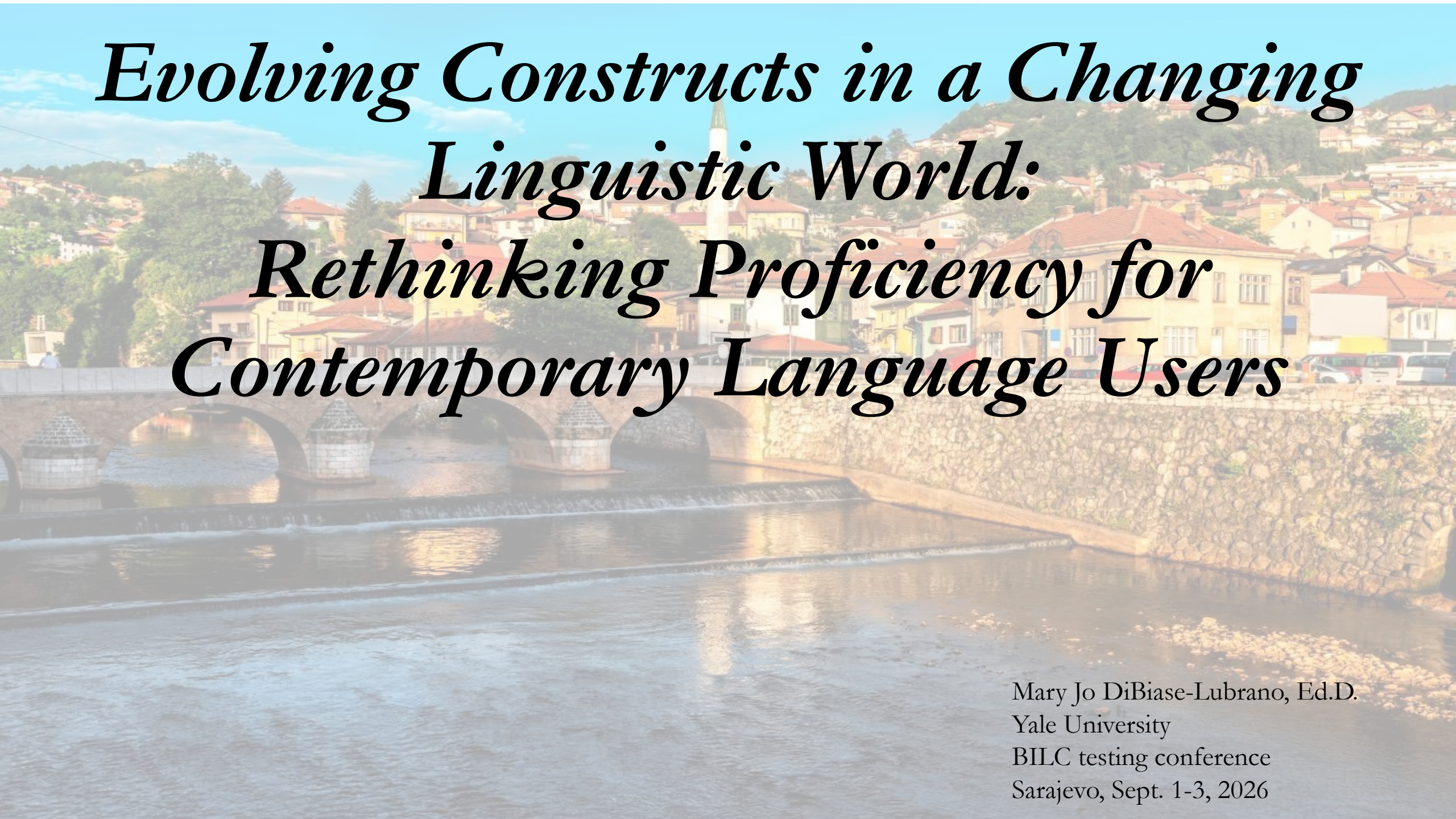


The background of the slide is a scenic photograph of a river flowing through a town. A multi-arched stone bridge spans the river in the middle ground. The town's buildings, with red-tiled roofs, are built into a hillside in the background. The sky is a clear, bright blue. The title text is overlaid on this image in a large, elegant, black serif font.

Evolving Constructs in a Changing Linguistic World: Rethinking Proficiency for Contemporary Language Users

Mary Jo DiBiase-Lubrano, Ed.D.
Yale University
BILC testing conference
Sarajevo, Sept. 1-3, 2026

Overview

1. The Historical Construction of Proficiency

2. The New Linguistic Landscape

3. Emerging Threats to Construct Validity

4. Towards New Constructs

Conclusions

1. The Historical Construction of Proficiency

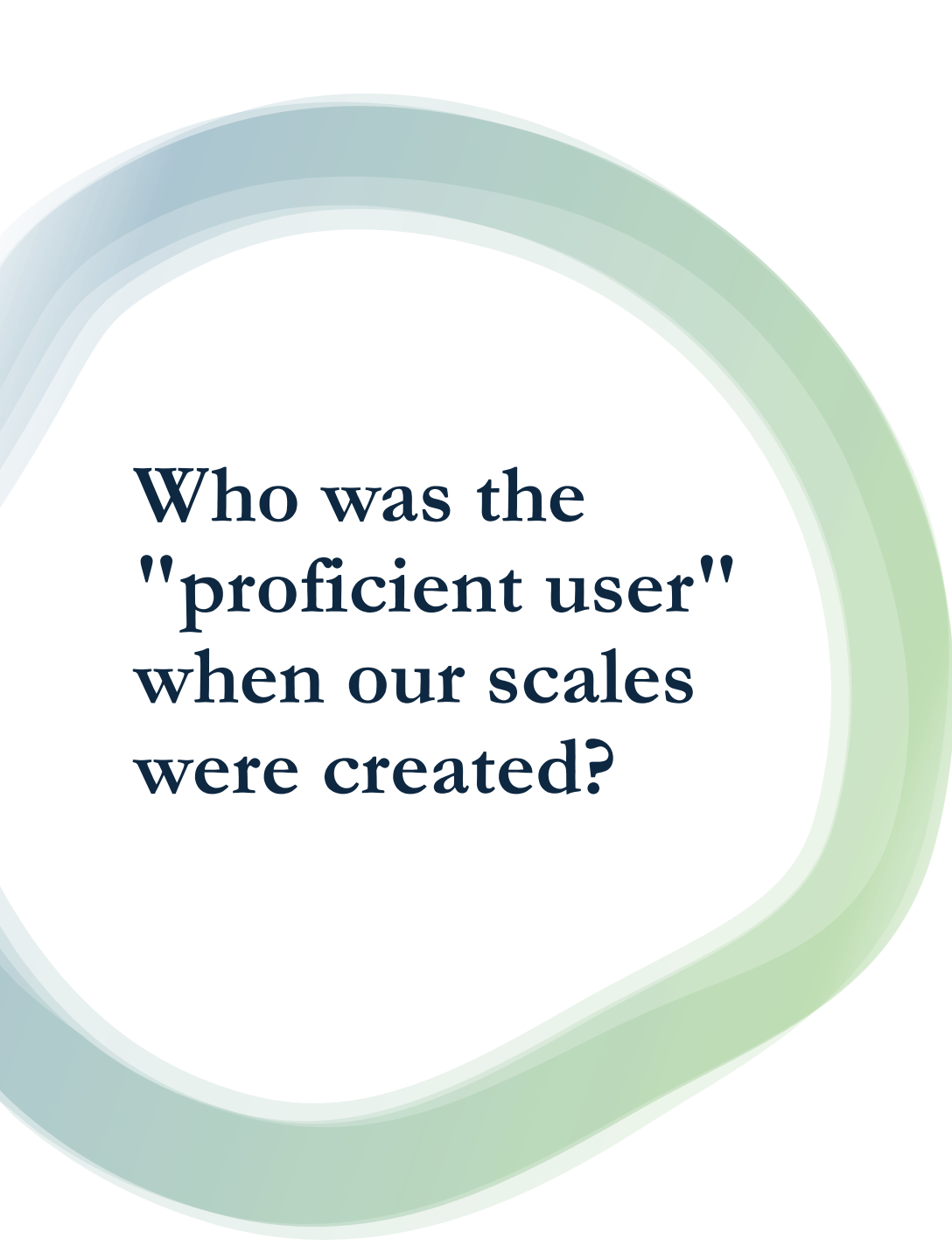
1. The Historical Construction of Proficiency

Origins of proficiency frameworks

Native-speaker influence

Monolingual assumptions

Why these constructs were revolutionary at the time



**Who was the
"proficient user"
when our scales
were created?**

Traditional scales:

- relatively bounded languages,
- stable speech communities,
- monolingual norms,
- progression toward native-speaker-like competence,
- language use within predictable domains.

Those assumptions made sense in the twentieth century -

Expansion is evolutionary rather than revolutionary.

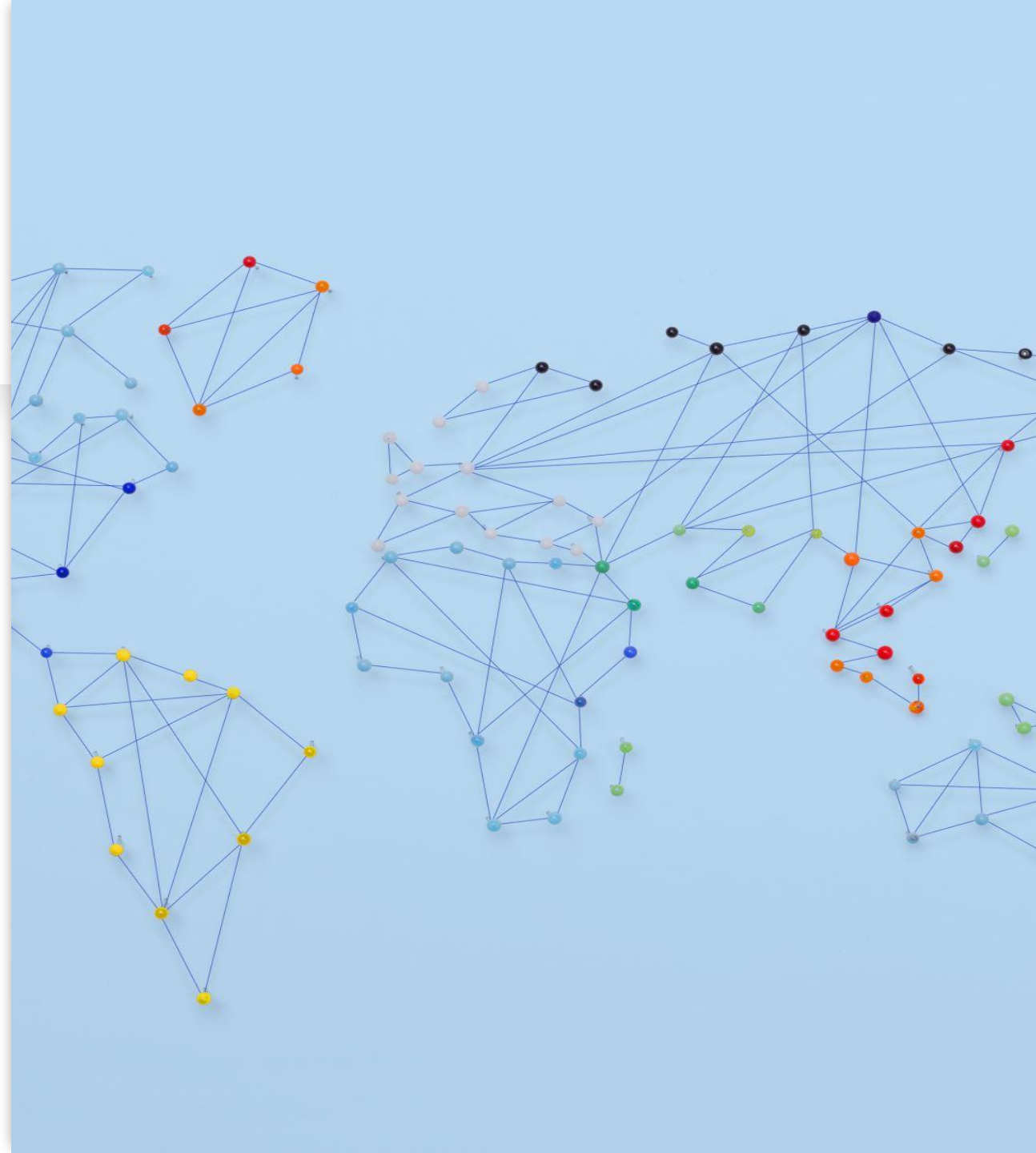
- **Hymes – Communicative Competence:** expand beyond grammar.
Today, expand beyond proficiency = multilingual, digital
- **Bachman – Communicative Language Ability** = strategic competence
central to successful communication.

2. The New Linguistic Landscape

The New Linguistic Landscape

Post-9/11 Globalization

- interdependent world – global mobility
- developing technology - AI assisted
- changing access to information: digital
- increasing global migration – transnational identities



Today's proficient language user

- operates across multiple languages simultaneously;
- engages in translanguaging practices;
- interacts through AI-mediated communication;
- participates in global digital communities;
- uses language collaboratively rather than individually;
- moves across linguistic and cultural contexts daily.



The construct of language proficiency emerged from assumptions about language users that no longer characterize large segments of today's test-taking population.

Theoretical Foundations for Revisiting Constructs

Hymes (1972) - Communicative competence

Bachman (1990) - Communicative language ability

Messick (1989) - Validity depends on interpretation

Kane (2006, 2013) - Argument-based validation

Douglas (2000) - Target Language Use determines construct

- *Has the target language use domain changed enough that our construct requires revision?*



3. Emerging Threats to Construct Validity

Emerging Threats to Construct Validity (Messick, Kane)

- Construct underrepresentation
- Construct irrelevance
- Misalignment between scales and real-world language use
- Examples from current testing practice – the case of listening

The world changed faster than our descriptors

Three shifts:

1. Multilingual as the norm (not the exception)
2. Varieties/accents/registers
3. Functioning including intercultural stance, identity, and agency

So.....

What should proficiency mean in this new linguistic world?



4. Toward New Constructs

Toward New Constructs

Four dimensions:

1. Adaptability
2. Mediation
3. Plurilingual competence
4. Digital communicative competence

Europe vs. (or “and”?) USA

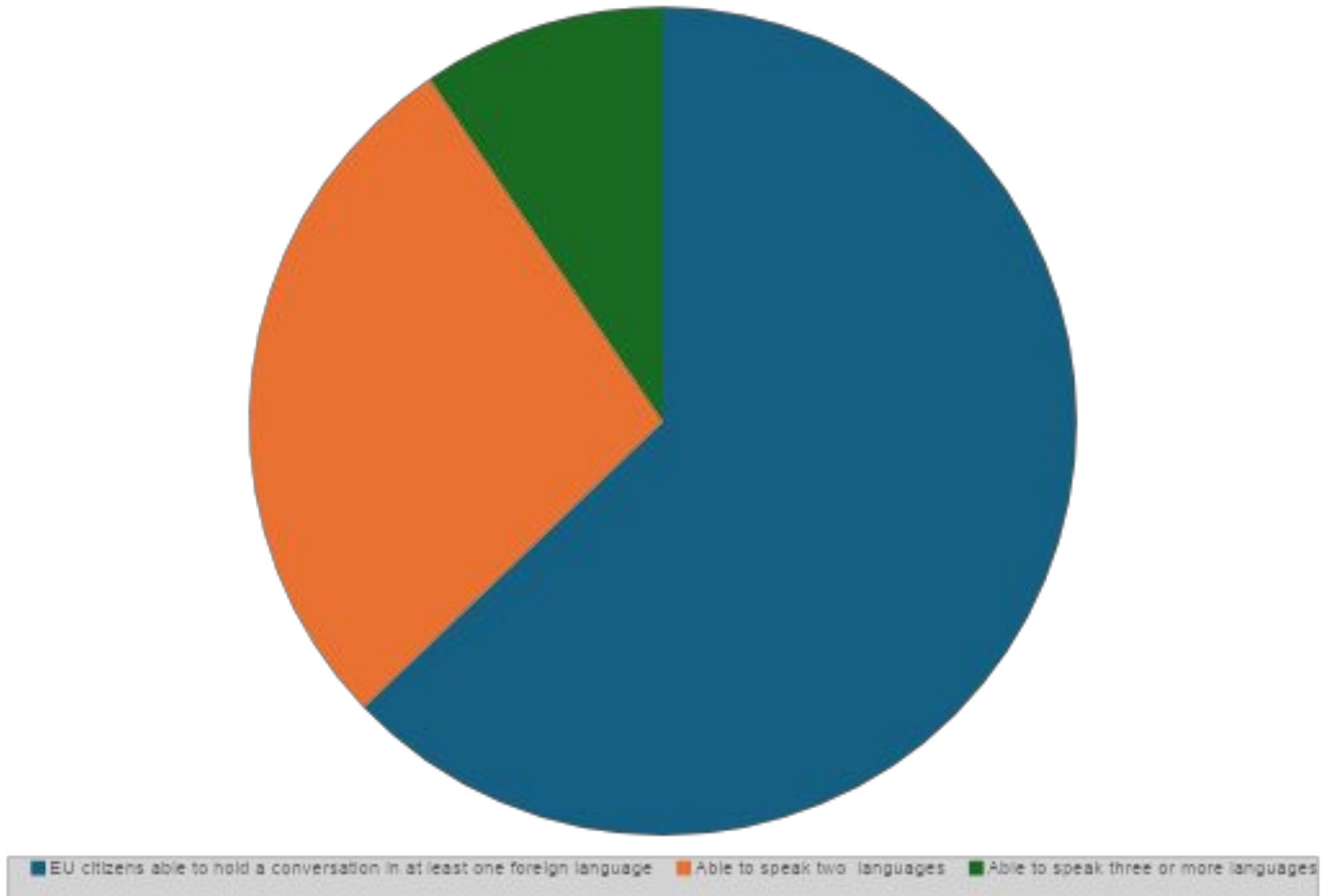
Europe contains dozens of countries, ethnic groups, and language traditions within a relatively small geographic area.

As a result:

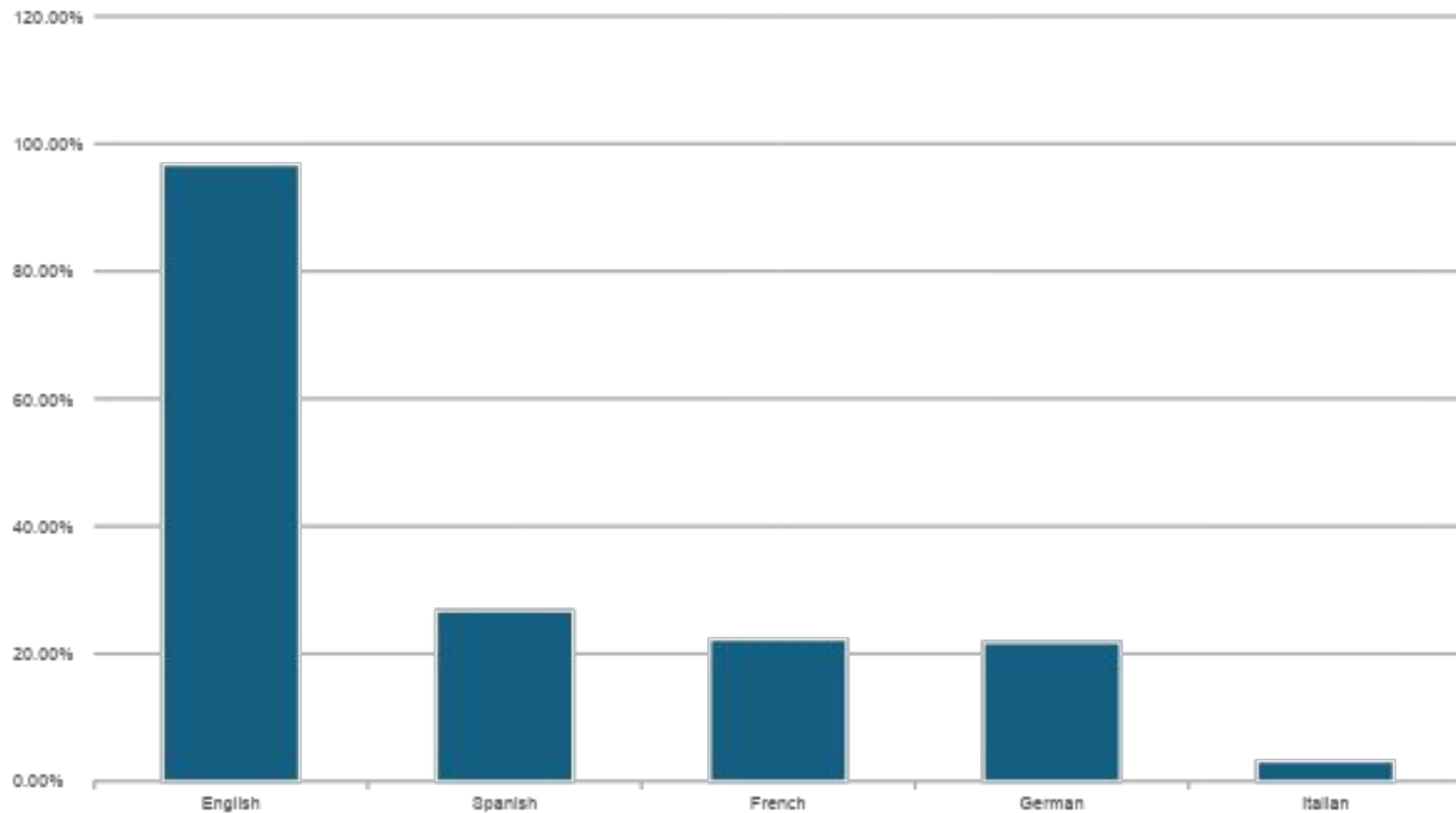
- More than one language spoken.
- Language learning encouraged and supported.
- One other language taught in K-12.
- The European Union promotes linguistic diversity and recognizes multiple official languages.



Europeans Who Speak Other Languages According to EU surveys:



Most Common Language Studied



English as a Lingua Franca

Although no longer the language of a large EU member state after Brexit, English remains:

- The most commonly taught language.
- The dominant language.
- Most spoken second language.

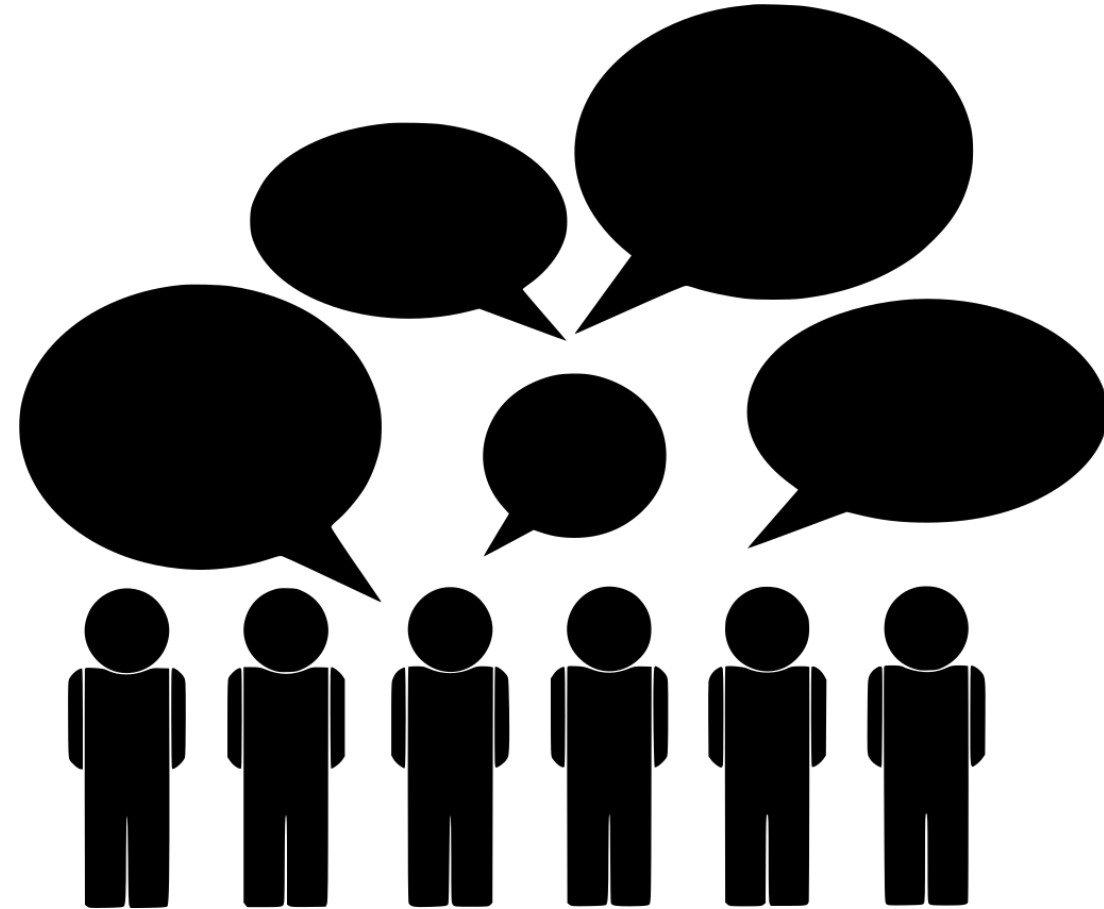
Misalignment?

If our construct story remains monolingual and single-standard, we risk misalignment.

Example from Yale: The Class of 2029: A Multilingual Community

- 1,633 students
- 77 languages spoken (other than English)

(Admissions Office data 2026-27)



In the U.S. (ASC data): the testing implication parallels Europe

What's your first language?

Number of Languages Spoken



Nearly 3 out of 4 speak more than one language

79% English as L1 - 27% monolingual English - 37% L2+ - 3% L3+ - 18% L4+

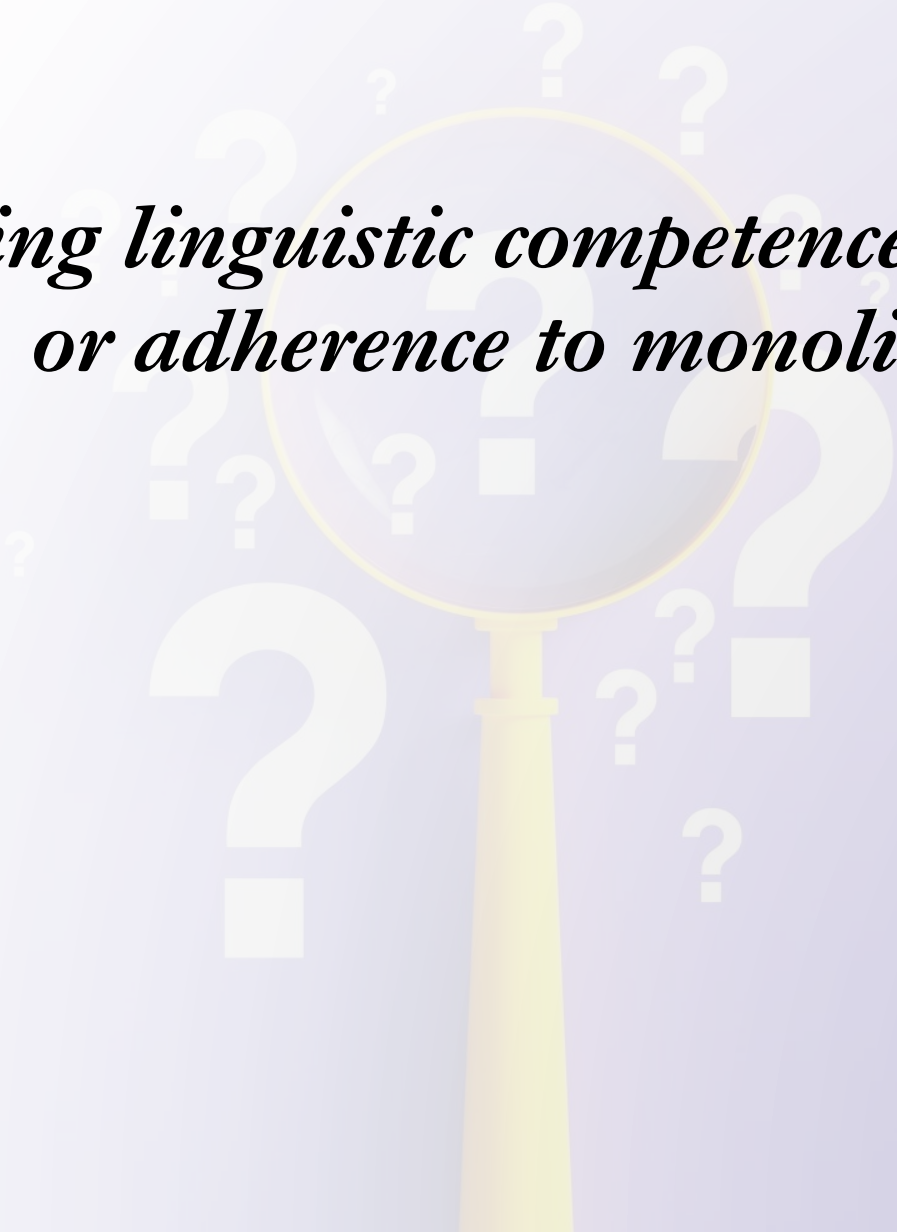
Construct update: proficiency-as-profile

Proficiency (for high-stakes)...

1. Linguistic resources
2. Variation & sociolinguistic competence
3. Interactional/pragmatic competence
4. Sociocognitive capacities

Question

Are we measuring linguistic competence, communicative effectiveness, or adherence to monolingual norms?



Let's think about “advanced proficiency” (L2+/3)

“Advanced is not one thing; our scales often treat it as if it were.”
(Ortega & Byrnes, 2008)

“Variation is not error; it is competence conditioned by context.”
(Geeslin)

“A high score in functional grammar does not guarantee sophisticated intercultural functioning—and sometimes the reverse is true.” (Czerwionka, 2021)

Shifting the paradigm: From advanced proficiency to advancedness

- Malovrh, P. A. & Benati, A. (Eds.) (2018).
 - L2 phonology is amorphous
 - Ultimate attainment; theoretical challenges
- Menke, M. & Malovrh, P. A. (Eds.) (2021).
 - Autonomy of intercultural competence
 - Stance; expertise
 - Need for socio-cognitive analysis
 - Limitations of traditional proficiency tests

Where traditional scales tend to over-promise

Common implicit assumptions

- bounded “target language” competence (monolingual norm)
- linear ladder of growth
- stable context and stable norms
- “more accurate/complex/ fluent” $\approx$ “more proficient”

Instead, “advanced” ability is not just “more language”

Consequence in measurement terms: construct underrepresentation

If the scale focuses mostly on structural control/sustained ability and language tasks, it can miss:

- sociolinguistic competence (variation, register)
- intercultural communicative competence
- mediation/synthesis and genre/literacy capacities
- identity/agency and critical language awareness

In Scucci & Malohvr (2025), learners equate advancedness with grammar/pronunciation/comprehensibility, and raters show “native-speakerism”/bias patterns

A single speaking level may hide whether the candidate
can *manage socially meaningful variability*
(varieties/accents/registers) versus simply producing
high-control forms.

The “new normal” for test takers

Construct drift (definition)

When TLU realities shift but constructs/interpretations don't

- Risk: Level "X" used as a proxy for capacities not sampled
- Repertoires (often >2 languages), uneven by domain
- Routine variety/accents exposure
- Digital mediation in real communication
- Intercultural demands + stance/identity

Digital and AI-mediated communication

Modern language users increasingly need to:

- participate in hybrid meetings,
- collaborate asynchronously,
- communicate across multiple digital platforms,
- critically evaluate AI-generated text,
- edit machine translations.

Some researchers now discuss **digital communicative competence** as a distinct dimension of language use (Rodriguez-Pinero 2025).

Question

If candidates routinely work with AI outside the testing room, what is the construct we are trying to isolate inside it?

In a nutshell....

We didn't
change the
scale; the
world
changed

Old model (monolingual, stable exposure)

vs.

Current (multilingual repertoires,
transnational trajectories, AI/digital
mediation).



What changed? The multilingual test taker and the mediated world



Repertoires over single
codes



Domain-specific
proficiency



Increased strategic
competence



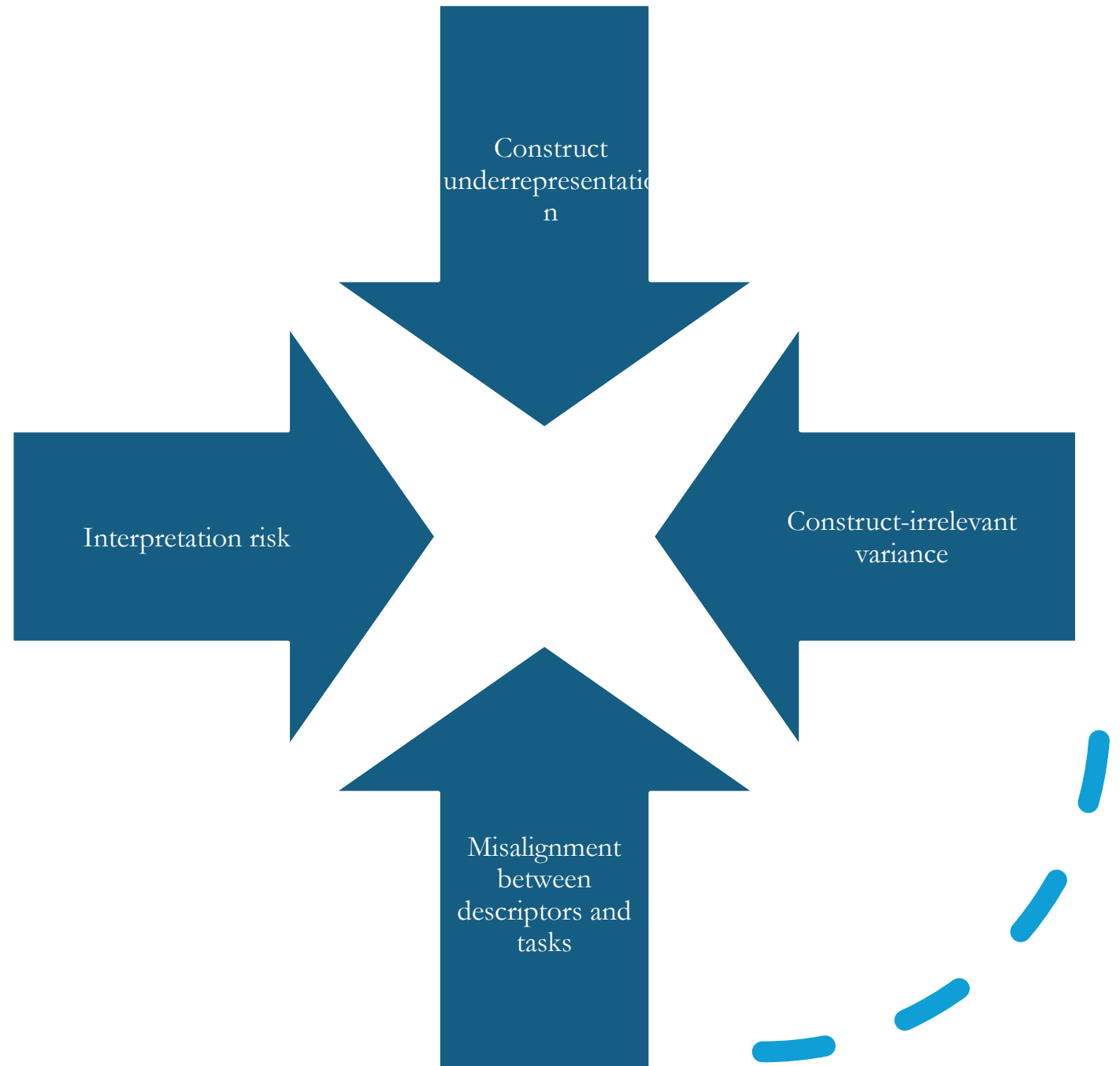
Digital mediation



Identity and
pragmatics



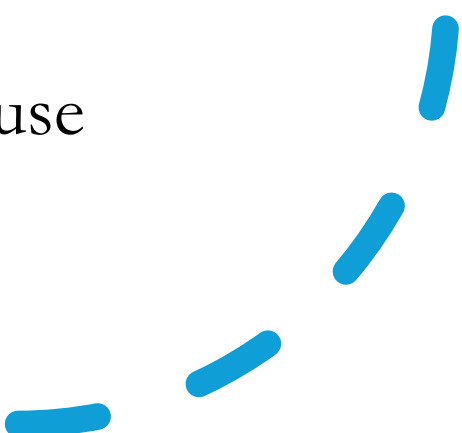
Level
descriptors
may no longer
mean what we
think they
mean anymore



A large orange circle is positioned on the left side of the slide, partially overlapping the text.

Multilingual interaction + mediation resources

Often missing from scales.....

- Meaning negotiation and repair
 - Strategic repertoire use
 - Pragmatic adaptability across contexts
 - Comprehension across varieties/accents
 - Mediation/synthesis
 - Ethical communication and source use
- 
- A series of blue dashed lines are located in the bottom right corner of the slide, forming a curved, upward-pointing shape.

The Case of Listening Comprehension

- Target Language Use (TLU) shift: what counts as “listening” for NATO personnel?

Does the change impact the STANAG listening construct and task design (without necessarily abandoning standardization)?

Large scale studies in academia (Part I)

Reading &
Listening -
Tschirner (2016):

Over 3000 participants from 21 different colleges

- Listening proficiency was lower at almost all levels of instruction and all languages.
- Listening proficiency was one sublevel lower than reading in 2nd/3rd year, and two sublevels lower in the 4th year

Large scale studies in academia (Part II)

(Rubio & Hacking, 2019)

- Findings from MSU, U of Minnesota, U of Utah: 10,000 students in 2nd & 4th semester
- After 2nd semester students are consistently below Intermediate range (L1) in all skills
- Listening is the weakest skill
- After 4 semesters, listening still lags behind in all languages!

What is/was the Target Language Use (TLU) in listening?

Past (roughly early–mid 2000s):

briefings, meetings, radio/phone calls

TV/radio news

face-to-face conversations

relatively stable genres, longer turns, fewer
channels

Listening TLU (2020s)

short-form video streams (TikTok/Reels/Shorts)

social audio/video

UGC news ecology

multimodal briefings

AI-mediated comms

Listening now often includes...

persuasion, misinformation, sarcasm

understanding what is *claimed* vs what is
shown

So the “task” is not only “What did they say?” but “What can we infer is reliable?”

According to studies on listening comprehension

...



Video can change comprehension processes



the role of visuals to protect construct validity.



Visuals can both help and introduce construct-irrelevant variance so task design and validation are critical. (Suvorov, 2015)

Suvorov (2013)

visuals often *carry*
the message

on-screen
text/captions may
be primary

editing changes
discourse cohesion
and inferencing
demands

Validity caution



Visual channel might
introduce visual literacy
effects



Caption availability and
policy creates
comparability issues



Short clips can inflate
reliance on background
knowledge



Platform content can
introduce cultural bias /
sensitive content

Some solutions

Define

define visuals' role
explicitly

Pilot

pilot for fairness

Use

use controlled but
authentic-looking
video stimuli

Balance

balance genres
(operational +
public/social)

What scholars say...

Gary Ockey —
(very aligned with
Suvorov's work).

Carsten Roever
(pragmatics) —

Naoko Taguchi
— pragmatic
comprehension
and processing

**Mark
Warschauer**
(CALL)

Implications for high-stakes testing

If we accept this argument, then...

For test design

- Broaden the evidence
- Increase authenticity

For scales/descriptors

- avoid monolingual end-states; emphasize effectiveness, appropriateness, adaptability.
- Avoid vague adverbs





Conclusions

Call to action (field agenda)

1

Convene construct
review working
group

2

Pilot profile
dimensions

3

Update rater
training

4

Publish guidance
for score users

A likely future direction

Instead of a single level (e.g., "ILR L3"), a profile might include:

- functional proficiency in each language
- domain-specific expertise
- mediation skills
- intercultural competence
- collaborative communication
- digital communication
- adaptability across multilingual settings

This approach aligns with a broader trend in educational measurement: moving away from a single score toward richer profiles of competence that capture the complexity of real-world performance.

Conclusion



The question facing our profession is not whether our scales remain useful.



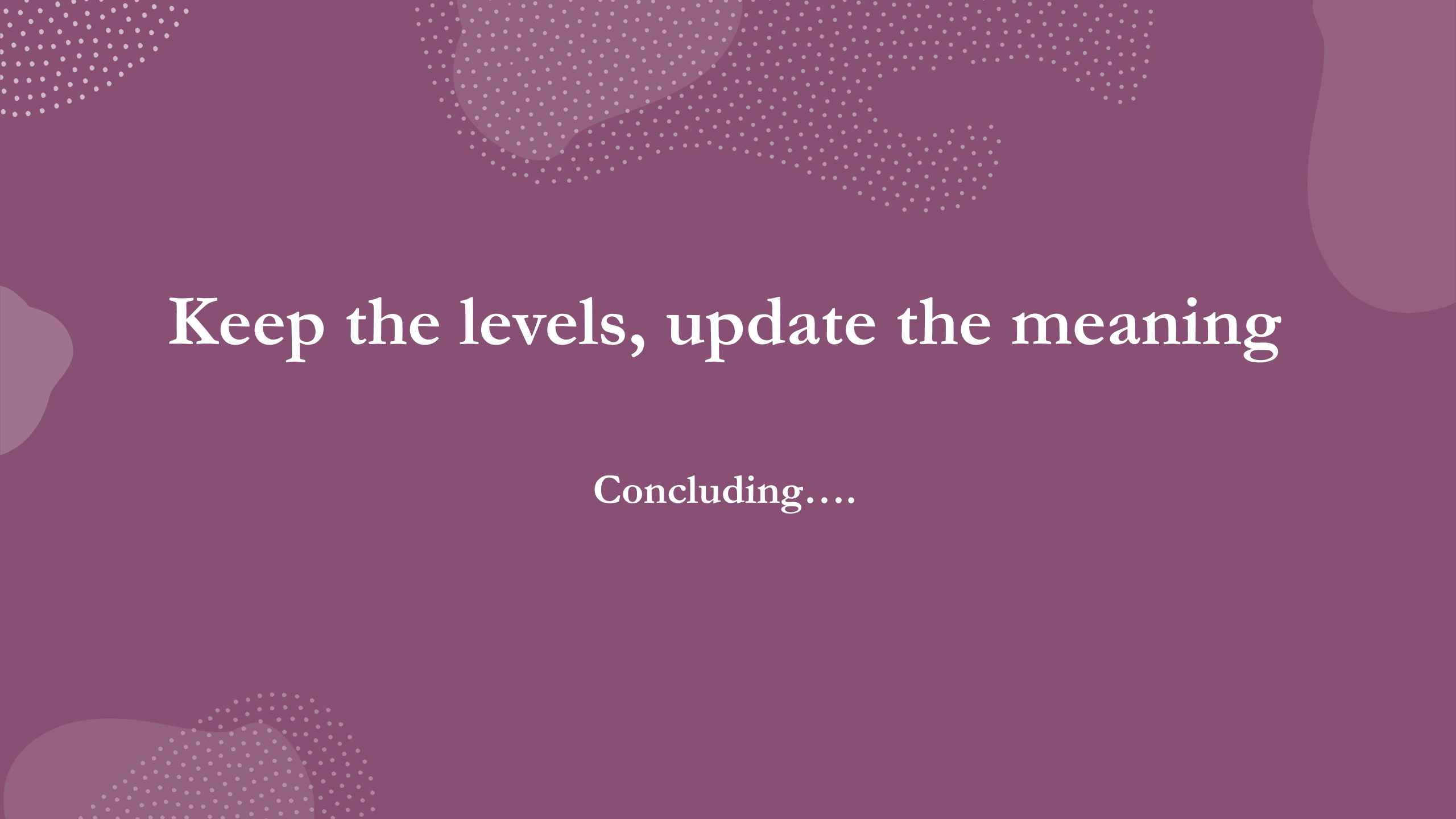
It is whether they remain sufficient.



The future of language testing may depend more on reimagining the construct those levels were designed to measure.



So.....



Keep the levels, update the meaning

Concluding....

Food for thought...

What would be the implications in
the STANAG world?



Generations of language testers...

1. preserve comparability with the past
2. remain relevant for the future

Question: **which one?**

Answer: both!

[illegible]

Selected References

- Archibald (2018). Advanced Level L2 phonology in (Malovrh & bennati, 2028) The Handbook of Advanced proficient in Second Language acquisition
- Cwerwionka, 2021. [Using the L2 to express intercultural competence: A measure of sophisticated language use in context](#)
- Forest (2010). Swender, E., Forest, L., & Watzke, J. (2010). *What proficiency level do high school students achieve?* Center for Applied Second Language Studies (CASLS), University of Oregon.
- Malovrh & Lee (2021) Malovrh, P. A., & Lee, J. F. (2021). The effects of study abroad on interlanguage development: A concept-oriented analysis of advanced Spanish majors' direct object expression. In P. A. Malovrh & M. R. Menke (Eds.), *Advancedness in Second Language Spanish: Definitions, challenges, and outcomes* (pp. 211–234). John Benjamins Publishing Company. [, 2]
- Malovrh, P. A. & Benati, A. (Eds.) (2018). *Advanced proficiency in Second Language Acquisition*. New York: John Wiley & Sons.
- Menke, M. & Malovrh, P. A. (Eds.) (2021). *Advancedness in Second Language Spanish: Definitions, Challenges and Possibilities*. Amsterdam: Benjamins.
- Ortega & Byrnes (2008). Ortega, L., & Byrnes, H. (Eds.). (2008). *The longitudinal study of advanced L2 capacities*. Routledge.
<https://doi.org/10.4324/9780203871652>
- Rodríguez-Piñero, A. I. (2025). *Being Competent in Foreign Languages in the Digital Age*. *RLA. Revista de Lingüística Teórica y Aplicada*, 63(2), 113–139.
- Scucchi, S. & Malovrh, P. A. (2025). Sociopragmatic variation and identity construction in L2 Spanish: An analysis of context and group membership. In M. Solon, M. Kanwit, & A. Gudmestad (Eds.), *Research at the Intersection of Second Language Acquisition and Sociolinguistics: Studies in Honor of Kimberly L. Geeslin* (pp. 203–229). New York: Routledge.
- Soneson (2021). Advanced users of Spanish: Relationships among modalities of speaking, reading, and listening. In P. A. Malovrh & M. R. Menke (Eds.), *Advancedness in Second Language Spanish: Definitions, challenges, and outcomes* (pp. 66–91). John Benjamins Publishing Company. doi.org
- Suvorov (2015). The use of eye tracking in research on video-based second language (L2) listening assessment: A comparison of context videos and