

Building an Interpretation and Use Argument for EPLIS: a validity study

Natalia de Andrade (DECEA/UNICAMP)

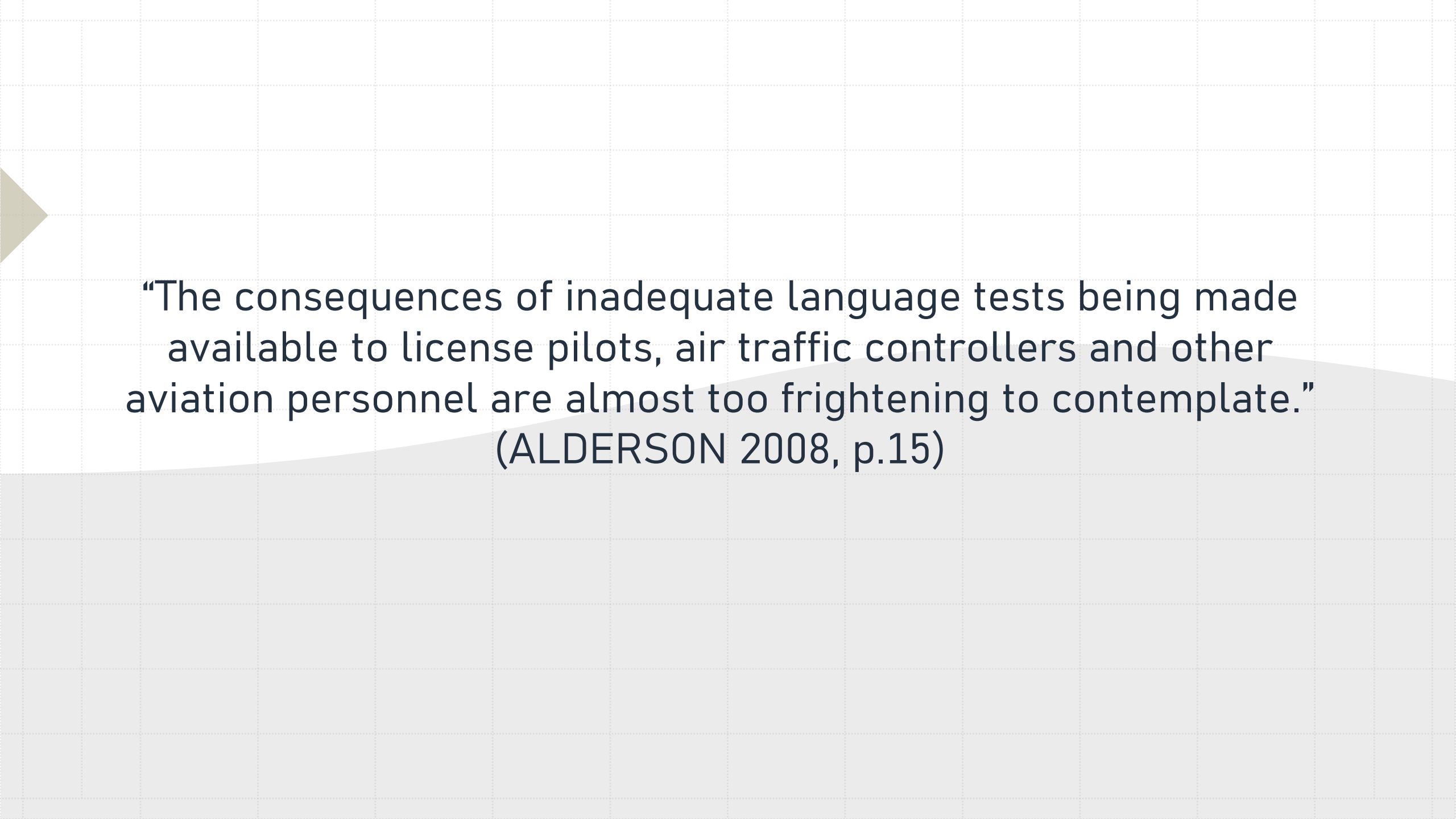
Supervisor: Dr. Matilde Scaramucci

YEARS



UNICAMP

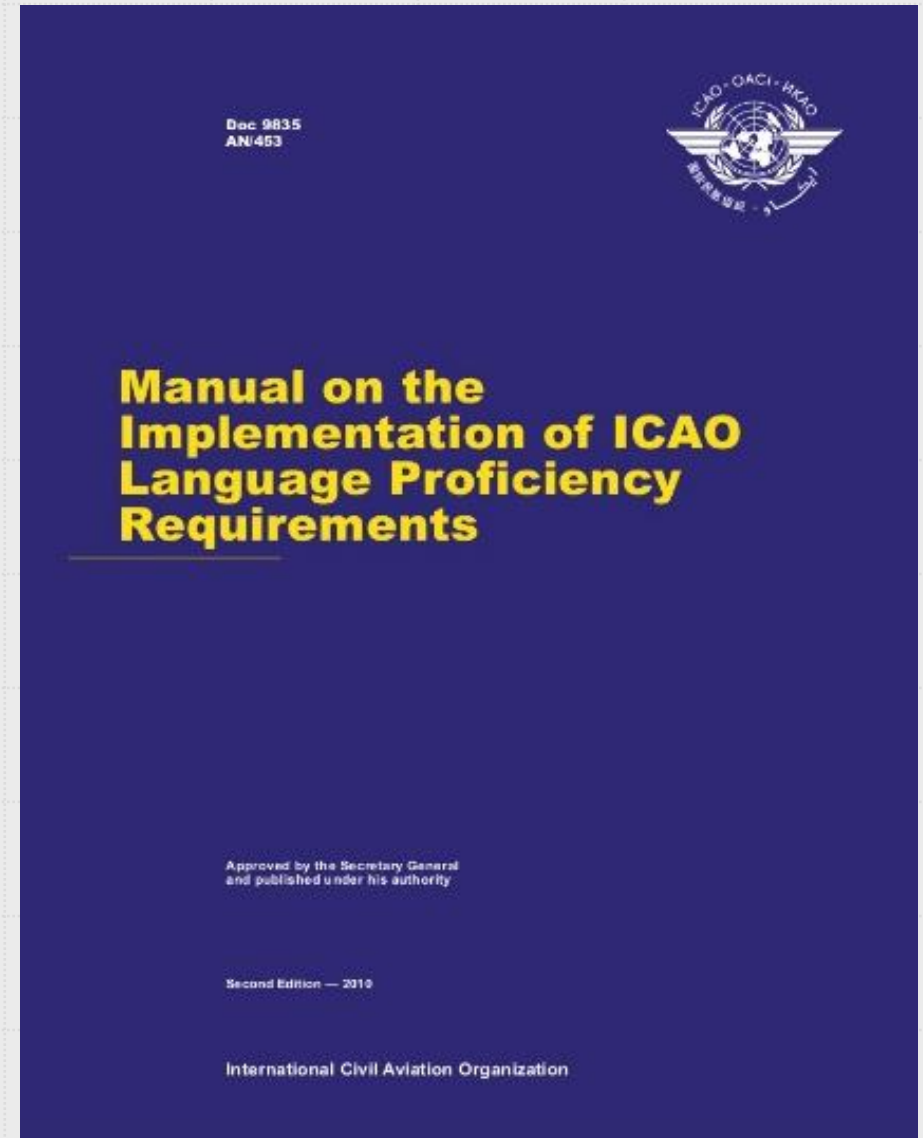




“The consequences of inadequate language tests being made available to license pilots, air traffic controllers and other aviation personnel are almost too frightening to contemplate.”
(ALDERSON 2008, p.15)

The international language policy for aviation (ICAO 2004; 2010)

- stipulates that all non-native English speaking pilots and controllers engaging in the international operation should take a specific purpose test and prove that they have the minimum required operational level of proficiency;
- six assessment criteria: Pronunciation, Structure, Vocabulary, Fluency, Comprehension, and Interactions;
- six levels of proficiency;
- **stipulates level 4 as the minimum for international operations.**





The national exam: EPLIS

- High-stakes test;
- Performance test;
- For Brazilian ATCO and AEO;
- Based on ICAO LPRs;
- Developed and maintained by the Brazilian Air Force;
- External validation process: 2010 and 2014

Research aims

Interpretation
and Use
Argument


Validity
Argument
focused on EPLIS
consequences

Argument-based approach to test validation

- It is aimed at investigating empirical evidence to support the warrants underlying the inferences made by the test developers;
- An interpretation and use argument indicates the proposed interpretations and uses of test results and serves as the first step in developing a validity argument for a test (Kane 2006, 2013);
- The type of evidence required for validation is determined by the claims being made;
- Kane (2013, p.1) “to validate an interpretation or use of test scores is to evaluate the plausibility of the claims based on the test scores”;
- “It is essential to analyze the claims made by the test developers with regards to the test use prior to start a validation study” (p.43);
- “The more ambitious the claims, the more evidence will be needed” (p.45).

Research Methodology

→ Document analysis



The screenshot shows the top part of the EPLIS website. At the top, there is a dark red banner with the text 'Exame de Proficiência em Inglês Aeronáutico' and 'EPLIS' below it. Below the banner is a blue navigation bar with the following menu items: 'Histórico', 'Validação', 'Informações', and 'Suporte'. Below the navigation bar is a main content area with a photograph of a person in a military uniform sitting at a computer workstation. To the right of the photograph, there is a text block that reads: 'Aqui você encontra informações sobre o Exame de Proficiência em Inglês Aeronáutico e pode solicitar alguma dúvida ou suporte.' Below the photograph, there is a 'Notícias' section with a date 'Ago/23' and the text 'ENTREVISTAS PARA O CURSO CTP019A'.



ENSINO

MCA 37-225

**MANUAL DOS
REQUISITOS DE PROFICIÊNCIA EM INGLÊS
AERONÁUTICO**

Manual do Candidato EPLIS



COMANDO DA AERONÁUTICA



PROTEÇÃO AO VOO

ICA 63-33

Intended Uses and Consequences

ICAO LPRs



U1- Personnel Licensing



C1- Ensure safety in the international airspace

EPLIS



U1- Personnel Licensing



C1- Ensuring safety in the Brazilian Airspace



EPLIS

U2- Placement for air
traffic shifts

C2- Mitigation measures

EPLIS

U3- Placement for
aviation English courses

C3- More focused
training programs

Interpretation and Use Argument

(Based on the framework by Chapelle et al, 2008)

Consequences: The Brazilian air space is likely to be safer.

↑ **IMPLICATION**

Test use: The controllers assessed as level 4, 5 or 6 in EPLIS are licensed to control international traffic.

↑ **UTILIZATION**

Target Score: The controllers are likely to obtain acceptable scores on other indicators of aviation English proficiency, including self-assessment and more experienced controllers' judgments about performance in real life communications.

↑ **EXTRAPOLATION**

Construct: The controllers' high level of performance in EPLIS can be explained by his high level of aviation English proficiency

↑ **EXPLANATION**

Expected score: The controller is likely to receive reliable scores consistently rated by different raters.

↑ **GENERALIZATION**

Observed scores: The air traffic controllers' performances are elicited and assessed to accurately determine whether the performance reached the operational level recommended by ICAO.

↑ **EVALUATION**

Observation: Observations of performances of air traffic controllers in EPLIS are appropriate.

↑ **DOMAIN DESCRIPTION**

Target domain: The language used in radiotelephony communications between Brazilian air traffic controllers and international pilots was analyzed and served as basis for task development.

Warrants, Assumptions, and Backing in the Interpretation and Use Argument of EPLIS

Inference in the Interpretive Argument	Warrant Supporting the Inference	Assumptions Underlying the Warrant	Analysis to obtain Backing for Assumptions	Possible rebuttals
<p>Domain Description</p>	<p>Observations of performance on the EPLIS tasks are representative of relevant knowledge and skills in situations representative of those in the target language use domain of radiotelephony communications between Brazilian air traffic controllers and international pilots.</p>	<p>1) Assessment tasks that are representative of the target language use domain can be identified; 2) Aviation English skills, knowledge, and processes needed for radiotelephony communications can be identified.</p>	<p>Domain analysis (subject matter expert consensus, document analysis)</p>	<p>Compliance with ICAO's policy that recommends tasks that do not relate to the target domain</p>
<p>Evaluation</p>	<p>Observations of performances of air traffic controllers on EPLIS are evaluated to provide observed scores reflective of targeted language abilities.</p>	<p>3) Both analytic and holistic rubrics for scoring performance are appropriate for providing evidence of targeted language abilities. 4) Task administration conditions are appropriate for providing evidence of targeted aviation English performance.</p>	<p>Rubric analysis (document analysis, subject matter expert consensus) Audit of test administration conditions and document analysis</p>	<p>Strategic competence is taken for granted in the rating scale developed by ICAO. Some Institutions are unable to comply with the test administration documentation</p>

Warrants, Assumptions, and Backing in the Interpretation and Use Argument of EPLIS

<p>Generalization</p>	<p>Observed scores are estimates of expected scores over the relevant parallel versions of tasks and test forms and across raters.</p>	<p>5) Enough tasks are included on EPLIS.</p> <p>(6) Task specifications are well defined for parallel task creation.</p> <p>(7) Ratings of different raters are consistent.</p>	<p>Generalizability and reliability studies</p> <p>Systematic task specification analysis</p> <p>Inter-rater reliability</p>	<p>Lack of statistical analysis as pointed out by AELTS</p>
<p>Explanation</p>	<p>Expected scores are attributed to the construct of aviation English proficiency for Brazilian air traffic controllers.</p>	<p>(8) Performance on EPLIS relates to performance on other aviation English assessment.</p> <p>(9) Strategies engaged by EPLIS tasks are construct relevant.</p>	<p>Concurrent correlational studies</p> <p>Discourse analysis of test takers' think-aloud protocol.</p>	<p>Other aviation English tests may not have the same construct.</p>

Warrants, Assumptions, and Backing in the Interpretation and Use Argument of EPLIS

Extrapolation	<p>The construct of aviation English proficiency for Brazilian air traffic controllers accounts for the quality of the performance of Brazilian air traffic controllers in real ATC communications with international pilots.</p>	<p>(10) The aviation English for Brazilian air traffic controllers assessed by EPLIS tasks accounts for other indicators of communicative language ability in radiotelephony communications in English within the Brazilian Air Space.</p>	<p>Test takers' self-assessment of their own aviation English, observation of real-life communications, discourse analysis of incident investigation, more experienced controllers' judgement</p>	<p>Strategic competence and operational knowledge might be critical factors to succeed in real life communications.</p>
Utilization	<p>Estimates of the quality of performance on EPLIS are useful for making decisions about controllers' licensing and follow-up ATC training.</p>	<p>(11) The meaning of the EPLIS scores is clearly interpretable by test users and test takers.</p>	<p>Score descriptors are provided to test takers and test users along with their test result.</p> <p>National Language Policy is developed based on EPLIS scores.</p>	
Implications/Consequences	<p>The consequences of using the EPLIS and the decisions that are made based on EPLIS scores are beneficial to the controllers and other stakeholders</p>	<p>(12) There hasn't been an accident involving language since EPLIS has been implemented.</p> <p>(13) The test raises awareness of the importance of English proficiency for radiotelephony communication</p>	<p>Analysis of accident/incident investigation reports.</p> <p>Expert interviews, follow-up questionnaires and interviews with controllers</p>	

Final remarks

- Aim of the IUA: present ways to gather data to serve as basis for the development of the validity argument of EPLIS;
- Sketching the IUA of a test makes one realize how much work is needed in order to improve both an assessment process and assessment policies.

→

References

- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- BRASIL. Departamento de Controle do Espaço Aéreo. Publicações. *Manual de implementação dos requisitos de proficiência em inglês (MCA 37-225)* 2018.
- Chapelle, C. A., Enright, M. E., & Jamieson, J. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language*. London: Routledge.
- Chapelle, C. A. (2020). *Argument-based validation in testing and assessment*. Thousand Oaks, CA: Sage.
- Douglas, D. (2000). *Assessing languages for specific purposes*: Cambridge University Press.
- Douglas, D. (2014). Nobody seems to speak English here today: Enhancing assessment and training in aviation English. *Iranian Journal of Language Teaching Research*, 2(2), 1-12.
- ICAO. (2010). *Manual on the Implementation of ICAO Language Proficiency Requirements (2nd ed.)*: International Civil Aviation Organization. International Civil Aviation Organization.
- Kane, M. (2006). Validation. In R. Brennen, (Ed.), *Educational Measurement (4th Edition)*, (pp 17-64). Westport, CT: Greenwood Publishing.
- Kane, M. E. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement* 50 (1), 1-73.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell Publishing.
- McNamara, T. (1996). *Measuring second language performance*. London: Longman.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.) *Educational measurement*. (3rd ed.) pp. 13-103. NY: Macmillan Publishing Co.



Thank you!

nandraderay@gmail.com

natalianar@decea.mil.br