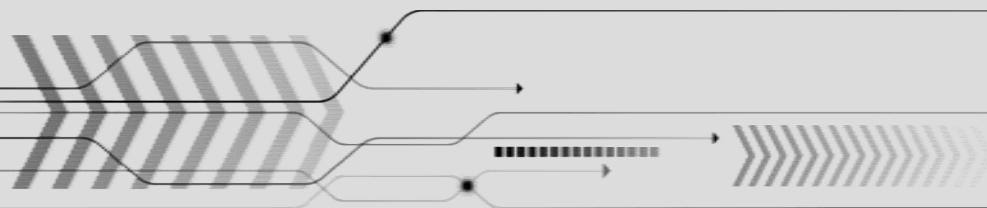# EPLIS VALIDITY EVIDENCES

AIM

To provide some validity evidences on the Aeronautical English Proficiency Exam (EPLIS) for the Brazilian Airspace System

OVERVIEW

VALIDITY

EVIDENCES

USOAP Universal Safety Oversight Audit Programme

Instituto de Controle
do Espaço Aéreo

FORÇA AÉREA BRASILEIRA
Asas que protegem o País

VALIDITY

VALIDITY

…it is important to point out that validity is not an all or none concept, but a matter of degree. Its inferential and ongoing nature makes that (Souza, Aragão 2023)
"the existing validity evidence becomes enhanced (or contravened) by new findings, and projections of potential social consequences of testing become transformed by evidence of actual consequences and by changing social conditions. Inevitably, then, validity is an evolving property and validation is a continuing process. (Messick 1989:1).

VALIDITY

Since evidence is always incomplete, validation is essentially a matter of making the most reasonable case to guide both current use of the test and current research to advance understanding of what the test scores mean" (Messick 1989:1).

EVIDENCES

External validation process (Unicamp, comprehensive report)

Topics covered: construct; specifications;
rater reliability;
Statistical analysis of the test and items; website.
Interviews, focal groups, questionnaires.

ICAO Aviation English Language Test Service
Test Assessment Report for SISCEAB English Language Proficiency
Exam (EPLIS)

EVIDENCES

Initial rating training course (proficiency, performance, analytic and holistic rating scales, ICAO categories in depth)
2 phases- theoretical/ simulation and
Practical

VALIDITY EVIDENCES

Doc 9835

Interlocutors are trained and monitored annually (remedial training, if necessary, readapted)

Raters are trained and their agreement rates are presented individually (agreement- disagremeent, ele/sme, dove/hawk)

	Audio A1							Audio A2							Audio A3	Audio A4							Audio A5							Audio A6							Audio A7	Audio A8	Audio A9	Audio A10							Audio A11						
	P	E	V	F	C	I	NP	P	E	V	F	C	I	NP	NP	P	E	V	F	C	I	NP	P	E	V	F	C	I	NP	P	E	V	F	C	I	NP	NP	NP	NP	P	E	V	F	C	I	NP	P	E	V	F	C	I	NP
T1	6	5	6	6	5	6	5	4	3	3	4	4	3	3	4	5	4	4	4	4	4	4	4	4	4	4	5	4	4	5	4	5	4	4	5	4	5	4	3	4	4	4	5	5	5	4	5	5	5	5	6	5	5
T2	6	6	6	6	6	6	6									5	5	4	4	4	4	4	3	3	4	4	4	4	3	5	5	5	5	5	4	5				5	5	5	5	6	5	5	6	6	6	6	6	5	5
T1	6	6	6	6	6	6	6	3	3	3	3	3	3	3	4	5	4	3	3	3	3	3	4	4	5	5	4	5	4	5	5	5	5	4	4	4	4	3	3	5	4	4	5	5	5	4	4	4	4	4	5	5	4
T2																																					5	4		5	5	5	5	6	5	5	4	4	4	4	4	5	4
T1																																																					
T2																																																					
T1	6	6	6	6	6	6	6	4	3	3	4	4	3	3	4	4	5	5	4	4	4	4	3	3	4	4	4	4	3	5	5	5	5	5	5	5	5	4	3	5	5	5	5	6	5	5	4	4	4	4	6	5	4
T2																																																					
T1	6	6	6	6	6	6	6	4	3	3	3	4	3	3	4	5	4	4	4	5	5	4	3	3	4	4	4	4	3	5	5	5	5	5	5	5	5	4	3	5	5	6	5	6	6	5	4	4	4	4	4	5	4
T2								4	3	3	4	4	3	3		4	5	4	4	4	4																			5	5	6	5	6	5	5	4	4	4	4	6	5	4
T1	6	6	6	6	6	6	6	4	3	4	4	4	4	3	4	4	4	4	4	4	4	4	3	4	4	4	4	4	3	5	5	5	5	5	5	5	5	4	3	4	4	5	5	6	5	4	5	4	4	4	6	5	4
T2								4	3	3	4	3	3	3		4	5	4	4	4	4		3	3	4	4	4	4	3											5	5	5	5	6	5	5							
T1	6	6	6	6	6	6	6	3	3	3	3	4	3	3	4	4	4	4	4	3	3	3	4	4	4	4	4	4	4	6	6	6	6	6	6	6	5	4	3	5	5	5	5	5	5	5	5	4	5	4	6	5	4
T2								4	3	3	4	4	3	3																5	5	5	5	5	5	5				5	5	5	5	6	5	5	4	4	4	4	6	5	4
T1	6	6	6	6	6	6	6	4	3	3	4	4	3	3	4	5	5	5	4	4	4	4	4	4	4	4	5	5	4	5	5	5	5	4	5	4	5	4	4	4	4	4	5	5	4	4	4	4	4	4	5	5	4
T2																							4	4	4	4	5	5	4	5	5	5	5	5	5	5			4	5	5	5	5	5	5	5	4	4	4	4	6	5	4

EVIDENCES

Individual performance

Audio A5	3	3	3	4	4	4	3
% categories	83.33% agreement						
Audio A6	5	5	5	5	5	5	5
% categories	66.67% agreement						
audio A7	5						
audio A8	4						
audio A9	3						
audio A10	5	5	5	5	6	5	5
% categories	100.00% agreement						

EVIDENCES

Recurrent training

Group performance – per audio

Variety of audios _ prototypiycal , border line, problematize

A14	4	3	3	3ou 4	3	3	3	KEY
	3	3	3	3	3	3	3	GROUP
	59%	100%	100%	90%	97%	100%	100%	%

EVIDENCES

The rating process is documented.

rating process -blind rating – at least two raters participate in the rating of tests, with a third expert rater consulted in case of divergent scores.

Holistic(interlocutor)
Analytic (rater)

EVIDENCES

Test design and items

Test developers – qualifications doc 9835
item writers (SME ELE)- item writing training (test construct and specifications)

Test construct - Specifications – Revised periodically

VALIDITY EVIDENCES

A brief analysis of language tasks used by air traffic controllers in radiotelephony communications in Brazil

Matilde V. R. SCARAMUCCI

Beatriz F. ARAGÃO

Instituto de Controle
do Espaço Aéreo

FORÇA AÉREA BRASILEIRA
Asas que protegem o País

VALIDITY EVIDENCES

Tasks by frequence

The tasks and subtasks were classified into daily, weekly, monthly, quarterly, biannual, annual or rare in relation to their occurrence in real life.

Tasks by relevance

Most operationally demanding tasks

Most linguistically relevant tasks (in terms of use of plain English)

VALIDITY EVIDENCES

Language functions: giving orders, informing something, explaining something, issuing clearance, accepting, denying, issuing instructions, warning, asking for information, giving reasons, querying the pilot, asking about possibility, asking about the nature of the emergency, requesting something, giving instructions and directions, giving options, prompting the pilot to do something and forwarding information.

VALIDITY EVIDENCES

Receptive or productive skills- language tasks used by pilots and controllers may require different skills (productive and /or receptive skills)

Phraseology or plain English- it has been noted that the distinction between phraseology and plain English is not always clear and obvious. The events may require a higher use of one or the other, but many times they are interchangeable and overlapping .

Air traffic control facility (TWR, APP and ACC)

Some statistical evidences

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	5	13.9	14.7	14.7
	B	2	5.6	5.9	20.6
	C	17	47.2	50	70.6
	D = KEY	10	27.8	29.4	100
	Total	34	94.4	100	
Missing	No answer	2	5.6		
Total		36	100.0		

Distractor analysis

Facility index - Descriptive Statistics Frequencies, scale of 0 to 1 or as a percentage.

Some statistical evidences

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	5	13.9	14.7	14.7
	B	2	5.6	5.9	20.6
	C	17	47.2	50	70.6
	D = KEY	10	27.8	29.4	100
	Total	34	94.4	100	
Missing	No answer	2	5.6		
Total		36	100.0		

Discrimination index Scale Reliability Analysis (-1 a +1an item with a high discrimination index, close to +1, shows that strong test takers are answering the item correctly . If the index is negative, it means the strong test takers are getting the item wrong. This indicates that there might be a problem with the key. Items with a value of 0.30 are considered suitable). = > 0.30

Some statistical evidences

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	5	13.9	14.7	14.7
	B	2	5.6	5.9	20.6
	C	17	47.2	50	70.6
	D = KEY	10	27.8	29.4	100
	Total	34	94.4	100	
Missing	No answer	2	5.6		
Total		36	100.0		

Discrimination index The item in the example provided had a discrimination index of 0.247. Because this item was considered difficult for this group of pretest takers, its discrimination might have been underestimated as strong and weak groups scored badly. Anyway, the item needs to be revised.

Some statistical evidences

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Item 767	22.3889	15.844	.542	.794
Item 777	22.3889	16.759	.129	.808
Item 787	22.4722	15.685	.436	.796
Item 790	22.7222	15.806	.271	.805
Item 791	23.0278	16.028	.247	.805

Internal consistency Cronbach Alpha coefficient (from –1 to +1) >0.7 (more than) The higher the Cronbach's Alpha value is, the higher the internal consistency (same construct) The Cronbach's Alpha for the test as a whole was calculated in .087, (a high level of reliability)

EVIDENCES

How Can Listening Contribute to Aviation Safety? EPLIS Paper 1 under the Spotlight
Paula Ribeiro e SOUZA, Beatriz Faria ARAGÃO 40–52

Applied Linguistics Papers Journal - "Aviation English Special Volume: Enhancing Efficiency in Aeronautical Communications – 9th GEIA Proceedings" Volume 27/1

USOAP

Contracting States – adherence to SARPS

8 different areas: legislation, organization, licensing, operations, airworthiness, accident investigation, air navigation services, aerodromes.

USOAP EVIDENCES

Primary aviation legislation and specific operating regulations (LEG) :23

Civil aviation organization (ORG); 13

Personnel licensing and training (PEL); 93

Aircraft operations (OPS); 126

Airworthiness of aircraft (AIR); 186

Aircraft accident and incident investigation (AIG);84

Air navigation services (ANS); 122

Aerodromes and ground aids (AGA). 143

USOAP
Protocol questions
AREA
QUESTION
6.655
PEL

Has the State implemented a system for formally evaluating and re-evaluating the ability of air traffic controllers, and aeronautical station operators to speak and understand the language used for radiotelephony communications at the level specified in Appendix 1 to Annex 1?
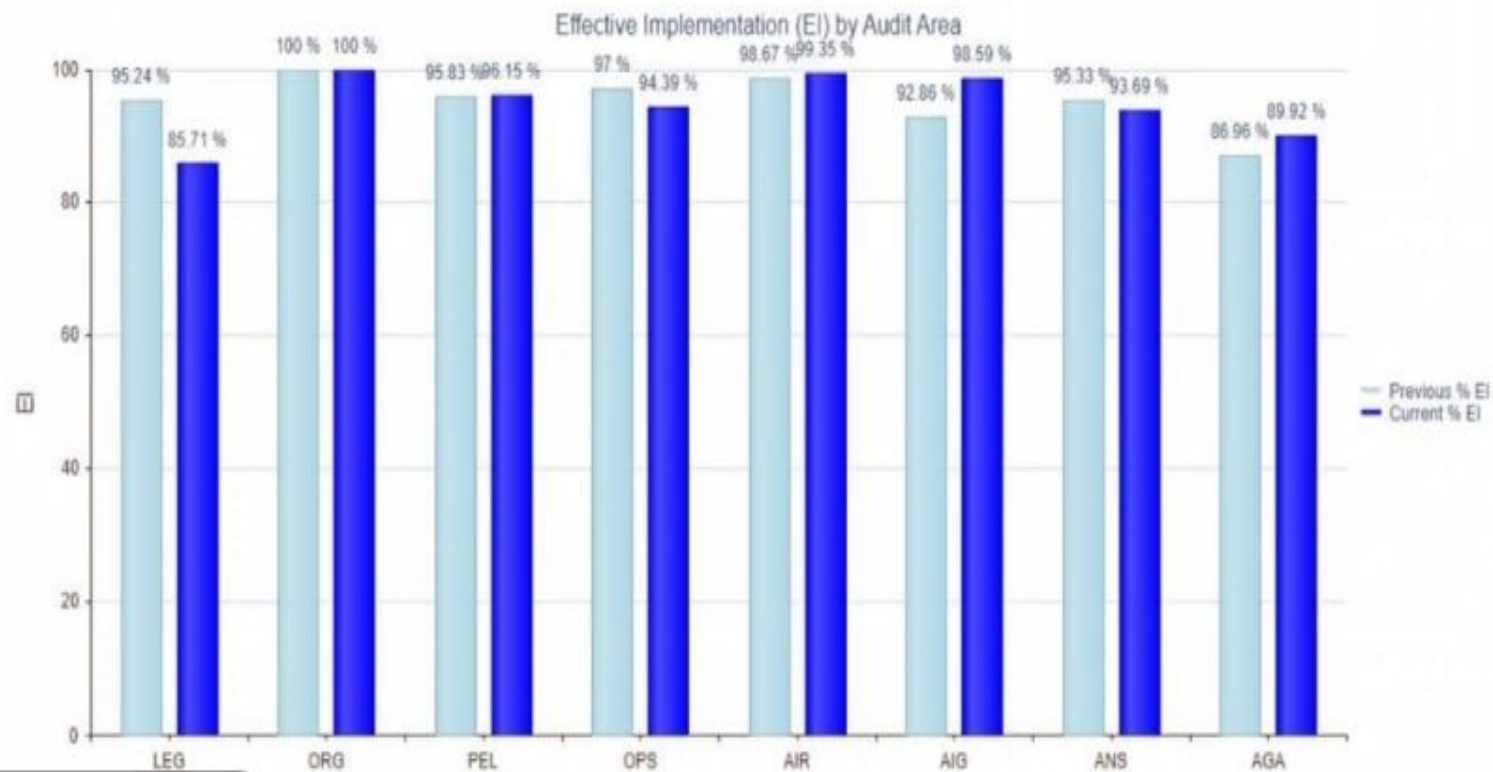
Check implementation of the system for endorsement of language proficiency ratings in licences.
   a) Check if the date of validity is clearly indicated on the licence.
   b) If the licence holder has been assessed at Level 4 or higher in English and another language, it should also be indicated on the licence.
   c) If a licence holder has failed to demonstrate a language proficiency of at least ICAO Operational Level (Level 4), it should also be clearly indicated on the licence.

source:DECEA

Thanks!

10 YEARS GEIA

Instituto de Controle do Espaço Aéreo — ICEA

Departamento de Controle do Espaço Aéreo — DECEA

FORÇA AÉREA BRASILEIRA
Asas que protegem o País