UNIVERSIDAD DE COSTA RICA

ELM Escuela de Lenguas Modernas

PELEx Programa de Evaluación en Lengua Extranjera

ALTE Associate Member

Implicaciones y desafíos del uso de Inteligencia Artificial en la evaluación por competencias en lenguas extranjeras

EL CASO DE COSTA RICA

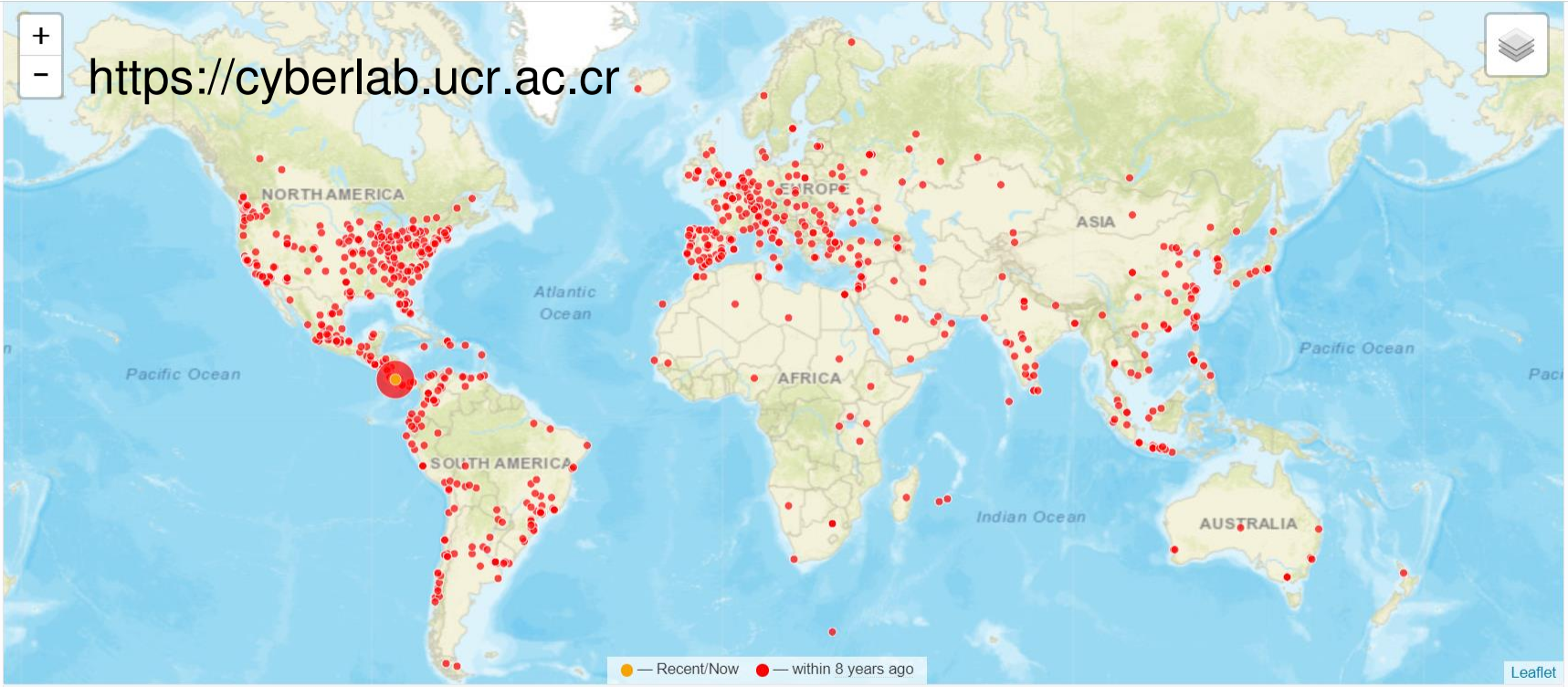
Allen Quesada Pacheco, Ph.D

2023

Visitor Traffic for Cyberlab.ucr.ac

clustrmaps.com/site/1iob

Documento sin título | Gmail: Email from G... | ten minutes is not e...



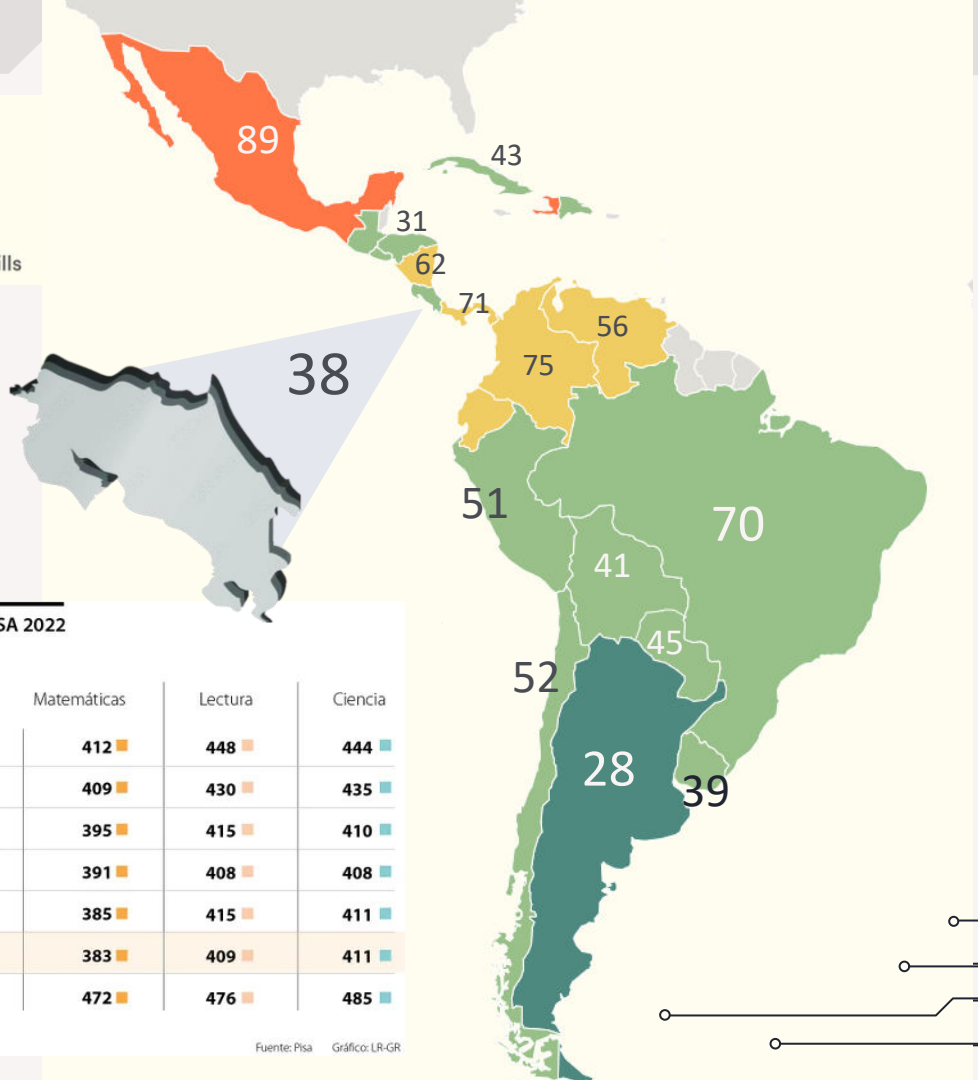
Recent visitors	Browser	OS	Date
New Visitor from San José, Costa Rica with 1 pageview	Chrome	Win10	December 5, 2023, 12:47 pm

Activar Windows
Ve a Configuración para activar Windows.

EF EPI

EF English Proficiency Index

A Ranking of 111 Countries and Regions by English Skills



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

RESULTADOS PISA 2022

AMÉRICA LATINA

	Matemáticas	Lectura	Ciencia
1 Chile	412	448	444
2 Uruguay	409	430	435
3 México	395	415	410
4 Perú	391	408	408
5 Costa Rica	385	415	411
6 **Colombia**	383	409	411
Promedio Ocde	472	476	485

Fuente: Pisa Gráfico: LR-GR

EF EPI

EF English Proficiency Index

A Ranking of 111 Countries and Regions by English Skills

525 543

Región principal: **Heredia*** **(566)**

Ciudad principal: **Heredia (565)**

Grupo de edad principal: **31-40 (nota media de 559)**

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

Costa Rica

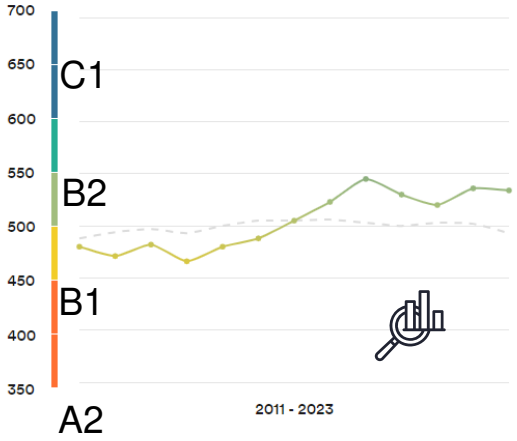
#38 de 113

Clasificación EF EPI: 534

Puntuación media global: 493

Posición en Latinoamérica: 3 de 20

Descargar hoja de datos

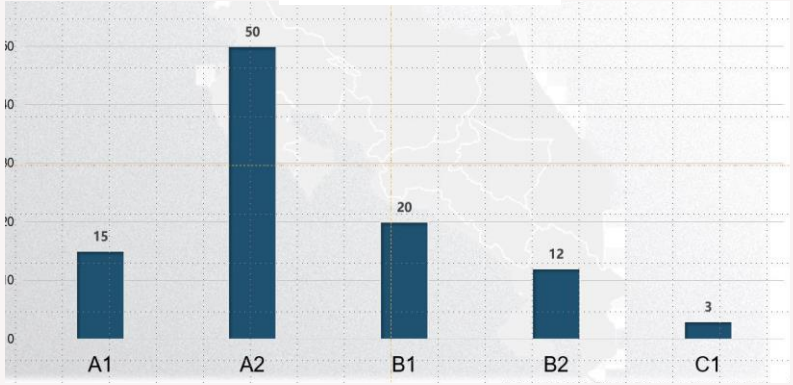


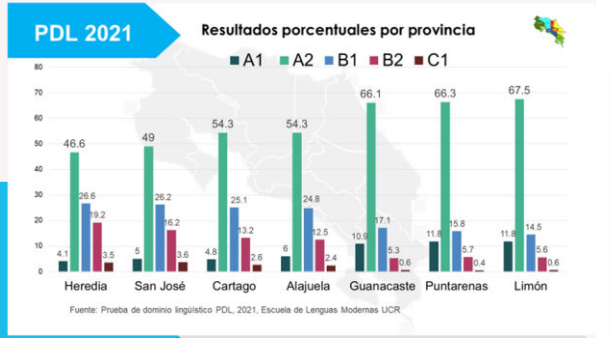
Ministerio de Educación Pública

English Proficiency Index

Reading

Listening





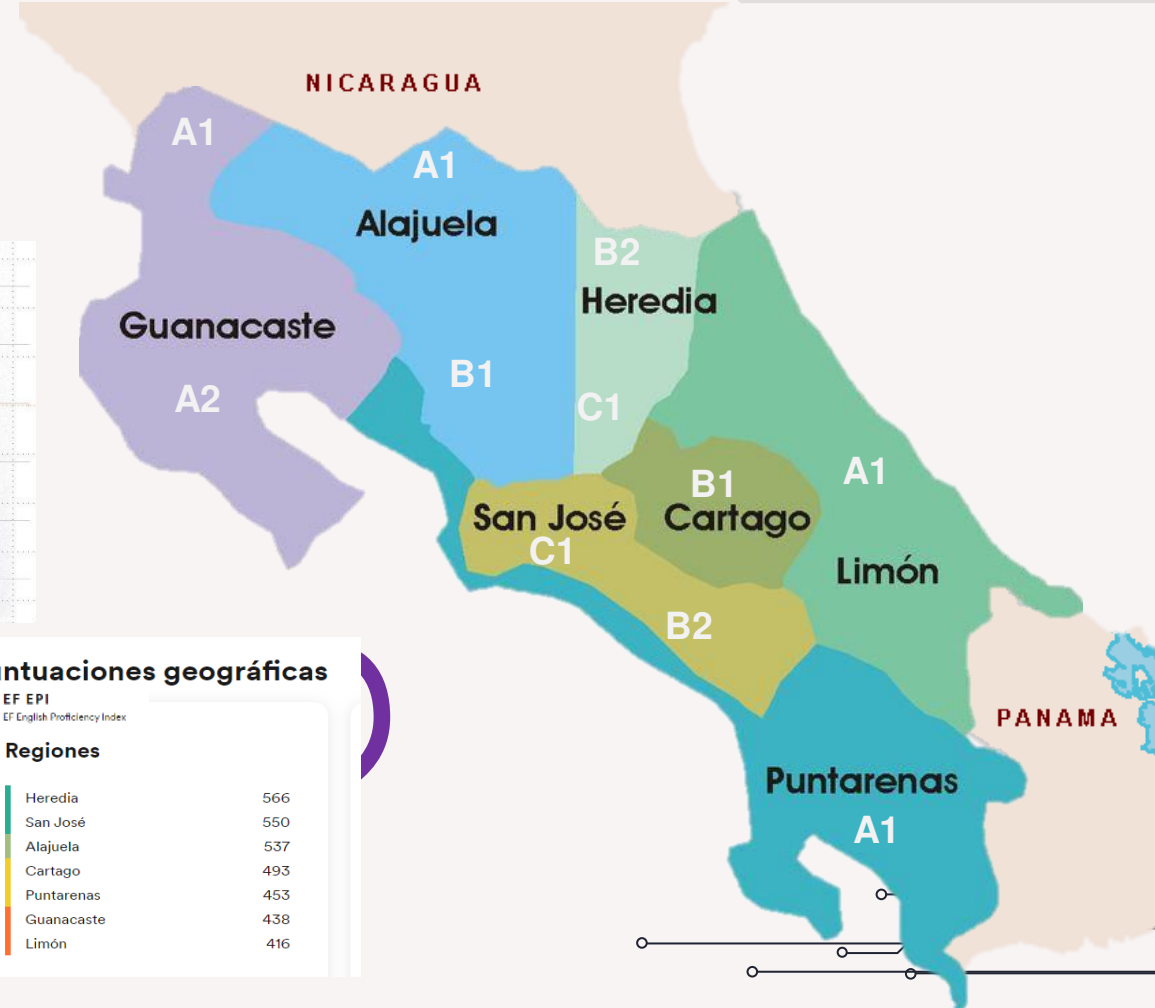
Puntuaciones geográficas

EF EPI

EF English Proficiency Index

Regiones

Heredia	566
San José	550
Alajuela	537
Cartago	493
Puntarenas	453
Guanacaste	438
Limón	416



Marcos de Referencias dominio de idiomas

El Marco Común Europeo de Referencia (MCER) es un marco de referencia de reconocimiento internacional para describir el dominio de un idioma (A1-C1).

Hay varios marcos de referencia con objetivos similares:

-Consejo Americano para la Enseñanza de Lenguas Extranjeras (ACTFL, American Council on the Teaching of Foreign Languages Proficiency Guidelines),

-Los exámenes de competencias lingüísticas canadienses (CLB, Canadian Language Benchmarks),

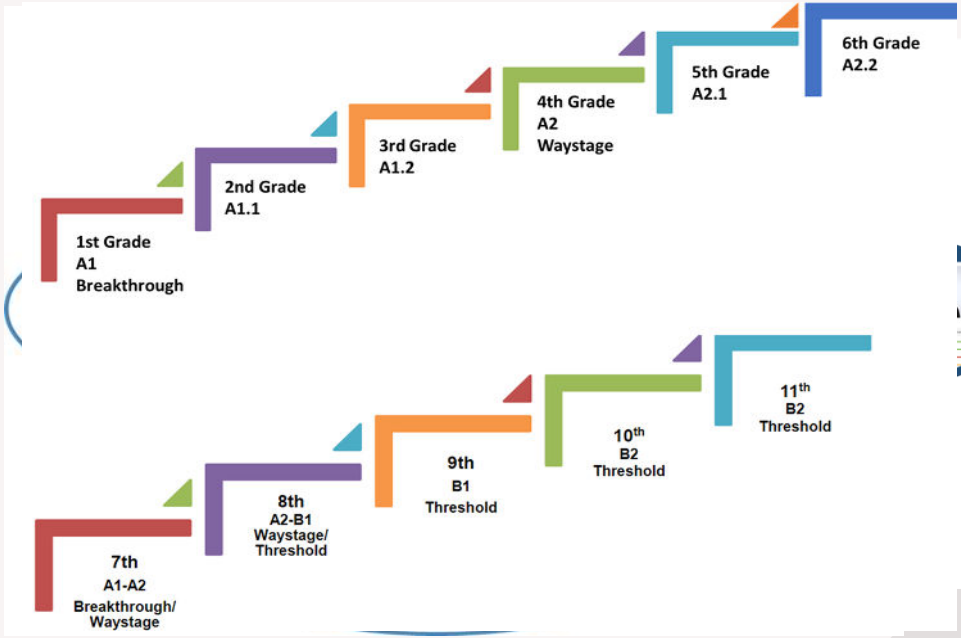
-La escala de competencia de la ILR (Interagency Language Roundtable) y

-WIDA (Wisconsin, Delaware, Arkansas Assessment language framework)

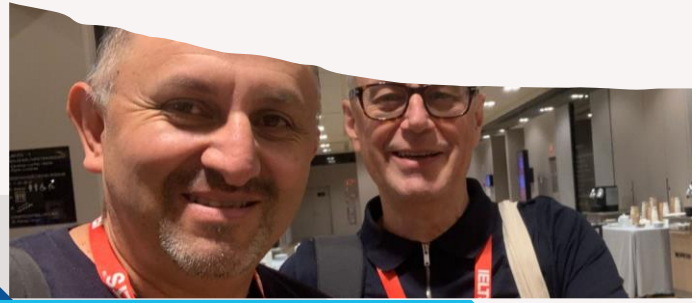
UNIVERSIDAD DE COSTA RICA

mep
Ministerio de Educación Pública

speechace




PELEX



turnitin

ALTE
Association of Language Testers in Europe

ECCI
Escuela de
Ciencias de la Computación e Informática

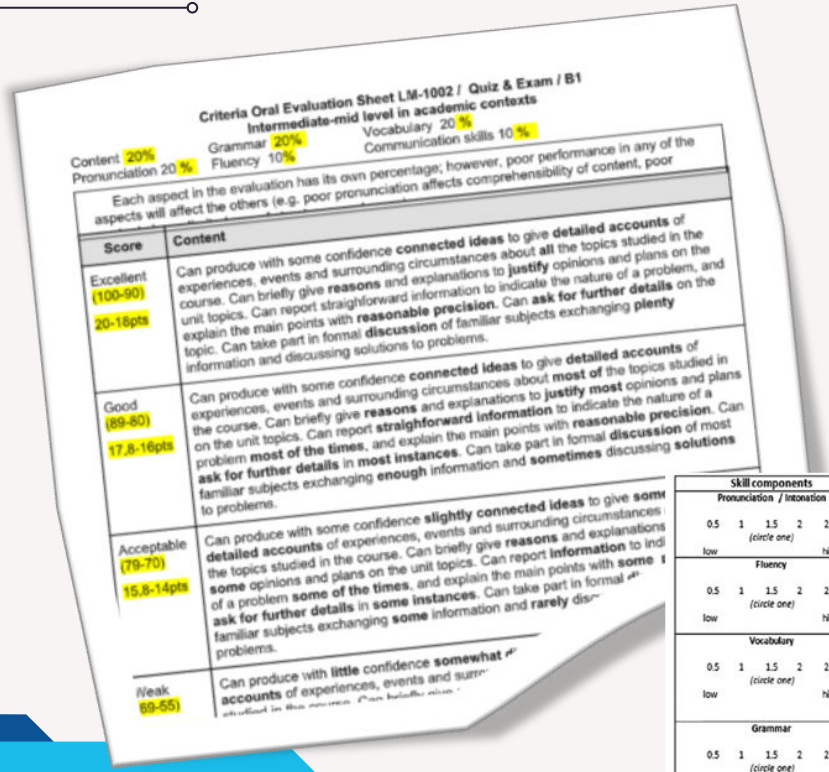
ILTA
INTERNATIONAL LANGUAGE TESTING ASSOCIATION

Laspau Affiliated with Harvard University

LAALTA
Latin American Association for Language Testing and Assessment

mep
Ministerio de Educación Pública

Rubrics



Criteria Oral Evaluation Sheet LM-1002 / Quiz & Exam / B1
Intermediate-mid level in academic contexts

Content 20% Grammar 20% Vocabulary 20 %
Pronunciation 20 % Fluency 10% Communication skills 10 %

Each aspect in the evaluation has its own percentage; however, poor performance in any of the aspects will affect the others (e.g. poor pronunciation affects comprehensibility of content, poor

Score	Content
Excellent (100-90) 20-18pts	Can produce with some confidence **connected ideas** to give **detailed accounts** of experiences, events and surrounding circumstances about **all** the topics studied in the course. Can briefly give **reasons** and explanations to **justify** opinions and plans on the unit topics. Can report straighforward information to indicate the nature of a problem, and explain the main points with **reasonable precision**. Can **ask for further details** on the topic. Can take part in formal **discussion** of familiar subjects exchanging **plenty** information and discussing solutions to problems.
Good (89-80) 17,8-16pts	Can produce with some confidence **connected ideas** to give **detailed accounts** of experiences, events and surrounding circumstances about **most of** the topics studied in the course. Can briefly give **reasons** and explanations to **justify most** opinions and plans on the unit topics. Can report **straighforward information** to indicate the nature of a problem **most of the times**, and explain the main points with **reasonable precision**. Can **ask for further details** in **most instances**. Can take part in formal **discussion** of most familiar subjects exchanging **enough** information and **sometimes** discussing **solutions** to problems.
Acceptable (79-70) 15,8-14pts	Can produce with some confidence **slightly connected ideas** to give **some** **detailed accounts** of experiences, events and surrounding circumstances the topics studied in the course. Can briefly give **reasons** and explanations **some** opinions and plans on the unit topics. Can report **information** to indi of a problem **some of the times**, and explain the main points with **some** **ask for further details** in **some instances**. Can take part in formal familiar subjects exchanging **some** information and **rarely** dis problems.
Weak 69-55)	Can produce with **little** confidence **somewhat** **accounts** of experiences, events and surr

Name: _____
Movie: _____
Topic: _____
Index card quiz ( ) Impromptu speech ( ) Prepared Speech ( ) Exam I ( ) Exam II ( )
Grade: _____

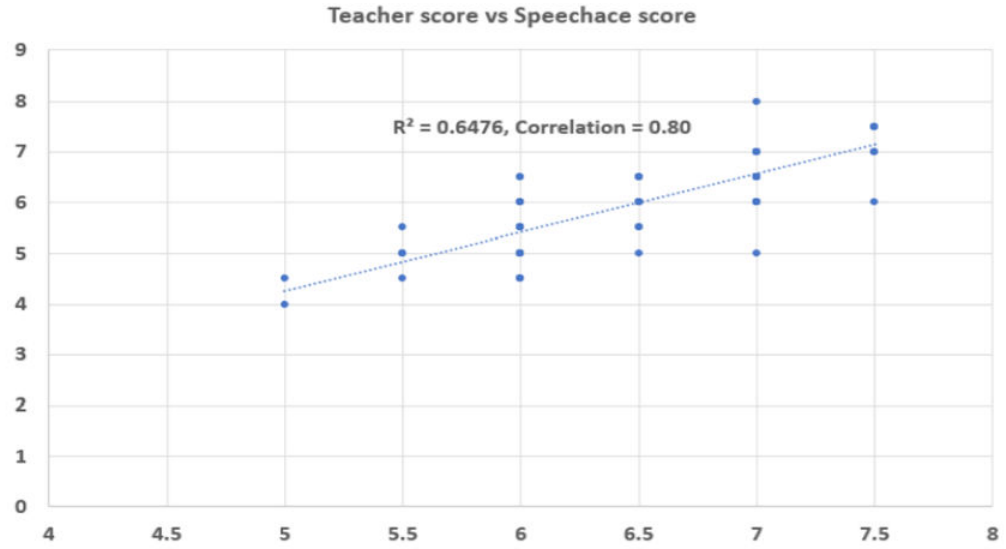
Content 50%	Language 50%		
	Pronunciation 20%	Grammar 20%	Vocabulary 10%
5 (25%)	5 (10%)	5 (10%)	5 (5%)
6 (30%)	6 (12%)	6 (12%)	6 (6%)
7 (35%)	7 (14%)	7 (14%)	7 (7%)
8 (40%)	8 (16%)	8 (16%)	8 (8%)
9 (45%)	9 (18%)	9 (18%)	9 (9%)
10 (50%)	10 (20%)	10 (20%)	10 (10%)

Errors in vocabulary or lack of it will affect content. / Errors in grammar and pronunciation will affect content. / Poor or inappropriate content will affect all the other areas.

Grammar & Vocabulary

Pronunciation

Skill components	Descriptors	Comments
Pronunciation / Intonation 0.5 1 1.5 2 2.5 *(circle one)* low high	Pronunciation / intonation is correct __ Always __ Most of the time __ Much of the time __ Sometimes __ Rarely	
Fluency 0.5 1 1.5 2 2.5 *(circle one)* low high	__ Uses native-like flow of speech __ **Uses fluent connected speech** __ Uses fluent connected speech, occasionally disrupted by search for correct form of expression __ Speech is connected but frequently disrupted by search for correct form of expression __ Uses simple sentences	
Vocabulary 0.5 1 1.5 2 2.5 *(circle one)* low high	__ Uses sophisticated vocabulary in a variety of contexts __ **Uses varied and descriptive language, possibly including native-like phrasing and / or idiomatic expressions** __ Uses vocabulary sufficient to communicate in most social and academic contexts __ Uses vocabulary sufficient to express needs and feelings and responds in familiar contexts __ Uses only basic vocabulary with possible use of first language	
Grammar 0.5 1 1.5 2 2.5 *(circle one)* low high	**Uses appropriate tenses, pronouns, gender and number agreement, negation, articles, prepositions and adjective placement** __ Always __ Most of the time __ Much of the time __ Sometimes __ Rarely	



speechace

Scores within

±0.5 IELTS

points of human grader

UCR

0.8

Correlation with human grader

ar Windows
Ve a Configuración para activar Window

TEYL/CAT Overview

1409

51%

8050

49 %

Table

Element	Field	Description
syllable_score_list[]	stress_level	Expected (Correct) lexical stress level based on the Lexicon. **Values:** • 0: unstressed • 1: primary stress • 2: secondary stress
syllable_score_list[]	predicted_stress_level	Detected lexical stress level based on the user audio. **Values:** • 0: unstressed • 1: primary stress • 2: secondary stress
syllable_score_list[]	stress_score	Floating point number between 0 and 100 indicating the correctness of the user's stress level.
syllable_score_list[]	pitch_range[]	[begin_pitch,end_pitch] recorded for this syllable in Hertz.
word_intonation_list[]	syllable_intonation_list[]	[pitch_change_from_previous, pitch_change_in_current] **pitch_change_from_previous:** indicates whether the pitch of a word rises, falls, or more or less unchanged. If we can't recognize the syllable (due to error from our side or due to user not saying the syllable), the value is null. If we can recognize the syllable but couldn't determine the pitch (due to error from our side or due to user reducing the sound to unvoiced), the value is REDUCED. **pitch_change_in_current:** is null unless one of the following cases occurs: If the pitch of the syllable falls, but the starting pitch of the syllable is higher than the ending pitch of previous syllable, secondary intonation is RISE while primary intonation is FALL. If the pitch of the syllable rises, but the starting pitch of the syllable is lower than the ending pitch of the previous syllable, secondary intonation is FALL while primary intonation is RISE.

Table

Field	Description
duration	total length of speech
articulation	total length of articula and non-speech event silence on very first se segment.
speech_rate	speaking rate in syllab
articulation_rate	articulation rate in syll
syllable_count	Count of syllables in th
word_count	Count of words in this
correct_syllable_count	Count of correctly spo
correct_word_count	Count of correctly spo
syllable_correct_per_minute	correct_syllable_count
word_correct_per_minute	correct_word_count /
all_pause_count	count of all pauses (fil minimum pause thresh
all_pause_duration	total duration of all pa
all_pause_list[]	a list of all the pauses extents of 10 msecs
mean_length_run	mean length of run in s
max_length_run	max length of run in sy
segment_metrics_list[]	A list of segments with metrics for each segm

Test design

No.	Question Type	Weight	Max score	Rubric
1	Read aloud	10%	B2	50% Pronunciation 50% Fluency
2	Task achievement Ask a question	15%	B2	50% Pronunciation & Fluency 50% Task achievement
3	Task achievement Describe 4 picture story	20%	B2	50% Pronunciation & Fluency 50% Task achievement
4	Task achievement Share an opinion and rationale	25%	B2	50% Pronunciation & Fluency 50% Task achievement
5	Open ended	30%	B2	25% Pronunciation 25% Fluency 25% Vocabulary 25% Grammar

There are only two seasons in Costa Rica, the rainy season and the dry season. People usually enjoy going to see a volcano or they can go to the beach on sunny days. I like both seasons. We can have a lot of fun!

Read aloud the text on the right (Lea en voz alta)

Preparation timer 30

Ask one question about the picture. (Haga una pregunta de la imagen)

Preparation timer 30

Describe the story in the pictures. (Cuente una historia de las imagines)

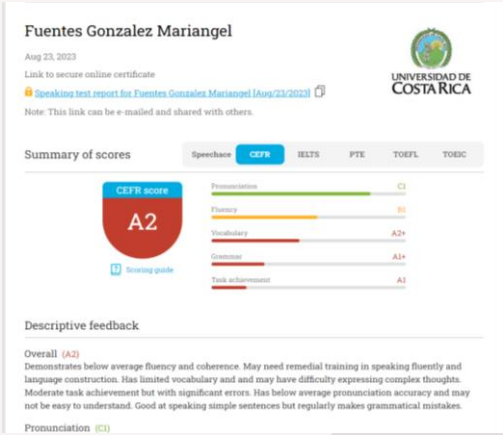
Preparation timer 30

Give an opinion on the following topic: Do you believe recycling is important? Why or why not? (Que opinas de reciclar, y por que?)

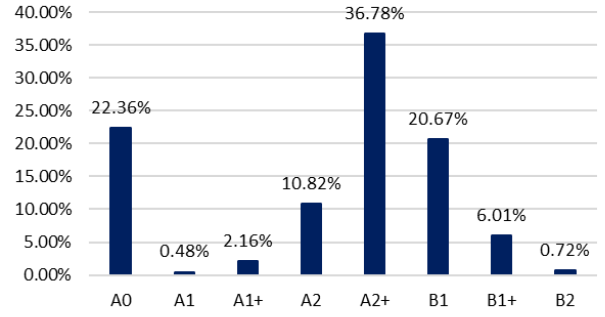
Preparation timer 30

Talk about your future vacation. Where to go? Who with? How will you get there? (Hable sobre sus vacaciones en el futuro)

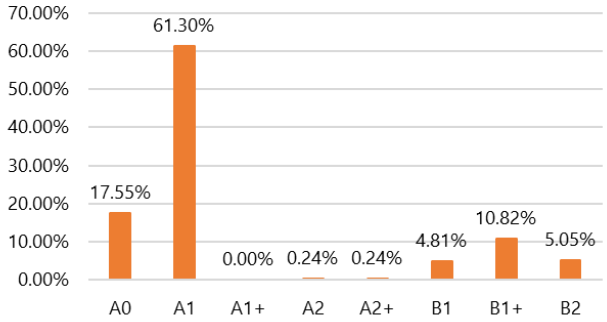
Preparation timer 30



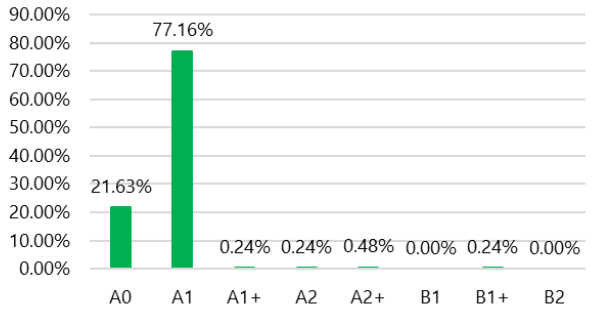
Q1 – Read Aloud



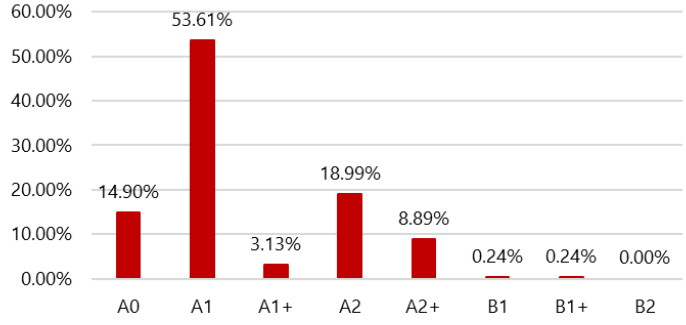
Q2 – Ask a question



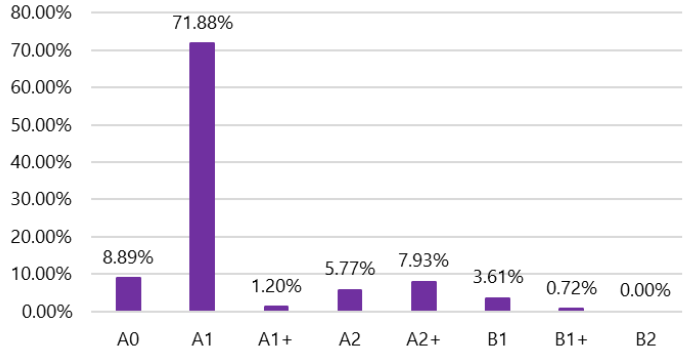
Q3 – Describe 4 pic story



Q4 – Share an opinion

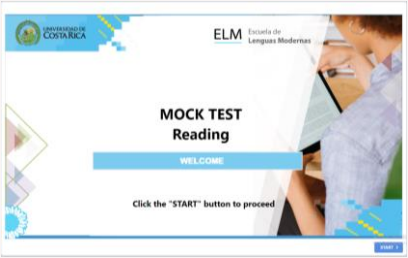


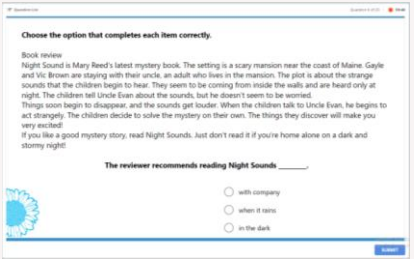
Q5 – Open ended

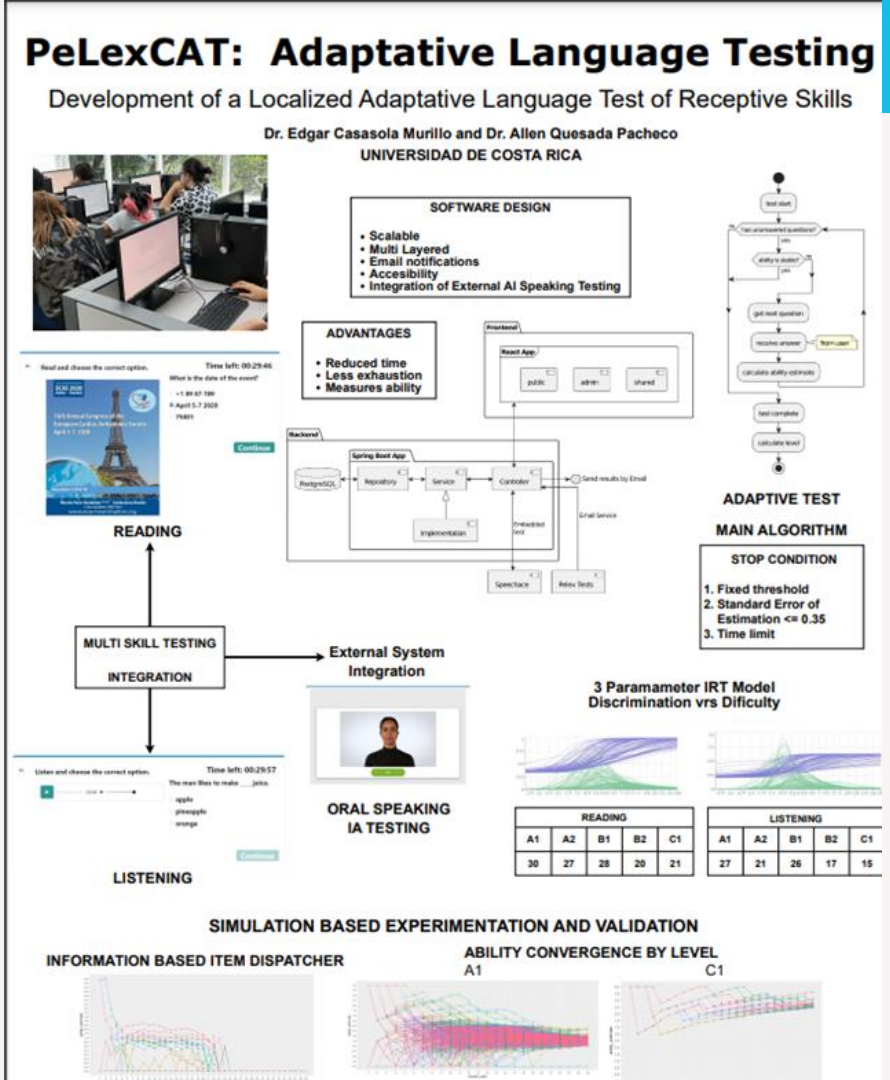


Prueba Inglés Adaptativa

- Escucha Lectura



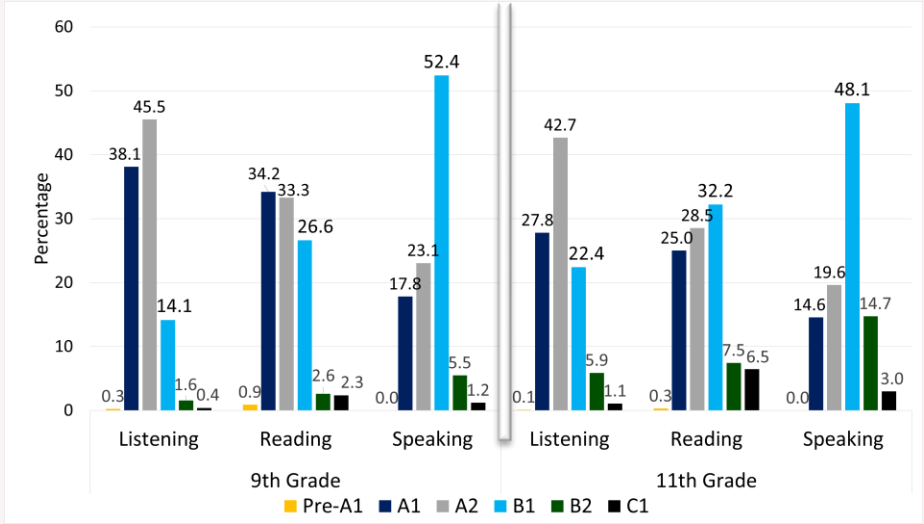




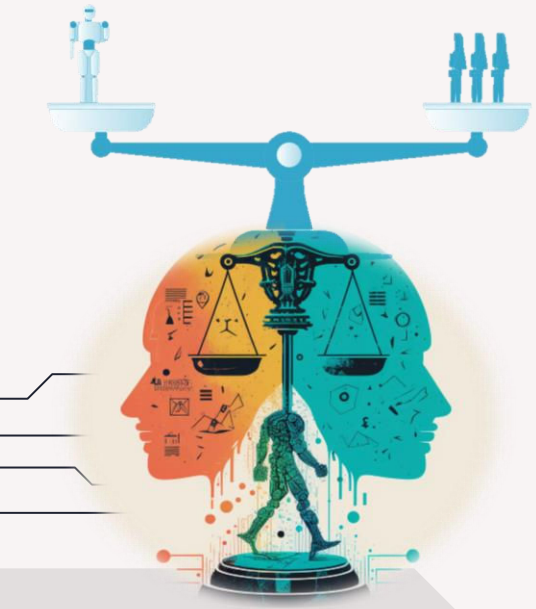
TEYL

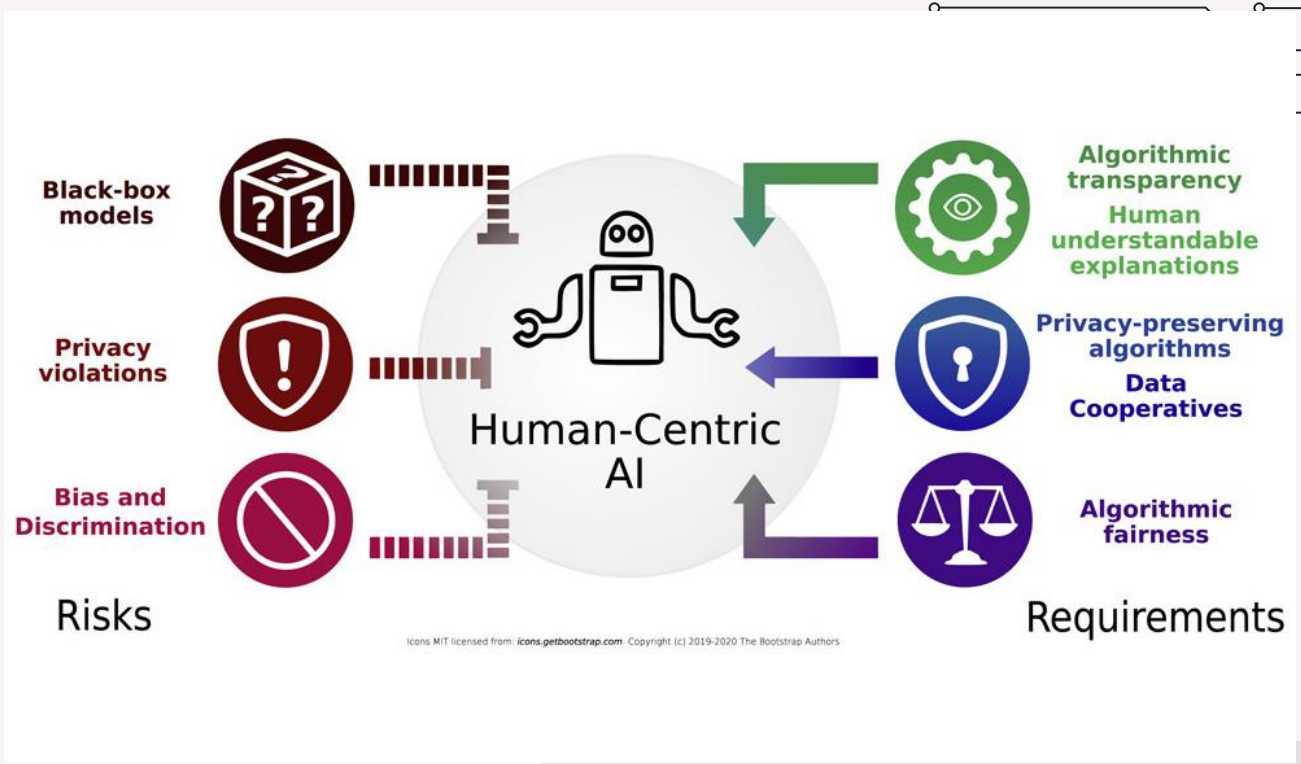
Banda	Reading	Listening	Speaking
Pre-A1	9.27	12.98	10.90
A1	56.94	58.31	60.19
A1+	16.58	18.57	13.80
A2	17.20	10.14	15.20
Total	100	100	100

Secondary



Ética en la IA





Lepri, B., Oliver, N., & Pentland, A. (2021). Ethical machines: The human-centric use of artificial intelligence. *iScience*, *24*(3), 102249. doi:10.1016/j.isci.2021.102249

LA INTELIGENCIA GENERATIVA (IG) EN LA EVALUACIÓN DE IDIOMAS

La IG puede aportar los siguientes beneficios a la evaluación de idiomas:

•**Personalización**: La IG puede utilizarse para crear evaluaciones personalizadas para cada estudiante.

•**Eficiencia**: La IG puede automatizar muchas tareas relacionadas con la evaluación, lo que puede liberar tiempo a los profesores para centrarse en otras actividades.

La IG se puede utilizar para aplicaciones específicas en la evaluación de idiomas, como:

•**Generación de preguntas**: La IG se puede utilizar para generar preguntas de evaluación, lo que puede ayudar a crear evaluaciones más desafiantes y relevantes.

•**Corrección de exámenes**: La IG se puede utilizar para corregir exámenes, lo que puede liberar tiempo a los profesores para centrarse en otras tareas.

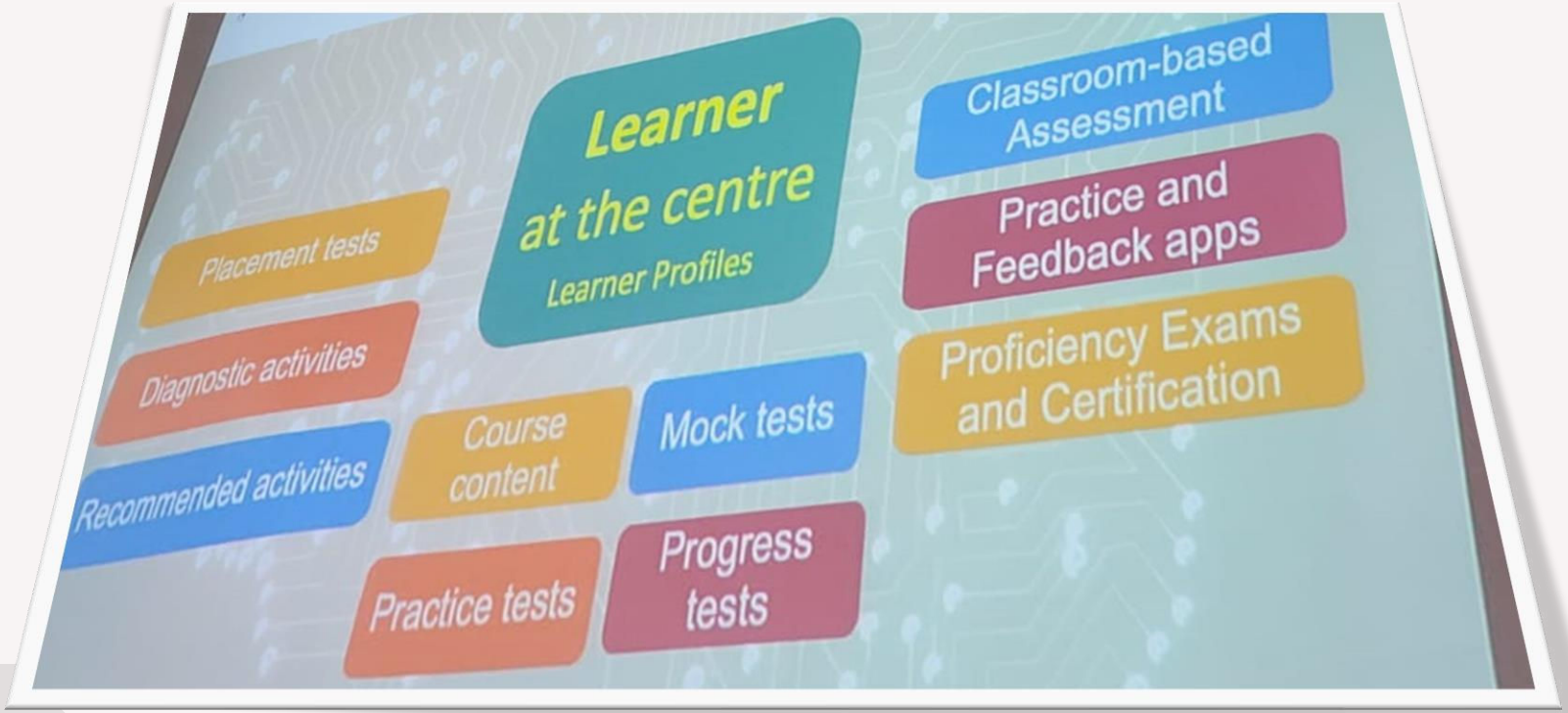
Retos y desafíos

La IG también plantea algunos retos y desafíos para la evaluación de idiomas, como:

•**Precisión**: La precisión de las herramientas de IG es todavía una preocupación, ya que pueden cometer errores que pueden afectar a la validez de las evaluaciones.

•**Sesgo**: Las herramientas de IG pueden estar sesgadas en función de los datos en los que se han entrenado, lo que puede afectar a la equidad de las evaluaciones.

INTEGRACIÓN DE LA EVALUACIÓN Y EL APRENDIZAJE



IMPACTO

Esta estrategia DE MONITOREO permite al MEP determinar la ubicación lingüística de la población estudiantil de acuerdo al MCER y sugerir recomendaciones curriculares para la mejora de cada modalidad escolar con un instrumento de evaluación personalizado (temas, nivel, y complejidad) con ayuda de la tecnología (IA/IG).

Gracias!

allenquesada@gmail.com

https://pelex.ucr.ac.cr