

**O'ZBEKISTON RESPUBLIKASI**  
**MAKTABGACHA VA MAKTAB TA'LIMI VAZIRLIGI HUZURIDAGI**  
**IXTISOSLASHTIRILGAN TA'LIM MUASSASALARI AGENTLIGI**  
**IS'HOQXON IBRAT NOMIDAGI NAMANGAN DAVLAT CHET**  
**TILLARI INSTITUTI**

**INGLIZ TILI O'QITISH METODIKASI KAFEDRASI**

**INGLIZ TILI O'QITISH VA BAHOLASHNING XALQARO**  
**STANDARTLARI**

fanidan

**O' Q U V – U S L U B I Y**  
**MAJ M U A**



**2025/2026 o'quv yili kunduzgi ta'lim shakli, 3-kurslari uchun**

Bilim sohasi:	100000 – Ta'lim
Ta'lim sohasi:	110000 – Ta'lim
Ta'lim yo'nalishi:	60111800 – Xorijiy til va adabiyot (ingliz tili)

Namangan-2025/2026

Ushbu o'quv uslubiy majmua O'zbekiston Respublikasi Oliy va o'rta maxsus ta'lim vazirligining 2017-yil 1-martdagi 107-sonli buyrug'i asosida tayyorlandi. Is'hoqxon Ibrat nomidagi Namangan davlat chet tillari institutining 2025-yil 29-avgustdagi 1-sonli kengashida tasdiqlandi.

**Tuzuvchi:** **M.Abdullayeva** - Ingliz tili o'qitish metodikasi kafedrasida katta o'qituvchisi, (PhD)

**Taqrizchilar:** A.Amanov - Ingliz tili o'qitish metodikasi kafedrasida mudiri, dotsent  
M.Yuldasheva - Ingliz tili o'qitish metodikasi kafedrasida Katta o'qituvchisi

**O'quv uslubiy majmua Namangan davlat chet tillari instituti Til va tarjima fakulteti ilmiy kengashida ko'rib chiqilgan va tasdiqqa tavsiya qilingan.**

2025-yil 28-avgust 1-sonli majlis bayoni.

**Fakultet dekani:**  **S.Misirov**



## MUNDARIJA

<b>№</b>	<b>MAVZULAR NOMI</b>
<b>1</b>	<b>O'QUV MATERIALLAR</b> asosiy matn; topshiriqlar variantlari; qo'shimcha maqolalar;
<b>2</b>	<b>FAN O'QUV DASTURI</b>
<b>3</b>	<b>GLOSSARIY</b>
<b>4</b>	<b>ILOVALAR</b>
	testlar;
	tarqatma materiallar;

## **Introduction: Why Do We Need Standards?**

- English today functions as a **global lingua franca** in business, science, education, diplomacy, and travel. Learners in Uzbekistan, China, Brazil, or Germany may all aim for proficiency — but how do we compare their levels fairly?
- **Problem before standards:** Different countries and institutions used **their own criteria** for “beginner,” “intermediate,” or “advanced,” which caused confusion. For example, a “B2” student in France may not be equivalent to an “intermediate” student in Japan.
- **Solution:** International standards provide:
  - A **shared frame of reference** for describing skills.
  - Consistency in **curricula, textbooks, and exams**.
  - Fairness and comparability across borders.
- **Quote to use:** *“Without standards, assessment is local. With standards, assessment is global.”*

## **Historical Background of Standards in Language Teaching**

- **Early 20th century:** Language learning was dominated by the **Grammar-Translation Method**. Proficiency was judged by knowledge of grammar and translation ability.
- **Mid-20th century:** Rise of **audio-lingual** and **communicative approaches** emphasized speaking skills. But still, there was no **universal agreement** on what “proficiency” meant.
- **1960s–1970s:** Standardized English proficiency tests such as **TOEFL (USA)** and **Cambridge English exams (UK)** emerged. However, these were designed for specific contexts (academic study or migration).
- **1990s:** The **Council of Europe** created the **Common European Framework of Reference for Languages (CEFR)** to unify descriptions of proficiency across Europe — and later, worldwide.

- **Today:** CEFR, ACTFL, TESOL Standards, and large-scale exams (IELTS, TOEFL, PTE, Cambridge) shape how English is taught and assessed internationally.

## Major International Standards and Frameworks

### CEFR (Common European Framework of Reference for Languages)

- Developed by the Council of Europe in 2001; updated in 2020 with the **Companion Volume**.
- **Six levels:**
  - **A1–A2:** Basic User (can introduce oneself, order food, handle simple tasks).
  - **B1–B2:** Independent User (can discuss everyday topics, travel, study in English-medium programs).
  - **C1–C2:** Proficient User (can understand complex texts, use language flexibly, function at near-native level).
- **Key Feature:** *Can-do descriptors* (e.g., “Can ask for directions and understand simple responses” at A2).
- **Practical Use:** Most international exams (IELTS, TOEFL, Cambridge) are mapped to CEFR levels. Textbooks and curricula are also aligned to CEFR.

### ACTFL Proficiency Guidelines (USA)

- Developed by the **American Council on the Teaching of Foreign Languages**.
- Levels: Novice, Intermediate, Advanced, Superior, Distinguished. Each subdivided into Low, Mid, High.
- Emphasis: **what learners can do in real-life tasks** rather than what grammar they know.
- Example:
  - *Novice Mid:* Can introduce self, use memorized phrases.
  - *Intermediate High:* Can handle everyday conversations and narrate past events.
  - *Superior:* Can discuss abstract topics and support opinions.

## TESOL Standards (International)

- Developed by **TESOL International Association**.
- Focuses on:
  - Teacher knowledge (linguistics, pedagogy, intercultural competence).
  - Effective teaching practices (curriculum design, assessment literacy).
  - Professional growth and ethics.
- Useful not only for learners but also for **teacher training** and **quality assurance** in ELT programs.

## International Exams and Benchmarks

- **IELTS** – measures listening, reading, writing, speaking; widely used for migration and university entry.
- **TOEFL iBT** – academic English test, especially for US universities.
- **PTE Academic** – computer-based test with fast results.
- **Cambridge English Exams** – progressive levels: KET (A2), PET (B1), FCE (B2), CAE (C1), CPE (C2).
- All of these are now **aligned with CEFR** to allow global comparability.

## Principles of International Assessment

1. **Validity** – The test measures what it claims (e.g., speaking test measures speaking, not writing).
2. **Reliability** – Results are consistent across time, examiners, and tasks.
3. **Fairness** – No cultural or social bias; tasks should be accessible to all learners.
4. **Transparency** – Learners know the criteria (rubrics, band descriptors).
5. **Balance:**
  - **Formative assessment:** ongoing, in-class (peer feedback, quizzes).
  - **Summative assessment:** standardized exams (IELTS, TOEFL).

## Benefits of International Standards

- **Comparability:** Students in different countries can be compared on the same scale.
- **Mobility:** Universities, employers, and immigration offices accept CEFR-linked scores.

- **Transparency:** Learners know their progress and goals (“I am B1 now; I need B2 to study abroad”).
- **Professional growth for teachers:** Standards guide teaching, curriculum design, and training.
- **Policy-making:** Governments use standards to reform education systems (e.g., Uzbekistan aligning curricula to CEFR).

### Challenges and Criticisms

- **Cultural bias:** Some tests include references unfamiliar to learners (e.g., Western holidays, idioms).
- **Over-standardization:** Teachers may “teach to the test,” neglecting creativity or local needs.
- **Accessibility:** International exams can be expensive; not all learners can afford them.
- **Adaptation:** Local education ministries must balance **international benchmarks** with **national educational priorities**.

### Practical Applications for English Teaching

- **Lesson planning:** Teachers design lessons around CEFR “can-do” descriptors.
  - Example: *B1 Writing can-do: “Can write a personal letter describing experiences.”*
- **Assessment tasks:** Oral exams, writing tasks, and projects designed to reflect real-world communication.
- **Student self-assessment:** Learners use checklists based on CEFR to monitor their progress.
- **Curriculum design:** Ministries and universities use CEFR/ACTFL to build syllabi.

### Future Directions in Standards and Assessment

- **Digital and AI-based assessment:** Automated scoring of speaking and writing (e.g., Duolingo English Test).
- **Adaptive testing:** Computer adjusts difficulty based on learner performance.
- **Inclusive assessment:** Accommodations for learners with disabilities and multilingual learners.
- **Micro-credentials & lifelong learning:** Short courses and online badges recognized internationally.

### References (Suggested Reading)

- Council of Europe. (2020). *CEFR Companion Volume*.
- ACTFL. (2012). *ACTFL Proficiency Guidelines*.
- TESOL International Association. (2018). *TESOL Standards for ESL/EFL Teachers*.
- Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment*. Routledge.
- Weir, C. J. (2005). *Language Testing and Validation*. Palgrave Macmillan.

### Recommended Articles and Resources:

#### 1. *The Role of Language Standards and Systems in Learning and Teaching English*

- **Author:** Nguyen Thi Ngoc Thu (2023)

#### 2. *Language Standards Frameworks and Language Proficiency Assessment*

- **Author:** M. Gottlieb (2022)

#### 3. *Ask an Expert: World Language Standards*

- **Journal:** Foreign Language Annals (2022)

#### 4. *The CEFR in Language Learning, Teaching, and Assessment*

- **Author:** E. Piccardo (circa 2021)

### Lecture 2: *Theoretical Foundations of Language Assessment*

#### Introduction: Why Study Assessment Theory?

- Assessment is not just about testing; it is a **theoretical field** that shapes how we measure and interpret language learning.
- A good assessment system ensures:
  - **Validity** (we test what we intend to test),
  - **Reliability** (results are consistent),
  - **Fairness** (equal for all learners),
  - **Practicality** (useful in real classrooms).
- Understanding these foundations helps teachers **design better classroom assessments** and interpret results from **international exams**.

## Historical Perspectives

### Structuralist & Behaviorist Approaches (1950s–1970s)

- **Focus:** Language as a set of discrete skills (grammar, vocabulary, pronunciation).
- **Assessment method:** *Discrete-point testing* (Lado, 1961) → one item per language point.
- **Example:** Multiple-choice test: “Choose the correct form of the verb.”
- **Criticism:** Did not capture **communicative competence**.

### Communicative Revolution (1970s–1990s)

- Inspired by **Hymes (1972)**: *communicative competence* = not just grammar, but sociolinguistic, discourse, and strategic competence.
- **Assessment method:** *Integrative tests* (e.g., cloze tests, dictations, interviews).
- **Example:** Interview tasks, role-plays, writing an email.
- Led to **performance-based testing** (tasks reflect real-world communication).

### Socio-cognitive Approach (2000s–present)

- Proposed by **Weir (2005)**, further by **O’Sullivan (2011)**.
- Views assessment as interaction between **test-taker, task, and context**.
- **Example:** IELTS speaking tasks simulate real communication; scoring considers fluency, coherence, and interaction.
- Ensures **validity** through theory-based task design.

## Core Theoretical Principles of Language Assessment

### Validity

- **Definition:** The degree to which a test measures what it claims.
- Types of validity:
  - **Content validity** – coverage of syllabus (e.g., vocabulary test matches taught units).
  - **Construct validity** – alignment with theory of language ability (e.g., communicative test covers interaction, not just grammar).
  - **Criterion validity** – correlation with external measures (e.g., TOEFL score vs. GPA).
- **Source:** Messick (1989) – unified theory of validity.

### Reliability

- **Definition:** Consistency of test scores across raters, tasks, or time.
- **Example:** Two teachers marking the same essay should give similar scores (inter-rater reliability).
- **Methods:** Statistical analysis, rubrics, standardization training.

### Practicality

- A test must be feasible in terms of **time, cost, and resources**.
- **Example:** National exams must balance comprehensiveness with efficiency (e.g., Uzbekistan CEFR-aligned national tests).

### Washback Effect

- Tests influence **teaching and learning**.
- **Positive washback:** Assessment encourages communicative teaching (e.g., task-based exams).
- **Negative washback:** Teaching “to the test” with memorization and drills.
- **Source:** Alderson & Wall (1993).

### Fairness & Ethics

- A test must not disadvantage learners due to culture, gender, or disability.
- **Example:** Using culture-specific idioms in listening may disadvantage non-native learners.

- Increasing focus on **inclusive assessment** (accommodations for special needs learners).

## Frameworks of Language Ability for Assessment

### Bachman & Palmer's Model (1996)

- Language ability = **language knowledge** + **strategic competence**.
- Influences large-scale tests like TOEFL and IELTS.

### CEFR (2001, 2020)

- Provides **can-do descriptors** across six levels (A1–C2).
- Supports **task-based assessment**.
- **Example:** At B1 speaking: *“Can enter unprepared into conversation on familiar topics.”*

### ACTFL Guidelines (2012)

- Emphasize **real-world performance**.
- Useful in American and international education contexts.

## Examples of Assessment Tasks by Theory

Theory	Example Task	Focus
Structuralist	Grammar multiple-choice	Accuracy
Communicative	Role-play (ordering food)	Meaning, fluency
Socio-cognitive	Group discussion with rating scale	Interaction, strategies
CEFR	Writing a travel blog at B2	Task authenticity
ACTFL	Oral proficiency interview	Real-world use

## Challenges in Language Assessment

- **Cultural bias** in test design.
- Balancing **standardization** with **authenticity**.
- Overemphasis on summative exams vs. need for formative classroom assessment.
- Ethical issues: high-stakes testing and learner anxiety.

## Future Directions

- **AI and digital testing** (automated essay scoring, adaptive testing).
- **Inclusive assessment** for learners with disabilities and multilingual backgrounds.
- **Alternative assessment:** portfolios, self-assessment, peer-assessment.
- Ongoing research into **dynamic assessment** (based on Vygotsky's Zone of Proximal Development).

### **Conclusion**

- Language assessment is grounded in theories of language and learning.
- Key principles: **validity, reliability, fairness, practicality, washback.**
- Modern assessment integrates **communicative competence** and **real-life tasks.**
- Teachers need **assessment literacy** to apply these principles in classrooms.

### **Key References**

- Alderson, J. C., & Wall, D. (1993). *Does washback exist?* *Applied Linguistics*, 14(2), 115–129.
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford University Press.
- Brown, H. D., & Abeywickrama, P. (2018). *Language Assessment: Principles and Classroom Practices*. Pearson.
- Hymes, D. (1972). *On Communicative Competence*.
- Lado, R. (1961). *Language Testing: The Construction and Use of Foreign Language Tests*. McGraw-Hill.
- Messick, S. (1989). *Validity*. In R. Linn (Ed.), *Educational Measurement* (3rd ed.).
- Weir, C. J. (2005). *Language Testing and Validation*. Palgrave Macmillan.

### **Articles for further reading:**

1. **Assessment for learning or assessment for scoring? Washback of analytic rubrics in Vietnamese EFL classrooms.**

Hoang Yen Phuong, Thanh Thao Le & Trut Thuy Pham

## **2. Ethical Issues in Language Testing from a Validity Perspective**

Shuling Gao, Jiayi Liu

## **3. Responsible AI for Test Equity and Quality: The Duolingo English Test as a Case Study**

Jill Burstein, Geoffrey T. LaFlair, Kevin Yancey, Alina A. von Davier, Ravit Dotan

## **4. Exploring AI-Enabled Test Practice, Affect, and Test Outcomes in Language Assessment.** Jill Burstein, Ramsey Cardwell, Ping-Ling Chuang, Allison Michalowski, Steven Nydick.

### ***Lecture 3: Types of Language Assessment***

#### **Introduction: Why Different Types of Assessment?**

- Language learning is multifaceted: grammar, vocabulary, reading, writing, speaking, listening, interaction, cultural competence.
- No single test can measure all aspects, so teachers and institutions use different types of assessment depending on goals.
- Assessments differ by:
  - Purpose (diagnosis, feedback, certification).
  - Timing (before, during, or after instruction).
  - Method (formal, informal, performance-based).
- Knowing the types helps teachers choose the right tool for the right context.

## **Major Types of Language Assessment**

### **Placement Assessment**

- Purpose: To place learners in the appropriate level/course before instruction begins.
- Features: Tests general ability, not specific course content.
- Examples:
  - Online placement test at a university (e.g., CEFR-based).
  - Diagnostic vocabulary + grammar quiz at the start of a semester.
- Key Point: Should be reliable, quick, and reflect learners' overall proficiency.

### **Diagnostic Assessment**

- Purpose: Identifies learners' strengths and weaknesses in specific areas.
- When: Beginning or early stages of a course.
- Examples:
  - Pronunciation screening test.

- Vocabulary checklist for academic English.
- Reading comprehension sub-skills test (skimming, scanning, inference).
- Benefit: Helps teachers adapt instruction to learners' needs.

### **Formative Assessment**

- Purpose: Provides ongoing feedback during the learning process.
- Features: Low-stakes, frequent, supportive.
- Examples:
  - Teacher feedback on a draft essay.
  - Peer review in speaking tasks.
  - Online quizzes with immediate feedback.
- Key Idea: *Assessment for learning* (AfL).
- Benefit: Encourages self-monitoring and motivation.

### **Summative Assessment**

- Purpose: Measures achievement at the end of a unit, course, or program.
- Features: High-stakes, formal, standardized.
- Examples:
  - Final exam in a language course.
  - National English proficiency test.
  - IELTS, TOEFL, Cambridge exams.
- Key Idea: *Assessment of learning* (AoL).
- Caution: Risk of negative washback if it drives teaching too narrowly.

### **Achievement Assessment**

- Purpose: Evaluates what learners have achieved in relation to course objectives.
- When: End of a unit or semester.
- Examples:
  - Unit vocabulary and grammar test.
  - Midterm exam covering taught materials.
- Key Point: Directly linked to syllabus, not general proficiency.

### **Proficiency Assessment**

- Purpose: Measures general language ability, regardless of specific instruction.
- Examples:
  - IELTS, TOEFL, PTE Academic.
  - CEFR-aligned national exams.
- Use: Admission to universities, immigration, job applications.
- Key Point: Independent of curriculum; reflects real-world communicative competence.

### **Performance-Based Assessment**

- Purpose: Measures language in real-life tasks.
- Examples:
  - Role-play ordering food in a restaurant.
  - Group discussion on a global issue.
  - Oral presentations and portfolios.
- Benefit: Encourages meaningful use of language.
- Challenge: Requires clear rubrics and trained raters for reliability.

## **Alternative Assessment (Non-traditional)**

- Forms: Portfolios, journals, projects, self-assessment, peer-assessment.
- Examples:
  - Student keeps a reflective journal in English.
  - E-portfolio with recordings, essays, presentations.
  - Peer evaluation of speaking tasks.
- Advantages: Promotes autonomy, lifelong learning skills.
- Limitation: More subjective; requires teacher training in rubric design.

## **High-Stakes vs. Low-Stakes Assessment**

- High-stakes: Major consequences for learners (e.g., university entry, job). → IELTS, TOEFL.
- Low-stakes: Minimal consequences, mainly feedback (e.g., weekly quizzes, class activities).
- Teaching note: Balance both types to reduce learner anxiety.

## **Formal vs. Informal Assessment**

- Formal: Planned, structured, documented (exams, quizzes, standardized tests).
- Informal: Spontaneous teacher feedback, observation, questioning in class.
- Example: Correcting a student's speaking error (informal).
- Teaching tip: Both types complement each other.

## **Comparative Table of Assessment Types**

Type	Purpose	Timing	Example
Placement	Group learners	Before course	CEFR placement test
Diagnostic	Identify strengths/weaknesses	Early stage	Pronunciation screening
Formative	Feedback, improve learning	During course	Draft essay feedback
Summative	Measure outcomes	End of course	Final exam
Achievement	Course-based success	End of unit/term	Midterm vocabulary test
Proficiency	General ability	Anytime	IELTS, TOEFL
Performance-based	Real-life tasks	During/after course	Role-play, presentation
Alternative	Learner-centered	Ongoing	Portfolios, journals
High-stakes	Decisions, certification	End point	University entry exam
Informal	Spontaneous support	Anytime	Teacher's oral feedback

### Practical Classroom Examples

- Placement: A CEFR-aligned online quiz to divide learners into A2, B1, B2 groups.
- Diagnostic: Listening test focusing on *gist* vs *detail* comprehension.
- Formative: Peer feedback circles on oral presentations.
- Summative: End-of-semester writing exam graded with rubrics.

- Performance-based: Debate on “The Role of Technology in Education.”
- Alternative: Students compile a digital portfolio showing progress across the semester.

### **Challenges and Considerations**

- Balancing validity (measuring the right skills) with reliability (consistent scoring).
- Ensuring fairness across diverse student populations.
- Managing washback effect: encouraging positive learning, avoiding “teaching to the test.”
- Resource constraints: Some methods (e.g., portfolios) require more time and training.

### **References (Recent and Classical)**

- Brown, H. D., & Abeywickrama, P. (2018). *Language Assessment: Principles and Classroom Practices* (2nd ed.). Pearson.
- Fulcher, G. (2021). *Language Testing and Assessment: Routledge Introductions to Applied Linguistics*. Routledge.
- Gao, L., & Liu, J. (2023). Ethical Issues in Language Testing from a Validity Perspective. *Open Journal of Modern Linguistics*, 13(3).
- Li, J. (2025). Test fairness of the in-house College English examination. *Frontiers in Education*.
- Nguyen, T. P. (2025). A literature review of washback effects of assessment. *Language Testing in Asia*, 15(2).
- Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Palgrave Macmillan.

### **Lecture 4: Test Design and Development in Language Assessment**

## **Introduction: Why Do We Need Careful Test Design?**

- **Assessment is not just a set of questions — it is a systematic process.**
- **Poorly designed tests can lead to:**
  - Invalid results (measuring the wrong construct),
  - Unreliable scores (inconsistent results),
  - Unfair consequences (bias against learners).
- **Well-designed tests:**
  - Reflect the construct of language ability (what we want to measure),
  - Are aligned with learning objectives and standards (CEFR, ACTFL),
  - Provide useful information for learners, teachers, institutions.

## **Theoretical Foundations in Test Design**

- **Validity (Messick, 1989):** The central concept — tests must measure intended constructs.
- **Reliability:** Scores must be consistent across items, time, raters.
- **Practicality:** Balance between quality and resources (time, cost, technology).
- **Fairness:** Avoid bias; ensure accessibility for all learners.
- **Washback (Alderson & Wall, 1993):** The test should promote *positive* effects on teaching and learning.

## **Steps in Test Design and Development**

### **Defining Purpose and Constructs**

- Ask: *Why are we testing?* → Placement, diagnosis, achievement, proficiency, research.
- Define the construct = the specific aspect of language ability measured.

- Example: Reading comprehension (skimming, scanning, inference).
- Example: Speaking interactional competence (turn-taking, fluency).

### **Test Specifications (Specs)**

- A blueprint for the test:
  - Target population (school learners, university, migrants).
  - Proficiency level (CEFR B1, ACTFL Intermediate).
  - Skills tested (listening, reading, writing, speaking).
  - Task types (multiple-choice, essays, interviews, role plays).
  - Time allocation.
- Example: IELTS Writing Task 2 spec: 250-word essay, 40 minutes, argumentative/discursive.

### **Item and Task Writing**

- Discrete-point items: focus on one element (e.g., grammar multiple-choice).
- Integrative items: combine skills (e.g., cloze test, dictation).
- Performance tasks: real-world activities (e.g., group discussion).
- **Principles for good items:**
  - Clear instructions.
  - Avoid cultural bias.
  - Only one correct answer (for objective items).
  - Aligned with construct.

### **Trialing and Piloting**

- **Pre-test items with a similar group of learners.**
- **Collect data on:**

- Item difficulty (p-values).
- Discrimination index (does it separate strong vs weak students?).
- Reliability coefficients (Cronbach's alpha).
- **Revise or discard problematic items.**

### **Administration and Scoring**

- Administration: Decide format (paper-based, computer-based, adaptive).  
Ensure security and standardized procedures.
- **Scoring:**
  - Objective items → automatically scored.
  - Subjective tasks (essays, oral interviews) → need rubrics and rater training.
- Example: IELTS Speaking uses a band descriptor with 4 criteria: fluency, lexical resource, grammar, pronunciation.

### **Analysis and Validation**

- **After administration, analyze results:**
  - Reliability statistics (Cronbach's alpha, inter-rater reliability).
  - Validity evidence (content, construct, criterion).
  - Test fairness (Kunnan's Test Fairness Framework).
- Revise test for future use.

### **Models of Test Development**

- **Classroom context:**
  - **Teacher creates short quizzes, midterms, final exam → guided by syllabus objectives.**
- **Large-scale context:**

- National exams (Uzbekistan CEFR-aligned test, TOEFL, IELTS).
- Requires extensive piloting, psychometric analysis, ongoing validation.
- **Cycle Model (Hughes, 2003):**

**Define purpose** →

**Design specs** →

**Write items** →

**Trial/pilot** →

**Analyze results** →

**Revise** →

**Operational test.**

### **Examples of Test Development in Practice**

- Placement Test: Online adaptive CEFR quiz → quickly adjusts difficulty.
- Achievement Test: Unit vocabulary/grammar exam aligned with syllabus.
- Performance Assessment: Portfolio of essays + oral presentations.
- National Test: Uzbekistan's CEFR-based English proficiency exam for university students (summative, standardized).

### **Current Trends in Test Design**

- AI and computer-adaptive testing: Tailors difficulty to learner performance (Duolingo English Test).
- Automated scoring systems: Use natural language processing to assess writing and speaking.
- Inclusivity in design: Accommodations for learners with special needs (extra time, simplified instructions).

- Learning-oriented assessment (LOA): Integrates formative elements into test design (feedback-rich tasks).

### **Challenges in Test Development**

- Balancing authenticity with practicality.
- Ensuring cultural neutrality (avoid idioms only familiar to certain groups).
- Managing test security in digital platforms.
- Training teachers and raters to maintain scoring reliability.

Designing and developing a language test is not just about writing a few questions. It is a systematic process that requires theoretical grounding, careful planning, and ongoing evaluation. A poorly designed test may produce results that are invalid, unreliable, or unfair, which can harm learners and misinform teachers. By contrast, a well-designed test reflects the construct of language ability, is aligned with learning goals, and provides useful feedback for teaching and learning.

### **Foundations of Test Design**

Every test should rest on key principles of assessment:

- Validity – Does the test measure what it intends to measure?
- Reliability – Are the results consistent across time, tasks, and raters?
- Fairness – Are all learners treated equally, without bias or disadvantage?
- Practicality – Is the test feasible in terms of time, cost, and administration?
- Washback – What effect does the test have on teaching and learning?  
Positive washback encourages meaningful learning; negative washback promotes rote memorization.

These principles guide each stage of the test development cycle.

### **Stages of Test Design and Development**

## **Defining Purpose and Construct**

The first step is to decide why the test is being created. Is it for placement, diagnosis, achievement, proficiency, or certification? Next, the construct—the aspect of language ability being measured—must be defined clearly. For example, a listening test might focus on *gist comprehension*, *detail recognition*, or *inference-making*.

## **Writing Test Specifications (Specs)**

Test specifications are the blueprint of the assessment. They outline:

- The target learners (age, proficiency, context).
- The skills to be assessed (reading, writing, listening, speaking).
- The task types (multiple-choice, essays, role plays).
- The length and timing of the test.
- The scoring method (points, rubrics, bands).

Clear specifications ensure that future test versions are consistent and comparable.

## **Item and Task Development**

Once specs are ready, items and tasks can be created. Good items should:

- Be clear and unambiguous.
- Avoid cultural bias or irrelevant difficulty.
- Reflect authentic language use.
- Match the intended construct.

Examples:

- Objective item: Grammar multiple-choice (“She \_\_\_ to school yesterday.” a) go b) goes c) went d) gone).
- Integrative item: Cloze passage or dictation.

- Performance task: Oral presentation or debate evaluated with a rubric.

### **Trialing and Piloting**

Before official use, the test must be piloted with a similar group of learners. This stage provides data on:

- Item difficulty (are items too easy/too hard?).
- Discrimination index (do items separate stronger from weaker learners?).
- Reliability coefficients (Cronbach's alpha, inter-rater consistency).

Based on pilot results, problematic items are revised or removed.

### **Administration and Scoring**

At this stage, the test is delivered under controlled conditions. Administration should ensure security and fairness. Scoring methods differ by item type:

- Objective items are scored automatically.
- Subjective tasks (essays, interviews) require clear rubrics and trained raters. For example, IELTS Writing and Speaking are graded with band descriptors across multiple criteria.

### **Analysis, Validation, and Revision**

After test administration, results are analyzed:

- Reliability statistics check internal consistency.
- Validity evidence is gathered (content validity, construct validity, criterion validity).
- Fairness checks ensure no group was disadvantaged.

Based on findings, the test is revised for future use. Test design is therefore cyclical, not linear.

### **Practical Applications**

- Classroom context: Teachers may design short quizzes, midterms, and finals guided by course objectives.
- Institutional context: Universities use placement and proficiency tests for admissions and grouping.
- Large-scale context: National or international exams (IELTS, TOEFL, CEFR-based tests) involve extensive piloting, statistical analysis, and global standardization.

### **Emerging Trends in Test Design**

- Computer-Adaptive Testing (CAT): Items adjust in difficulty depending on learner responses (e.g., Duolingo English Test).
- AI-based scoring: Automated systems for essays and speech analysis.
- Inclusive design: Accommodations for learners with special needs.
- Learning-Oriented Assessment (LOA): Integrating feedback into test design to support ongoing learning.

### **References**

- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in Practice*. Oxford University Press.
- Brown, H. D., & Abeywickrama, P. (2018). *Language Assessment: Principles and Classroom Practices* (2nd ed.). Pearson.
- Fulcher, G. (2021). *Language Testing and Assessment*. Routledge.
- Kunnan, A. J. (2018). *Evaluating Language Assessments*. Routledge.
- Weir, C. J. (2005). *Language Testing and Validation*. Palgrave Macmillan.

### **Lecture 5: Assessing Receptive Skills: Reading & Listening**

## **1. Introduction**

Language learning involves two broad categories of skills: receptive skills (reading and listening) and productive skills (speaking and writing). Receptive skills are the ways learners receive, decode, and understand input. Assessment of receptive skills is crucial because:

- It shows learners' comprehension abilities, not just production.
- It reflects how students can handle real-life tasks (reading news, listening to lectures).
- It informs teachers about areas of weakness in vocabulary, grammar, inference, or strategy use.

### **Assessing Reading Skills**

#### **Nature of Reading**

**Reading is more than recognizing words; it involves:**

- Decoding: understanding vocabulary and grammar.
- Comprehension: grasping meaning at sentence, paragraph, and text levels.
- Interpretation: inferring, predicting, and critically evaluating texts.

#### **Reading Sub-Skills Commonly Assessed**

- Skimming – reading for gist (e.g., identifying the main idea of a news article).
- Scanning – reading for specific information (e.g., train schedule).
- Intensive reading – close reading for detail and accuracy.
- Inferencing – reading between the lines to understand implied meaning.
- Critical reading – evaluating arguments and identifying bias.

#### **Types of Reading Assessment Tasks**

- Multiple-choice questions (MCQs): Test literal and inferential understanding.

- Short-answer questions: Encourage direct engagement with text.
- Matching tasks: Match headings with paragraphs or information with sections.
- Cloze tests: Fill-in-the-blank exercises for vocabulary and cohesion.
- Summary writing: Demonstrates ability to extract main ideas.

### **Example**

*Text:* “Global warming has led to rising sea levels.”

- Q1: What is one effect of global warming? (Detail)
- Q2: Is the author’s tone positive, neutral, or negative? (Inference)

### **Assessing Listening Skills**

#### **Nature of Listening**

Listening is an active process of:

- Perception: recognizing sounds and words.
- Parsing: grouping into meaningful phrases.
- Comprehension: interpreting the message in context.

#### **Listening Sub-Skills Commonly Assessed**

- Listening for gist – understanding the general idea.
- Listening for detail – focusing on specific information (dates, names, figures).
- Note-taking – extracting key points from lectures.
- Inferential listening – deducing speaker’s attitude, intention, or implied meaning.
- Interactive listening – understanding in conversations, turn-taking, clarification.

## **Types of Listening Assessment Tasks**

- True/False or MCQs: Test understanding of main ideas and details.
- Information transfer: Complete a chart, table, or diagram while listening.
- Gap-fill tasks: Fill missing words from a transcript (tests perception and meaning).
- Note-taking tasks: Based on academic lectures or talks.
- Response tasks: Learner reacts to listening input (e.g., answer comprehension questions, summarize orally).

### **Example**

*Audio:* “The train to Tashkent will depart at 6:45 from platform 3.”

- Q1: What time does the train leave? (Detail)
- Q2: Which platform is mentioned? (Scanning for detail)

## **Common Principles in Assessing Receptive Skills**

**Validity** – Test should measure comprehension, not just memory.

- Example: Instead of asking students to *repeat sentences*, ask them to *paraphrase meaning*.

**Authenticity** – Materials should reflect real-life use.

**Example:** News articles, podcasts, emails, announcements.

**Variety of Tasks** – Include different sub-skills.

- Example: A test should assess both gist and detail.

**Fairness** – Avoid cultural bias in texts and recordings.

- Example: References only familiar to one group can distort results.

**Practicality** – Tests should be feasible in time and resources.

## Challenges in Assessing Receptive Skills

- Invisible process: Teachers cannot directly observe comprehension. They infer it from tasks.
- Test-taking strategies: Students may guess answers without understanding.
- Background knowledge effect: Learners familiar with a topic perform better, independent of language ability.
- Listening conditions: Noise, audio quality, or speed can affect scores unfairly.

## Classroom Applications

- Reading: Use authentic materials (menus, blogs, emails). Design pre-reading questions, comprehension checks, and post-reading discussions.
- Listening: Use short audios/videos. Assess comprehension through note-taking, summaries, or group discussions.
- Encourage self-assessment: learners reflect on what they understood and where they struggled.

## References

- Alderson, J. C. (2000). *Assessing Reading*. Cambridge University Press.
- Field, J. (2019). *Listening in the Language Classroom*. Cambridge University Press.
- Brown, H. D., & Abeywickrama, P. (2018). *Language Assessment: Principles and Classroom Practices* (2nd ed.). Pearson.
- Fulcher, G. (2021). *Language Testing and Assessment*. Routledge.

## Lecture 6: Assessing Productive Skills: Speaking & Writing

Productive skills – **speaking** and **writing** – are central to language learning because they show how well learners can actively use language in real contexts. Assessing them is complex, since performance depends on not just knowledge of

grammar or vocabulary, but also fluency, creativity, coherence, and appropriateness.

## 2. Key Principles of Productive Skills Assessment

**Validity** – tasks must measure real communication ability.

**Reliability** – scoring should be consistent (rubrics help).

**Practicality** – tests should be feasible in time/resources.

**Authenticity** – tasks should resemble real-world language use.

**Positive Washback** – assessment should improve learning, not narrow it.

## 3. Assessing Speaking

### 3.1 Common Speaking Tasks

**Interviews** (structured/semi-structured Q&A).

**Role-plays** (e.g., ordering food, job interview).

**Discussions** (pair/group problem-solving).

**Presentations** (prepared or spontaneous).

**Storytelling / Picture description.**

### 3.2 Example Speaking Prompts

“Describe your favorite place and explain why it’s special.”

“Work with your partner: plan a holiday with a limited budget.”

“You are at a job interview. The interviewer will ask about your skills.”

### 3.3 Speaking Analytic Rubric (example, 1–5 scale each)

<b>Criteria</b>	<b>5 (Excellent)</b>	<b>3 (Satisfactory)</b>	<b>1 (Weak)</b>
<b>Fluency</b>	Smooth, natural flow	Some hesitation	Frequent pauses, broken speech
<b>Accuracy</b>	Few/no errors	Some errors, meaning clear	Many errors, meaning unclear
<b>Pronunciation</b>	Clear, natural intonation	Understandable, some issues	Difficult to understand
<b>Vocabulary</b>	Wide, precise use	Adequate, limited range	Very limited, repetitive
<b>Interaction</b>	Initiates & responds actively	Responds but limited initiative	Minimal or no interaction

### 3.4 Challenges in Speaking Assessment

Test anxiety may affect performance.

Different raters may score differently → need for **training & rubrics**.

Balancing **fluency vs. accuracy**.

## 4. Assessing Writing

### 4.1 Common Writing Tasks

**Guided writing:** letters, short messages.

**Functional writing:** emails, reports, applications.

**Creative writing:** essays, stories.

**Academic writing:** research papers, reviews.

### 4.2 Example Writing Tasks

“Write an email to your university professor requesting an extension.”

“Write a short essay: *Should smartphones be allowed in classrooms?*”

“Describe a memorable event in your life.”

#### 4.3 Writing Analytic Rubric (example, 1–5 scale each)

<b>Criteria</b>	<b>5 (Excellent)</b>	<b>3 (Satisfactory)</b>	<b>1 (Weak)</b>
<b>Content</b>	Relevant, fully developed	Adequate but limited	Irrelevant, underdeveloped
<b>Organization</b>	Clear, logical, cohesive	Some organization	Lacks structure
<b>Vocabulary</b>	Wide range, appropriate	Adequate range	Very limited, repetitive
<b>Grammar</b>	Accurate use of structures	Some errors	Frequent errors
<b>Mechanics</b>	Correct spelling, punctuation	Some errors	Many errors

#### 4.4 Challenges in Writing Assessment

Time pressure vs. natural drafting/revision.

Cultural differences in writing styles.

Ensuring originality (avoiding plagiarism)

#### 5. Holistic vs. Analytic Scoring

**Holistic** → one overall score (quick but less diagnostic).

**Analytic** → separate scores for fluency, accuracy, etc. (more reliable, diagnostic).

Major standardized exams (IELTS, TOEFL, CEFR-based exams) use analytic descriptors.

## 6. Classroom Assessment Techniques

**Portfolios:** collection of speaking/writing over time.

**Peer & Self-Assessment:** checklists, reflection journals.

**Project-based tasks:** debates, essays, group presentations.

**Observation checklists:** teacher tracks progress in participation.

### **Speaking (in class):**

Did the student contribute actively?

Was the student understandable?

Did they use new vocabulary?

### **Writing (in class):**

Did the writing answer the task question?

Was it organized into clear paragraphs?

Were key grammar/vocabulary items used correctly?

## **Key Books**

Brown, H. D., & Abeywickrama, P. (2018). *Language Assessment: Principles and Classroom Practices* (2nd ed.). Pearson.

Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment: An Advanced Resource Book*. Routledge.

Hughes, A. (2003). *Testing for Language Teachers* (2nd ed.). Cambridge University Press.

Weigle, S. C. (2002). *Assessing Writing*. Cambridge University Press.

Luoma, S. (2004). *Assessing Speaking*. Cambridge University Press.

### **Online Resources**

Council of Europe CEFR Descriptors

British Council – Assessing Speaking & Writing

IELTS Speaking & Writing Band Descriptors

## **Lecture 7. Item Analysis and Test Evaluation**

### **1. Introduction**

In language testing, simply designing and administering a test is not enough. We must ensure that the test **measures what it intends to measure** and provides

reliable, fair, and useful results. This process is supported by **item analysis** and **test evaluation**.

**Item analysis** → a systematic study of individual test items (questions) to determine their quality and effectiveness.

**Test evaluation** → a broader process of examining the overall test to judge its reliability, validity, and usefulness for learning.

Both processes ensure that assessments are **fair, accurate, and beneficial** for students and teachers.

## 2. Purposes of Item Analysis

**Identify strengths and weaknesses of test items** (are they too easy, too hard, or confusing?).

**Improve reliability and validity** of the test.

**Provide feedback to teachers** on student learning and test design.

**Guide future test development**, avoiding repeated mistakes.

## 3. Key Procedures in Item Analysis

### Item Difficulty (Facility Value, FV)

Measures how easy or difficult a test item is.

Formula:

$$FV = \frac{\text{Number of correct responses}}{\text{Total number of students}}$$
$$FV = \frac{\text{Total number of students}}{\text{Number of correct responses}}$$

**Interpretation:**

FV close to 1.0 → item too easy

FV close to 0.0 → item too difficult

Optimal range: **0.3 – 0.7**

Item Discrimination (Discrimination Index, D)

Measures how well an item differentiates between high- and low-performing students.

Formula:

$$D = \frac{U - L}{N}$$

where  $U$  = number correct in upper group,  $L$  = number correct in lower group,  $N$  = number in a group.

**Interpretation:**

$D \geq 0.40$  = excellent

$D = 0.20$ – $0.39$  = acceptable

$D < 0.20$  = poor, needs revision

Distractor Analysis

For multiple-choice tests, wrong options (distractors) should attract students who don't know the correct answer.

Good distractors: chosen by some weaker students.

Poor distractors: never chosen → useless.

4. Test Evaluation

Aspects of Test Evaluation

**Validity** – Does the test measure what it claims?

Content validity, construct validity, criterion-related validity.

**Reliability** – Are results consistent and stable?

Test–retest, inter-rater, internal consistency.

**Practicality** – Is the test feasible in terms of time, resources, and scoring?

**Authenticity** – Are tasks realistic and similar to real-life language use?

**Washback** – Does the test encourage positive learning outcomes?

Methods of Evaluation

**Quantitative:** statistical analysis (difficulty, discrimination, reliability coefficients).

**Qualitative:** teacher and student feedback, expert judgment.

Classroom Applications of Item Analysis

After a test, teachers can review **which items most students missed**.

Revise unclear questions or unfair distractors.

Use results to identify **learning gaps**: if many failed one grammar item, reteach that point.

Share findings with students to **increase transparency**.

Example: Suppose 30 students answered a multiple-choice question.

24 got it correct →  $FV = 24/30 = 0.8$  (too easy).

Upper group (top 10 students): 10 correct

Lower group (bottom 10 students): 6 correct

$D = (10-6)/10 = 0.4$  (good discrimination).

Distractor: one option never chosen → should be revised.

## Resources & References

### Books

Brown, H. D., & Abeywickrama, P. (2018). *Language Assessment: Principles and Classroom Practices* (2nd ed.). Pearson.

Hughes, A. (2003). *Testing for Language Teachers* (2nd ed.). Cambridge University Press.

Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment*. Routledge.

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.

### Online Resources

Council of Europe – CEFR Assessment

British Council – Language Testing

ETS Research – Item Analysis Methods

## Lecture 8. Large-Scale and Standardized Testing (IELTS, TOEFL, CEFR)

### 1. Introduction

In modern language education, large-scale standardized tests play a crucial role in measuring learners' proficiency across the world. Unlike classroom-based assessments, these tests are designed to provide **valid, reliable, and comparable results** across different contexts and populations.

Three of the most recognized systems are:

**IELTS (International English Language Testing System)**

**TOEFL (Test of English as a Foreign Language)**

**CEFR (Common European Framework of Reference for Languages)**

These frameworks not only assess learners but also set international standards for education, employment, and migration.

## 2. Characteristics of Large-Scale Standardized Tests

**Standardization** – same conditions, tasks, scoring criteria worldwide.

**Reliability** – consistent results regardless of examiner or location.

**Validity** – aligned with real-world language use.

**High Stakes** – used for university admissions, work permits, immigration.

**Benchmarking** – results comparable across countries and institutions.

## 3. IELTS (International English Language Testing System)

### Overview

Jointly managed by British Council, IDP: IELTS Australia, and Cambridge Assessment English.

Accepted by 12,000+ institutions worldwide.

Two formats: **Academic** and **General Training**

Test Structure

**Listening** – 4 sections, ~30 minutes.

**Reading** – 3 passages (Academic/General), 60 minutes.

**Writing** – 2 tasks, 60 minutes (Task 1: report/letter, Task 2: essay).

**Speaking** – face-to-face interview, 11–14 minutes.

Scoring

0–9 **band scale** (increments of 0.5).

Band descriptors focus on fluency, coherence, vocabulary, grammar, pronunciation, task achievement.

4. TOEFL (Test of English as a Foreign Language)

Overview

Administered by ETS (Educational Testing Service).

Widely accepted in North America and globally.

Delivered as **TOEFL iBT** (Internet-based Test).

Test Structure

**Reading** – 3–4 academic passages, 60–80 minutes.

**Listening** – conversations/lectures, 60–90 minutes.

**Speaking** – 6 tasks (independent + integrated), 20 minutes.

**Writing** – 2 tasks (integrated + independent essay), 50 minutes.

## Scoring

Each section scored **0–30**, total score **0–120**.

Reports emphasize both overall proficiency and sectional performance.

## CEFR (Common European Framework of Reference for Languages)

### Overview

Developed by the Council of Europe.

Provides a **framework** for describing language proficiency across languages, not a single exam.

Used in Europe and increasingly worldwide for curriculum, teaching, and testing.

### Proficiency Levels

**A1, A2** – Basic user

**B1, B2** – Independent user

**C1, C2** – Proficient user

### Application

Basis for national exams, curricula, and international certificates (e.g., Cambridge English, Goethe-Institut).

Influences descriptors in IELTS, TOEFL, and many other exams.

Comparison of IELTS, TOEFL, and CEFR

Feature	IELTS	TOEFL iBT	CEFR
Origin	UK, Australia	USA (ETS)	Europe (Council of Europe)
Format	Paper & Computer	Internet-based	Framework, not a test
Skills Tested	L, R, W, S	L, R, W, S	All skills (descriptive)
Scoring	0–9 bands	0–120 points	A1–C2 levels
Acceptance	UK, EU, Commonwealth, etc.	USA, Canada, worldwide	Europe/global benchmark

## 7. Challenges of Large-Scale Tests

**High stakes pressure** → test anxiety.

**Cultural bias** → some tasks favor certain backgrounds.

**Test-preparation industry** → risk of focusing on “test skills” instead of real communication.

**Accessibility issues** → cost, availability, technology barriers.

## 8. Classroom Implications

Teachers can use **IELTS/TOEFL descriptors** for classroom rubrics.

CEFR levels can guide **curriculum design** and **lesson planning**.

Practice tests help learners become familiar with exam demands.

Awareness of assessment criteria helps learners focus on **skills, not just scores**.

Resources & References: Hughes, A. (2003). *Testing for Language Teachers* (2nd ed.). Cambridge University Press.

## Lecture 9. Teaching Certifications and Professional Language Assessments

### (TESOL, TEFL, TKT)

#### Introduction

In addition to language proficiency tests such as IELTS and TOEFL, there are **professional certifications and assessments** designed for **teachers of English as a foreign or second language**. These certifications:

Provide recognition of **teaching competence**.

Help educators pursue **international teaching opportunities**.

Ensure **standardized training and assessment** for English language teachers worldwide.

The three most widely known are:

**TESOL (Teaching English to Speakers of Other Languages)**

**TEFL (Teaching English as a Foreign Language)**

**TKT (Teaching Knowledge Test)** by Cambridge English

#### 2. TESOL (Teaching English to Speakers of Other Languages)

An umbrella term covering both teaching English as a second language (ESL) and as a foreign language (EFL). Also refers to certification programs offered at universities and institutions.

#### Key Features

Often postgraduate programs (Diplomas, MAs).

Combines **theory and practice** in language teaching.

Focuses on **methodology, linguistics, assessment, and classroom practice**.

Widely recognized in North America, Asia, and the Middle East.

Career Value

Opens teaching opportunities in **schools, universities, and language institutes**.

Strong professional development pathway for long-term careers.

TEFL (Teaching English as a Foreign Language)

Overview

Focused on teaching English in countries where English is **not the native language** (e.g., Uzbekistan, China, Spain). Usually refers to **short-term certificate programs** (120–150 hours).

Key Features

Practical classroom-oriented training.

Can be online or in-person.

Covers teaching methods, lesson planning, and classroom management.

Accredited TEFL certificates are often required for teaching abroad.

Career Value

Gateway for young teachers to travel and work internationally.

Especially popular for **entry-level teaching positions**.

TKT (Teaching Knowledge Test)

Overview: Developed by **Cambridge English**. Modular assessment for teachers of English, focusing on **teaching knowledge**, not English proficiency.

Modules: **Module 1** – Language and background to language learning & teaching.

**Module 2** – Lesson planning and use of resources.

**Module 3** – Classroom management and teaching skills.

Additional modules: **TKT CLIL** (Content and Language Integrated Learning), **TKT Young Learners**.

Key Features

Paper-based test, flexible scheduling.

No teaching practice required.

Widely recognized for **professional development** and career advancement.

Comparison of TESOL, TEFL, and TKT

<b>Feature</b>	<b>TESOL</b>	<b>TEFL</b>	<b>TKT (Cambridge)</b>
<b>Focus</b>	Broad (ESL & EFL contexts)	EFL teaching abroad	Teaching knowledge (theory-based)
<b>Duration</b>	Months–years (Diploma, MA)	120–150 hrs (short-term)	Flexible modules (paper exam)
<b>Level</b>	Advanced, postgraduate	often Beginner intermediate	to For teachers at all experience levels
<b>Recognition</b>	Global, academic institutions	Schools, centers	International (Cambridge brand)
<b>Practical Training</b>	Includes teaching practice	Sometimes includes practicum	No teaching practice required

Benefits of Teaching Certifications

**Professional credibility** – internationally recognized qualification.

**Improved teaching skills** – methodology, assessment, classroom management.

**Global mobility** – opens doors to teaching positions abroad.

**Career progression** – foundation for advanced studies (e.g., DELTA, MA TESOL).

Challenges & Considerations

**Cost and accessibility** – some programs are expensive.

**Accreditation** – not all TEFL/TESOL providers are recognized.

**Mismatch between training and local contexts** – methods may need adaptation.

Resources & References

**Books:** Richards, J. C., & Farrell, T. S. C. (2005). *Professional Development for Language Teachers*. Cambridge University Press.

Harmer, J. (2015). *The Practice of English Language Teaching* (5th ed.). Pearson.

Brown, H. D. (2007). *Teaching by Principles: An Interactive Approach to Language Pedagogy*. Pearson.

## **Online Resources**

TESOL International Association

Cambridge English – TKT

TEFL.org – Accredited TEFL Courses

**Lecture 10.** Validity in Language Assessment

## Introduction

In language testing and assessment, **validity** is one of the most essential concepts. It answers the question: *Does the test measure what it is supposed to measure?* Without validity, even the most reliable test scores cannot be considered meaningful.

Today's lecture will explore:

1. The concept and types of validity
2. Historical perspectives on validity
3. Validity in modern language assessment frameworks
4. Practical ways to ensure validity in classroom and large-scale testing

### 1. Defining Validity

**Validity** refers to the degree to which evidence and theory support the interpretations of test scores for their intended purposes (Messick, 1989).

It is not a property of the test itself but of the inferences made from test scores.

A test can be reliable but still not valid. For example, if a grammar test only asks students to memorize rules, it may reliably produce results but not reflect actual communicative ability.

### 2. Types of Validity

#### a) Content Validity

Ensures the test covers the full range of the construct being measured.

Example: A reading comprehension test should include different text types, not just one genre.

#### b) Construct Validity

Concerns whether the test actually measures the theoretical construct it claims to measure.

Example: If we say a speaking test measures “oral proficiency,” tasks should involve interaction, fluency, and accuracy—not just pronunciation drills.

#### c) Criterion-Related Validity

Compares test performance with an external criterion.

Two forms:

**Concurrent validity:** correlating with an existing test (e.g., comparing a new vocabulary test with IELTS scores).

**Predictive validity:** showing that test results can predict future performance (e.g., university placement tests predicting academic success).

#### d) Face Validity

The extent to which the test appears valid to students, teachers, and stakeholders.

While not technically a scientific measure, high face validity promotes test acceptance and reduces anxiety.

### 3. Historical Perspectives

**Classical view:** Validity was treated as separate categories (content, criterion, construct).

**Unified view (Messick, 1989):** Validity is a single construct supported by multiple sources of evidence. He emphasized social consequences of testing as part of validity.

**Modern frameworks:** (Kane, 2006) propose an argument-based approach, where validation is seen as building a logical argument for score interpretation.

#### Ensuring Validity in Practice

##### a) Test Design Stage

Align test tasks with learning objectives and curriculum.

Use test specifications and blueprints to map skills assessed.

Pilot test items with a representative group.

#### b) Administration and Scoring

Provide clear instructions.

Use rubrics and trained raters for productive skills.

Avoid construct-irrelevant factors such as complicated instructions or cultural bias.

#### c) Post-Test Analysis

Conduct **item analysis** to check if questions function as intended.

Use statistical techniques (e.g., correlation, factor analysis) to support construct validity.

Collect feedback from teachers and students on test fairness and authenticity.

#### Validity in Large-Scale Assessments

Tests like **IELTS, TOEFL, and CEFR-based exams** invest heavily in validity research.

They use **trial versions, item response theory, differential item functioning (DIF) analysis,** and expert panels to ensure fairness and accuracy.

Validity arguments are documented in technical manuals and research reports.

st think of validity at every stage—design, administration, and interpretation.

Resources:

Messick, S. (1989). *Validity*. In R. Linn (Ed.), *Educational Measurement* (3rd ed.). Macmillan.

Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in Practice*. Oxford University Press.

Brown, H. D., & Abeywickrama, P. (2018). *Language Assessment: Principles and Classroom Practices* (2nd ed.). Pearson.

Kane, M. T. (2006). *Validation*. In R. Brennan (Ed.), *Educational Measurement* (4th ed.). Praeger.

## **Lecture 11. Alternative and Classroom-Based Assessment**

### Introduction

Traditional language assessment often relies on standardized tests and formal examinations. While these provide valuable information, they may not always reflect learners' real-world language ability or classroom performance. This is where **alternative and classroom-based assessment** comes in.

In today's lecture, we will cover:

The rationale for alternative assessment

Types of alternative and classroom-based assessment

Advantages and challenges

Practical strategies for teachers

### 1. Rationale for Alternative Assessment

Traditional testing emphasizes *product* (final answers), while alternative assessment focuses on the *process* of learning.

Encourages **authenticity** by aligning tasks with real-life language use.

Promotes **learner autonomy, reflection, and motivation**.

Provides richer evidence of learning for both teachers and students.

## 2. Types of Alternative and Classroom-Based Assessment. Portfolio Assessment

A systematic collection of student work over time (essays, projects, recordings).

Shows progress and achievement.

Encourages self-reflection and ownership of learning.

### b) Self-Assessment

Students evaluate their own performance using checklists, rubrics, or reflective journals.

Builds **metacognitive skills** and learner independence.

### c) Peer Assessment

Students assess each other's work through structured rubrics.

Promotes collaboration, critical thinking, and awareness of assessment criteria.

### d) Project-Based Assessment

Learners complete extended tasks such as presentations, debates, or research projects.

Assesses integrated skills (reading, writing, listening, speaking).

### e) Performance-Based Assessment

Observing students performing authentic tasks (role-plays, interviews, discussions).

Provides direct evidence of communicative competence.

#### f) Observation & Anecdotal Records

Teachers document learner behavior, participation, and progress during class activities.

Useful for formative purposes and feedback.

#### g) Learning Journals & Logs

Students keep diaries about their learning experiences.

Encourages reflection, goal setting, and language awareness.

### 3. Advantages of Alternative Assessment

- Promotes authenticity and real-world relevance.
- Supports formative assessment and continuous feedback.
- Encourages learner-centered classrooms.
- Provides a holistic picture of learner progress.

#### Suggested Readings & Resources

- Brown, H. D., & Abeywickrama, P. (2018). *Language Assessment: Principles and Classroom Practices* (2nd ed.). Pearson.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.
- Richards, J. C., & Lockhart, C. (2007). *Reflective Teaching in Second Language Classrooms*. Cambridge University Press.
- O'Malley, J. M., & Valdez Pierce, L. (1996). *Authentic Assessment for English Language Learners*. Addison-Wesley.

## **Lecture 12. Technology and Language Assessment**

## Introduction

Technology has transformed the landscape of language assessment, providing new opportunities for test design, administration, scoring, and feedback. From computer-based testing to artificial intelligence and online learning platforms, technology offers innovative ways to assess learners' language abilities more efficiently and authentically.

In this lecture, we will cover:

1. The role of technology in modern assessment
2. Types of technology-enhanced assessment
3. Benefits of using technology in assessment
4. Challenges and considerations
5. Future directions in technology and language assessment

### 1. The Role of Technology in Modern Assessment

- Facilitates large-scale standardized testing (e.g., TOEFL iBT, IELTS on Computer).
- Enables adaptive testing that adjusts difficulty to student responses.
- Provides instant scoring and feedback through automated systems.
- Supports authentic assessment tasks such as online discussions, collaborative projects, and simulations.
- Expands access to testing for remote and diverse learners.

### 2. Types of Technology-Enhanced Assessment

#### a) Computer-Based Testing (CBT)

Traditional paper-based tests delivered on computers. Examples include IELTS Computer-Based Testing and TOEFL iBT.

#### b) Computer-Adaptive Testing (CAT)

Uses algorithms to adjust the difficulty of items based on student responses. Provides a more accurate measure of proficiency in less time. Example: Duolingo English Test.

### c) Online Formative Assessment Tools

Platforms like Kahoot, Quizlet, and Google Forms allow teachers to conduct quizzes, polls, and surveys in real time, providing immediate feedback.

### d) Automated Scoring Systems

Artificial Intelligence and Natural Language Processing (NLP) are used to score essays (e.g., ETS e-rater) and speaking tasks (e.g., Pearson Versant Test).

### e) Learning Analytics

Collecting and analyzing data from LMS platforms (Moodle, Canvas, Coursera) to track learner progress and predict performance.

### f) Gamified Assessment

Incorporates game elements (points, badges, leaderboards) into assessment to enhance motivation and engagement.

### g) Mobile-Assisted Language Assessment (MALA)

Mobile apps and smartphones enable flexible, on-the-go assessment of skills, including vocabulary quizzes, pronunciation practice, and chatbots for speaking.

## 3. Benefits of Using Technology in Assessment

- Efficiency: Faster administration, scoring, and reporting.
- Accessibility: Provides opportunities for learners in remote or underserved areas.
- Authenticity: Simulates real-world language use (e.g., online chats, multimedia tasks).
- Personalization: Adaptive tests tailor difficulty to the learner's level.
- Immediate feedback: Supports learning by providing results instantly.

## 4. Challenges and Considerations

- Digital divide: Not all learners have equal access to technology.
- Security and fairness: Risks of cheating and technical malfunctions.
- Reliability of automated scoring: AI-based tools may misinterpret context or creativity.

- Teacher training: Educators need digital literacy skills to use technology effectively.

- Washback effect: Overreliance on technology may shift focus from pedagogy to tools.

## 5. Future Directions in Technology and Language Assessment

- Increasing use of Artificial Intelligence and machine learning for more sophisticated assessments.

- Virtual Reality (VR) and Augmented Reality (AR) to simulate immersive language environments.

- Blockchain technology for secure certification and credentialing.

- Integration of multimodal assessment (combining text, speech, video, and gestures).

- Expansion of personalized learning pathways supported by real-time analytics.

## Suggested Readings & Resources

- Chapelle, C. A., & Voss, E. (2017). 21st Century Language Assessment. In E. Shohamy, I. Or, & S. May (Eds.), *Language Testing and Assessment*. Springer.

- Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency: The Interface Between Learning and Assessment*. Continuum.

- Luoma, S. (2004). *Assessing Speaking*. Cambridge University Press.

- Weigle, S. C. (2002). *Assessing Writing*. Cambridge University Press.

- Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment: An Advanced Resource Book*. Routledge.

## **Lecture 13. Ethical Issues and Fairness in Language Testing**

### Introduction

Language testing is not just a technical process but also an ethical responsibility.

Tests have a significant impact on learners' educational opportunities, career

prospects, and even migration status. Therefore, fairness, justice, and ethics are central concerns in language assessment.

### 1. The Importance of Ethics in Language Testing

- Language tests influence high-stakes decisions such as university admission, employment, and immigration.
- Unethical or unfair testing can create barriers and reinforce inequality.
- Ethical testing ensures respect for learners' rights, dignity, and diversity.
- Ethics guide test design, administration, interpretation, and consequences.

### 2. Key Ethical Issues

- **Test misuse**: Using test results for purposes they were not designed for.
- **Test bias**: Items that favor certain groups over others due to cultural, gender, or linguistic background.
- **Access and equity**: Ensuring that all learners, regardless of socio-economic status, have fair opportunities to take the test.
- **Informed consent**: Test takers should understand the purpose and consequences of the test.
- **Privacy and confidentiality**: Protecting test takers' personal data and results.

### 3. Fairness and Bias in Testing

- Fairness means providing equal opportunity for all learners to demonstrate their ability.
- Sources of bias:
  - Linguistic bias (unfamiliar vocabulary, idioms)
  - Cultural bias (references only familiar to certain groups)
  - Gender bias (tasks stereotyped for male/female roles)
- Fairness also requires reasonable accommodations for learners with disabilities (extra time, accessible formats).

#### 4. Principles of Responsible Testing

- **Validity and reliability**: Ethical tests must be scientifically sound.
- **Transparency**: Clear information on test purpose, scoring, and interpretation.
- **Accountability**: Test developers and administrators should be responsible for consequences.
- **Washback**: Ethical testing promotes positive washback (better teaching and learning practices).
- **Inclusivity**: Considering diverse learner backgrounds when designing assessments.

#### 5. Practical Strategies for Teachers and Test Designers

- Avoid cultural or gender-specific references that disadvantage learners.
- Provide practice materials and clear test instructions.
- Train raters to reduce subjectivity in scoring productive skills.
- Regularly review test items for bias and fairness.
- Consider the impact of test results on learners' motivation and opportunities.
- Use multiple forms of assessment (not just one test) for high-stakes decisions.

#### Suggested Readings & Resources

- Shohamy, E. (2001). *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. Pearson Education.
- Kunnan, A. J. (2018). *Fairness and Validation in Language Assessment*. Routledge.
- McNamara, T., & Ryan, K. (2011). Fairness Versus Justice in Language Testing: The Place of Ethics. *Language Testing*, 28(4), 443–459.
- Hamp-Lyons, L. (2000). Social, Professional, and Individual Responsibility in Language Testing. *System*, 28(4), 579–591.

### **Lecture 14. Research in Language Testing**

## Introduction

Language testing is both a professional practice and a field of academic research. Research in language testing provides the foundation for developing valid, reliable, and fair assessments. It also informs policy, classroom practice, and innovations in testing. In this lecture, we will examine the role of research in shaping language assessment theory and practice.

We will cover:

1. The scope of research in language testing
2. Key research areas
3. Research methods
4. Impact of research on testing practice
5. Future directions in language testing research

### 1. The Scope of Research in Language Testing

- Encompasses theoretical, empirical, and applied investigations.
- Addresses issues of test design, administration, scoring, interpretation, and impact.
- Serves multiple stakeholders: test developers, teachers, policymakers, and learners.
- Integrates insights from applied linguistics, psychometrics, educational measurement, and sociology.

### 2. Key Research Areas

- **Validity and reliability studies**: Ensuring that tests measure what they claim to measure.
- **Construct definition**: Investigating what it means to know a language and how to operationalize it.
- **Test fairness and ethics**: Examining bias, accessibility, and social consequences.
- **Performance assessment**: Research on speaking and writing tasks, rating scales, and rater training.
- **Technology-enhanced assessment**: Exploring computer-adaptive testing, AI

scoring, and digital literacy.

- **Washback studies**: Investigating how tests influence teaching and learning.
- **Impact studies**: Assessing broader social, educational, and policy consequences of testing.

### 3. Research Methods

- **Quantitative methods**: Statistical analyses (e.g., factor analysis, Rasch modeling, correlation studies).
- **Qualitative methods**: Interviews, case studies, classroom observations.
- **Mixed-methods research**: Combining quantitative and qualitative approaches for comprehensive findings.
- **Corpus-based studies**: Using language corpora to inform test design and task authenticity.
- **Validation studies**: Building an argument for score interpretation and use.

### 4. Impact of Research on Testing Practice

- Research informs the development of international standardized tests (IELTS, TOEFL, PTE).
- Classroom-based assessment practices are improved through action research.
- Findings influence language education policy and curriculum design.
- Research contributes to innovation in test design, including task-based and scenario-based assessments.

### 5. Future Directions in Language Testing Research

- Expanding research on **AI and automated scoring** to ensure fairness and transparency.
- Greater focus on **multimodal assessment** (text, speech, visuals, interaction).
- Addressing challenges of **plurilingual and intercultural competence assessment**.
- Exploring the role of **assessment in online and hybrid learning contexts**.
- Strengthening research on **social justice, equity, and inclusivity** in testing.

## Conclusion

Research in language testing is essential for advancing both theory and practice. It ensures that tests remain valid, fair, and relevant to real-world needs. By engaging with research, teachers and test developers contribute to the ongoing improvement of language assessment in education and society.

## Suggested Readings & Resources

- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language Assessment in Practice*. Oxford University Press.
- Chalhoub-Deville, M. (2008). The Intersection of Assessment and Instructed SLA: A Framework. *Language Teaching*, 41(3), 235–264.
- Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment: An Advanced Resource Book*. Routledge.
- Kunnan, A. J. (2014). *The Companion to Language Assessment*. Wiley.

## **Lecture 15. Course Review and Reflection**

## Introduction

As we reach the final lecture in this course on Language Assessment, it is important to pause, reflect, and review what we have learned. Course review and reflection allow students to consolidate their knowledge, evaluate their own learning progress, and prepare for applying these insights in their future teaching or research practices.

In this lecture, we will cover:

1. Summary of key topics covered in the course
2. Reflection on personal learning
3. Practical applications of course knowledge
4. Challenges and future directions
5. Final thoughts and closing remarks

### 1. Summary of Key Topics Covered in the Course

- Theoretical foundations of language assessment (validity, reliability, fairness).
- Types of language assessment (formative, summative, diagnostic, alternative).
- Assessing receptive and productive skills (listening, reading, speaking, writing).
- Large-scale standardized testing (IELTS, TOEFL, CEFR).
- Classroom-based and alternative assessments (portfolios, projects, peer/self-assessment).
- The role of technology in language assessment (computer-adaptive tests, AI scoring).
- Ethical issues and fairness in testing.
- Research and future trends in language assessment.

### 2. Reflection on Personal Learning

- Which concepts were the most useful or eye-opening?
- How has your understanding of assessment evolved throughout the course?
- What skills and knowledge do you feel more confident about now?
- What areas would you like to explore further?

### 3. Practical Applications of Course Knowledge

- Designing classroom assessments that are valid, reliable, and fair.
- Applying rubrics, rating scales, and feedback strategies effectively.
- Using technology tools to enhance assessment.
- Conducting action research to improve assessment practices.
- Balancing formative and summative assessment in teaching contexts.

### 4. Challenges and Future Directions

- Recognizing ongoing challenges: test bias, digital divide, ethical dilemmas.
- Anticipating future developments: AI, multimodal assessment, and global standards.
- Commitment to continuous professional development in assessment literacy.

### 5. Final Thoughts and Closing Remarks

Language assessment is not only about measuring performance but also about supporting learning. As future teachers, researchers, or practitioners, your responsibility is to use assessment ethically, creatively, and reflectively. Remember that good assessment practices foster growth, confidence, and opportunity for learners.

This course has aimed to give you both the theoretical foundation and practical strategies needed to engage with language testing responsibly. Moving forward, keep questioning, reflecting, and striving for fairness and innovation in your assessment practices.

#### Reflection Activity

Take a few minutes to write down:

- Three key insights you gained from this course.
- Two areas you still want to develop further.

**FAN O'QUV  
DASTURI**

<b>Fan/modul kodi:</b> ITO‘BXS304		<b>O‘quv yili:</b> 2025-2026		<b>Semestr: 5</b>		<b>ECTS – Kreditlar: 4</b>	
<b>Fan/modul turi:</b> Majburiy		<b>Ta’lim tili: Ingliz</b>				<b>Haftadagi dars soatlari:</b> 4	
	Fanning nomi	Semestr	ECTS – Kreditlar	Auditoriya mashg‘ulotlari (soat)		Mustaqil ta’lim (soat)	Jami yuklama (soat)
				Ma’ruza	Seminar		
	Ingliz tili o‘qitish va baholashning xalqaro standartlari	5/2025	4	30	30	60	120
		Jami	4	30	30	60	120

### I. FANNING MAZMUNI

Ingliz tili o‘qitish va baholashning xalqaro standartlari kursi ma’ruza va seminar mashg‘ulotlaridan iborat bo‘lib, ushbu fanning maqsadi talabalarni chet tilini o‘qitish va baholash, shuningdek hozirgi davrda ta’lim oluvchilarda chet tili o‘rganish va o‘rgatish muhimligi tushunchasi; chet tilini bilish darajasini aniqlab beradigan CEFR, IELTS, TOEFL va boshqa xalqaro baholash tizimi haqida ma’lumot berish, o‘z ustida mustaqil ishlash ya’ni self-study approach va ijtimoiy moslashuv vositasi sifatida foydalanish ehtiyojini shakllantirish va rivojlantirishdan iborat. Ushbu fanda bo‘lajak ingliz tili o‘qituvchilari nimalarga ahamiyat berish, xalqaro baholash tizimlari, til o‘rganuvchilarning bilimlarini baholash va uni amalga oshirish jarayoni haqida bilimlarni egallaydilar.

Fanning vazifasi umum e’tirof etilgan meyorlarga ko‘ra o‘rganilayotgan ingliz tilini C1 darajada egallashlari uchun zaruriy bilimlarni integrallashgan tarzda o‘rgatish va muloqot malakalarini rivojlantirishdan iborat.

#### **Ta’lim natijalari (TN):**

#### **Fanni o‘zlashtirish natijasida talaba:**

**TN1-** Ingliz tili o‘qitish va baholashning xalqaro standartlari tushunchasi tasavvuriga ega bo‘ladi;

**TN2** - Baholash turlarini (formativ, summativ, diagnostik va boshqalar) farqlay olish ko'nikmasiga ega bo'ladi;

**TN3** – Kasbiy faoliyatida xalqaro baholash dasturlari (CEFR, TOEFL va IELTS) talablari asosida ta'lim natijalarining sifatini baholay oladi. Ta'lim natijalarini baholash uchun turli usul va vositalaridan foydalana oladi.

## II. FAN TARKIBIGA QUYIDAGI MAVZULAR KIRADI:

### 5-semestr uchun

	<b>Mashg'ulot shakli: Ma'ruza (M)</b>	<b>Soat</b>
<b>M 1</b>	Introduction to International standards for English language teaching and assessment	2
<b>M 2</b>	<i>*Theoretical Foundations of Language Assessment</i>	2
<b>M 3</b>	Types of Language Assessment	2
<b>M 4</b>	Test Design and Development	2
<b>M 5</b>	Assessing Receptive Skills: Reading & Listening	2
<b>M 6</b>	Assessing Productive Skills: Speaking & Writing	2
<b>M 7</b>	Item Analysis and Test Evaluation	2
<b>M 8</b>	<i>*Large-Scale and Standardized Testing (IELTS, TOEFL, CEFR)</i>	2
<b>M 9</b>	Teaching Certifications and Professional Language Assessments (TESOL, TEFL, TKT)	2
<b>M 10</b>	Validity in Language Assessment	2
<b>M 11</b>	<i>*Alternative and Classroom-Based Assessment</i>	2
<b>M 12</b>	Technology and Language Assessment	2
<b>M 13</b>	<i>*Ethical Issues and Fairness in Language Testing</i>	2
<b>M 14</b>	Research in Language Testing	2
<b>M 15</b>	Course Review and Reflection	2
	<b>Jami:</b>	<b>30</b>

**\*[https://catalog.apps.asu.edu/catalog/courses/courselist?subject=LIN&term=2244&utm\\_source=](https://catalog.apps.asu.edu/catalog/courses/courselist?subject=LIN&term=2244&utm_source=)**

**Ma'ruza mashg'ulotlari** multimedia qurilmalari bilan jihozlangan auditoriyada akademik guruhlar uchun o'tiladi.

**5-semestr uchun**

	<b>Mashg'ulot shakli: Seminar (S)</b>	<b>Soat</b>
<b>S1</b>	Introduction to International standards for English language teaching and assessment	2
<b>S2</b>	*Theoretical Foundations of Language Assessment	2
<b>S3</b>	Types of Language Assessment	2
<b>S4</b>	Test Design and Development	2
<b>S5</b>	Assessing Receptive Skills: Reading & Listening	2
<b>S6</b>	Assessing Productive Skills: Speaking & Writing	2
<b>S7</b>	Item Analysis and Test Evaluation	2
<b>S8</b>	*Large-Scale and Standardized Testing (IELTS, TOEFL, CEFR)	2
<b>S9</b>	Teaching Certifications and Professional Language Assessments (TESOL, TEFL, TKT)	2
<b>S10</b>	Validity in Language Assessment	2
<b>S11</b>	*Alternative and Classroom-Based Assessment	2
<b>S12</b>	Technology and Language Assessment	2
<b>S13</b>	*Ethical Issues and Fairness in Language Testing	2
<b>S14</b>	Research in Language Testing	2
<b>S15</b>	Course Review and Reflection	2
	<b>Jami</b>	<b>30</b>

\*[https://catalog.apps.asu.edu/catalog/courses/courselist?subject=LIN&term=2244&utm\\_source=](https://catalog.apps.asu.edu/catalog/courses/courselist?subject=LIN&term=2244&utm_source=)

**Seminar mashg'ulotlar** multimedia qurilmalari bilan jihozlangan auditoriyada har bir akademik guruhga alohida o'tiladi. Mashg'ulotlar interfaol usullari yordamida o'tiladi. Ko'rgazmali qurollar va multimediali vositalar qo'llash orqali o'tiladi.

### III. MUSTAQIL TA'LIM VA MUSTAQIL ISHLAR

Mustaqil ta'lim uchun tavsiya etiladigan mavzular:

	<b>Mashg'ulotlar shakli: Mustaqil ta'lim (MT)</b>	Soat
<b>MT1</b>	The Evolution of Language Testing – A historical perspective on language assessment and its development over time.	4
<b>MT2</b>	Validity and Reliability in Language Assessment – Examining key principles and their importance in test design.	4
<b>MT3</b>	The Washback Effect: How Testing Influences Teaching and Learning – Investigating positive and negative impacts on curriculum design.	4
<b>MT4</b>	Assessing Speaking: Challenges and Best Practices – Evaluating different methods for scoring oral proficiency.	4
<b>MT5</b>	Diagnostic vs. Proficiency vs. Achievement Tests: Key Differences – Understanding the purposes of different assessment types.	4
<b>MT6</b>	Language Testing in Multilingual and Multicultural Contexts – Challenges of assessing English proficiency in diverse settings.	4
<b>MT7</b>	Developing Rubrics for Writing and Speaking Assessments – Creating and analyzing effective scoring rubrics.	4
<b>MT8</b>	Ethical Issues in Standardized Language Testing – Addressing test fairness, bias, and accessibility concerns.	4
<b>MT9</b>	Alternative Assessments in Language Testing – Exploring portfolio-based, self-assessment, and peer-assessment strategies.	4
<b>MT10</b>	Comparing High-Stakes English Proficiency Tests (IELTS vs. TOEFL vs. PTE vs. CAE) – Analyzing test structures, scoring, and real-world implications.	4

<b>MT11</b>	CEFR vs. ACTFL: A Comparative Analysis of Language Proficiency Frameworks – Understanding the differences in international standards.	4
<b>MT12</b>	The Role of AI and Automated Scoring in Language Assessment – A critical review of AI-powered assessment tools (e.g., Duolingo English Test, ETS SpeechRater)	4
<b>MT13</b>	Language Assessment for Young Learners vs. Adults – Adapting-testing methods to different age groups.	4
<b>MT14</b>	Technology-Enhanced Language Assessment – Exploring online testing platforms, gamification, and remote proctoring.	4
<b>MT15</b>	The Future of English Language Assessment – Predicting trends and innovations in language testing.	4
	<b>Jami:</b>	<b>60</b>

Mustaqil ta’lim mavzulari talabalar tomonidan o‘zlashtiriladi. Mavzular yuzasidan talabalar mustaqil ish, taqdimotlar tayyorlash va himoya qilishi tavsiya etiladi.

*Amaliy mashg‘ulotlarining barcha mavzularini to‘la o‘zlashtirgan hamda mustaqil ishni bajargan talabalarga yakuniy nazoratda ishtirok etishga ruxsat etiladi. Talaba semestr oxirida yakuniy nazorat topshiradi.*

#### **IV. TA’LIM TEXNOLOGIYALARI VA METODLARI:**

- Kahoot o‘yinga asoslangan ta’lim platformasi;
- interfaol keys-stadilar;
- mantiqiy fiklash, tezkor savol-javoblar;
- guruhlarda ishlash;
- taqdimotlar qilish;
- individual loyihalar;

- jarayonga asoslangan metod;
- jamoa bo'lib ishlash va himoya qilish uchun loyihalar.

## VI. KREDITLARNI OLISH UCHUN TALABLAR

Oraliq nazorat shakllarida berilgan vazifa va topshiriqlarni bajarish, yakuniy nazorat bo'yicha yozma ishni topshirish.

Talabalar bilimni baholash 100 ballik tizimda amalga oshiriladi. Oraliq nazorat semestrlarda 2 tadan o'tkaziladi. Fan bo'yicha yakuniy nazorat test shaklida o'tkaziladi

### FANDAN BAHOLASH MEZONLARI

Talabalar bilimni baholashda nazorat turlari bo'yicha baholar umumiy bahoga foiz hisobida quyidagicha taqsimlanadi:

**Davomat – 10 foiz;**

**Taqdimot/loyiha – 10 foiz;**

**Mustaqil ta'lim – 10 foiz;**

**Oraliq nazorat(lar)ga – 20 foiz;**

**Yakuniy nazorat – 50 foiz.**

Yuqoridagi baholash tizimida fanning xususiyatidan kelib chiqqan holda, kafedra tomonidan ma'ruza, amaliy, seminar hamda laboratoriya mashg'ulotlari uchun belgilanadi.

Davomat, taqdimot, mustaqil ta'limga ajratilgan foizlar **(30 foiz) joriy nazorat**, ma'ruza mashg'ulotlaridagi **oraliq nazorati** uchun **(20 foiz)** va **yakuniy nazorat** uchun **(50 foiz)** olinadi.

Ma'ruza dars mashg'ulotlari mavjud bo'lmagan fanlardan umumiy baholarni foizlarda taqsimlanganda amaliy mashg'ulotlari uchun **50 foiz** va yakuniy nazoratlar uchun 50 foiz miqdorida olinadi.

## **VI. Asosiy va qo‘shimcha o‘quv adabiyotlar hamda axborot manbalari**

### **Asosiy adabiyotlar**

1. Brown, H. D., & Abeywickrama, P. (2018). *Language assessment: Principles and classroom practices* (2nd ed.). Pearson Education.
2. Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR)*. Cambridge University Press.

### **Qo‘shimcha adabiyotlar**

1. Cohen, A. D. (2014). *Assessing language skills in the classroom*. Heinle & Heinle.
2. Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Routledge.
3. Ginsburg, A. B., & Brown, A. J. (Eds.). (2017). *High-stakes testing in education: Value, fairness, and consequences*. Routledge.
4. O’Sullivan, B., & Weir, C. (2011). *Language testing: Theories and practices*. Palgrave Macmillan.
5. Rea-Dickins, P., & Germaine, K. (2012). *Cambridge English teacher: Testing, assessment, and evaluation*. Cambridge University Press.
6. Educational Testing Service (ETS). (Latest edition). *The official guide to the TOEFL test*. McGraw-Hi

### **Axborot manbalari**

1. [www.teachingenglish.org.uk](http://www.teachingenglish.org.uk)
2. [www.onestopenglish.com](http://www.onestopenglish.com)
3. [www.businessenglishonline.net](http://www.businessenglishonline.net)
4. [www.elgazette.com](http://www.elgazette.com)
5. [www.tesol.org](http://www.tesol.org)
6. [www.tefl.com](http://www.tefl.com)

**Mazkur fan o'quv dasturi Namangan davlat chet tillar instituti tomonidan ishlab chiqilgan va tasdiqlangan.**

- Ingliz tili o'qitish metodikasi kafedrasining 2025-yil 27-avgustdagi 1-sonli majlisida muhokama qilingan va tasdiqqa tavsiya etilgan.

- Til va tarjima fakulteti kengashining 2025-yil 28-avgustdagi 1-sonli majlisida ma'qullangan va tasdiqqa tavsiya etilgan.

- NamDCHTI o'quv-uslubiy kengashining 2025-yil 29-avgustdagi 1-sonli majlisida muhokama qilingan va tasdiqlangan.

**Fan/modul uchun mas'ul:**

M.Abdullayeva - Namangan davlat chet tillari instituti Ingliz tili o'qitish metodikasi kafedrasida katta o'qituvchisi

**Taqrizchilar:**

A.Amanov – Namangan davlat chet tillari instituti Ingliz tili o'qitish metodikasi kafedrasida dotsenti

M.Yo'ldasheva – Namangan davlat chet tillari instituti Ingliz tili o'qitish metodikasi kafedrasida o'qituvchisi

Izoh: Ushbu fan o'quv dasturi Arizona State University (USA) ta'lim dasturlari va Ibrat farzandlari Youtube darslari asosida takomillashtirildi.



# **GLOSSARI**

## **Glossary (Alphabetical Order)**

- Alternative Assessment – Non-traditional forms (portfolios, peer/self-evaluation).
- Computer-Adaptive Testing (CAT) – A system that adjusts difficulty based on responses.
- Construct Validity – How well a test represents the theoretical concept of ability.
- Content Validity – The degree to which a test covers the syllabus or construct.
- Criterion-related Validity – Correlation of test scores with external measures.
- Diagnostic Assessment – Identifies strengths/weaknesses at the start of learning.
- Ethical Testing – Assessment that is valid, fair, transparent, and learner-centered.
- Face Validity – The degree to which a test appears valid to learners and teachers.
- Fairness – Equal opportunity for all learners regardless of background.
- Formative Assessment – Ongoing, low-stakes assessment during learning.
- High-Stakes Test – Exam with significant consequences (e.g., IELTS, TOEFL).
- Item Analysis – Statistical study of test items to check difficulty and discrimination.
- Low-Stakes Test – Informal quiz or activity with minimal consequences.
- Negative Washback – When testing narrows learning to test-preparation only.
- Performance-based Assessment – Real-world tasks (presentations, role-plays).
- Positive Washback – When testing encourages meaningful learning.
- Reliability – Consistency of test results across time, tasks, and raters.
- Summative Assessment – Final evaluation at the end of instruction (high-stakes).
- Validity – The extent to which a test measures what it intends to measure.
- Washback – The influence of tests on teaching and learning (positive or negative).

**ILOVALAR**

## TESTS:

1. Which principle ensures consistency of test results?

- a) Validity
- b) Reliability
- c) Practicality
- d) Washback

2. Which principle refers to the usefulness of tests in real-life contexts?

- a) Practicality
- b) Reliability
- c) Validity
- d) Transparency

3. Which concept describes the influence of tests on teaching and learning?

- a) Washback
- b) Authenticity
- c) Reliability
- d) Face validity

4. Which type of validity checks whether test tasks represent the content domain?

- a) Construct validity
- b) Content validity
- c) Face validity
- d) Criterion validity

5. Which validity type measures correlation with external outcomes?

- a) Predictive validity
- b) Construct validity
- c) Content validity
- d) Face validity

6. Which statement is true?

- a) A reliable test is always valid
- b) A valid test must also be reliable

- c) Validity and reliability are unrelated
- d) Reliability guarantees fairness

7. Which method checks consistency of rater scoring?

- a) Parallel forms reliability
- b) Inter-rater reliability
- c) Split-half method
- d) Predictive validity

8. Which assessment is done during the learning process?

- a) Summative
- b) Formative
- c) Diagnostic
- d) Proficiency

9. Which assessment identifies strengths and weaknesses before instruction?

- a) Diagnostic
- b) Summative
- c) Formative
- d) Achievement

10. Which assessment is high-stakes and measures end results?

- a) Formative
- b) Summative
- c) Diagnostic
- d) Alternative

11. Which is a receptive skill?

- a) Speaking
- b) Reading
- c) Writing
- d) Presenting

12. A role-play task best assesses:

- a) Grammar

- b) Speaking
- c) Reading
- d) Listening

13. Cloze tests are often used for assessing:

- a) Vocabulary
- b) Writing
- c) Listening
- d) Pronunciation

14. An item with facility value (FV) = 0.20 is:

- a) Too easy
- b) Too difficult
- c) Optimal
- d) Reliable

15. Which index shows how well an item distinguishes between high and low achievers?

- a) Reliability
- b) Discrimination index
- c) Validity
- d) Practicality

16. Which step comes first in test design?

- a) Item analysis
- b) Defining purpose
- c) Piloting
- d) Scoring rubric

17. Which is an example of alternative assessment?

- a) Portfolios
- b) Multiple-choice test
- c) Final exam
- d) Grammar quiz

18. Peer assessment helps learners:

- a) Compete with others
- b) Develop critical awareness
- c) Avoid responsibility
- d) Focus only on scores

19. Which is NOT an alternative assessment tool?

- a) Self-reflection journal
- b) Speaking portfolio
- c) Computerized grammar quiz
- d) Peer feedback

20. Which test is computer-adaptive?

- a) IELTS
- b) Duolingo English Test
- c) TOEFL PBT
- d) Cambridge PET

21. AI scoring systems are mainly used for:

- a) Speaking and writing
- b) Reading
- c) Listening
- d) Vocabulary

22. Gamified assessment includes features such as:

- a) Random items
- b) Badges and leaderboards
- c) Summative reports
- d) Standardized rubrics

23. IELTS and TOEFL are examples of:

- a) Diagnostic tests
- b) Large-scale standardized tests
- c) Classroom tests
- d) Portfolios

24. CEFR describes proficiency in:

- a) 4 levels
- b) 5 levels
- c) 6 levels
- d) 7 levels

25. Which is the highest CEFR level?

- a) B2
- b) C1
- c) C2
- d) A2

26. Which is an ethical principle in testing?

- a) Transparency
- b) Competition
- c) Confidentiality breach
- d) Bias

27. Which of the following creates unfairness in testing?

- a) Clear rubrics
- b) Cultural bias
- c) Reliability
- d) Transparency

28. Fairness means:

- a) Giving the same tasks to everyone
- b) Ensuring equal opportunity
- c) Ignoring learner diversity
- d) Making tests easy

29. Washback is positive when:

- a) Teachers only focus on test practice
- b) Testing improves teaching
- c) Learners memorize answers

- d) Assessment is unfair

30. Which research method uses interviews and observations?

- a) Quantitative
- b) Qualitative
- c) Experimental
- d) Statistical

31. Which is a future direction in language testing research?

- a) Avoiding technology
- b) AI and multimodal assessment
- c) Eliminating CEFR
- d) Paper-only tests

32. Which assessment is most learner-centered?

- a) Summative
- b) Alternative assessment
- c) Diagnostic
- d) Proficiency

33. Which CEFR level is 'Can understand main points of clear standard speech'?

- a) A1
- b) A2
- c) B1
- d) C1

34. Which reliability type checks consistency across time?

- a) Test-retest reliability
- b) Inter-rater reliability
- c) Internal consistency
- d) Face validity

35. The TOEFL iBT focuses primarily on:

- a) General conversation
- b) Academic English

- c) Grammar drills
- d) Literature

36. Face validity is about:

- a) How the test appears to stakeholders
- b) Reliability of test scores
- c) Difficulty level
- d) Test length

37. A high-stakes test example is:

- a) Weekly quiz
- b) IELTS
- c) Peer feedback
- d) Homework

38. Which principle balances cost and effort?

- a) Practicality
- b) Validity
- c) Reliability
- d) Transparency

39. Which is a productive skill?

- a) Writing
- b) Reading
- c) Listening
- d) Vocabulary

40. Who emphasized unified validity theory?

- a) Messick
- b) Weir
- c) Hymes
- d) Bachman

41. Rater training is essential for:

- a) Multiple-choice tests

- b) Speaking/writing tests
- c) Listening tests
- d) Vocabulary tests

42. Which is a low-stakes test?

- a) National exam
- b) Classroom quiz
- c) IELTS
- d) TOEFL

43. Which test uses bands 1–9?

- a) TOEFL
- b) IELTS
- c) PTE
- d) Cambridge KET

44. Which is NOT part of CEFR?

- a) A1
- b) B2
- c) C2
- d) Level D

45. Adaptive testing advantage:

- a) Inefficiency
- b) Efficiency and precision
- c) Cultural bias
- d) Costly only

46. Which assessment promotes reflection?

- a) Self-assessment
- b) Summative
- c) Diagnostic
- d) Achievement

47. Language portfolios mainly show:

- a) Progress over time
- b) Test scores only
- c) Grammar drills
- d) Cultural bias

48. An item that all students answer correctly is:

- a) Too easy
- b) Too difficult
- c) Reliable
- d) Valid

49. Which test is UK-focused?

- a) TOEFL
- b) IELTS
- c) Duolingo
- d) PTE

50. A summative assessment occurs:

- a) During instruction
- b) At the end of instruction
- c) Before learning
- d) As self-reflection

### **Short Answer Questions**

1. Explain the difference between formative and summative assessment with examples.
2. What is construct-irrelevant variance in testing, and why is it problematic?
3. List three challenges of large-scale standardized tests such as IELTS/TOEFL.
4. How does technology (e.g., AI scoring) improve language testing? Give one benefit and one risk.
5. Why is fairness considered an ethical responsibility in assessment?

## C. Practical Task

Design a peer assessment rubric for a 3-minute speaking presentation at CEFR B1 level. Include at least three criteria (e.g., fluency, accuracy, vocabulary).

### Case Studies

#### Case Study 1: Validity in Practice

A university in Uzbekistan develops a new English test. It only includes grammar multiple-choice questions. Students with good communicative skills but weak grammar scores fail the test.

- Task: Identify which type of validity is missing. Propose ways to improve test design.

#### Case Study 2: Technology and Ethics

An institute adopts AI-based essay scoring. Students complain that the system penalizes creativity and favors formulaic writing.

- Task: Discuss fairness issues. Should human raters be included? Suggest a hybrid assessment model.

#### Case Study 3: Washback Effect

In a high school, the national exam focuses heavily on reading comprehension. Teachers spend most lessons drilling past exam papers. Students rarely practice speaking or writing.

- Task: Identify the type of washback. Recommend three classroom strategies to balance assessment and learning.

#### Case Study 4: Alternative Assessment

A teacher introduces digital portfolios where students upload essays, voice recordings, and reflection logs. Some students thrive, while others struggle with technology.

- Task: Evaluate the advantages and limitations of this method. Suggest solutions for students with limited digital literacy.

