to consider is that since the new exam is aiming to cover the Listening modality, the platform on which the exam will be assembled must be able to run audios with good quality for the listening comprehension tasks.

As can be seen, the innovation of the PNELE provides an exciting opportunity for the country since it requires that financial resources will be used in a different way for those used on the Bachillerato exam. This may incur additional resources, especially at the beginning to set up systems. Later it will be more balanced. Since political willingness for such an investment is mandatory, the technical staff at DGEC needs to be able to reassure that funds will be available to pursue this innovative and challenging approach to assessment, especially if PISA is a benchmark for how assessment should be carried out in the country. Aligned to this direction, different stakeholders have reaffirmed the importance given to English; in fact, the teaching of English has been declared a national priority. Support for a Bilingual Costa Rica has been characterized as a "must" by the national authorities. It is indeed noteworthy that the previous administrations have

supported an English competency national policy. President Alvarado, for example, affirmed in August 2018 that, "The challenge is to forge a true inter sectorial alliance to extend the command of English and other languages as a state policy that ensures universal access opportunities for inhabitants of all ages and from all territories" (ABi, 2018). In sum, within the potential direction of PISA, the national advisors, technical support of specialists in the field, in consideration of the enormous political support and importance of the new National English exam, would like to contribute to the national initiative by assessing competences in English in a new and innovative way. This assessment will support the efforts of the curricular reform and will hopefully provide a positive washback effect to support the country in becoming a bilingual Spanish-English country.

In addition, the computer delivery will facilitate accurate data collection and alleviate some of the complications experienced with PBA, since all the students' answers will be compiled in a file extracted from the delivery system. Furthermore, if the test were delivered by paper, considerable resources would have to be devoted to data entry for analysis, which is not cost effective and prone to human error. Time and other resources would be better spent in performing analyses of the data and in creating systems for automatizing processes such as rating performance samples in a near future.

The *Prueba Nacional Estandarizada de Lenguas Extranjeras: Inglés* will provide concrete information on the level of mastery of the English language, and more importantly, will reveal what linguistic abilities the new citizens have in an ever-changing world, where the use of technology is required to be successful implementing real world competencies. The aspiration is that new citizens will become more responsible for their own learning as aligned to their own dreams of being able to communicate in a foreign language in the workplace and in their personal lives.

The purpose of the conceptual framework document

This conceptual framework is intended to outline considerations for guiding research, development, and validation for the PNELE. The framework provides a space to articulate the sociocultural, political, and economic context of the test, especially in consideration of the OECD's influence in the Costa Rican's national exam's (Bachillerato) history and direction in the country. The document serves to guide the revision of the previous national assessment, Bachillerato, for English as a second language, which assesses examinees' reading comprehension skills at the end of the Diversified Education. One of the main goals of this framework is to inform stakeholders about the elements that ground the design of a new high-stakes English exam. In support of this idea, the definition of classroom and large-scale assessment to inform the educational community about the similarities, but mostly the differences about these two types of assessments. In the same vein, the theoretical approach to construct definition is unpacked within the theoretical framework.

The context of the assessment

Evidence presented in previous sections of the current document maintain that there is an explicit need to re-think the English language exam in Costa Rica. The previous Bachillerato English exam assessed only one language modality: reading. In addition, this exam is conceptualized as an achievement test aligned to a curriculum that is now outdated with regard to contextual considerations, approach to teaching, and theoretical conceptualization of proficiency for English Language Learners (ELLs). Another important element to consider is the exam design. Bachillerato uses a norm-reference approach to analyzing examinees' performance, where examinees are provided with a single score linked to a pass/fail decision. Data from the test has not been used to provide explanations about what examinees could or could not do or what they might need to improve.

The mandate for a new National Exam of English proficiency lies in the following points. First, the OECD has recommended that the Bachillerato exams need to be revised to reflect the curricular reform. Second, the NEC overarching approach to English language instruction is Action-Oriented Approach (Piccardo, 2014), where ELLs are considered social agents empowered by their own learning process; thus the new exam needs to reflect this position. In addition, this approach emphasizes activities and tasks that emulate authentic language use situations, so that ELL's can perform language tasks that examinees are likely to encounter in the real world. And third, classroom assessment in the new curriculum was changed to be aligned with AOA where its mission is to foster relatable macro-contexts for students, so that teachers could design test tasks to assess the examinees' knowledge, skills, and abilities (KSAs) based on their performance within a specific language use domain using different language modalities or competences.

Highlights of the 2016 National English Curriculum

Speaking English fluently is one of the competencies a 21st century learner must develop to have access to better life opportunities. The Costa Rican educational system is committed to achieving the goal of having bilingual citizens in two or more languages by means of a comprehensive, articulated curriculum from kindergarten through high school. Such an achievement will foster a historic change in the formation of our country, thereby recognizing English proficiency as a resource for creating opportunities, enhancing employability and driving competitiveness. To achieve such an ambitious goal, the National English Curriculum for Secondary Education has been sequenced so that learners reach a minimum level of English proficiency of either B1 or B2 (depending on the study plans, the B2 curriculum for bilingual high schools) when completing secondary education at the Diversified Education, based on the levels described in the CEFR.

The proficiency standards outlined in the CEFR will be implemented in the English language curriculum throughout Costa Rica's grade levels. This implementation began in 2017, starting with the first grade of primary school and seventh grade of the secondary

education. The goal is for learners to progress steadily, aiming to reach level A2 by the conclusion of the Second Cycle and level B1 upon completion of Diversified Education by 2021.

To further detail the CEFR proficiency levels and their corresponding performance indicators for each educational cycle, the framework outlines what learners should be capable of achieving in their communication using the target language. Aligning curriculum, teaching methods, learning strategies, and assessments with classroom practices is envisioned to facilitate the achievement of this goal. Moreover, adjustments to the national exam are intended to address this new approach.

Highlights of the English Bachillerato Exam

Since 1982, the Costa Rican government has used different models to administer various versions of the national exam. The first attempt at designing an assessment system that would certify high schools' graduates in English took place in 1960, when these exams were paper-based administered. By 1970, the national exams assembly process was entrusted to teachers, which resulted in every school creating their own version of the exams, without robust research and analysis that grounded the exam design in a validation framework. In 1973, the Ministry of Public Education suspended the national exams, in response to claims of inequity, lack of standardization and low promotion rates. In 1985, a new public debate arose concerning the need to establish curricular control in secondary education. However, it was until 1988 that Costa Rica adopted the Bachillerato exams, which included a second language exam, in either English or French.

Given the need to provide empirical evidence that supported the validity of the Bachillerato exam, back in 1994, a group of national assessment advisors created a theoretical framework meant to foster the development of Bachillerato exams in nine subjects: namely, Civic Education, Social Studies, Mathematics, Spanish, Biology, Chemistry, Physics, English, and French. This framework adopted the norm-referenced

approach in the interpretation of examiner scores. This approach is used to determine the relative position of a student's score within a group of examinees. A standardized scale of 0 to 100 percentile was established, and a minimum cut score was set. Over the years, the DGEC has been in charge of guiding test development, designing test tasks, administering the exam, scoring responses, analyzing the data, and reporting the results of the Bachillerato exams.

The English Bachillerato exam has been designed as an achievement test, aligned with the topics and objectives specified in the previous English curriculum. However, given the implementation of the new National English Curriculum that has considered international standards in its conceptualization, the Ministry of Public Education, through the DGEC, has decided to create a new English proficiency exam that will also align with the levels of the CEFR, and with the KSAs specified in the policy "Educating for a new citizenship," which describes the proposed exit profile for students at the end of the Diversified Education.

The next section of this framework addresses the design of the new English exam as a reflection of the KSAs required to meet competencies in different language use domains.

Design of the Prueba Nacional Estandarizada de Lenguas Extranjeras Inglés

Given the mandate that the Costa Rican citizenship must become competent at the B1 level of English as a second language, the National English advisory at the Curricular Division has joined forces with other stakeholders to create and implement a new curriculum designed to teach students to communicate in specific language use domains using the four language modalities. On the other hand, the National English advisory at DGEC conducted research to create and administer a new English exam designed to measure the semantico-grammatical and pragmatic language knowledge, in three target language use domains across different language modalities and delivered by computer.

Since the main goal of this exam is to measure the examinees receptive skills in different domains, the PNELE is designed to measure the Listening and Reading language modalities. Furthermore, to achieve such an ambitious goal, the overarching framework that will ground the design of this new exam will be a Learning-Oriented Assessment (LOA) approach (Purpura, 2020) because it takes into consideration several features of the assessment event that other frameworks do not account for. On the other hand, the Meaning-Oriented Conceptualization of Language Knowledge (MOCLK), and the performance levels according to CEFR. The LOA approach will be the framework that operationalizes the construct around the contextual, elicitation and proficiency dimensions of the assessment.

Theoretical Framework

In this section of the document, it will be unveiled the theoretical grounding of the exam as well as the expectations of the new end-of-cycle National Exam of English Proficiency in Costa Rica. As the name of the new exam indicates, the main purpose is to assess the L2 proficiency of examinees at the end of the Diversified Education. With respect to this claim, Carr (2011) states that a proficiency test assesses the language ability level of examinees "without respect to a particular curriculum" (p.8). Therefore, while the Bachillerato exam was aligned with the objectives and topics of the previous curriculum in Costs Rica, it was not designed to measure the L2 proficiency of students. For the new exam though, alignment comes through the performance levels of the CEFR proposed by the curriculum. Thus, the purpose of the new exam is to target what test-takers can do when exposed to simulated real-life tasks. The PNELE results are thus used to make decisions with reference to the CEFR standards at the proficiency level. In this respect, the English exam will elicit language performance from examinees in two language modalities (reading and listening) and in three language use domains (socio-interpersonal, socio-transactional and academic).

Assumptions regarding the redesign of the 2021 National Exam of English Proficiency

The mandate for a reconceptualization of the English Bachillerato exam, which only assesses reading comprehension, demanded that the national English advisors at the DGEC conduct training sessions with international experts about large-scale assessment specifically in the area of second language assessment, as well as theoretical discussions around the topic of the redesign of the English exam. Based on this training, it was agreed that for the design of the new English exam, there must be a robust research process associated with all aspects of the NEEP development and validation of scores.

The design assumptions consider the following processes:

- First, considering the purpose of the test and the types of decisions to be made, another variable aligned with the type of exam is to consider the framework to interpret results either using norm-reference or criterion-reference testing aligned with the design and interpretation of the scores of the exam.

- Second, the exam will define the construct with reference to a model of L2 proficiency and then will operationalize the construct by choosing a framework that acknowledges the synergies among the critical variables mandated by the PNELE.

- Third, the exam will be delivered by computer to take advantage of innovative design options and to facilitate the collection and scoring of data.

- Fourth, the scoring system will contemplate aspects such as developing score systems, scoring in Classical test theory, scores transformations, Item-response theory, among other technical considerations, that will be also used as parameters to design the item banking.

- Fifth, reporting scores will contemplate including not only the exam scores but also indicating the corresponding level on the Common European Framework of Reference (CEFR) and detailing the skills and abilities (Can dos) the test-takers will demonstrate in the PNELE.

- The exam data will be submitted to a range of analyses as part of a coherent validation argument.

In sum, the redesign of the English exam must consider empirical and scientific evidence to support the validity of the instrument. These validation efforts are especially important given that PNELE is a high-stakes exam that will be used to make critical decisions about language learners, English teachers, school administrators, regional and national advisors, and in general the country's educational direction.

Conceptualizations of assessment

This section defines L2 assessment in terms of both classroom-based and large-scale assessment. It being by defining the different elements used in evaluation (measurement, assessment, testing), as well as two important factors related to testing: validity and reliability. In like manner, it discusses the description, similarities, differences, and relevance of classroom-based and large-scale assessment in the improvement of the students' L2 acquisition to exhibit proficiency. Finally, the main goal of the English exam is unpacked to wrap up this section of the framework.

Kizlik (2014) stated that, in the field of education, it is extremely important to understand the concepts of evaluation elements such as measurement, assessment, and testing. Kizlik (2014) concluded that,

> Assessment in educational settings may describe the progress that students have made towards a given educational goal at a point in time. However, it is not concerned with the explanation of the underlying reasons and does not offer recommendations for actions. In other words, assessment is the process of organizing measurement data into interpretable forms. (p. 35)

Aligned with the same idea, assessment refers to all the processes and products used to describe the nature and the extent of students' learning. Moreover, assessment is a process meant to gather information derived from objectives or goals. Another conclusion is that assessment is a broad term that includes testing; this means that testing is another form of assessment. Based on Kizlik (2014) "tests are assessments, but not all assessments are tests". For instance, in the classroom, students are given a test at the end of a lesson or unit to obtain information about their learning progress. Another example of assessment is testing students' achievement or proficiency which are also assessed at the end of a school year.

As mentioned at the beginning of this section, there are two relevant factors in relation to testing which are called validity and reliability. Thus, validity describes the extent to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests (AERA, APA, & NCME, 1999). Reliability describes the stability of measures over factors that should be irrelevant to the measure (time, forms, raters, etc). Implied in these conceptions is a sense that the measure itself is in some ways external to the inferences or actions. As stated by Brookhart (2004), the test functions as a "dipstick" into the "oil tank" of the student's achievement. The validity goal is a meaningful inference about student performance and/or effective use of that information for a specified purpose.

In this conceptualization of validity and reliability, there are theories about the validity and the reliability of test scores, (Fulcher & Davidson, 2009, pp. 23-25). Figure 2 includes the differences in validity and reliability between the classroom and large-scale assessments. The following features about the concepts of assessments must be considered by any tester in the field of evaluation.

Table 2 Contrasting Large-Scale and Classroom Assessment Concepts

Concepts in Large Scale Assessment	Concepts in Classroom Assessment
Validity: ● The measure is external to the inferences made and actions taken. ● Students are "subjects" upon whom observations are made. ● The validity goal is a meaningful inference about student performance and/or effective use of that information for a specified purpose.	**Validity:** ● Inferences made and actions taken are internal to the measurement process. ● Students are observers jointly with teachers; "those measured" make the inferences and take the actions in the formative assessment process. ● Students' awareness of and benefit from assessment information are part of the "information" itself. ● The validity goal is an understanding of how students' work compares to "ideal" work (as defined in the learning objectives) and/or effective use of that information for further learning.
Validity: ● The measurement context is construct irrelevant. ● Content specifications describe a domain. ● Administration can be standardized. ● Scores can be equated or linked across contexts and forms of assessment.	**Validity:** ● The measurement context is construct relevant. ● Assessment is part of instruction. A good assessment is an "episode of genuine learning." ● Content specifications reflect both the domain (learning objectives) and instruction (modes, activities). ● Teacher beliefs, teacher instructional practices, and teacher understanding of both the subject matter and students (including cultural and linguistic differences) are relevant validity concerns.
Reliability: ● Reliability is consistency over irrelevant factors.	**Reliability:** ● Reliability is sufficiency of information.

● Occasions, time, items and/or tasks are facets of error variance. The reliability goat is stable ranking of students on a score scale (NRT) or stable categorization of students.	● The reliability goal is stable information about the gap between students' work and "ideal" work (as defined in students' and teachers' learning objectives).

Source: Brookhart, S. (2003).

Now that the validity and reliability concepts have been defined in this section, for the purpose of the design of the *Prueba Nacional Estandarizada de Lenguas Extranjeras: Inglés,* it is important to conduct research to gather information that grounds the validity arguments underlying the design of the new English exam. Similarly, to guarantee the exam's reliability, it is necessary to design protocols and guidelines to make sure that the standardization of the test administration is consistent for all the examinees.

Given the importance of classroom-based and large-scale assessment in the educational system, it is important that all the stakeholders involved in the educational community be informed about the characteristics that ground both types of assessments. Fulcher & Davidson (2009) stated that "Testing and assessment are part of modern life. Schoolchildren around the world are constantly assessed, whether to monitor their educational progress, or for governments to evaluate the quality of their school systems" (Fulcher & Davidson 2009, p. 30). In the same way, they also say that "understanding what we test, how we test and the impact that the use of tests has on individual, and societies cannot be overstated" (p. 22). Other experts, cited by Sian Morgan (2008), claim that "The challenge for exam providers is to align all level of assessment, large-scale, institutional and classroom, and link these to policy-related goals; so that coherent and comprehensive plan can be implemented within an education system" (p.76) These authors also suggest to have "a broader view of testing impact" and at the same time they warn testers to diminish the negative impact of large-

scale assessment on the target population but to maximize positive impact on them (Saville & Salamoura, 2014, p.57).

According Stroynoff & Chapelle, there are different types of tests according to the decision level the exam is designed for. Following the same idea, Figure 3 presents the continuum that enlists the type of tests starting with low-stakes and following gradually to high-stakes.

Figure 2 Test Decisions about Individuals

Source: Stroynoff & Chapelle, cited in Purpura, 2017

In sum, there have been different important changes in the National English Curriculum; therefore, there is an explicit need to design a different type of exam at the national level. Based on this need, DGEC authorities have decided to design a computer-based test that uses the CEFR performance levels as the standard.

In the next section, this framework describes the theoretical approach to construct definition, as well as a brief description of its dimensions.

Theoretical Approach to Construct Definition

In the previous section the differences between classroom and large-scale assessment were discussed. It was documented that the processes and the stakes involved in large scale evaluations contemplate variables such as the rigor in the test design, the robust research aligned to the creation of test tasks and most importantly the decisions derived from the scores. Now that this has been established, it is important to highlight that for the design of the PNELE, it is necessary to select a framework that contextualizes the assessment event. There are different approaches to define test constructs; however, for the purpose of the new English exam the LOA (Purpura 2004, 2009; and more recently Purpura and Turner, forthcoming) is the framework that accounts for the synergies among instruction, learning   and assessment. LOA provides the theoretical ground for the basis of meaningful interpretation of performance consistencies as well as for assessment design, operationalization, interpretation, and use (Purpura, 2016, p.193).

Several frameworks have been proposed to guide the design, development, and validation of assessment protocols. Mislevy et al (2003) proposed evidence centered design as a framework for specifying assessment tasks as they relate to the quality of evidence in the response for measuring test claims. Bachman and Palmer (2010) specified a framework of test task characteristics designed to control the effect of test method on the measurement of performance consistencies. Moving beyond the specification of how tasks are designed to elicit language performance, Turner and Purpura (2016) contextualized the elicitation of L2 proficiency with a broad framework that accounted for the synergies among assessment, learning and instruction. More specifically, this framework not only specified the elicitation of language performance as indicators of proficiency, but also specified potential moderators of performance through the specification of other dimensions in the assessment event, namely the contextual dimension, the socio-cognitive or learning dimension, the affective dimension, the socio-interactional dimension in interactive speaking tests, and in assessments that provide examinees with information or assistance, that is, the

instructional dimension of assessment. As the new curriculum recognizes how each of these dimensions play a role in assessment, the new assessment will use this learning-oriented framework to guide the design development and validation of the new national test.

Turner and Purpura's (2016) and Purpura's (2020) LOA framework accounts for seven related dimensions: contextual, elicitation, proficiency, socio-cognitive, affective, social-interactional, and instructional. These dimensions correspond to the notion that L2 learning is an individual cognitive process as well as a social one when considered in the classroom context that makes it a highly intricate socio-cognitive and sociocultural process (Turner & Purpura, 2015, p. 6). Its final goal is facilitating best L2 practices for teaching (Turner & Purpura, 2015), and for any purposes in assessment. Each dimension is described as follows and will ultimately need to be specified in the assessment design.

1. **The Contextual Dimension:** This pertains to the context of a particular situation or task that a learner is encountering. It offers a comprehensive description of a situation or hypothetical scenario, intending to closely represent a real-world task. The goal is to engineer the task with an authentic context that aligns with what the student will face. As outlined by Purpura (2020), the contextual dimension emphasizes the connection between an assessment task and a real-world scenario.

   There are different elements or characteristics that must be taken into account in this dimension, and these elements should be reflected in the task or scenario that the student is going to undertake. Depending on the competency being assessed, these characteristics may vary.

   In the Contextual dimension of test design, it is essential to consider specific features to effectively shape the context of the task. These features play a crucial role in engineering the overall context for the task. Purpura (2020) describes the following features:

Topic of Communication: This entails the overarching theme or subject matter of the task. For instance, it could revolve around vacation plans, family dynamics, or the context of grocery shopping.

Real-Life Communication Goal: This aspect describes the linguistic objectives within a given context. For instance, students may be tasked with asking for specific information, engaging in information exchange, or persuading someone to take particular action.

Roles: It identifies the participants involved in the tasks.

Communicative Events: This refers to specific situations or instances where communication transpires between individuals. Such events encompass the exchange of information, ideas, messages, or expressions among participants.

Scenario Setting: This component elucidates the location and circumstances surrounding the designated task. Understanding where and how the task unfolds contributes to a more realistic scenario.

Language Use Domain: This revolves around the specific context or environment in which language is utilized for communication. It considers the practical application of language in real-life situations and can be categorized based on social, professional, or academic contexts. These language use domains are described as follows:

Table 3 Language Use Domain

Language Use Domain	
Socio-Interpersonal	The language user can maintain interpersonal connections with peers, family, school faculty, community members, among others by addressing topics such as talking about personal experiences, exchanging information with faculty members, and getting to know the social conventions to interact in the community context, among others.

Socio-Transactional	The language user can conduct transactions in diverse social contexts to clarify or obtain information about a good or service, from peers, family, school faculty and community members.
Academic	It is concerned with the learning/training context (generally of an institutional nature) where the aim is to acquire specific knowledge or skills.

Adapted from: Bachman, L., & Palmer, A. (2010)

The following task sample exemplifies each of the characteristics described above in a specific task.

Table 4 Task Specifications

CEFR level: A1
Competency: Listening

Topic of communication	Places around town
Real-Life communication goal	To ask for information to look for a specific place in town.
Roles	Two strangers (A man and a woman)
Communicative event	Interacting, asking for and giving information, locating places and buildings.
Scenario setting	A casual conversation in the street
Language Use Domain	Socio-transactional

In essence, the key components within the Contextual Dimension facilitate the design of tasks and assessments that resemble real-life situations. By considering sociocultural

contexts, real-world topics, scenarios, and competencies, learners can develop tasks that closely resemble authentic communication experiences.

2. **The Elicitation Dimension:** It focuses on the methods used to draw out responses from examinees to assess their language proficiency. This dimension serves as a blueprint, guiding the creation of tasks that effectively measure the desired competencies or modalities. (Purpura, 2016)

The choice of elicitation methods depends on the specific language skills being targeted and the goals of the language assessment, ensuring alignment with the intended language proficiency levels and objectives. Key considerations within this dimension include:

Delivery Procedures:
- Refers to how the test is presented to examinees.
- Includes options like paper-delivered or digitally delivered tests.

Types of Items:
- Involves a variety of items, categorized into two main types: Selected-Response (SR) and Constructed-Response (CR) tasks (Purpura and Dakin), 2019). SR tasks present items where examinees choose an answer from two or more options. CR tasks require examinees to produce a single word, a sentence, or language varying in quantity.

Task Scoring Procedures:
- Involves systematic and standardized methods used to evaluate and assign scores to individuals' language task performances.
- Designed to ensure objectivity, reliability, and fairness in the assessment process.

By understanding and incorporating these aspects, language assessments can be effectively designed and administered, promoting accurate evaluations of examinees' language proficiency.

3. **Proficiency Dimension:** It refers to the level of competence that a learner has acquired in a particular subject. Particularly, proficiency in learning a second language (L2) refers to the specific language-related knowledge, skills, and abilities (KSAs) that learners need to effectively communicate in a particular language context or domain. It is a key measure of how well someone can use the language in real-life situations.

   Proficiency can be assessed using models like L2 proficiency models, standards such as the Common European Framework of Reference (CEFR) descriptors, or the outcomes of specific units or courses. There are various theoretical models that describe L2 proficiency. In the context of the PNELE, the Meaning-Oriented Conceptualization of Language Knowledge (MOCLK) will be utilized to assess the examinee´s performance in the L2 knowledge. According to Purpura (2016), this model highlights the importance of both literal and implied meanings in communication as essential tools for using language effectively in real-life situations. Additionally, the MOCLK considers grammatical forms and their semantic meanings as crucial resources for expressing and comprehending propositions.

   Grammar Definition

   Traditionally, a definition of grammar refers to the system of rules and structures that govern the organization and use of language, in foreign language acquisition (FLA). These rules dictate how words are combined to form meaningful sentences and how those sentences are structured. Grammar encompasses various aspects, including syntax (word order and sentence structure),

morphology (word formation and inflection), phonology (sound patterns and pronunciation rules), and semantics (meaning). In the context of FLA, learners typically acquire grammar through exposure to the target language, instruction, and practice. This process involves developing an understanding of the rules and patterns of the language and applying them in communication.

Over time, various scholars have proposed different interpretations of grammar which have shaped language learning, teaching, and assessment methodologies. Linguists have presented two main perspectives: the syntactocentric view and the communicative perspective of language. The syntactocentric approach considers formal grammar as "a systematic way of accounting for and predicting an 'ideal speaker' or hearer's knowledge of language" (Purpura, 2004). This definition underscores a structured understanding of grammar and its components, analyzing sentences from the smallest units—sounds forming words, words combining into phrases, phrases forming clauses, and ultimately clauses composing sentences. Essentially, this perspective focuses on the grammatical form and the rules governing it.

On the other hand, the communicative perspective views language as a system of communication wherein a speaker or writer uses grammatical forms to convey several meanings (Purpura, 2004). Under this viewpoint, grammar is seen as one of the tools for achieving communicative goals; however, it serves to convey meaning not only within individual sentences but also extends to broader contexts. Essentially, this perspective emphasizes the overarching message being conveyed and the potential interpretations it may elicit.

The critical construct of PNELE is the mastery of the semantico-grammatical and pragmatic knowledge of English as a foreign language in the socio-interpersonal, socio-transactional, and academic target use domains upon completion of Diversified Education in two language modalities: Listening and Reading. To

achieve such an ambitious objective and make inferences from the results about grammatical ability based on the test posited by MEP, the Meaning-Oriented Conceptualization of Language Knowledge, along with other elements, provides a theoretical basis for the operationalization of the construct.

Meaning-Oriented Conceptualization of Language Knowledge

The fundamental characteristic of successful communication lies in the ability to express, comprehend, co-construct dynamically, and repair varied meanings in a wide range of language use contexts (Purpura 2004). A second language user utilizes language to get things done or interact in everyday activities such as socializing at school, communicate in social media, request information about a product in the supermarket, among other activities. In other words, language users utilize their linguistics resources to convey a wide variety of meanings in socio-interpersonal, socio-transactional, and academic contexts. From this point, understanding and conveying meaning play a fundamental role in the assessment of a second language.

According to Purpura, the Meaning-Oriented Conceptualization of Language Knowledge (MOCLK) consists of two interconnected mental constructs: semantico-grammatical knowledge and pragmatic knowledge, both interrelated at the communication level. The semantico-grammatical knowledge refers to how the language user utilizes both the grammatical knowledge of forms and their associated semantic meanings and their ability to use these forms together to communicate literal propositional meanings or topical meanings.

In Figure 3, the MOCLK is a theoretical framework proposed by Dr. James Purpura. It emphasizes the importance of meaning in language learning and use. In MOCLK, language knowledge is viewed as more than just knowing grammar rules and vocabulary; it involves understanding and effectively using language to convey and interpret meanings in various contexts. This conceptualization

highlights the interconnectedness of different aspects of language knowledge, including semantics, grammar, pragmatics, and discourse, to achieve communicative competence.

The MOCLK emphasizes the dynamic and multifaceted nature of language use, where meaning is central to effective communication. Next, the MOCLK is described, step by step, to understand how this conceptualization of language knowledge will help operationalize the construct of the PNELE, starting with the conceptualization of semantico-grammatical knowledge.

Figure 3 Meaning-Oriented Conceptualization of Language Knowledge.

Meaning Oriented Conceptualization of Language Knowledge

Semantico- Grammatical Knowledge

Pragmatic Knowledge

Knowledge of Grammatical Forms & Meanings

Knowledge of Grammatical Forms

Knowledge of forms & their associated literal sentential & discourse levels

Phonological & graphological, Lexical Morphosyntactic Cohesive Interactional Informational management Interactional forms

Propositional or Topical Knowledge

Ability to combine forms & meaning into propositions to express, interpret, co-construct, & negotiate literal propositional content or topical meaning in messages & texts

Use of S-G resources, along with topical knowledge, contextual factors, & socio-cognitive abilities to express, interpret, co-construct or negotiate. Literal propositional or topical content

Linguistic meaning Utterance/sentential meaning Subject-matter literality Context-free literality

Functional Knowledge

Ability to use propositions to express/interpret agency or communicative intentionality.

Use of S-G resources, along with topical knowledge, contextual factors, & socio-cognitive abilities to express, interpret, co-construct or negotiate one´s agency & intentionality in performing direct or indirect communicative acts or language functions.

Propositional or Topical Knowledge

Ability to combine forms & meaning into propositions to express, interpret, co-construct, & negotiate literal propositional content or topical meaning in messages & texts.

Use of S-G resources, along with topical knowledge, contextual factors, & socio-cognitive abilities to express, interpret, co-construct or negotiate the following contextually linked implied meaning:

-situational
-sociolinguistic
-psychological
-sociocultural
-literary
-rhetorical
-interactional

Source: Purpura, J.E. Assessing Meaning-Teachers College, Columbia University, New York, NY, USA.

Knowledge of Grammatical Forms and Meanings

As it was discussed before, the MOCKL posited the knowledge of grammatical forms and meanings with a focus on how these grammatical resources are used accurately and meaningfully. As Purpura points out, although assessments solely focused on grammatical structures are no longer deemed comprehensive indicators of proficiency, numerous L2 evaluations end up primarily assessing semantico-grammatical understanding, encompassing both forms and meanings (Purpura, 2018). This delineates language competence in terms of comprehension and utilization of grammatical structures and their corresponding meanings, emphasizing accuracy and meaningful application, exemplified by tests such as cloze exercises or graded paragraphs where only forms and meanings are evaluated.

Knowledge of grammatical forms considers linguistic features at the (sub) sentential level; for example, Phonological: Phonetics (sound system), Graphological: handwriting especially for the purpose of character analysis, and at the discourse level such as cohesive: related to logical connectors, Informational management, and Interactional forms, as can be seen in Figure 2. In terms of assessment, these forms have been measured in terms of accuracy or precision, range or complexity (Purpura 2004).

Thus, within this initial pillar of the MOCLK, Figure 4 presents how grammar considers forms and meanings, while pragmatism is a separate area, yet a highly related component. On the other hand, grammatical knowledge combined with strategic competence makes up grammatical ability, which considers the ability to carry out grammatical knowledge accurately and meaningfully when taking exams or performing other linguistic tasks, as posited by Purpura. The ability to access grammatical knowledge to understand and make oneself understood assists the examinee in implementing their grammatical ability in language use. In this sense, to make inferences about the grammatical abilities of examinees in the PNELE, it is important to understand what grammar refers to and to specify the components of grammatical knowledge for the purpose of measuring the proposed construct.

Figure 4 Purpura's Meaning-Oriented Conceptualization of Language Knowledge
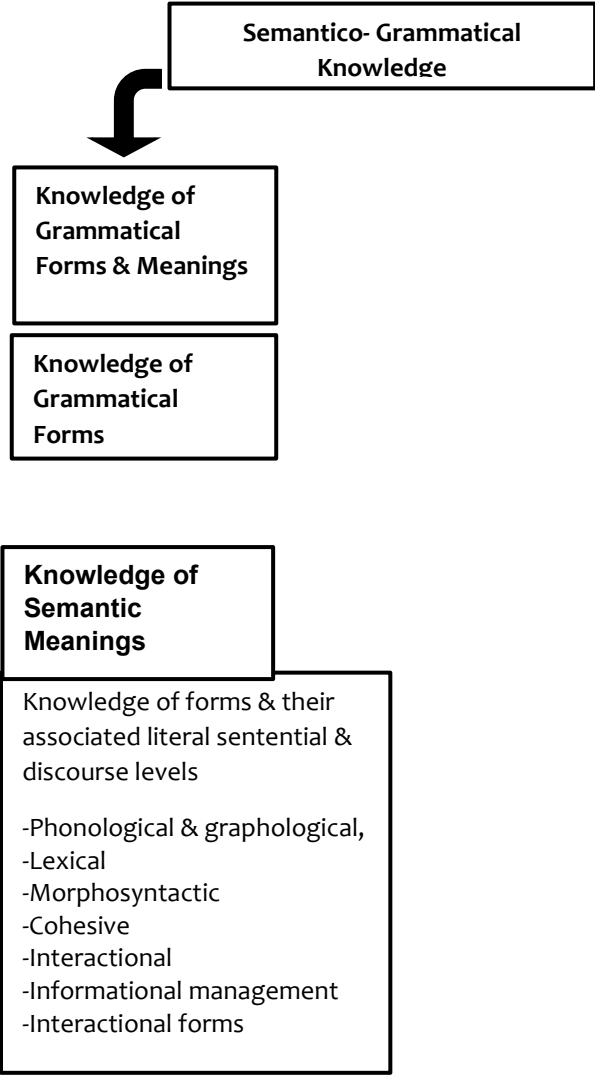
```
                        ┌─────────────────────────┐
                        │  Semantico- Grammatical │
                        │        Knowledge        │
                        └─────────────────────────┘
        ┌───────────────────────┐
        │  Knowledge of         │
        │  Grammatical          │
        │  Forms & Meanings     │
        └───────────────────────┘
        ┌───────────────────────┐
        │  Knowledge of         │
        │  Grammatical          │
        │  Forms                │
        └───────────────────────┘


        ┌───────────────────────┐
        │  Knowledge of         │
        │  Semantic             │
        │  Meanings             │
        └───────────────────────────────┐
        │  Knowledge of forms & their    │
        │  associated literal sentential &│
        │  discourse levels              │
        │                                │
        │  -Phonological & graphological,│
        │  -Lexical                      │
        │  -Morphosyntactic              │
        │  -Cohesive                     │
        │  -Interactional                │
        │  -Informational management     │
        │  -Interactional forms          │
        └────────────────────────────────┘
```

Source: Purpura, J.E. Assessing Meaning-Teachers College, Columbia University, New York, NY, USA.

According to Purpura, grammatical knowledge, in turn, is subdivided into two components: grammatical forms and meanings. Grammatical forms refer to linguistic forms at the sub-sentential, sentential, and suprasentential levels as derived from the aforementioned syntactocentric approach (Purpura, 2004). On the other hand, Purpura defines grammatical meanings as "the knowledge of the meaning associated with an utterance as the sum of its parts and how these parts are arranged in syntax (literal

meaning) as well as how these parts are used to convey the speaker's intended meaning in context (intended meaning)." (Purpura, 2016)

Based on the definition of grammar, and the following Figure 5, Purpura proposes the following framework as an illustrative guide for defining the grammatical construct. If a continuum is drawn, or rather, an imaginary line that goes from situations with little context to situations that require higher context to understand the message conveyed, through this path it can be observed how language users utilize their strategic competence to convey grammatically correct messages or accurate messages, then provide meaningfulness to the message or understand the meaning implied and finally based on sociocultural constructs conveyed appropriateness in their interactions (Purpura, 2004). Even though the PNELE is not meant to measure all the categories in this grammatical framework, it does provide a starting point with suggested grammatical points that could be used to measure each component.

Figure 5 Components of grammatical and pragmatic knowledge. (Purpura, 2004, page 90)

Grammatical knowledge ⟷		Pragmatic knowledge
Grammatical form (accuracy)	**Grammatical meaning (meaningfulness)**	**Pragmatic meanings (appropriateness/ conventionality/naturalness/ acceptability)**
SENTENTIAL LEVEL	**SENTENTIAL LEVEL**	**SENTENTIAL OR DISCOURSE LEVELS**
Phonological/ graphological form • segmental forms • prosodic forms (stress, rhythm, intonation, volume) • sound-spelling correspondences • writing systems	**Phonological/ graphological meaning** • minimal pairs • interrogatives, tags • emphasis/contrast • homophony (they're, there) • homography (the wind, to wind)	**Contextual meaning** • interpersonal **Sociolinguistic meaning** • social identity markers (gender, age, status, group membership) • cultural identity markers (dialect, nativeness)
Lexical form • orthographic forms • syntactic features and restrictions (nouns) • morphological irregularity • word formation (compounding, derivational affixation) • countability and gender restrictions • co-occurrence restrictions (*depend on, in spite of) • formulaic forms	**Lexical meaning** • denotation and connotation • meanings of formulaic expressions • meanings of false cognates • semantic fields (attributes of words denoting physical attractiveness) • prototypicality (words denoting physical attractiveness) • polysemy (head of person/bed/table) • collocation (table and chair)	• social meanings (power, politeness) • register variation and modality (registers in speaking, writing) • social norms, preferences, and expectations • register variation and genres (academic, ESP) **Sociocultural meaning** • cultural meanings (cultural references, figurative meanings, metaphor)
Morphosyntactic form • inflectional affixes (-ed) • derivational affixes (un-) • syntactic structures (tense, aspect) • simple, compound and complex sentences • voice, mood, word order	**Morphosyntactic meaning** • time/duration • reversive (pack/unpack) • interrogation, passivization • cause–effect, factual/counterfactual	• cultural norms, preferences, and expectations (naturalness, frequency and use of apologies, formulaic expressions, collocations) • modality differences (speaking, writing)
DISCOURSE OR SUPRASENTENTIAL LEVEL	**DISCOURSE OR SUPRASENTENTIAL LEVEL**	**Psychological meaning** • affective stance (sarcasm, deference, importance, anger, impatience, irony, humor, criticism, understatement)
Cohesive form • referential forms (personal, demonstrative, comparative) • substitution and ellipsis • lexical forms (repetition) • logical connectors (therefore) • adjacency pairs	**Cohesive meaning** • possession, reciprocity • spatial, temporal or psychological links • informational links to avoid redundancy • additive, contrast, causal	**Rhetorical meaning** • coherence • genres • organizational modes
Information management form • prosody • emphatic "do" • marked word order (clefts) • given/new organization • parallelism	**Information management meaning** • emphatic meaning • focal meaning • contrastive meaning • foregrounding	
Interactional form • discourse markers (oh, ah) • communications management strategies (turn-taking, repairs, fillers, paraphrase, word coinage)	**Interactional meaning** • disagreement, alignment, hedging • keeping the conversation moving, interruption • repair by clarification	
Low to high context		**High context**

Source: Purpura, J.E. Assessing Meaning-Teachers College, Columbia University, New York, NY, USA.

Propositional or Topical Knowledge
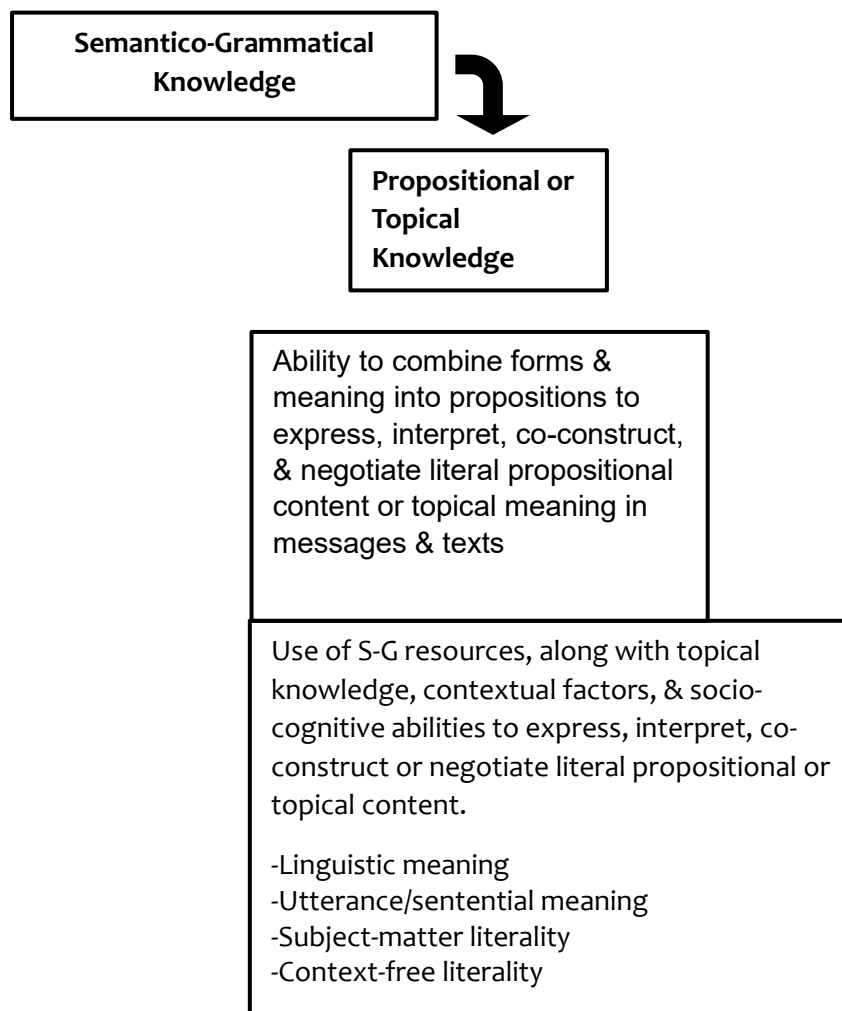
The other mental asset posited by the MOCKL, within the Semantico-grammatical component, is the propositional or topical knowledge. In this knowledge the language user comprehends the grammatical forms and their semantic meanings, employing these forms to convey and understand meanings at a literal level through propositions, as depicted in Figure 6. The language user integrates elements of grammar, vocabulary, and cohesion to articulate propositional content and interpret messages at a propositional level (Purpura, 2004).

Figure 6 Purpura's Meaning-Oriented Conceptualization of Language Knowledge

```
┌─────────────────────────────┐
│   Semantico-Grammatical     │ ⟶
│        Knowledge            │
└─────────────────────────────┘
              ┌──────────────────────┐
              │ Propositional or     │
              │ Topical              │
              │ Knowledge            │
              └──────────────────────┘
```

Ability to combine forms & meaning into propositions to express, interpret, co-construct, & negotiate literal propositional content or topical meaning in messages & texts

Use of S-G resources, along with topical knowledge, contextual factors, & socio-cognitive abilities to express, interpret, co-construct or negotiate literal propositional or topical content.

-Linguistic meaning
-Utterance/sentential meaning
-Subject-matter literality
-Context-free literality

For example, a reading passage describing the scenario of a car accident, and it is mentioned the expression 'break a leg', automatically the reader interprets it as an action that might result in a severe fracture in someone's leg. However, if the context were a school play and the sentence indicates 'The teacher asked us to break a leg at the annual day dance.' In the latter context though, 'break a leg' is used as an idiomatic expression meaning good luck. Purpura argues that language users must resort to contextual clues, including the speaker's or writer's intentions, to interpret the meaning of an utterance in relation to a real or possible situation. (Purpura, 2004)

Building on the previous example, the communication of messages and their meanings are heavily influenced by the intention of the interlocutor, as elucidated by Purpura in the MOCKL. Disregarding this propositional content when assessing L2 proficiency is like having language ability without anything to say. Therefore, within the framework of the PNELE, this propositional knowledge will be assessed through comprehension tasks, encompassing linguistic competence in both reading and listening comprehension.

Functional Knowledge

Following the discussion of the MOCKL, in terms of pragmatic knowledge, the language user utilizes the linguistic resources that have been discussed so far, such as grammatical forms and their semantic meanings, context, among others, that in turn constitute the foundational resources to convey meaning. In this section of the framework, Purpura posits that L2 users can express meanings beyond what is explicitly stated by resorting to elements such as indexical, presuppositions, situational, and cultural implicatures.

In terms of functional knowledge, Purpura defines it as "how utterances or sentences and texts are related to the communicative goals of language users", in other words, L2 users can get things done in communication (Purpura, 2016). This competence enables
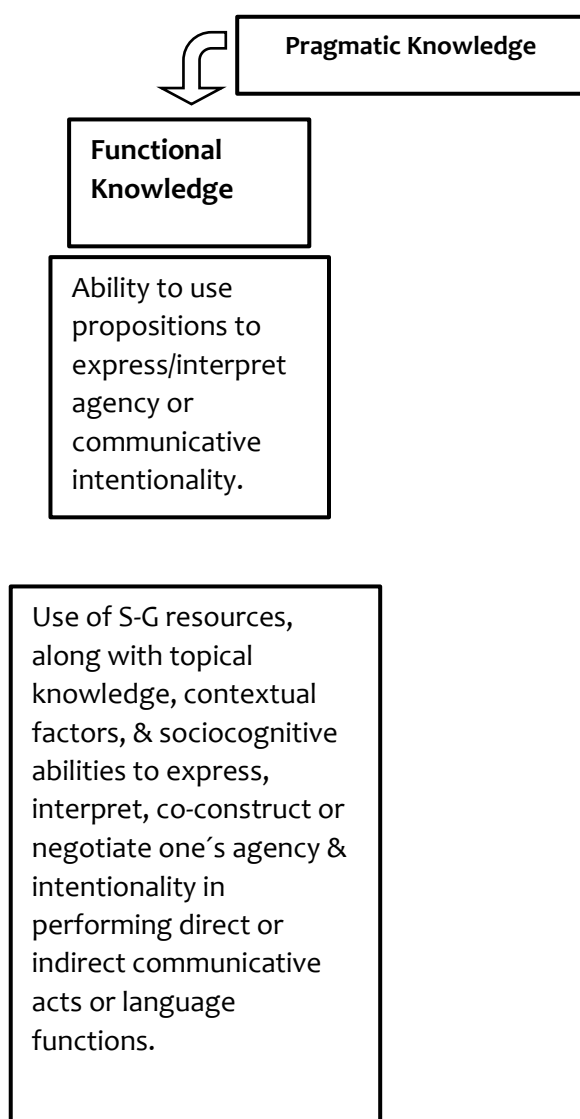
language users to get the message across and constitutes a core competence that, in turn, has been operationalized to generate the CEFR descriptors according to Purpura.

In Figure 7, Purpura breaks down the functional knowledge characteristics and the resources individuals can resort to express or interpret messages in different language use contexts.

Figure 7 Purpura's Meaning-Oriented Conceptualization of Language Knowledge

Pragmatic Knowledge

Functional Knowledge

Ability to use propositions to express/interpret agency or communicative intentionality.

Use of S-G resources, along with topical knowledge, contextual factors, & sociocognitive abilities to express, interpret, co-construct or negotiate one´s agency & intentionality in performing direct or indirect communicative acts or language functions.

In the following example, a woman (Lorena) is communicating her intention by pointing out that someone is cutting in line or entering a line at any position other than the end. Lorena's intention can be interpreted in many ways depending on the context of the situation. Lorena is resorting to her functional linguistic competence to convey her message, which can possibly be interpreted to raise awareness about the importance of respecting the line, among other possible meanings. On the other hand, according to Purpura, the ability to understand and comprehend functional meanings depends on the user's functional and implicational knowledge, which is a key competence for L2 learners (Purpura, 2016). Based on this premise, the other woman (Silvia) understood the implied message and provided an explanation to Lorena.

Figure 8 Sample

Source: DGEC.

Finally, implicational knowledge is defined as "communication depends on the participants' shared presuppositions, experiences, and situational associations, much of

what occurs in language use is unstated or implied. As a result, these same messages embody other layers of meaning, referred to as implied or implicational pragmatic meanings," according to the definition proposed by Purpura, 2018.
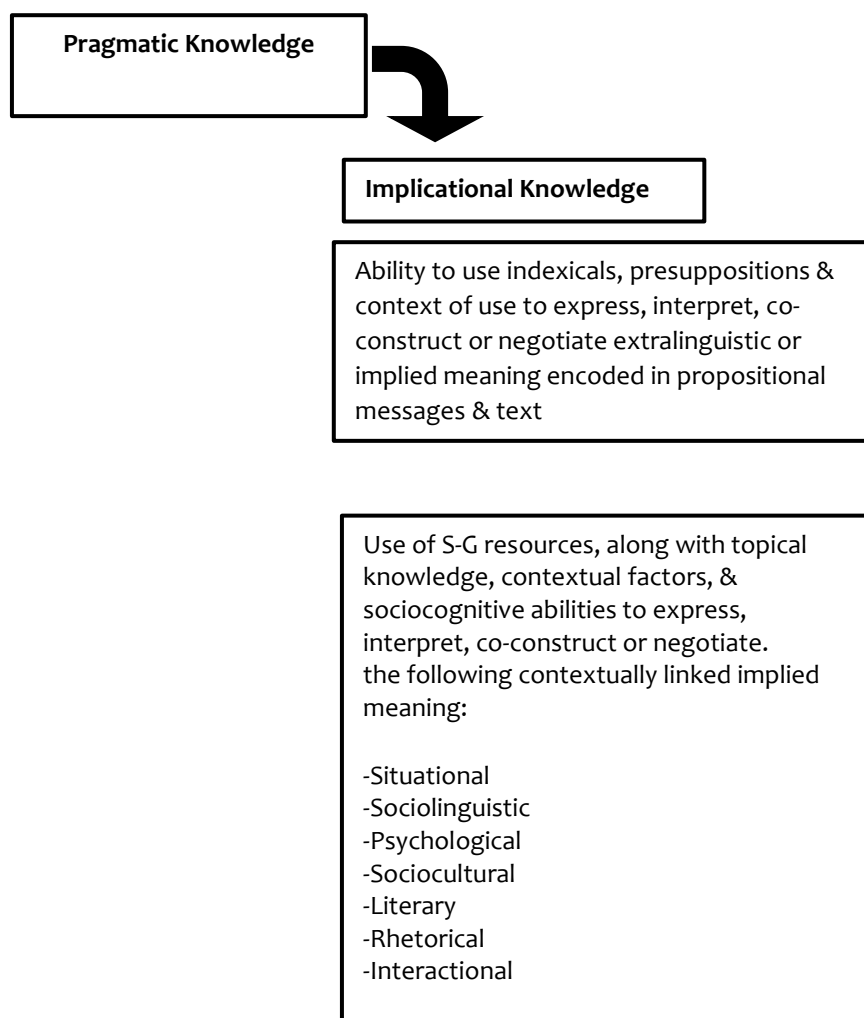
Implicational knowledge

The last section of the MOCKL describes the knowledge structures that a language user can resort to convey and understand meaning in a higher level of complexity. Purpura states that "More interestingly, pragmatic knowledge also involves knowledge structures that enable learners to simultaneously encode, onto these same utterances or texts, a wide range of meanings that are implied by shared presuppositions, experiences, and associations with reference to the communicative situation" (Purpura 2018, page 19).

The language users' ability to convey and understand this kind of message lies in the fact that they must use verbal and non-verbal resources along with contextual factors. In other words, language users need to resort to their implicational knowledge ability to understand these messages.

Figure 9, presents the linguistic resources required for a language user to affirm that has the ability to decode the meaning implied in a message. The layers of meanings proposed in this section of the model are mentioned as well.

Figure 9 Purpura's Meaning-Oriented Conceptualization of Language Knowledge

```
┌─────────────────────────────┐
│   Pragmatic Knowledge       │
└─────────────────────────────┘
                    ┌──┐
                    ↓
        ┌──────────────────────────────┐
        │   Implicational Knowledge     │
        └──────────────────────────────┘
        ┌──────────────────────────────┐
        │ Ability to use indexicals,    │
        │ presuppositions &             │
        │ context of use to express,    │
        │ interpret, co-               │
        │ construct or negotiate        │
        │ extralinguistic or            │
        │ implied meaning encoded in    │
        │ propositional                 │
        │ messages & text               │
        └──────────────────────────────┘

        ┌──────────────────────────────┐
        │ Use of S-G resources, along   │
        │ with topical                  │
        │ knowledge, contextual         │
        │ factors, &                    │
        │ sociocognitive abilities to   │
        │ express,                      │
        │ interpret, co-construct or    │
        │ negotiate.                    │
        │ the following contextually    │
        │ linked implied                │
        │ meaning:                      │
        │                               │
        │ -Situational                  │
        │ -Sociolinguistic              │
        │ -Psychological                │
        │ -Sociocultural                │
        │ -Literary                     │
        │ -Rhetorical                   │
        │ -Interactional                │
        └──────────────────────────────┘
```

Source: Purpura, J.E. Assessing Meaning-Teachers College, Columbia University, New York, NY, USA.

Within this section of the framework, below, according to Purpura (2016), seven types of implied pragmatic meanings are detailed, encoded in talk and text:

- Situational meanings: based on understandings of the local context of situation (i.e., how to communicate meanings specific to a given situation) – e.g., acceptable, appropriate, natural, and/or conventional use of indirect function interpersonal references or associations, figures of speech, proverbs, and situational and formulaic implicatures.

- Sociolinguistic meanings: based on understandings of the social norms, assumptions, preferences, and expectations within a specific speech community (i.e. How to communicate with a given person in a given social context) – e.g., acceptable, appropriate, and conventional use of social deixis (group identity markers), politeness (relative power, degree of imposition, social distance), registers, varieties, etc.

- Sociocultural/intercultural meanings: based on understandings of the convergent or divergent assumptions, norms, values, preferences, and expectations across different demographic and linguistic cultures (how to communicate within a given culture or across cultures) – e.g., acceptable, appropriate, and conventional use of topic, humor, gratitude, regret, and criticism; avoidance of taboos; etc.

- Psychological meanings: based on understandings of affective stance (how to communicate mood, attitudes, feelings, emotionality, and other dispositions) – e.g., acceptable, appropriate, or conventional use of humor or sarcasm or the conveyance of anger, deference, patience, affection, self-importance, etc.

- Literary meanings: based on understandings linked to aesthetic imagination, fantasy, embellishment, exaggeration, and figures of speech – e.g., appropriate, creative, and original use of literary conventions.

- Rhetorical meanings: based on understandings of textual structuring practices, genres, discourse modes, and coherence – e.g., acceptable, appropriate, and conventional use of organizational patterns.

- Interactional meanings: based on understandings of conversational structuring practices, sequencing practices, turn-taking practices, and repair practices – e.g., acceptable, appropriate, natural, and conventional practices associated with conversational norms, assumptions, and expectations.

Following Lorena's example, in terms of assessment, a tester can inquire about the relationship between the characters in the image: family, acquaintances, coworkers, among others. Language tasks can also ask examinees to identify the tone of each

character, or rather, the psychological meaning, for example sarcastic, upset, enthusiastic, among others. By providing a more detailed context, it could also be inferred where this situation is taking place: at the bank, supermarket, among others. Finally, to measure the construct of the PNELE, only situational, sociolinguistic, psychological, and sociocultural meanings will be taken into consideration.

4. **The Socio-Cognitive Dimension:** It specifies the role in understanding and fostering the intersection of social interactions and cognitive processes in the learning environment. Turner and Purpura (2016) state that this component of the LOA framework is transcendental for the "understanding of how L2 learners learn and the effect this has on how instruction and assessment are conceptualized and implemented, how performance evidence is interpreted, and how inferences from evidence are used to provide feedback and, if needed, learning assistance" (p. 264).
   It emphasizes the importance of assessing not only the individual's cognitive abilities but also their ability to engage meaningfully with others, share knowledge, and contribute to a collaborative learning process.
   Purpura (2016) considers the following resources important to complete a task:

   ● Cognitive architecture: It includes attention, memory power and background knowledge needed to perform well during a task. It serves as a performance moderator.
   ● Cognitive functionality: It considers aspects such as processing, reasoning, executive function and strategies that might affect the performance of the examinee during a task. It also contemplates self-regulation.

5. **<u>The Instructional Dimension</u>**: It serves as a cornerstone, bringing together assessment practices and instructional strategies to create a relationship that enhances a meaningful learning process. A key principle is that the test itself

should not only measure learning outcomes but also have a teaching function. This involves incorporating elements like assistance and feedback directly into the assessment process.

According to Turner and Purpura (2015), feedback and self-regulation are important components of LOA. Feedback serves as a tool to help learners identify areas for improvement during the learning process, while self-regulation encourages learners to take responsibility for their own learning. This self-regulation involves using metacognitive techniques to set goals and organize the learning process effectively.

Additionally, the information presented to examinees may have an instructional impact either incidentally or by intentional design. This intentional alignment of assessment and instruction ensures a more comprehensive and effective educational experience.

6. **The Social-Interactional Dimension** This is another important moderator of performance in tests where reciprocal communication is involved. It relates to the examinees' ability to engage in talk or interaction during the test, and the effect this has on performance. If refers to all the communication activities that take place during the assessment process. As said by Turner & Purpura (2015), the structure of communication and feedback greatly influences students' processing and integration of the L2. When language is measured, especially the speaking ability, integrated assessment is needed, due to the fact that an oral response always follows an oral or written stimulus in the case of a dialogic process, or a researched and planned work if this process is monologic; thus, elements of interaction such as turn-taking need to be considered when planning that assessment process. Turner and Purpura (2016) add that:

"LOA is particularly interested in exchange patterns that provide a positive evaluation of a learner's learning or performance… related to a learning goal, or

a negative evaluation... followed by scaffolded assistance... in repairing some aspect of communication or learning with hints or by a more elaborated learning intervention" (p. 266).

In Dean's work, "The Interactional Dimension of LOA: Within and Beyond the Classroom", written for the journal *Working* Papers *in Applied Linguistics & TESOL* (2014, pp. 50.52), she states 3 major ideas of how this dimension affects an assessment, and of course, these ideas will be referenced in the design of the 2021 National Exam of English proficiency. The author emphasizes, first of all, the necessity of creating authentic scenarios which emulate real interactions.

A second idea is focusing on the use of dialogic conversations because unlike simple conversations, they incorporate specific contextual values, long-term learning goals, targeted pedagogical questioning, critical thinking, and argumentation. The third idea stated by Dean is promoting interpersonal interaction using technology, which integrates simulated peer and teacher interactions.

7. <u>**The Affective Dimension:**</u> The affective dimension deals with the learner's affective dispositions that are triggered during a task. It involves the learners' emotions, engagement, motivation, attitudes and beliefs surface during the process of examination.

In addition, Turner and Purpura (2016) state that "characteristics such as the learners' emotions, their beliefs about learning and competence, their personality characteristics..., their attitudes towards learning and performance..., and their motivation" (p. 266) play a determining role in the body of emotions with which examinees will face an assessment device. Stabler-Havener (2014, p. 53) discusses how the learners' emotions, motivation, attitudes and beliefs about learning, as well as their personality traits such as introversion and extroversion influence the affective dimension of both the learning and the assessment process.

Overall, two key emotional elements that mostly interfere with the affective dimension of an assessment process are motivation and anxiety. Davies et al. (1999) in the Dictionary of language testing (1999) state different principles of the terms "motivation" and "anxiety" when they appear in the context of an assessment; indeed, they use the term "test anxiety" instead of "anxiety". In regard to motivation, they say that some evidence has been gathered to consider that "increased motivation will lead to better test performance" (p. 123). Another issue is that the degree of motivation for the test taker may be inclined by the stakes the test offers: the higher the consequences, the bigger the motivation. This also will add a degree of reliability to the examination, since "the higher the test taker's level of motivation, the truer the reflection of ability shown by the performance, and hence the lower the amount of error," and, therefore, "results from… test takers… not motivated are likely to be less reliable than those where they are strongly motivated." (p. 123)

Finally, with the intention of determining the pertinence of setting the accurate elements and perspective of the affective dimension PENELE, it is important to anticipate protocols to reduce anxiety. This means that actions will need to start months before the assessment date, to familiarize teachers, proctors, and examinees with the test format. Also, we will need to prepare mock tests in an online format so that stakeholders can know what to expect.

References

Arikan, A. (2015). The CEFR and reading: A document analysis. *Procedia, Social and BehavioralSciences (199),*501-504, Elsevier Ltd.

Bachman, L. & Palmer, A. (1996). *Language testing in practice.* Oxford: Oxford University Press

Bachman, L., & Palmer, A. (2010). *Language Assessment in Practice.* Oxford: Oxford University Press.

Brown, G. & Yule, G. (1983). *Teaching the Spoken Language. (2nd ed.)* Cambridge: Cambridge University Press.

Burnkart, G. (1998). *Spoken language: What it is and how to teach it.* Washington DC: Center for Applied Linguistics.

Butler, F.; Eignor, D.; Jones, S.; McNamara, T. & Suomi, B. (2000). *TOEFL 2000 Speaking Framework: A Working Paper.* RM-00-6. Princeton, New Jersey: Educational Testing Service

Buck, G. (2001). *Assessing Listening. Cambridge,* Cambridge University Press.

Look at how I edited this. Use the new APA manual to revise this. Look it up online.

Carr, Nathan. Designing and Analyzing Language Tests. 2011. Oxford University Press.

Carrol, B (2018). A learning-oriented assessment perspective on scenario-based assessment.

Working Papers in Applied Linguistics & TESOL, Vol. 17, No. 2, pp. 28-35 The Forum 28 Teachers College, Columbia University

 Chapelle, C. (2008). The TOEFL validity argument. Building a validity argument for the Test of English as a Foreign Language®. 319-352.

Common European Framework of Reference for Languages: Learning, Teaching, Assessment. (2003). Cambridge, U.K.: Press Syndicate of the University of Cambridge.

Council of Europe. (2001). The Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Structured overview of all CEFR scales. Cambridge: Cambridge University Press.

Degand, L. & Simon, A. (2009). Identifying basic discourse units in speech: theoretical and empirical issues. Discourse, 4, 1-19. DOI: https://doi.org/10.4000/discours.5852

Derakhshan, A; Khalili, A. & Beheshti, F. (2016). Developing EFL Learner's Speaking Ability, Accuracy and Fluency English. *Language and Literature Studies, 6(2), 177-186.* doi:10.5539/ells.v6n2p177

Douglas, D. (1997). Testing speaking ability in academic contexts: Theoretical considerations. *TOEFL Monograph Series*, 8. Princeton: Educational Testing Service.

East, M. (2017). Research into practice: The task-based approach to instructed second language acquisition, *Language Teacher, 50(3)*, 412-424. Doi: 10.1007/SO26244481700009X.

Ekbatani, G. (2010). Measurement and Evaluation in Post-Secondary ESL. New York : Routledge.

ETS Standards for Quality and fairness 2014).

Foreign Language Teaching Methods. Carl Blyth, Editor. 2010. COERLL, The University of

Texas at Austin. http://coerll.utexas.edu/methods

Franco, C. & Galvis, H. (2013). The role of situational context and linguistic context when testing EFL vocabulary knowledge in a language teacher education program: A preliminary approach. *Colombian Applied Linguistics Journal, 15(1), 85-99.* https://doi.org/10.14483/udistrital.jour.calj.2013.1.a06.

Gass, S. M., & Mackey, A. (2012). The Routledge handbook of second language acquisition. London, New York: Routledge.

Hughes, R. & Szczepek, B. (2016). Teaching and Researching Speaking. (3rd ed.). *Applied Linguistics in Action*. New York: Routledge.

Introductory Guide to the Common European Framework of Reference (CEFR) for English

Language Teachers (2013). Cambridge University Press.

Jucker, A. & Ziv, Y. (1998). Discourse Markers: Descriptions and theory. Amsterdam: John Benjamins.

Kandefer, M. & Shapiro, S.C. (2008). A categorization of contextual constraints. *In Biologically inspired cognitive architectures: Papers from the AAAI Fall Symposium, 88-93.* Virginia: AAAI Press.

Lindsay, C. and Knight, P. (2006) Learning and Teaching English. Oxford: OUP.

McCarthy, M. (1998). Spoken Languages and Applied Linguistics. Cambridge: Cambridge University Press.

Ministerio de Educación Pública (MEP). (2015). Informe de Diagnóstico: Programas de Estudio de Inglés para Preescolar, Primaria y Secundaria, San José: MEP.

Ministerio de Educación Pública (MEP). (2015). Política Curricular Educar para Una Nueva Ciudadanía. San José: MEP.

Ministerio de Educación Pública (MEP). (2016). Programas de Estudio de Inglés. Tercer Ciclo de la Educación General Básica y Educación Diversificada. San José: MEP.

Ministerio de Educación Pública. (6 de febrero de 2018). Reglamento de Evaluación de los Aprendizajes 2018 N° 40862-MEP. La Gaceta.

Miselvey, R., Almond, R., & Lukas, J. (2003). *A brief introduction to evidence-centered design research report RR-03-15*. Princeton, NJ. Educational Testing Service.

Nazara, S. (2011). Students' Perception on EFL The Development. *Journal of English Teaching), 1(1), 28-43.* DOI: http://ejournal.uki.ac.id/index.php/jet/article/view/50

News Figueras & José NOIJONS (EDS.) (2009). Linking to the cefr levels: research perspectives. European Association for Language Testing and Assessment (EALTA).(P. 93)

Oh, S. R. (2018). Investigating Test-takers' Use of Linguistic Tools in Second Language Academic Writing Assessment. Ed.D., Teachers College, Columbia University https://doi.org/10.7916/D8B00HDQ

Piccardo, Enrica (2014). From Communicative to Action-Oriented: A Research Pathway

Purpura, J. (2004). *Assessing Grammar.* Cambridge: Cambridge University Press.

Purpura, J. (2016). *Assessing Meaning.* Teachers College, Columbia University, New York, NY, USA.

Purpura, J. (2016). Second and Foreign Language Assessment. The modern Language Journal, pp. 190-204.

Sarem, S. & Shirzadi, Y. (2014). A Critical Review of the Interactionist Approach to Second Language Acquisition. *Journal of Applied Linguistics and Language Research* 1(1), 62-74

Tanaka, H. (2000). Turn-Taking in Japanese Conversation: A Study in Grammar and Interaction. Pragmatics & Beyond New Series, 56. Amsterdam: John Benjamins.

Wolf, M. K. & Butler, Y. G. (2017). English Language Proficiency Assessments for Young Learners. New York: Routledge.