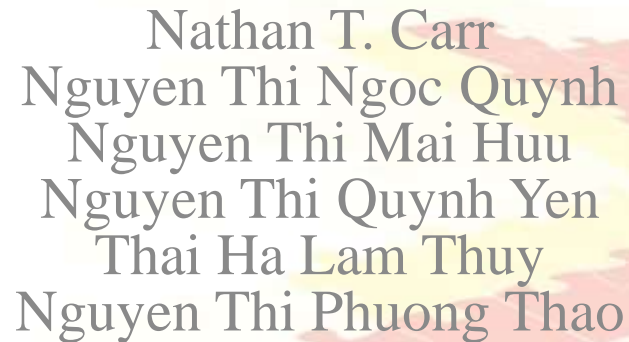




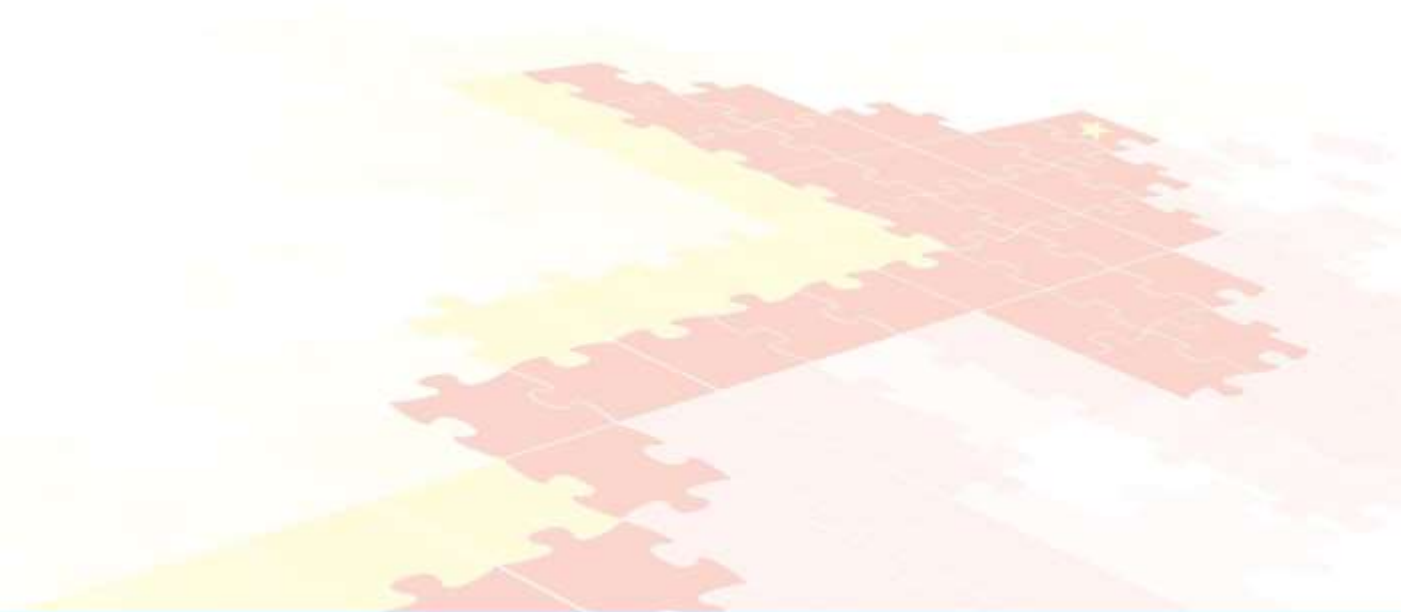
Systematic Support for a Communicative Standardized Proficiency Test in Vietnam



Nathan T. Carr
Nguyen Thi Ngoc Quynh
Nguyen Thi Mai Huu
Nguyen Thi Quynh Yen
Thai Ha Lam Thuy
Nguyen Thi Phuong Thao

Content

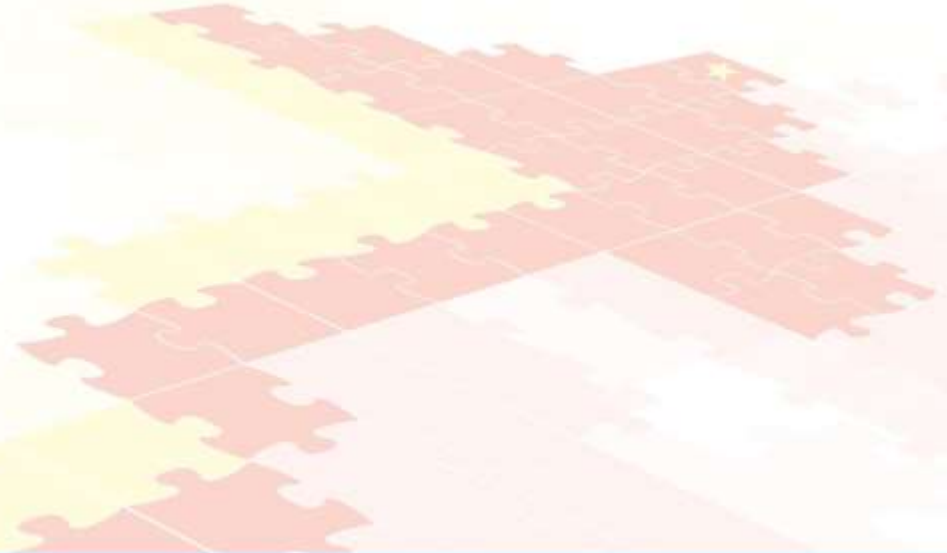
1. Background of language testing and assessment of Vietnam
2. Assessment Use Argument (AUA)



BACKGROUND

DIFFERENT PRODUCTS OF TESTS

- VSTEP.3-5
- VSTEP.2
- VSTEP.1 PLUS
- VSTEP.2 PLUS
- VSTEP.3 PLUS



BACKGROUND

ABOUT VSTEP.3-5

- Nationally released on March 11th 2015
- A research product of language testing experts from ULIS-VNU
- The first-ever Vietnamese standardized test of English proficiency
- Based on CEFR and VN-CEFR

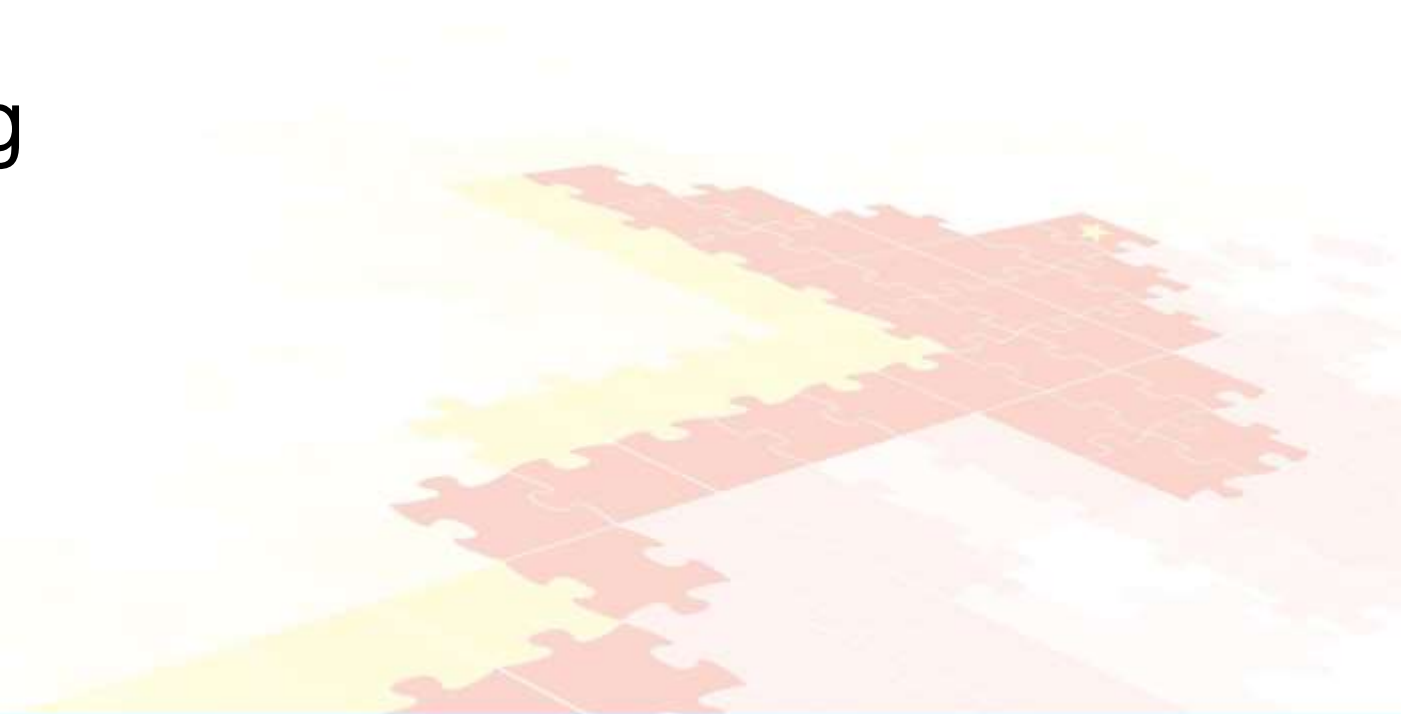
BACKGROUND

ABOUT VSTEP.3-5

SCORE	LEVEL	CEFR
0-3.5	NOT RATED	NOT RATED
4.0 – 5.5	3	B1
6.0 – 8.0	4	B2
8.5-10	5	C1


VSTEP test

- Reading
- Listening
- Writing
- Speaking



BACKGROUND

WHO CAN TAKE VSTEP.3-5?

- Any adult learners
 - In VNU
 - Admission to a program
 - Graduation from university
 - Every two months
- 

Vietnamese Standardized Test of English Proficiency

The logo for the Vietnamese Standardized Test of English Proficiency (VSTEP). The word "VSTEP" is written in a bold, blue, sans-serif font. A thick blue line starts from the top of the 'V', loops over the top of the 'S', 'T', and 'E', and then loops under the 'P' before ending at the bottom of the 'P'.

VSTEP

Validity – Sustainability – Transparency – Equity - Practicality

More validation
research should
be conducted.



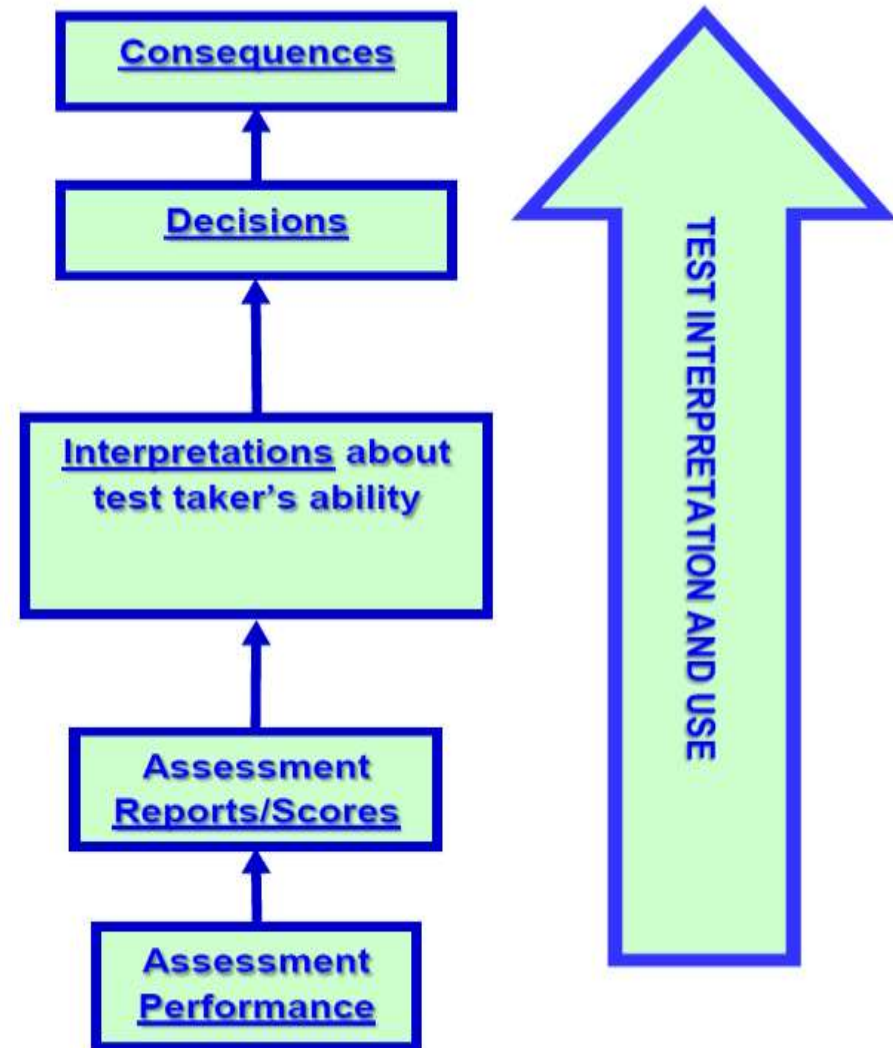
Assessment Use Argument (AUA)

- Bachman (2002) and Bachman & Palmer (2010)
- Validation framework
- Focus on test uses



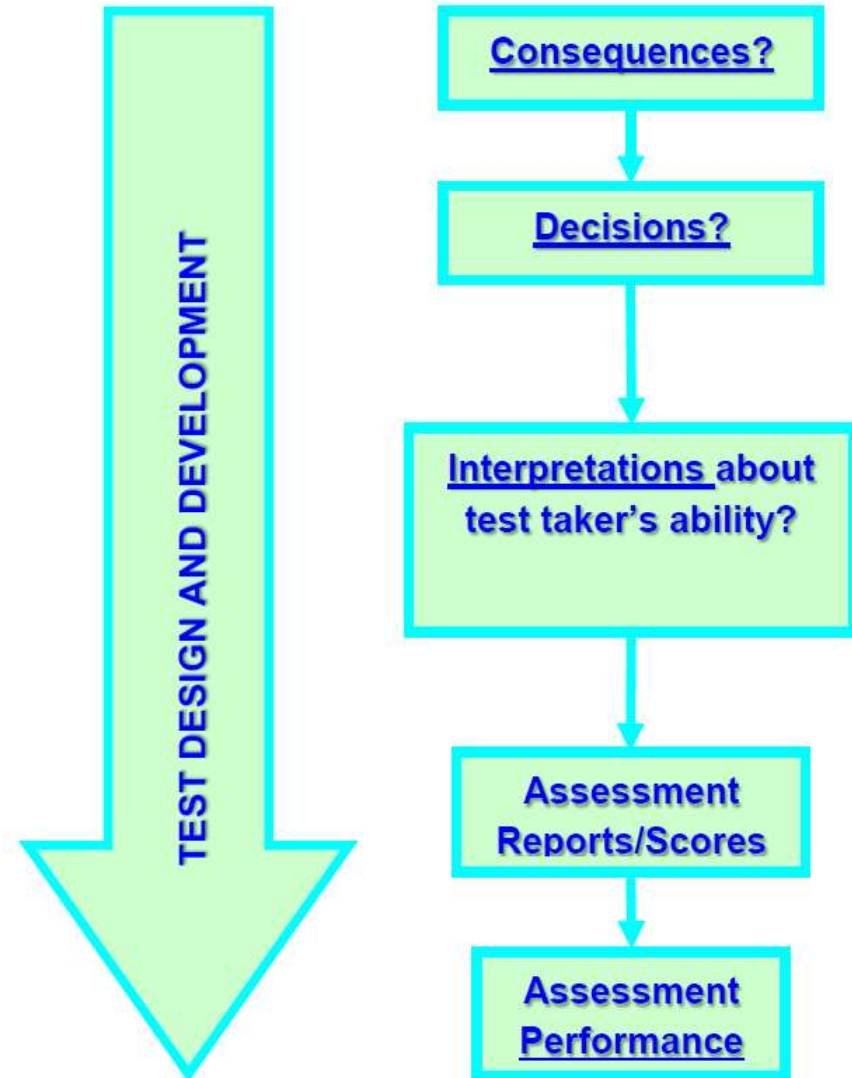
Assessment Use Argument (AUA)

- Provides the conceptual framework for linking our intended consequences and decisions to the test taker's performance.

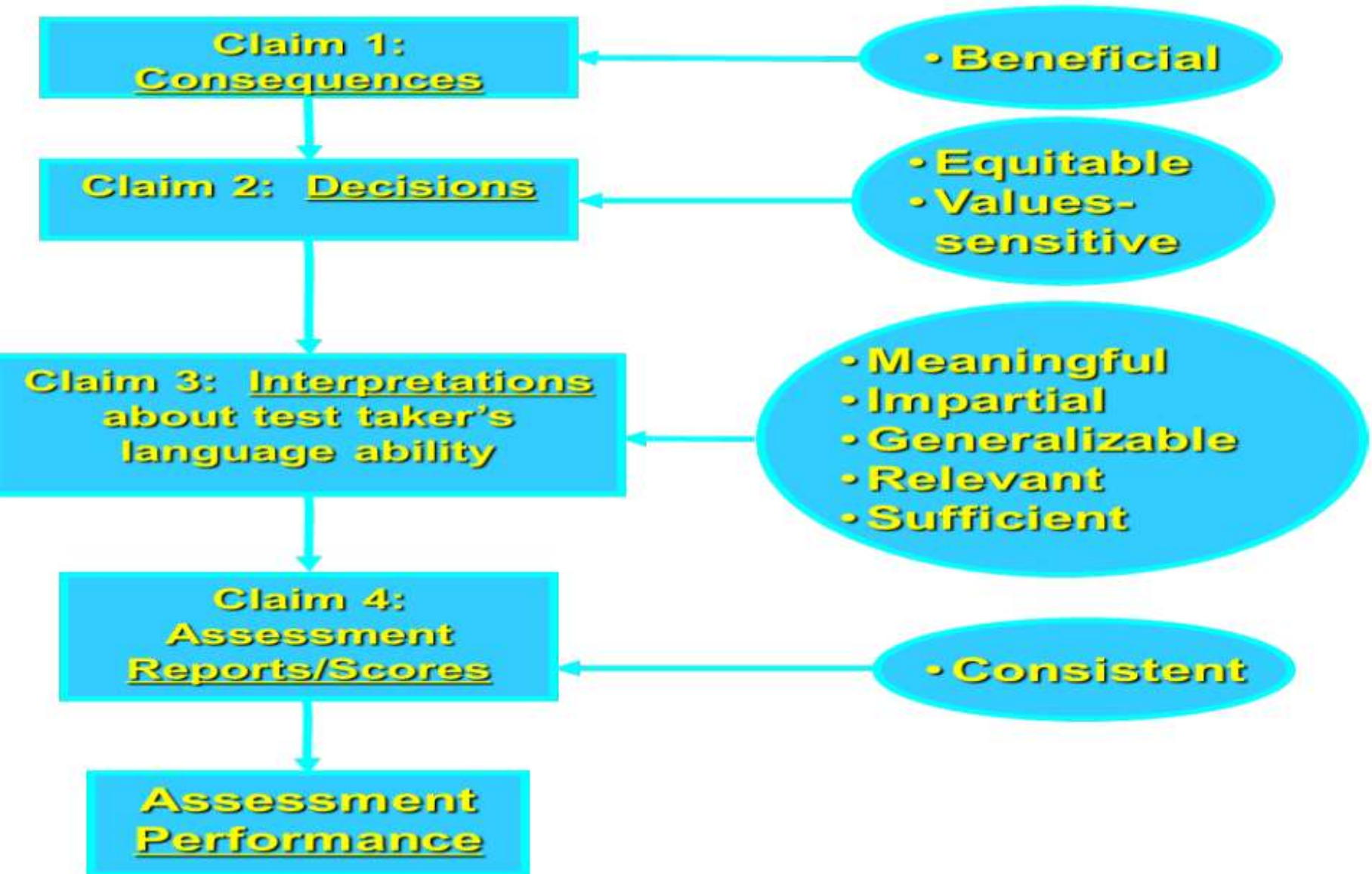


Assessment Use Argument (AUA)


- Provides the rationale and basic for justifying the decisions we make in designing and developing an assessment. And thus provides a guide for designing and developing language assessments.




Validation Project for VSTEP using AUA Framework



This Presentation: Filling in the Key Blanks in the AUA

- While not an exhaustive list, the following areas of investigation and reporting are seen as high- and medium-priority topics.
 - These points will receive the lion's share of attention and resources in the next one to three years.
- 


Claim 1: Consequences

- The consequences of using the VSTEP and of the decisions that are made are beneficial to stakeholders.
 - Main stakeholder groups:
 - People who take the VSTEP
 - Parents of test takers
 - English teachers
 - Employers
 - Schools
 - Policy makers
- 

Plans for Investigating and Reporting on Claim 1 (1/2)

- Studies on the clearness and interpretability of score reports for test takers, parents, teachers, and employers
- Studies on the washback of including speaking and writing on the VSTEP
- Technical report/bulletin to be available on VSTEP website explaining:
 - Confidentiality procedures
 - Score reports and timelines for issuing them
 - Constructs assessed, validity, and reliability

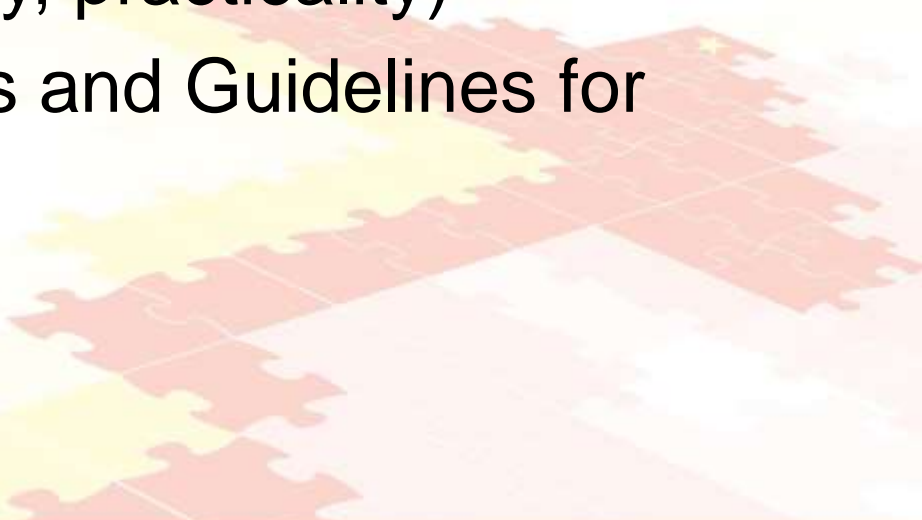
Plans for Investigating and Reporting on Claim 1 (2/2)

- Studies investigating cut scores, and evaluating standard setting procedures
 - Studies equating various forms of the VSTEP in order to maintain consistent difficulty levels
- 

Claim 2: Decisions

- The decisions that are made on the basis of score interpretations take into consideration existing educational and societal **values** and relevant laws, rules, and regulations, and are **equitable** for those stakeholders who are affected by the decision.

Plans for Investigating and Reporting on Claim 2 (1/2)

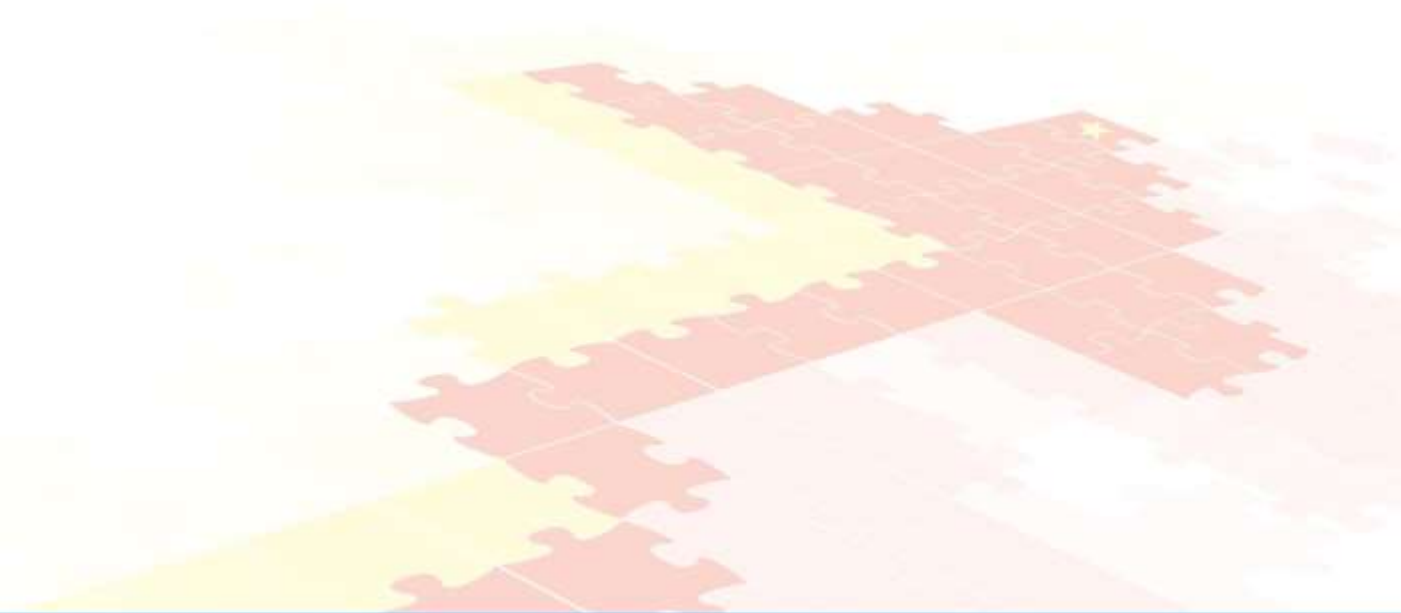
- Study evaluating how the VSTEP meets relevant points contained in:
 - Relevant laws, rules, and regulations
 - VSTEP core values (validity, sustainability, transparency, equity, practicality)
 - ILTA Code of Ethics and Guidelines for Practice
- 

Plans for Investigating and Reporting on Claim 2 (2/2)

- Standard setting activities using a variety of approaches
 - Multiple approaches should converge to yield optimal cut scores
 - Methods include Angoff, Hofstee, borderline groups, contrasting groups, paper selection
- Technical reports/bulletins to be available on VSTEP website explaining scoring, how cut scores were set, equating, etc. in lay language

Claim 3: Interpretations

- Interpretations about the abilities assessed by the VSTEP are meaningful, impartial, generalizable, relevant, and sufficient.



Plans for Investigating and Reporting on Claim 3A: Meaningfulness (1/3)

- Study using logical comparison, showing correspondence between each component of the construct definition and its support in various reference documents:
 - Needs analysis report
 - CEFR
 - English Profile
 - Theories of reading, writing, speaking, and listening

Plans for Investigating and Reporting on Claim 3A: Meaningfulness (2/3)

- Study showing logical comparison, showing correspondence between
 - Tasks, situations, topics, functions, etc. to be sampled from, as listed in the specifications, vs.
 - Support in the various reference documents (i.e., needs analysis report, CEFR)

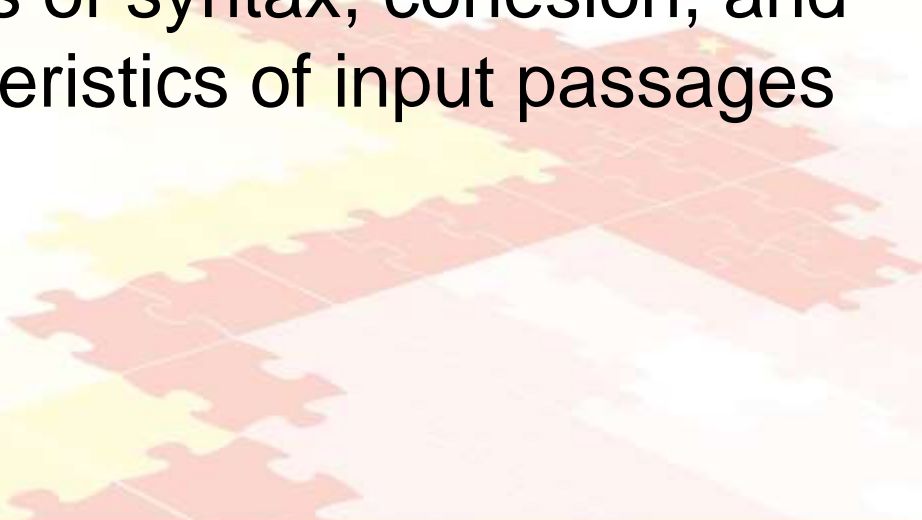
Plans for Investigating and Reporting on Claim 3A: Meaningfulness (3/3)

- Studies showing that scores can be interpreted as indicators of the ability to be assessed:
 - Factor analytic studies
 - Task analyses of rubrics, passages, items, and prompts, including test content analysis (especially of passages)

Plans for Investigating and Reporting on Claim 3B: Impartiality

- DIF analyses comparing male/female, rural/urban, academic background (e.g., high school vs. above high school education), and other salient groups
- Technical report to be available on website detailing sensitivity review procedures (panel, checklist).
- Multi-group structural equation modeling studies investigating factorial invariance across test taker groups

Plans for Investigating and Reporting on Claim 3C: Generalizability

- Studies comparing test task characteristics vs. characteristics of common and/or prototypical TLU domain tasks
 - Bachman & Palmer TTC framework
 - Automated analysis of syntax, cohesion, and vocabulary characteristics of input passages
- 

Plans for Investigating and Reporting on Claims 3D & 3E: Relevance and Sufficiency(1/2)

- Writing and speaking: Study describing:
 - Standard setting process, using multiple approaches
 - Comparison of test taker responses to CEFR descriptors
 - Analysis of test taker responses (syntax, vocabulary) vs. the English Profile

Plans for Investigating and Reporting on Claims


3D & 3E: Relevance and Sufficiency(2/2)

- Reading and listening: Study describing:
 - Standard setting process, using multiple approaches
 - Task analysis comparing reading and listening items with CEFR descriptors
 - Content analysis comparing reading and listening text syntax and vocabulary with English Profile descriptors

Claim 4: Assessment Records

- Assessment records (scores, performance descriptions) are **consistent** across different VSTEP tasks, different aspects of the VSTEP procedure, and across different groups of test takers.

Plans for Investigating and Reporting on Claim 4 (1/3)

- Brief description of test administration procedures, taken from test administrators manual, to be posted on VSTEP Website
 - Description of rater training and certification programs, to be provided in technical report posted on VSTEP Website
- 

Plans for Investigating and Reporting on Claim 4 (2/3)

- Reading and listening:
 - Cronbach's alpha, SEM, Φ , CRT CI for
 - Item analysis for improving dependability
 - Rasch (WINSTEPS)
- Speaking and writing:
 - Inter-rater reliability (correlations)
 - Generalizability studies
 - Rasch (Facets)
- Results to be reported in annual score report summary

Plans for Investigating and Reporting on Claim 4 (3/3)

- Equating of forms using Rasch
- Comparison of anchor testlets across administrations
 - *t*-tests or ANOVA across administrations to compare means
 - Frequency polygons to compare shapes of score distributions
 - Comparison of factor structure across administrations to test for invariance

References (1/2)

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford: Oxford University Press.

References (2/2)

- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.), 111-153. Westport, CT: Praeger.
- Weir, J. C., & Milanovic, M. (2005). *Examining listening: Research and practice in assessing second language listening*. Cambridge: Cambridge University Press.
- Weir, J. C., & Khalifa, H. (2009). *Examining reading: Research and practice in assessing second language learning*. Cambridge: Cambridge University Press

Thank you

