

The ACTFL Listening Proficiency Test (LPT): Evidence of Validity

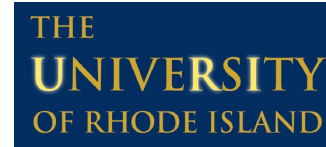
April 12, 2024

Prepared by
Erwin Tschirner
ILR Testing Committee



PRESENTED BY

- **Erwin Tschirner, Ph.D.,**
Gerhard Helbig Professor of German as a Foreign Language, Emeritus, University of Leipzig;
Adjunct Professor of Modern Languages, University of Rhode Island
- **Thomas Wagner, Ph.D., CAPT, USN (ret.),**
Director of Contracts and Compliance, ACTFL
- **Kim Sallee,**
Vice President of Assessments, Language Testing International (LTI)



UNIVERSITÄT LEIPZIG



OUTLINE



Test Specifications

- Overview



Construct



Evidence of Validity

- Reliability
- Convergent Validity



Q & A

ABOUT ACTFL & LTI

ACTFL

- A non-profit membership organization of more than 13,000 language professionals
- Developed the *ACTFL Proficiency Guidelines*, international standards for proficiency.
- Develops valid and reliable assessments in all modalities in over 120 languages for learners in Grade 3 through adult.
- Use of the ACTFL Proficiency Guidelines and Issuing of Official ACTFL Ratings

Language Testing International (LTI)

- Exclusive licensee of ACTFL Assessments since 1992
- Delivers over 1 million tests per year through our end-to-end secure delivery platform in Academic, Government and Corporate markets, as well as to individuals
- Testing in over 120 languages in approximately 60 countries

TEST SPECIFICATIONS

LPT HIGHLIGHTS



- Extensively validated & recognized; additional points of accreditation by the American Council on Education for college credit and by the Council of Europe
- Real-life passages, tasks, and procedures
- 10 passages at two sublevels: Provides test takers with sufficient opportunity to obtain a reliable rating.
- 4 different question types (global, detail, selective, inference): Covers a broad range of listening goals
- Questions are asked before listening: Provides listener intent
- Questions are in the target languages: Enhances the construct
- Test takers are able to take notes on screen/within the test: Prevents having to rely on memory

LANGUAGES AVAILABLE



Arabic



Chinese (Mandarin)



English



French



German



Italian



Japanese



Korean



Portuguese



Russian



Spanish

TEST SPECIFICATIONS

- 25 individual texts at five sublevels (five texts per sublevel)
 - ILR 1, 1+, 2, 2+, 3
- 3 items per text in the target language, 4 choices per item
- Six limited ranges, two complete ranges (one adaptive)
- Limited ranges: 50 (2 sublevels) or 75 min. (3 sublevels; semi-adaptive complete range: 50-75 min (G); and complete range: 2 hours (H)
- Limited ranges: Rating is based on 10 passages and 30 items or 15 passages and 45 items
- Semi-adaptive complete range: 10 to 15 texts and 30 to 45 items depending on proficiency level
 - 10 texts for ILR 1+ to 2+ and 15 texts for ILR 0-1 and 3
- Complete range: 25 texts and 75 items

TEST FORMATS

- A (ILR 0 to 1+): ILR 1 and 1+ passages and items
- B (1 to 2+): ILR 1+ and 2 passages and items
- C (1+ to 2+): ILR 2 and 2+ passages and items
- D (2 to 3): ILR 2+ and 3 passages and items
- E (0 to 2): ILR 1, 1+, and 2 passages and items
- F (1+ to 3): ILR 2, 2+, and 3 passages and items
- G (0 to 3 adaptive): ILR 1+ to 2+ passages/items for most examinees; ILR 1 to 2+ items for examinees lower than 1+; 1+ to 3 items for examinees higher than 2+
- H (0 to 3): ILR 1, 1+, 2, 2+, and 3 passages and items

Step 1
Setup



Step 2
Sample Question



Step 3
Begin Test

Before we start...



Shut Down Programs

Before starting your test, shut down all other programs running on your computer.



Technical Issues

If you experience technical difficulties, relax. You can always sign back in within 2 hours.



Time is Precious

Your test will be 50 minutes long. The clock starts after clicking **Begin**.

Are you ready?

Begin >

Familiarize yourself with the topic and the questions below.

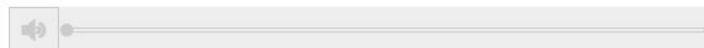
The recording will play in **12 seconds**.

1 of 2: David's Hobbies

Questions

- 1 How often does David play tennis?
- 2 With whom does David go to the movies?
- 3 What hobbies does David say are expensive?

Listening Passage



Take notes here:

[Next >](#)

Familiarize yourself with the topic and the questions below.

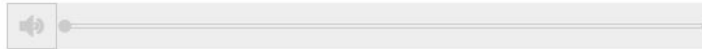
The recording will play in **12 seconds**.

1 of 2: David's Hobbies

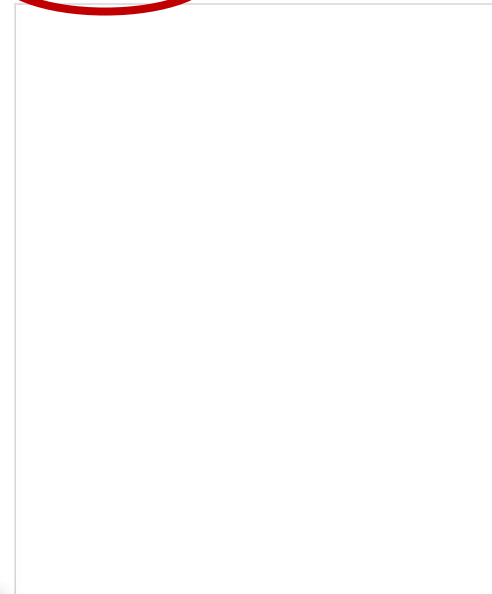
Questions

- 1 How often does David play tennis?
- 2 With whom does David go to the movies?
- 3 What hobbies does David say are expensive?

Listening Passage



Take notes here:



Next >

1 of 2: David's Hobbies


Time Remaining for this Selection: **1:53**

- 1 How often does David play tennis? David plays tennis...
 - ☐ once a week.
 - ☐ twice a week.
 - ☐ twice a month.
 - ☐ every day.
- 2 With whom does David go to the movies? He goes to the movies with...
 - ☐ friends.
 - ☐ Jenny and his brother.
 - ☐ his brother.
 - ☐ Jenny.
- 3 What hobbies does David say are expensive? David says it is expensive to...
 - ☐ go to restaurants.
 - ☐ go to the movies.
 - ☐ play tennis.

Take notes here:

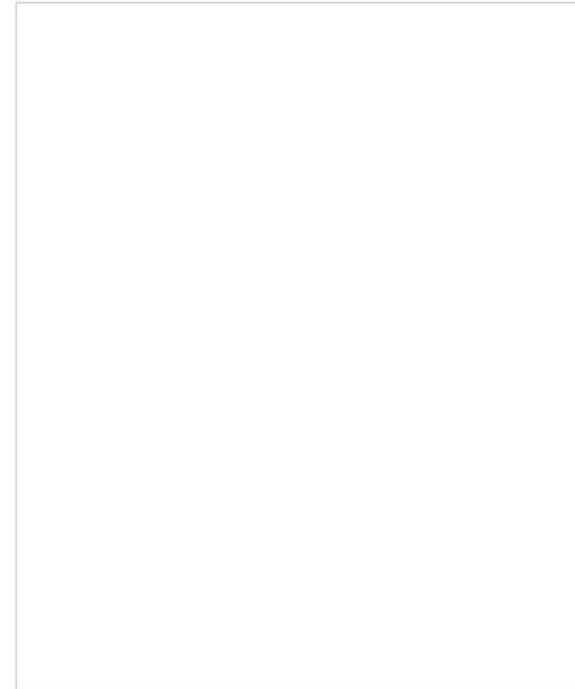
Next >

1 of 2: David's Hobbies

 Time Remaining for this Selection: **1:53**

- 1** How often does David play tennis? David plays tennis...
 - ☐ once a week.
 - ☐ twice a week.
 - ☐ twice a month.
 - ☐ every day.
- 2** With whom does David go to the movies? He goes to the movies with...
 - ☐ friends.
 - ☐ Jenny and his brother.
 - ☐ his brother.
 - ☐ Jenny.
- 3** What hobbies does David say are expensive? David says it is expensive to...
 - ☐ go to restaurants.
 - ☐ go to the movies.
 - ☐ play tennis.

Take notes here:



Next >

CONSTRUCT

Sentence-Level Processing

- Can understand straightforward instructions and simple statements and questions, particularly if delivered at a slower rate than normal, with frequent repetitions or paraphrase. Can understand explicitly stated simple messages when supported by redundancy or context.
- Understands connected phrases and sentences in simple speech. Understands some high frequency vocabulary. Understands some basic structures. May distinguish some different sounds and tones.
- Understanding is limited to self, family, or immediate environment, including meals, time, directions, simple transactions, or personal information. Generally understands short exchanges and statements.
- Demonstrates an extremely limited understanding of cultural norms and expectations.

Understand Single Sentences

- Where does this conversation take place?
- How much less do seniors and children pay for tickets?
- At what time can the children feed the elephants?

ILR 1

At the Zoo

Where does this conversation take place? It takes place...
(Global)

- a. at the information booth
- b. on the telephone
- c. at a ticket counter
- d. at the shop

How much less do seniors and children pay for tickets? They pay... (Detail)

- a. one dollar less
- b. two dollars less
- c. four dollars less
- d. the same

At what time can the children feed the elephants? They can feed them... (Detail)

- a. at 10 in the morning.
- b. at 11 in the morning.
- c. at 1 in the afternoon.
- d. at lunchtime.

Paragraph-Level Processing

- Can understand speech in everyday interactions. Can understand the main ideas and some details in straightforward descriptions, instructions, and narrations across timeframes.
- Can understand speech that conveys sequencing of connected phrases. Can readily understand high frequency vocabulary and sometimes understand references to major historical or cultural events. Generally understands information presented in almost all basic structures. Usually understands standard pronunciation, but not typically implication of stress, intonation, and tone (in tonal languages).
- Understands spoken language about personal life, such as family, activities, and work, or major current events. Understands most main points and some details.
- May demonstrate some understanding of common cultural norms, albeit inconsistently. Understands straightforward spoken language delivered at a normal rate, but complexity or high rates of speech impede understanding.

Understand Paragraphs

- What does the speaker hope to do in this speech?
- Where does the speaker feel that the current administration's biggest failure has occurred?
- What is the speaker most passionate about, regarding the town's future?

ILR 2

A Speech

What does the speaker hope to do in this speech? She hopes to... (Global)

- a. convince her audience to cast their votes for her in an upcoming election.
- b. highlight the failures of the incumbent before an upcoming election.
- c. discuss the many ills of the town in advance of an upcoming election
- d. focus on the dysfunctional school system prior to an upcoming election

Where does the speaker feel that the current administration's biggest failure has occurred? She feels that it has been especially egregious in its inability... (Detail)

- a. to remedy the school system's finances.
- b. to attract outside business to the town.
- c. to reverse a soaring crime rate.
- d. to provide townspeople with employment.

What is the speaker most passionate about regarding the town's future? She is most passionate about... (Inference)

- a. providing the town's children with the best possible education.
- b. lowering the high crime rate.
- c. creating a climate that will entice investors to revive the economy.
- d. tackling the spiraling unemployment that bedevils the town.

Extended Discourse Processing

- Can understand argumentation expressed through supported opinion, conjecture, and hypothesis. Can draw inferences.
- Able to understand cohesive discourse with clear relationships of ideas found in genres such as speeches, lectures, and briefings. Readily understands a broad range of vocabulary, including mid frequency terms and technical vocabulary in the area of professional specialty. Understands widely used idiomatic expressions and cultural references. Understands a variety of structures accurately.
- Can understand speech about one's own fields of competence and matters of general interest. Understands extended speech consistently.
- Understands common cultural norms and expectations in routine and professional speech, though may have difficulty with texts that rely on cultural knowledge. Understands speech delivered at a normal rate with ease, but may have some difficulty with faster than normal spoken language.

Extended Discourse

- What is the importance of attribution theory?
- What does the speaker likely believe is the most effective use of attribution theory?
- Why is the attribution of success or lack thereof to luck seen as potentially negative?

ILR 3

Psychology Seminar

What is the importance of attribution theory? (Global)

- a. It provides a critical component to understanding why people excel in some areas but not others.
- b. It is a valuable new tool for potential educators to endow their students with more confidence.
- c. It assists researchers in their goal of explaining why people have mental blocks to success.
- d. It gives psychologists special insights into why young people succeed or fail at various things.

What does the speaker likely believe is the most effective use of attribution theory? She likely believes that... (Inference)

- a. persuading people to attribute the results of their endeavors to effort can help them improve.
- b. convincing people that their innate abilities drive their ultimate success is essential.
- c. allowing people to self-assess their talents will help them overcome poor self confidence.
- d. educating future teachers about the theory will enable them to get students to do better.

Why is the attribution of success or lack thereof to luck seen as potentially negative? (Detail)

- a. Luck is not a factor that an individual can influence, so it can make people feel powerless.
- b. Luck is inextricably tied with effort, therefore one who perceives they have bad luck stops trying.
- c. Belief in luck will often play a decisive but irrational role in the strategies an individual adopts.
- d. Belief in luck can dissuade one from relying on their innate ability to overcome a challenge

TEST SPECIFICATION CATEGORIES

- Genre/Source
- Content/Setting
- Rhetorical Organization
- Vocabulary
- Listener Purpose (Item Types)
- Clarity of Speech
- Number of Speakers
- Passage Length
- Number and Types of Questions

DISCOURSE TYPE, DIFFICULTY, AND LENGTH

- **ILR 1 passages** are minimally connected and primarily organized in individual utterances and strings of utterances, dealing with a limited number of basic personal and social needs. They include face-to-face conversations, voice mails, straightforward announcements, simple instructions and directions. ILR 1 passages are heavily redundant with a great deal of restatement, paraphrasing, and contextual clues, e.g., written clues (passage title and items).
- **ILR 1+ passages** convey basic information and deal with highly familiar and predictable personal and social topics. They include face-to-face conversations, voice mails, straightforward announcements, simple instructions and directions. ILR 2+ passages exhibit a relatively high level of redundancy with rephrasing and paraphrasing, and contextual clues.

DISCOURSE TYPE, DIFFICULTY, AND LENGTH

- **ILR 1 passages** are understood utterance by utterance. They belong to the most common and most readily understood oral text types of a particular language containing the most frequent common basic words and phrases, common names, cognates and shared international vocabulary. ILR 1 passages contain 40 to 80 words.
- **ILR 1+ passages** predominantly contain high frequency vocabulary, cognates, and shared international vocabulary. ILR 1+ passages contain 100 to 140 words.

DISCOURSE TYPE, DIFFICULTY, AND LENGTH

- **ILR 2 passages** are authentic descriptive and narrative texts such as description of persons, places, and things and narrations about past, present, and future events. Their topics pertain to daily life events (family, work, leisure, sustenance, living/moving, traveling, personal memories). They consist of simple narratives, news stories of people's lives, personal past events, explanations, instructions, anecdotes or travelogue descriptions.
- **ILR 2 discourse** is connected and has a clear predictable structure and predictable content. ILR 2 discourse predominantly contains high-frequency vocabulary and uncomplicated structures. ILR 2 passages contain 160 to 180 words.

DISCOURSE TYPE, DIFFICULTY, AND LENGTH

- **ILR 3 passages** consist of complex extended discourse such as that found in academic and professional settings, lectures, speeches, and reports. They feature analysis, argumentation, supported opinion, and hypothesis dealing with professional, academic, and literary content.
- **ILR 3 passages** contain complex grammatical structures as well as precise, often specialized and low-frequency vocabulary and expressions, including idioms and colloquialisms. In addition, they may frequently contain cultural references. ILR 3 passages contain 250 to 300 words.

ITEM TYPES

- **ILR 1:** 2 Detail, 1 Selective
- **ILR 1+:** 2 Detail, 1 Global
- **ILR 2:** 2 Detail, 1 Global
- **ILR 2+:** 1 Detail, 1 Global, 1 Inference
- **ILR 3:** 1 Detail, 1 Global, 1 Inference

ITEM TYPES ALIGN WITH TEXT LEVELS: DETAIL

- **ILR 1:** Able to comprehend simple single facts. These facts are the easiest to understand aurally and do not necessarily have to be important for the passage as a whole. Distractors are viable passage-based options that are clearly false.
- **ILR 1+:** Able to comprehend simple single facts. These facts contribute to the gist of the passage. Still, their comprehension only requires understanding single-sentence-length utterances one by one as they are spoken. Distractors are viable passage-based options, which may contain words taken from the passage. Keys use synonyms or paraphrases that consist of highly frequent or shared international vocabulary.

ITEM TYPES: DETAIL

- **ILR 2:** Able to understand explicitly mentioned facts and thoughts. They go beyond single utterance-based facts. Their understanding is dependent on understanding the gist of the passage. They generally require understanding more than one utterance. Distractors focus on other relevant facts mentioned in the passage. Keys use synonyms or paraphrases that contain general vocabulary.
- **ILR 3:** Able to understand argument, finer points of detail and abstraction. They require understanding complete subsections of the passage. Keys and distractors focus on finer points of detail and abstraction that support the main argument of the passage. Key uses synonyms and paraphrases. Stem, key, and distractors commonly contain precise, specialized and low-frequency vocabulary and complex structure.

ITEM DIFFICULTY

Items align with their level with respect to function, vocabulary, and grammar.

- **ILR 1:** Most frequent common basic words and phrases, common names, cognates and shared international vocabulary; short, simple sentence-length utterances, predominantly in the present tense.
- **ILR 2:** Variety of frequent words and phrases, cognates, and shared international vocabulary; longer and more complex turns containing some subordinate clauses, prepositional phrases and other features of connected discourse.
- **ILR 3:** Precise, often specialized and low-frequency vocabulary and expressions, including idioms and colloquialisms; complex paragraph-length turns containing subordinate and prepositional clauses, gerunds and participial clauses referring to complex, abstract, and hypothetical argument and relationships.

VALIDITY ARGUMENT

ITEM & EQUIVALENT-FORM RELIABILITY: PROCESS

- Rigorous, standardized, quality assured development process
- Item writers and reviewers: native speakers w/ college degrees and considerable experience in teaching and test writing; provided with detailed manuals and check lists
- Multiple stages:
 - Passage and item development
 - Review 1: authenticity, style, correctness, and item bias and sensitivity check
 - Review 2: level appropriateness, compliance with instructions (manual), quality of items, item bias and sensitivity check
 - Professional recording of passages: acting specialists, professional recordings and postproduction in sound studios
 - Final quality assurance testing at LTI, incl. item sensitivity assurance
 - Rasch analysis of at least 200 examinees at all proficiency levels per test form

UNIVERSITÄT LEIPZIG

Item review based on Rasch analysis following the above protocol

ITEM & EQUIVALENT-FORM RELIABILITY: IRT

- Rasch analyses after 300-400 test administrations
- Nine anchor items for form equating
- Reports include item values and overall test values
 - N of administrations
 - Item difficulty (logits)
 - SEM (logits)
 - Infit and Outfit Statistics
 - Separation Indices
 - Misfitting items are flagged for inspection and/or revision

RELIABILITY RESULTS (BEFORE REVISIONS)

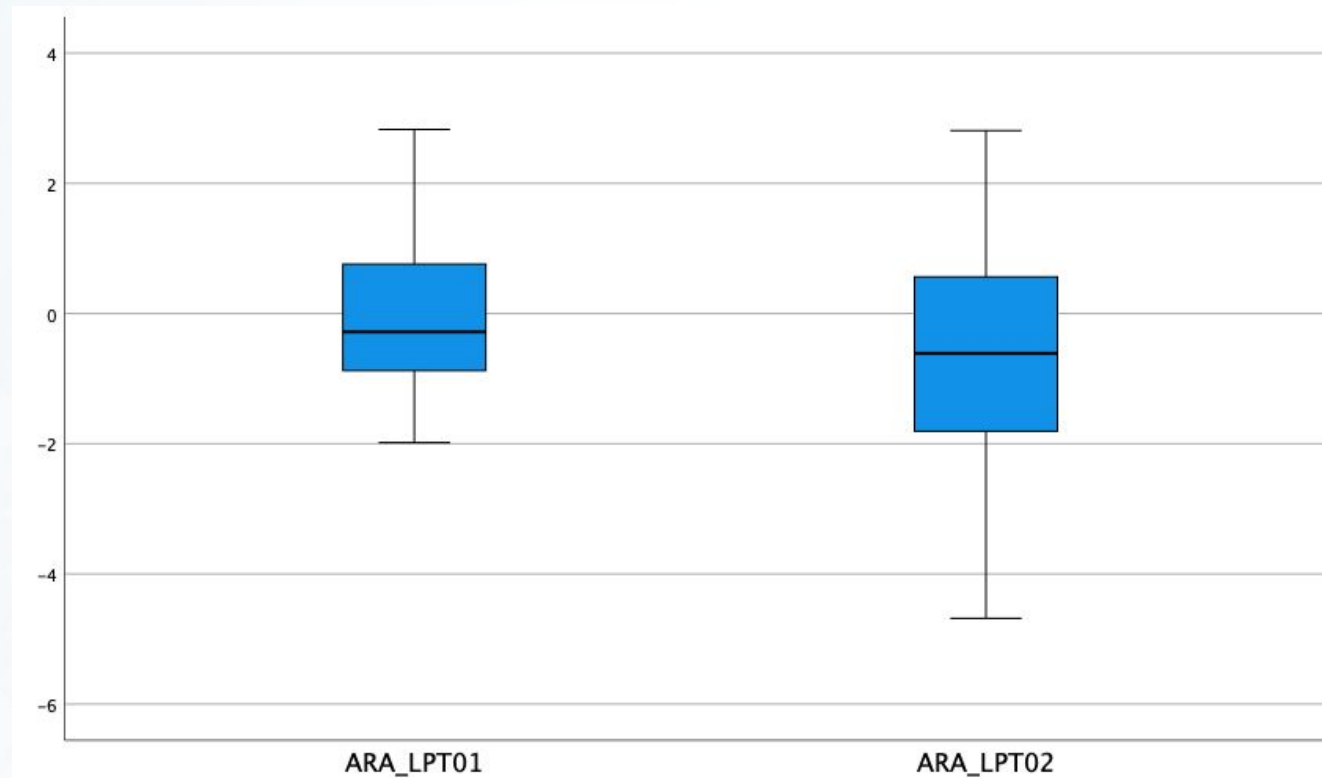
Language	Forms	<i>N</i>	Separation	Infit	Outfit
Arabic	3	280-609	.86-.90	.96-1.01	1.00-1.04
Chinese	2	401-652	.87-.88	1.00-1.02	1.01-1.06
French	4	303-1,064	.80-.87	.98-1.03	1.01-1.08
German	2	443-1,156	.85-.91	.99-1.00	.99-1.06
Italian	1	220	.82	.99	.98
Korean	1	183	.92	1.00	.99
Portuguese	2	147-256	.76-.81	.98-1.00	1.00-1.03
Russian	2	439-476	.88-.89	1.00-1.01	1.07-1.09
Spanish	6	388-1,658	.76-.85	.98-1.03	.99-1.08

SCALE RELIABILITY (VERSION H)

Language	N	Cronbach's Alpha
Arabic LPT 01	219	0.86*
Arabic LPT 02	42	0.87*
German LPT 01	21	0.95*
German LPT 02	15	0.91*
Spanish LPT 01	140	0.89*
Spanish LPT 02	14	0.86*
Spanish LPT 03	335	0.88*
Spanish LPT 04	123	0.88*
Spanish LPT 05	50	0.88*

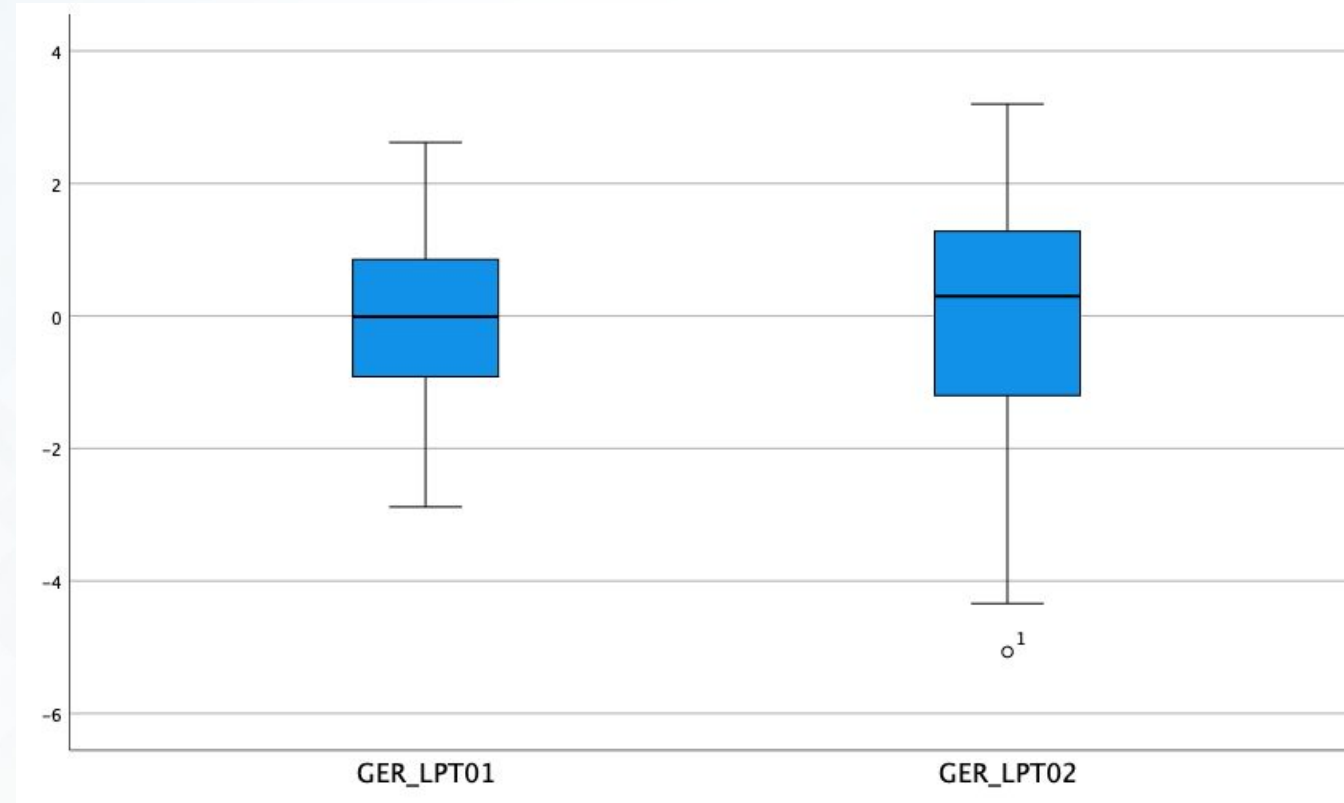
FORM EQUIVALENCE: LOGIT DISTRIBUTIONS

Arabic



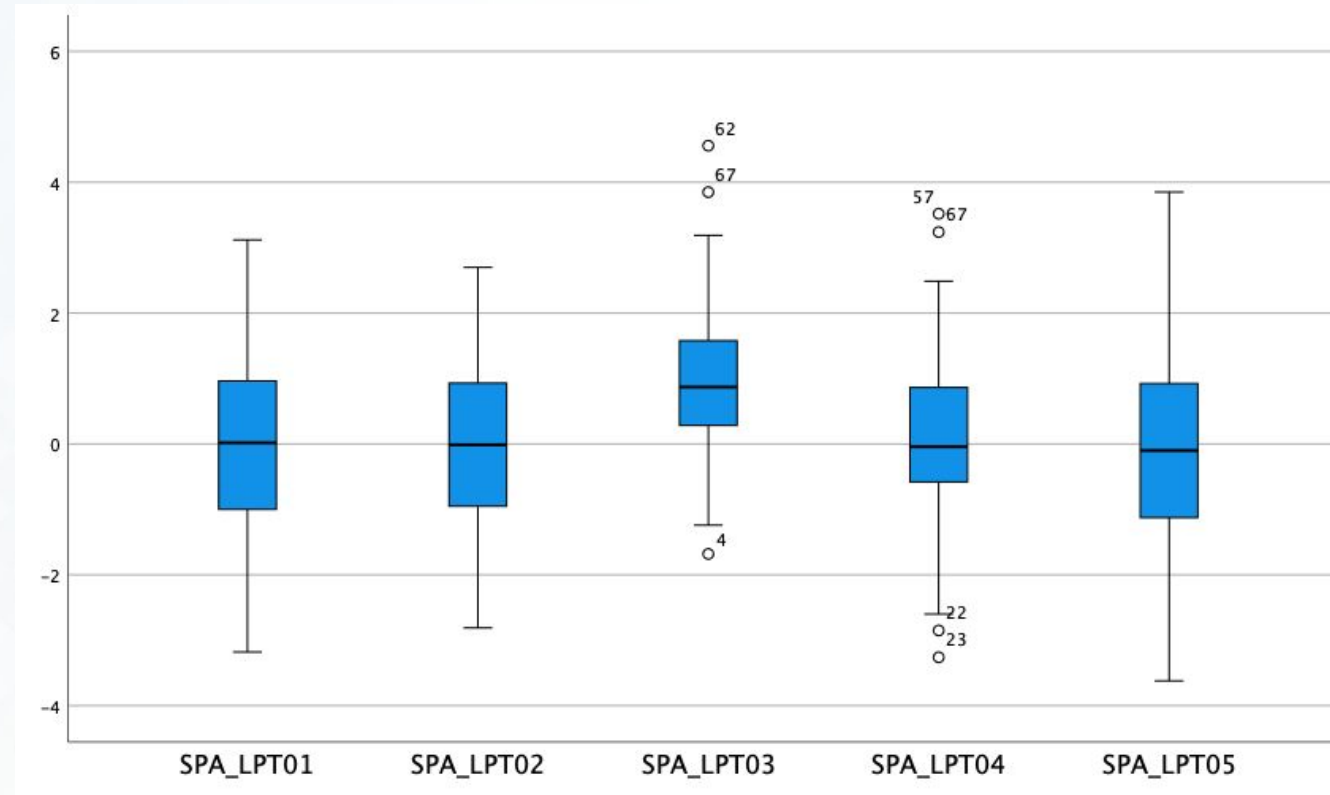
FORM EQUIVALENCE: LOGIT DISTRIBUTIONS

German



FORM EQUIVALENCE: LOGIT DISTRIBUTIONS

Spanish



CONVERGENT VALIDITY: NATO BAT-L

- Benchmark Advisory Test English: Based on the DLPT and ILR Scale (STANAG 6001)
- Split Test Design (N=88)
- Three groups of examinees
 - Lower proficiency: Form E and BAT levels 1-2 (75 min. each)
 - Mid proficiency: Form C and BAT levels 1-2 (50 and 75 min.)

Hi	<i>N</i>	Pearson's <i>r</i>	Spearman's <i>rho</i>	Goodman-Kruskall's <i>gamma</i>
	88	.842*	.833*	.898*

* $p < .01$

CONVERGENT VALIDITY: VOCABULARY SIZE

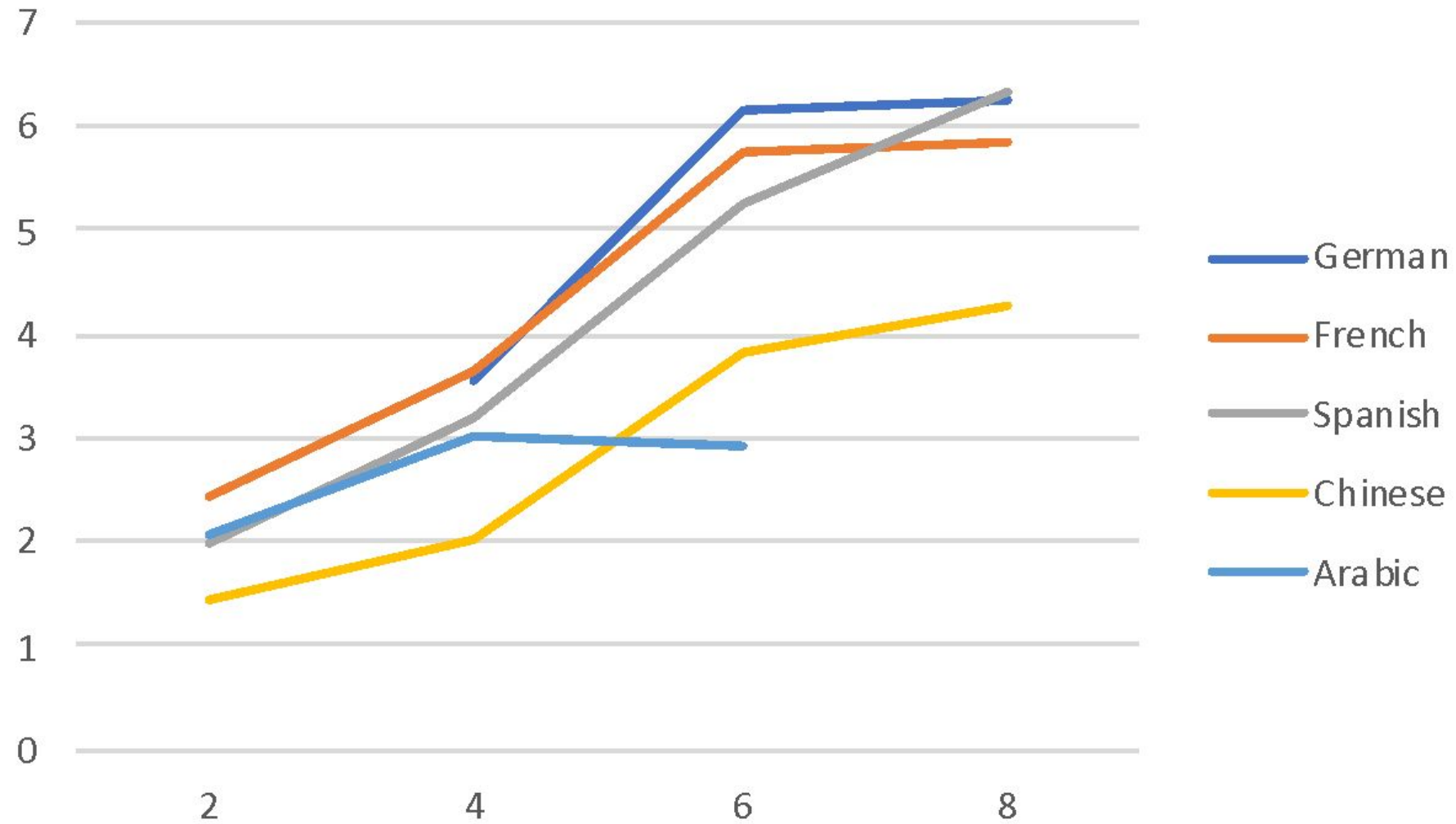
Vocabulary size as measured according to how many of the frequent words of a language are known is highly predictive of reading and listening proficiency. A high correlation between a listening proficiency test and vocabulary size, therefore, provides evidence of convergent validity.

	N	Pearson's <i>r</i>	Spearman's rho
LPT	607	0.80*	0.76*
RPT	1099	0.82*	0.84*

* $p < .01$

CONVERGENT VALIDITY: PROFICIENCY DEVELOPMENT OVER TIME

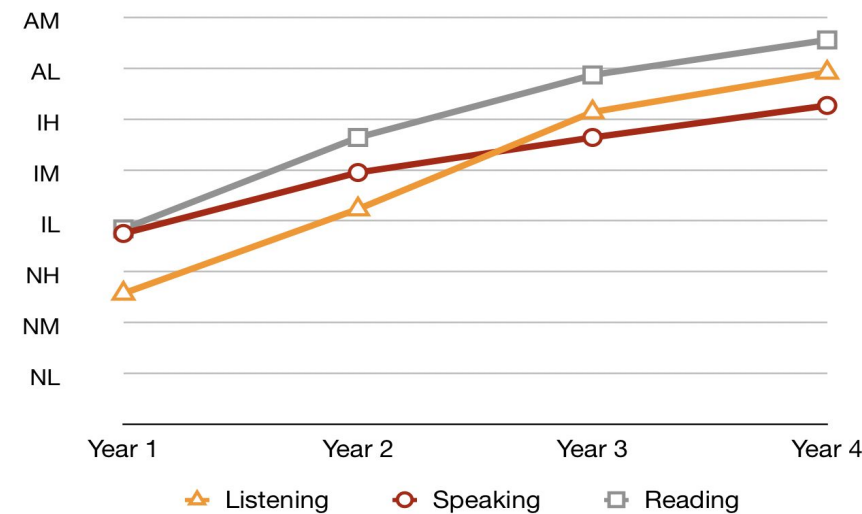
Listening



SPANISH PROFICIENCY (SONESON, 2019)

Spanish Mean Proficiency Ratings by Curriculum Year

	Year 1			Year 2			Year 3			Year 4		
	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N
Listening	2.57	1.13	87	4.23	1.07	192	6.14	1.08	161	6.92	0.95	91
Speaking	3.75	0.72	87	4.95	0.72	231	5.64	0.76	194	6.27	0.79	96
Reading	3.84	1.01	87	5.64	1.28	226	6.87	0.89	193	7.56	0.85	97



Q&A

For any additional questions,
please contact us at:

Dr. Tschirner: tschirner@uni-leipzig.de

Dr. Wagener: Twagener@actfl.org

Kim Sallee: Ksallee@languagetesting.com