

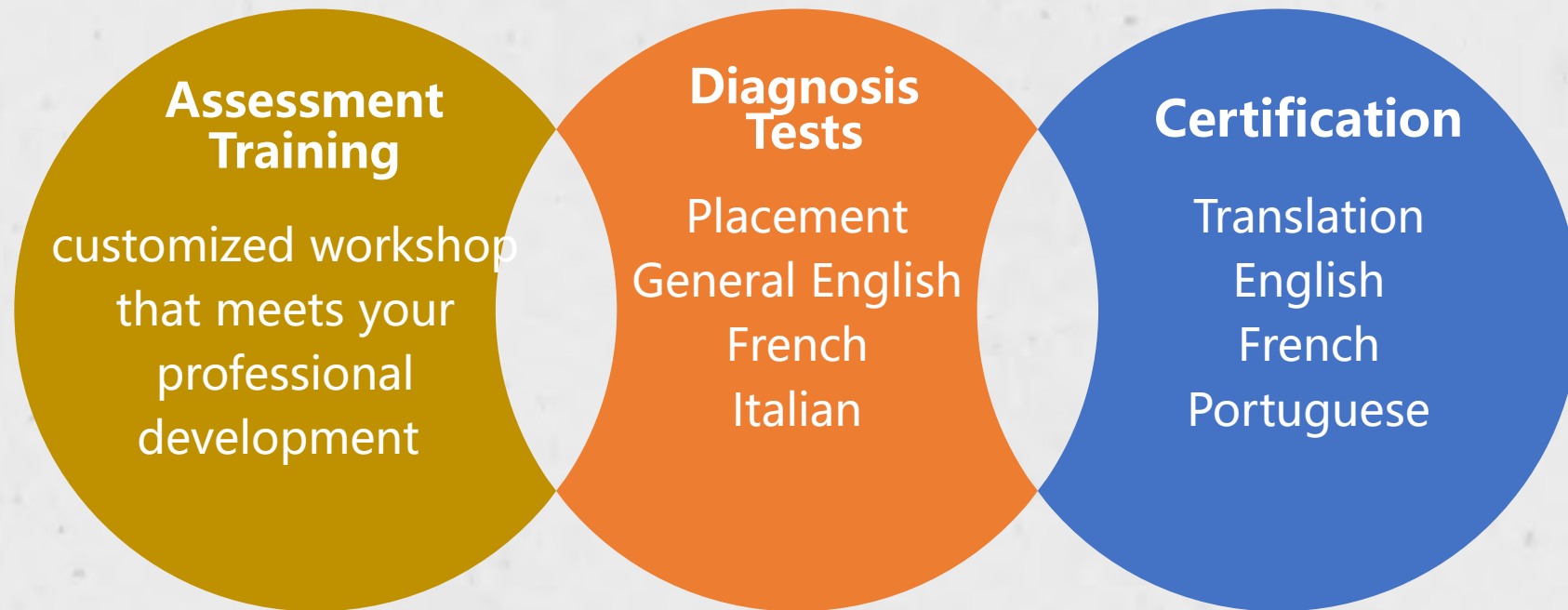
Dr. Allen Quesada-Pacheco  
allenquesada@gmail.com

July 5<sup>th</sup>, 2024





## Research-based Outreach Program

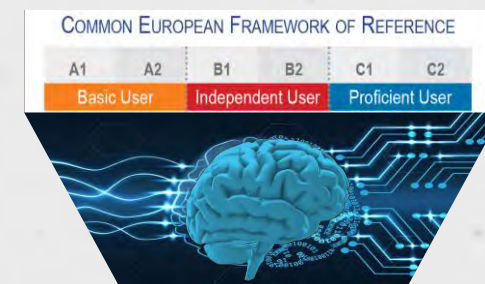


**informing, raising awareness of assessment-related issues**



# PELEx Reference framework

The Common European Framework of Reference (CEFR) is an internationally recognized framework of reference for describing language proficiency (A1-C1).



- American Council on the Teaching of Foreign Languages (ACTFL), American Council on the Teaching of Foreign Languages Proficiency Guidelines),
- The Canadian Language Benchmarks (CLB),
- The ILR (Interagency Language Roundtable) proficiency scale, and
- WIDA (Wisconsin, Delaware, Arkansas Assessment language framework)

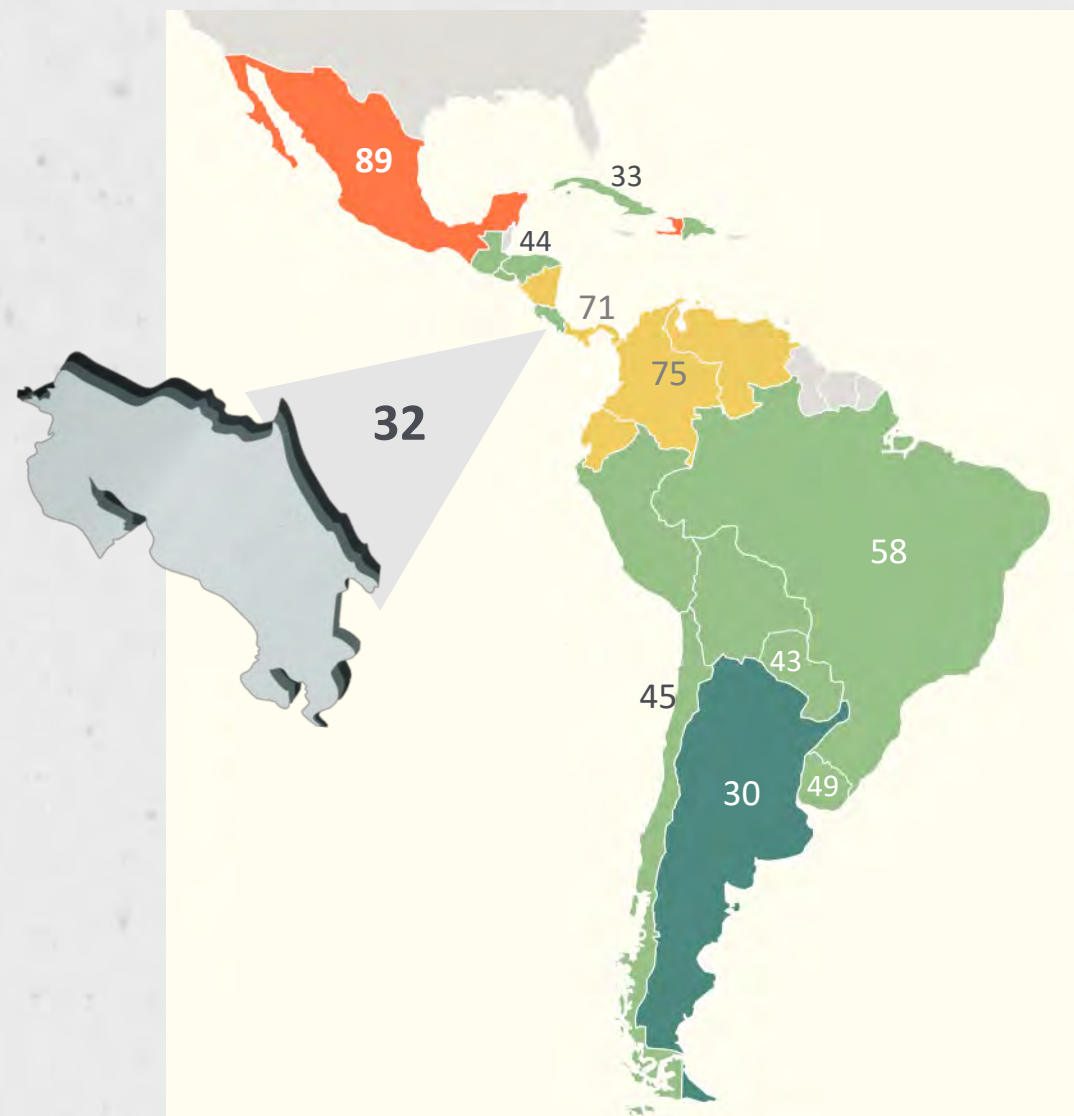


# EF EPI

## EF English Proficiency Index

A Ranking of 111 Countries and Regions by English Skills

- Very high
- High
- Moderate
- Low
- Very low







## English Proficiency Index



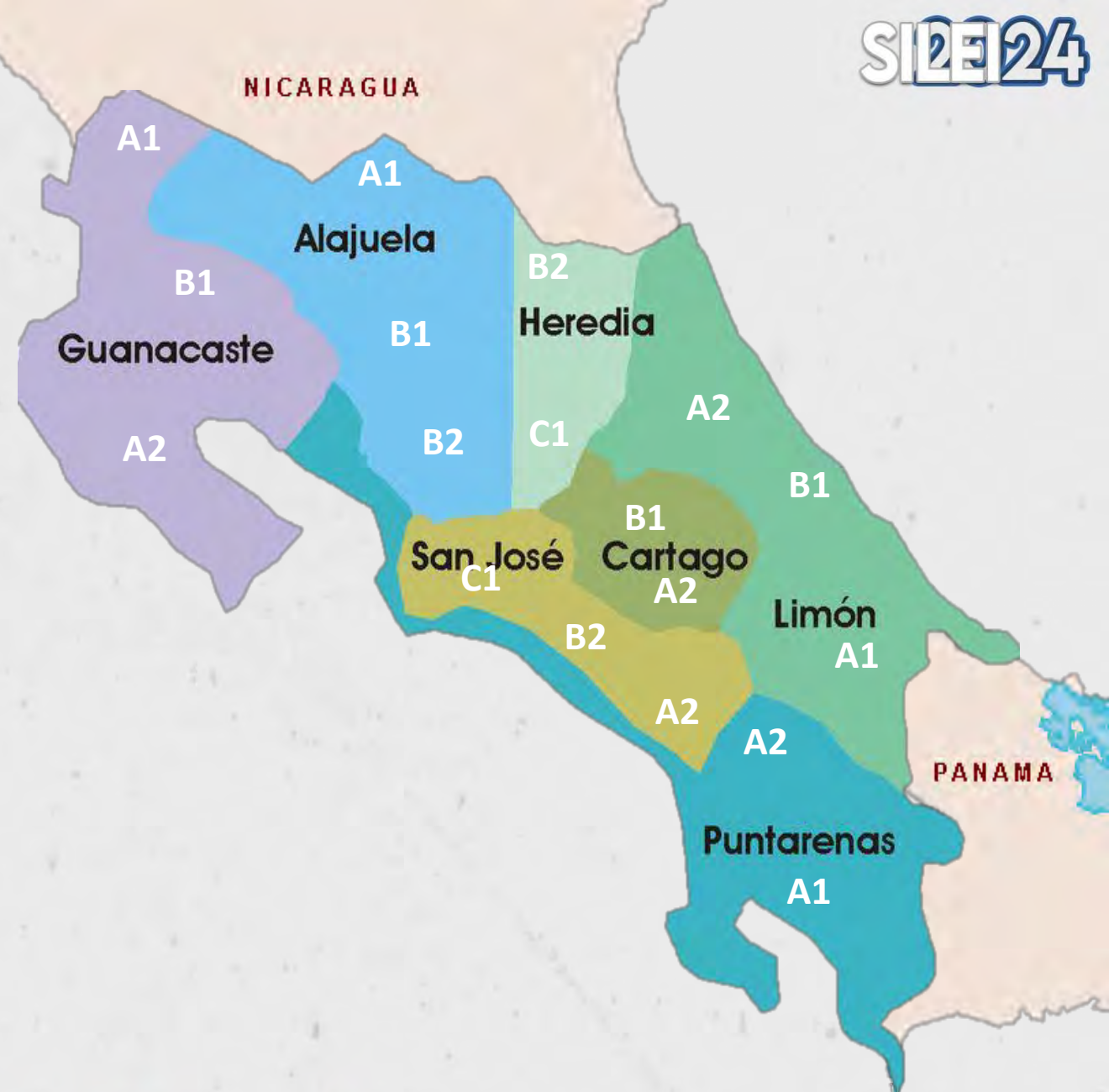
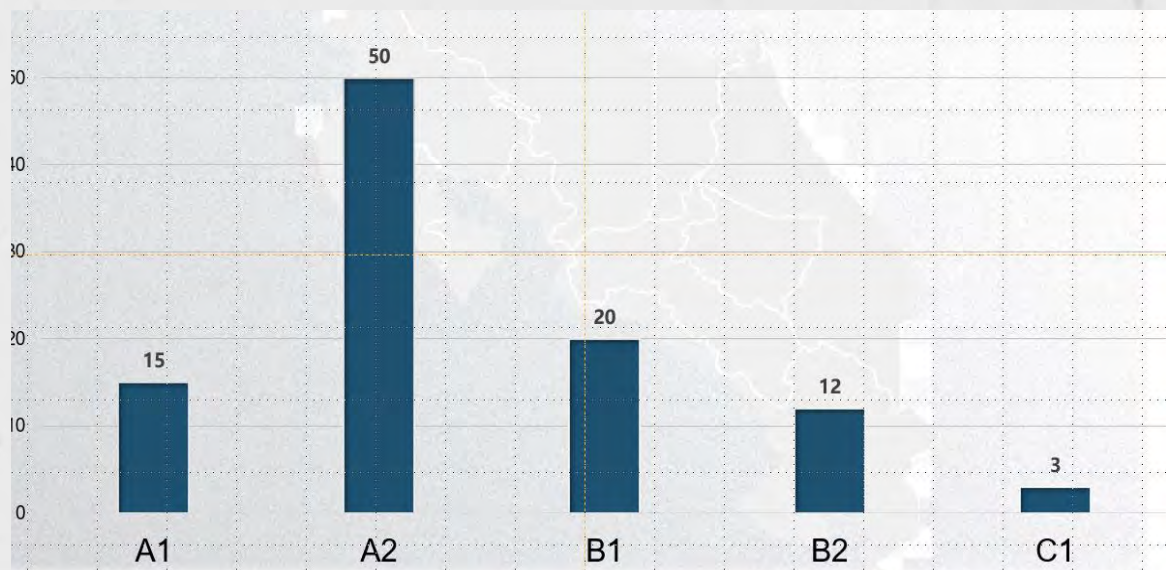
Reading



Listening



Speaking



# PELEx: Glocal Language Tests

**400.000**  
Thousand  
users

People we reached  
in 2017–24

**countries**  
**7**

We are present  
on the ground

**14**  
years

Building connections  
since 2010

**25**  
languages

Languages we assess  
In translation

# PELEx: Glocal Language Tests

4

Types of users

People we reach in  
primary  
Secondary  
University  
Companies

7

Provinces

We are present  
on the ground &  
online

3

Digital Tests

Building linear, adaptive  
& multistage tests  
since 2023

4

affiliations

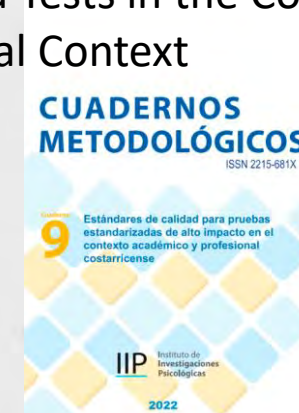
ILTA  
LAALTA  
ALTE  
CIECPE





Interinstitutional Squad of Quality Standards for High-Impact Standardized Tests in the Costa Rican Academic and Professional Context

ALTE Auditing System





MINISTERIO DE  
EDUCACIÓN PÚBLICA

01



PROGRAMA  
ESTADO DE LA NACIÓN

03

02



CONSEJO NACIONAL  
DE RECTORES



# Family of Tests



TEYL-UCR



EPT-UCR



ECT-UCR



French-UCR  
Italian-UCR  
Portuguese-UCR



GET-UCR



ELTA-UCR

# Development of PELEx Language Tests

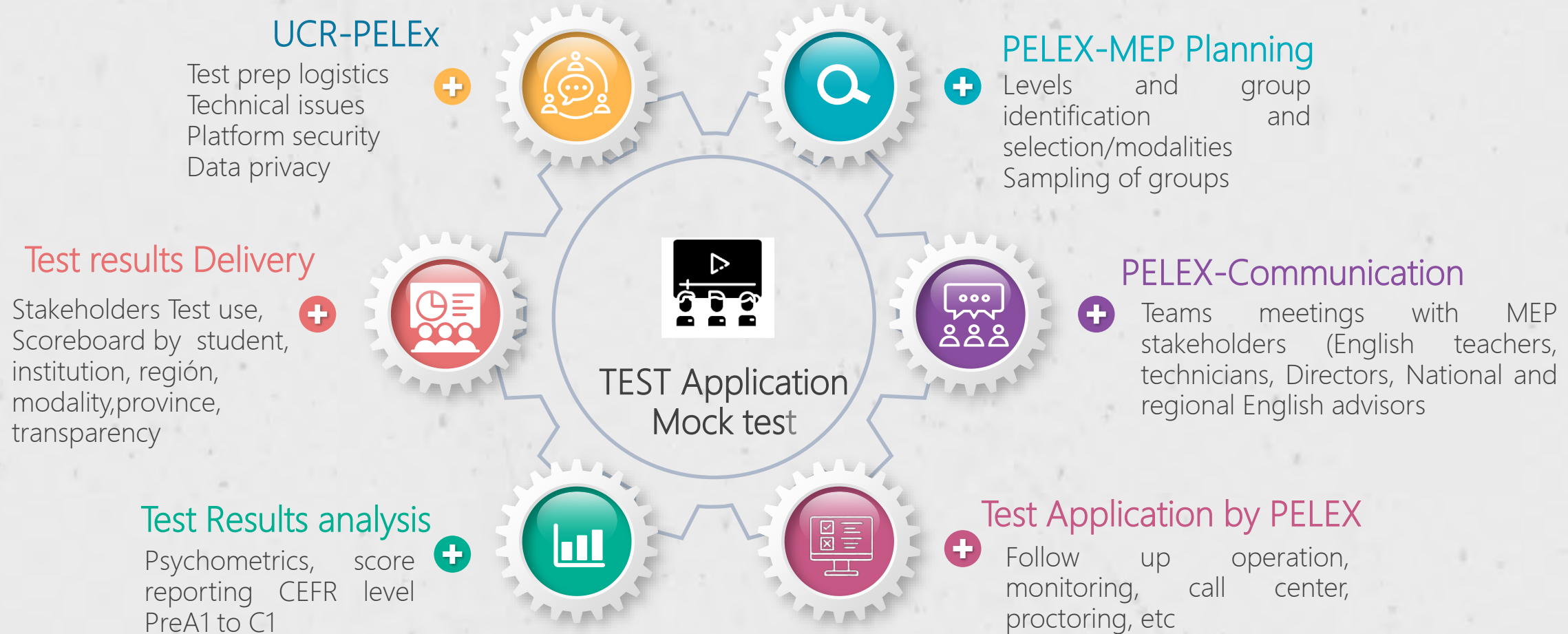
The UCR digital English Language tests are designed to comply with the basic principles of standardized assessment such as:

- a. **Practicality of the test.** (8000 students per day) and results in real time.
- b. **Reliability.** Results consistent with the needs of the students.
- c. **Validity.** Research conducted to find evidence of the validity of the content, its consequences or impact, among others.
- d. **Authenticity.** Authentic tasks that can be extrapolated to classroom contexts.
- e. **Impact.** The interpretation of results (assigning a CEFR band) useful for different stakeholders.

Our testing policy has proudly embraced a zero-paper approach, significantly minimizing our environmental impact



# PELEx FRAMEWORK



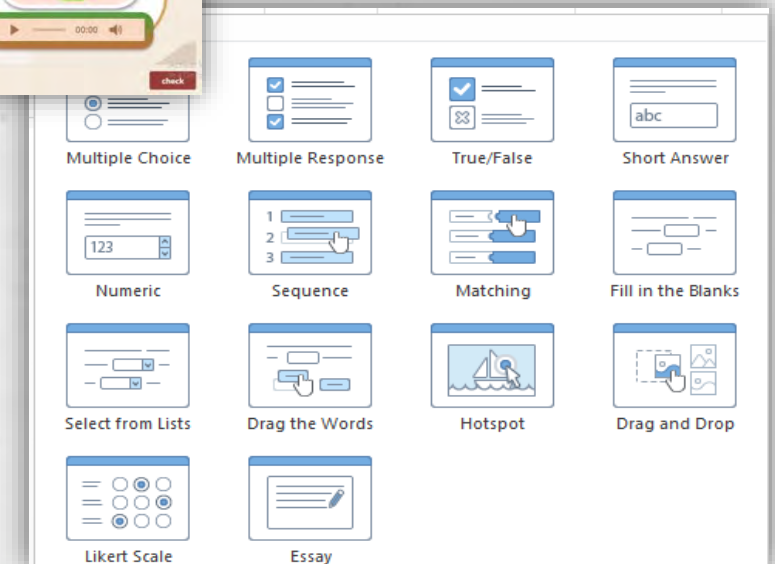
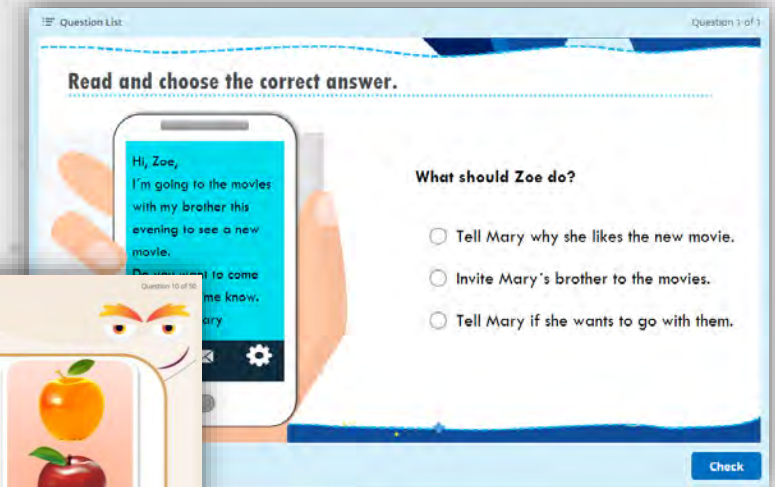
# Usability testing

considerations :

Authoring software	Navigation	Page layout
interface design	Colors	item type

The primary aim of good interface design is to reduce to a minimum construct-irrelevant variance that could be attributed to test method (Messick, 1989)

Construct irrelevant variance (CIV) is **the introduction of extraneous, uncontrolled variables that affect assessment outcomes and therefore its validity.**





# Reliability and Evidence of Validity

## I. Test construction

- a. Methodological and referential framework
- b. Purposes of Testing
- c. Construct and sub-constructs of the test
- d. Cut-off points and their respective criteria
- e. Technical-scientific validity studies related to the concurrence of Results with other tests



## III. Test Grading

- a. Scoring processes to ensure the accuracy of the result.
- b. Interpretation of scores
- c. Return of results to different populations, and uses and usefulness of the results

## II. Exam Administration

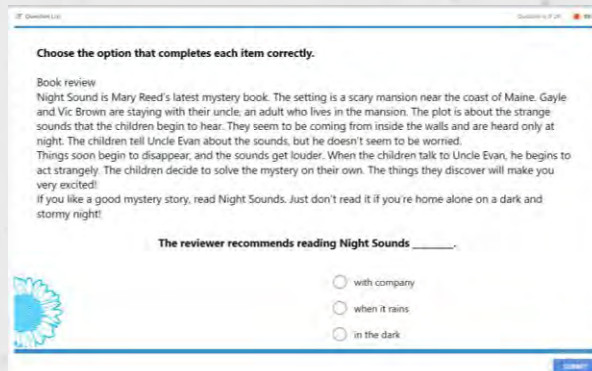
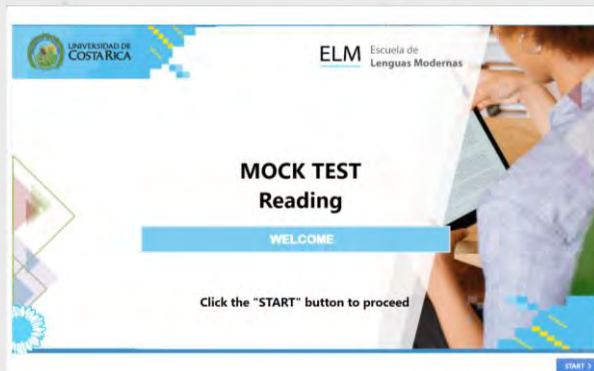
- a. Established procedures for the application and administration of the test
- b. If applicable, adaptation or modifications made



## IV. Analysis of Consistency

- a. Statistical and psychometric studies on the level or degree of Item discrimination
- b. Studies on the consistency of the items
- c. Consistency studies between judges for oral and written production (validation by expert judgment)
- d. Task consistency analysis

# Adaptive and MST TESTS



Adaptive testing based on a candidate's level of ability.

## PeLexCAT: Adaptive Language Testing

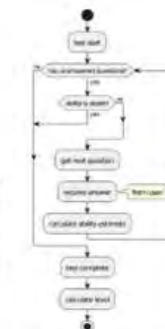
Development of a Localized Adaptive Language Test of Receptive Skills

Dr. Edgar Casasola Murillo and Dr. Allen Quesada Pacheco  
UNIVERSIDAD DE COSTA RICA



- SOFTWARE DESIGN**
- Scalable
  - Multi Layered
  - Email notifications
  - Accessibility
  - Integration of External AI Speaking Testing

- ADVANTAGES**
- Reduced time
  - Less exhaustion
  - Measures ability



**ADAPTIVE TEST  
MAIN ALGORITHM**

- STOP CONDITION**
1. Fixed threshold
  2. Standard Error of Estimation  $\leq 0.35$
  3. Time limit



READING

MULTI SKILL TESTING  
INTEGRATION

External System  
Integration



ORAL SPEAKING  
IA TESTING

LISTENING

3 Parameter IRT Model  
Discrimination vs Difficulty

READING					LISTENING				
A1	A2	B1	B2	C1	A1	A2	B1	B2	C1
30	27	28	20	21	27	21	26	17	15

SIMULATION BASED EXPERIMENTATION AND VALIDATION

INFORMATION BASED ITEM DISPATCHER



ABILITY CONVERGENCE BY LEVEL



UNIVERSIDAD DE  
COSTA RICA

Escuela de  
Ciencias de la  
Computación e  
Informática

PELEX

Programa de  
Evaluación en  
Lengua Extranjera

ELM Escuela de  
Lenguas Modernas

Made by Edgar Casasola Murillo &  
Alejandro Marcano Jimenez



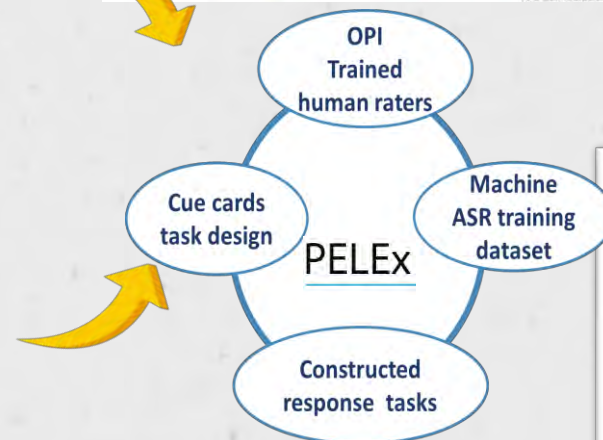
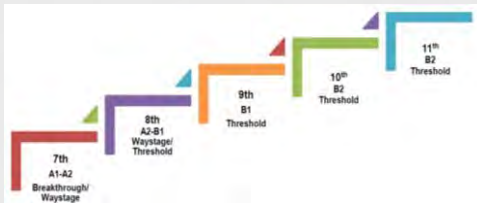


# Localized Automated Speaking Test for EFL Learners



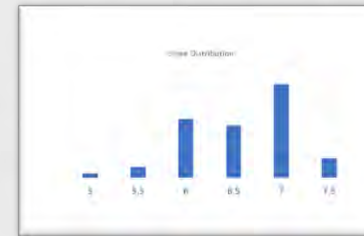
**SILE 24**

## Proficiency CEFR level for Primary and Secondary School in Costa Rica

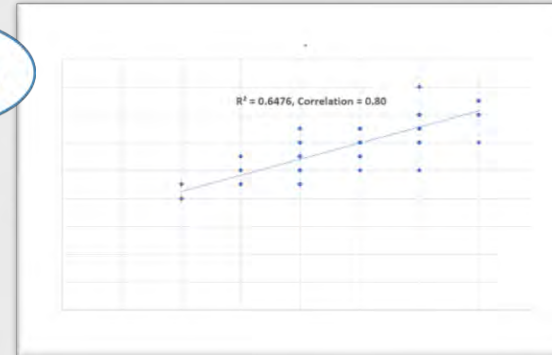


## Localized (CR ) Oral Task Development

- A. Read-aloud
- B. Ask a question
- C. Tell a picture-based story
- D. Share an opinión and rationale
- E. Open-ended question



## Reliability & Calibration



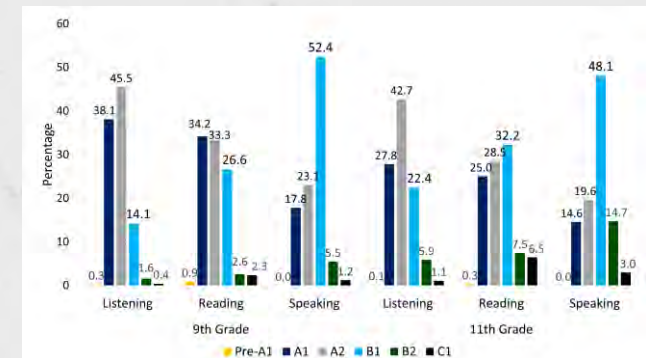
## High Precision Scoring: Machine-Human

## Speaking Application- August 2023 (primary & secondary)

### TEYL

CEFR	3-q Quiz scores	Q5 scores in 5-q test	5-q Quiz scores
A0	9%	9%	13%
A1	61%	72%	67%
A1+	13%	1%	10%
A2	9%	6%	6%
A2+	4%	8%	4%
B1	4%	4%	0%
B1+		1%	

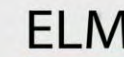
## Secundaria



**DGRI**  
Dirección General de Relaciones Internacionales



**DCIA**  
Dirección de Centros de Idiomas y de Autoacceso



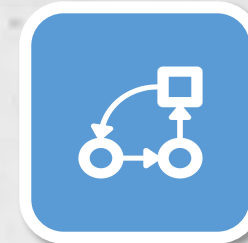
**Escuela de Lenguas Modernas**

**PELEX**

Programa de Evaluación en Lengua Extranjera

# Stakeholders Use of Test

**Scoreboard Preparation:** The results are compiled into detailed reports that may include individual scores, class or group performance, and aggregated data by institution, region, modality, and province



**Direct Communication:** The results are sent to relevant stakeholders, which may include educational institutions, regional education authorities, teachers, and administrators.

•**Stakeholder Meetings:** Meetings are often held with key stakeholders, such as language teachers, school directors, and regional advisors, to review the results.



•**Interpretation Sessions:** Sessions may be conducted to help stakeholders understand the data, trends, and implications of the test results.



**Curriculum Adjustments:** Changes or enhancements to the curriculum may be implemented to better align with the identified needs and performance levels.

# Remember

"No test is perfect. It's important to consider them as one of many tools for assessment as and for learning. They should be used in conjunction with other evaluations and observations to get a well-rounded understanding of an individual's language abilities."

"The PELEX language tests provide a comprehensive and environmentally sustainable assessment of language proficiency, offering detailed insights into individual and group performance that help guide educational improvements and professional development."

takeaway



