ФАРҒОНА ДАВЛАТ УНИВЕРСИТЕТИ ҲУЗУРИДАГИ ПЕДАГОГ КАДРЛАРНИ ҚАЙТА ТАЙЁРЛАШ ВА УЛАРНИНГ МАЛАКАСИНИ ОШИРИШ
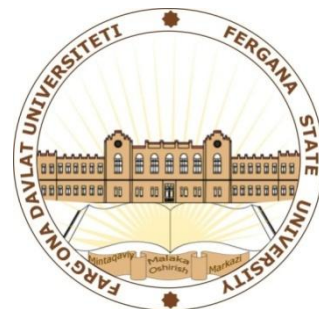
"ТИЛ КОМПЕТЕНЦИЯЛАРИНИ БАҲОЛАШ МЕХАНИЗМЛАРИ" МОДУЛИ БЎЙИЧА

Ў Қ У В – У С Л У Б И Й

М А Ж М У А

Бош илмий-методик марказ

2021

Модулнинг ўқув-услубий мажмуаси Олий ва ўрта махсус таълим вазирлигининг 2020 йил 7 декабрдаги 648-сонли буйруғи билан тасдиқланган ўқув дастури ва ўқув режасига мувофиқ ишлаб чиқилган ва ФарДУ Илмий кенгашининг 2020 йил «28» декабрдаги 2 -сонли қарори билан тасдиқланган.

Тузувчи: **Д.Усмонова - ФарДУ катта ўқитувчиси**

Тақризчи: **Н.Хошимова – Инглиз тили ўқитиш методикаси кафедраси катта ўқитувчиси, ф.ф.д., PhD**

М У Н Д А Р И Ж А

3

I. ИШЧИ ДАСТУР

Кириш

Дастур Ўзбекистон Республикасининг 2020 йил 23 сентябрда тасдиқланган "Таълим тўғрисида"ги Қонуни, Ўзбекистон Республикаси Президентининг 2017 йил 7 февралдаги "Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида"ги ПФ-4947-сон, 2019 йил 27 августдаги "Олий таълим муассасалари раҳбар ва педагог кадрларининг узлуксиз малакасини ошириш тизимини жорий этиш тўғрисида"ги ПФ-5789-сон, 2019 йил 8 октябрдаги "Ўзбекистон Республикаси олий таълим тизимини 2030 йилгача ривожлантириш концепциясини тасдиқлаш тўғрисида"ги ПФ-5847-сонли Фармонлари ҳамда Ўзбекистон Республикаси Вазирлар Маҳкамасининг 2019 йил 23 сентябрдаги "Олий таълим муассасалари раҳбар ва педагог кадрларининг малакасини ошириш тизимини янада такомиллаштириш бўйича қўшимча чора-тадбирлар тўғрисида"ги 797-сонли Қарорларида белгиланган устувор вазифалар мазмунидан келиб чиққан ҳолда тузилган бўлиб, у олий таълим муассасалари педагог кадрларининг касб маҳорати ҳамда инновацион компетентлигини ривожлантириш, соҳага оид илғор хорижий тажрибалар, янги билим ва малакаларни ўзлаштириш, шунингдек амалиётга жорий этиш кўникмаларини такомиллаштиришни мақсад қилади.

Дастур доирасида берилаётган мавзулар таълим соҳаси бўйича педагог кадрларни қайта тайёрлаш ва малакасини ошириш мазмуни, сифати ва уларнинг тайёргарлигига қўйиладиган умумий малака талаблари ва ўқув режалари асосида шакллантирилган.

Қайта тайёрлаш ва малака ошириш йўналишининг ўзига хос хусусиятлари ҳамда долзарб масалаларидан келиб чиққан ҳолда дастурда тингловчиларнинг инглиз тили фани доирасидаги билим, кўникма, малака

ҳамда компетенцияларига қўйиладиган талаблар такомиллаштирилиши мумкин.

Модулнинг мақсади ва вазифалари

"Тил компетенцияларини баҳолаш механизмлари" модулиниг мақсади: тингловчиларнинг халқаро эътироф этилган стандартларга мувофиқ хорижий тилларни билиш даражасини баҳолаш доирасидаги билим, кўникма, малака ҳамда компетенцияларини ривожлантириш.

"Тил компетенцияларини баҳолаш механизмлари" модулининг вазифалари:

- Тингловчиларга тил билиш даражасини бахолаш томойиллари ва усуллари буйича билим бериш;

- мавзуга оид олимлар фикрларини таҳлил килишга ургатиш;

- Баҳолаш мезонларини яратиш ва улардан самарали фойдаланишни амалиятта қўлланишга ўргатиш;

-. Альтернатив баҳолаш, Синфни тадқиқ қилиш, Фидбек бериш, Портфолио тузиш, Ўз ўзини баҳолаш учун материаллар яратишни ургатиш;

.-CEFR тизими бўйича баҳолашни ургатиш;.

-Тестлар яратиш усуллари ва принципларини билиш, тил кўникмалари бўйича тестлар яратиш кўникмасини шакллантириш.

Модул бўйича тингловчиларнинг билими, кўникма ва малакаларига қўйиладиган талаблар

"Тил компетенцияларини баҳолаш механизмлари" модулини ўзлаштириш жараёнида амалга ошириладиган масалалар доирасида **тингловчилар:**

- тил компетенцияларини баҳолашнинг асосий тамойиллари буйича тасаввурга эга булиш;

- альтернатив баҳолаш турларини билиш ва улардан фойдаланиш;

- тил кўникмаларини (ўқиш, гапириш, тинглаб-тушиниш, ёзув) бахолаш мезонларини яратиш;

- CEFR тизими бўйича баҳолашни қўлланиш;

- тил кўникмалари бўйича тестлар тайёрлаш;

- Мавзуга оид муаммоларини ўрганиш ва олимлар фикрларини таҳлил қилиш малакаларини эгаллаши лозим

"Тил компетенцияларини баҳолаш механизмлари"модулининг соатлар бўйича тақсимоти

№	Модул мавзулари	Тингловчининг ўқув юкламаси, соат				
			Аудитория ўқув юкламаси			Кўчма машғулот
		Ҳаммаси	Жами	жумладан		
				Назарий	Амалий машғулот	
1.	Тил билиш даражасини баҳолаш тамойиллари. Баҳолаш мезонларини яратиш ва улардан самарали фойдаланиш.	4	4	2	2	
2.	Баҳолаш ва материал тўплаш. Коммуникатив ва вазифага асосланган тил ўргатишда баҳолаш мезонларини қўллаш. Альтернатив баҳолаш. Синфни тадқиқ қилиш.	2	2		2	
3.	Фидбек бериш. Ўз ўзини баҳолаш. Баҳолашнинг асосий турлари. Мавзуга оид олимлар фикрлари таҳлили. Портфолио тузиш. CEFR тизими бўйича баҳолаш.	4	4		4	
4.	Тестлар яратиш усуллари. Тест принциплари.	4	4	2	2	
5	Аудио матнларни танлаш ва матн харитасини яратиш.	4	4		4	
6	Тинглаб тушуниш, ўқиш, ёзув, гаплашиш кўникмаси бўйича тестлар тахлил қилиш ва яратиш кўникмасини шакллантириш.	2	2		2	
	Жами:	**20**	**20**	**4**	**16**	**0**

НАЗАРИЙ МАШҒУЛОТ МАТЕРИАЛЛАРИ

1-Мавзу: Тил билиш даражасини баҳолаш тамойиллари. Баҳолаш мезонларини яратиш ва улардан самарали фойдаланиш.

Баҳолаш ва материал тўплаш. Коммуникатив ва вазифага асосланган тил ўргатишда баҳолаш мезонларини қўллаш. Альтернатив баҳолаш. Синфни тадқиқ қилиш. Фидбек бериш. Ўз ўзини баҳолаш. Баҳолашнинг асосий турлари. Мавзуга оид олимлар фикрлари таҳлили. Портфолио тузиш. CEFR тизими бўйича баҳолаш.

2-Мавзу:Тестлар яратиш усуллари. Тест принциплари.

Аудио матнларни танлаш ва матн харитасини яратиш. Тинглаб тушуниш, ўқиш, ёзув, гаплашиш кўникмаси бўйича тестлар тахлил қилиш ва яратиш кўникмасини шакллантириш.

АМАЛИЙ МАШҒУЛОТЛАР

1-Амалий машғулот. Тил билиш даражасини баҳолаш тамойиллари. Баҳолаш мезонларини яратиш ва улардан самарали фойдаланиш.

2-Амалий машғулот. Баҳолаш ва материал тўплаш. Коммуникатив ва вазифага асосланган тил ўргатишда баҳолаш мезонларини қўллаш. Альтернатив баҳолаш.Синфни тадқиқ қилиш.

3-4-Амалий машғулот. Фидбек бериш. Ўз ўзини баҳолаш. Баҳолашнинг асосий турлари. Мавзуга оид олимлар фикрлари таҳлили. Портфолио тузиш. CEFR тизими бўйича баҳолаш.

5-Амалий машғулот. Тестлар яратиш усуллари. Тест принциплари.

6-7-Амалий машғулот.Аудио матнларни танлаш ва матн харитасини яратиш.

8-Амалий машғулот. Тинглаб тушуниш, ўқиш, ёзув, гаплашиш кўникмаси бўйича тестлар тахлил қилиш ва яратиш кўникмасини шакллантириш.

АДАБИЁТЛАР РЎЙХАТИ

I. Ўзбекистон Республикаси Президентининг асарлари

1. Мирзиёев Ш.М. Буюк келажагимизни мард ва олижаноб халқимиз билан бирга қурамиз. – Т.: "Ўзбекистон", 2017. – 488 б.

2. Мирзиёев Ш.М. Миллий тараққиёт йўлимизни қатъият билан давом эттириб, янги босқичга кўтарамиз. 1-жилд. – Т.: "Ўзбекистон", 2017. – 592 б.

3. Мирзиёев Ш.М. Халқимизнинг розилиги бизнинг фаолиятимизга берилган энг олий баҳодир. 2-жилд. Т.: "Ўзбекистон", 2018. – 507 б.

4. Мирзиёев Ш.М. Нияти улуғ халқнинг иши ҳам улуғ, ҳаёти ёруғ ва келажаги фаровон бўлади. 3-жилд.– Т.: "Ўзбекистон", 2019. – 400 б.

5. Мирзиёев Ш.М. Миллий тикланишдан – миллий юксалиш сари. 4-жилд.– Т.: "Ўзбекистон", 2020. – 400 б.

II. Норматив-ҳуқуқий ҳужжатлар

6. Ўзбекистон Республикасининг Конституцияси. – Т.: Ўзбекистон, 2018.

7. Ўзбекистон Республикасининг "Таълим тўғрисида"ги Қонуни.

8. Ўзбекистон Республикаси Президентининг 2015 йил 12 июндаги "Олий таълим муасасаларининг раҳбар ва педагог кадрларини қайта тайёрлаш ва малакасини ошириш тизимини янада такомиллаштириш чора-тадбирлари тўғрисида" ги ПФ-4732-сонли Фармони.

9. Ўзбекистон Республикаси Президентининг 2017 йил 7 февралдаги "Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида"ги 4947-сонли Фармони.

10. Ўзбекистон Республикаси Президентининг "Олий таълим тизимини

янада ривожлантириш чора-тадбирлари тўғрисида"ги 2017 йил 20 апрелдаги ПҚ-2909-сонли Қарори.

11. Ўзбекистон Республикаси Президентининг "2019-2021 йилларда Ўзбекистон Республикасини инновацион ривожлантириш стратегиясини тасдиқлаш тўғрисида"ги 2018 йил 21 сентябрдаги ПФ-5544-сонли Фармони.

12. Ўзбекистон Республикаси Президентининг 2019 йил 27 майдаги "Ўзбекистон Республикасида коррупцияга қарши курашиш тизимини янада такомиллаштириш чора-тадбирлари тўғрисида"ги ПФ-5729-сон Фармони.

13. Ўзбекистон Республикаси Президентининг 2019 йил 17 июндаги "2019-2023 йилларда Мирзо Улуғбек номидаги Ўзбекистон Миллий университетида талаб юқори бўлган малакали кадрлар тайёрлаш тизимини тубдан такомиллаштириш ва илмий салоҳиятини ривожлантири чора-тадбирлари тўғрисида"ги ПҚ-4358-сонли Қарори.

14. Ўзбекистон Республикаси Президентининг 2019 йил 27 августдаги "Олий таълим муассасалари раҳбар ва педагог кадрларининг узлуксиз малакасини ошириш тизимини жорий этиш тўғрисида"ги ПФ-5789-сонли Фармони.

Ш. Махсус адабиётлар

1. Beresova, J. (2017). The impact of the CEFR on teaching and testing English in the local context. Theory and Practice in Language Studies, 7(11), 959-964.

2.Brown, H. D., & Abeywickrama, P. (2014). Language assessment. Principles and classroom practices. White Plains, NY: Pearson Education.

3.Cumming, A. and R. Berwick (Eds.). (2010) Validation in Language Testing. Clevedon, England: Multilingual Matters, Ltd.

4.Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), Handbook of second

language assessment: Volume 12 of handbooks on applied linguistics (pp. 413-428). Berlin, Germany: Walter de Gruyter GmbH & Co KG. doi: https://doi.org/10.1515/9781614513827-027

5.Henning, G. (2014). A guide to language testing: Development, evaluation, research. Newbury House Publishers.

6.Hughes, A. (2013). Testing for Language Teachers. 2nd ed. Cambridge: Cambridge University Press.

7.Isaacs, T., Zara, C., Herbert, G., Coombs, S. J., & Smith, C. (2013). Assessment of learning. In The SAGE key Concepts Series: Key concepts in educational assessment (pp. 12-17). London: SAGE Publications Ltd doi: 10.4135/9781473915077.n4

8.Knight, P., & Yorke, M. (2013). Assessment, learning and employability. Maidenhead: Society for Research into Higher Education & Open University Press.

9.Madsen H.S. (2014). Techniques in Testing. England: Oxford University Press.

10.McMilan, J.H. (2014). Fundamental Assessment Principles for Teachers and School Administrators. Practical Assessment, Research and Evaluation. Retrieved from https://pareonline.net/getvn.asp?v=7&n=8 [Accessed July 17, 2018]

11.McNamara, T. (2014). Language Testing. UK. Oxford: Oxford University Press.

12.Taylor, L. (2019). Developing assessment literacy. Annual Review of Applied Linguistics, 29, 21-36.

13.Weir, C. (2015). Language testing and validation: An evidence-based approach. Palgrave

IV. Интернет сайтлар

1. Ўзбекистон Республикаси Олий ва ўрта махсус таълим вазирлиги: www.edu.uz.

2. Бош илмий-методик марказ: www.bimm.uz

3. www. Ziyonet. uz

4. Ўзбекистон Республикаси Президентининг расмий сайти: www.gov.uz

5. Ўзбекистон Республикаси Олий ва ўрта махсус таълим вазирлигининг расмий сайти: http://www.edu.uz/modules/ wfchannel

6. www.tefl.net- Тил ўрганиш веб сайти

7. busyteacher.org -Тил ўрганиш веб сайти

II. МОДУЛНИ ЎҚИТИШДА ФОЙДАЛАНИЛАДИГАН ИНТРЕФАОЛ ТАЪЛИМ МЕТОДЛАРИ

Brainstorming	Brainstorming is a group creativity technique by which efforts are made to find a conclusion for a specific problem by gathering a list of ideas spontaneously
Prediction	A prediction (Latin præ-, "before," and dicere, "to say"), or forecast, is a statement about a future event. They are often, but not always, based upon experience or knowledge.
Group work	Group work is a form of voluntary association of members benefiting from cooperative learning, that enhances the total output of the activity than when done individually.
Pair work	Pair work is learners working together in pairs. One of the main motivations to encourage pair work in the English language classroom is to increase the opportunities for learners to use English in the class.
Reflection	Reflection is the process of reflecting on your experience in order to learn from that experience.
Comparison	Comparison or comparing is the act of evaluating two or more things by determining the relevant, comparable characteristics of each thing, and then determining which characteristics of each are similar to the other, which are different, and to what degree.
Discussion	the activity in which people talk about something and tell each other their ideas or opinions
Elicitation	Elicitation technique or elicitation procedure, any of various data collection techniques in social sciences or other fields to gather knowledge or information from people.

Project work	Project work is a series of activities that allows the students to study,do research and act by themselves using their abilities, interests, personal experience
True/False	Type the question in the form of a statement that students can answer with true or false.
Matching	going together well : suitably paired or used together.

ELICITATION

Lesson 2. Activity 1 (5 min)

Ask participants the following questions (if possible write key words on the white board):

- What is *testing*?
- What is *assessment*?
- What is evaluation?

Lesson 7. Activity 1 Different types of tests

Ask participants the following questions. Take answers to each question

one at a time and comment and expand as necessary:

~ Do you test your students' knowledge? If you do, how do you do it?

~ Why do you have tests with your students?

~ Are there any obligatory tests at your universities? If yes, what kinds of test are these?

☺(10 min) Ask participants when each of the tests is given and why. Draw a timeline to illustrate when different tests are given taking into account different purposes of the tests. Use the following questions:

~ When are the tests conducted ?

~ What is the purpose of each of the tests?

1)Placement test / proficiency test

2)Progress test

3) Achievement test

4)Proficiency test

Lesson 8. Lead-in activity

Objective: to uncover the current practice of testing and assessing receptive skills

Time: 5 min

►Procedure:

☺(5 min) Refer participants to the sessions on Teaching Reading and Authenticity.

Ask them what kind of reading subskills they remember. Elicit scanning, skimming, reading intensively. Ask participants the following question:

~ How do you test or otherwise assess your students' reading and listening

skills?

• Elicit random answers.

• Establish that in our context reading skills are regarded as a tool for testing other skills like writing, speaking, pronunciation, etc. through such tasks as 'read and retell the text','read aloud', 'read and learn by heart'. Reading skills such as scanning, skimming and reading intensively are not tested. Testing listening skills is not very popular because of the lack of resources. Tell participants that this session focuses on testing listening and reading skills.

PAIR WORK

Lesson 5. Activity 4

Time: 20 min

Materials: Handout 4

Procedure:

- Ask participants to work in pairs and fill in the table (handout 4) with the questions/problems and disscuss:
- Is any qualitative data necessary to research any of the written research questions? If yes, how this data can be got? (students' tests, questionnaires, etc.)
- What are the reasons of collecting quantitative data to research the question/problem?

- Ask participants to share filled table and report about the discussed answers to the stated questions.

  Possible answers: participants' answers

TRUE/ FALSE

Lesson 2. Activity 2, Handout 1b. True/False statements

A) This course is taught in semesters 5-6.

B) Students will be able to analyze, develop and plan assessment procedure.

C) Only theory of testing is introduced in this course.

D) Students will analyse and develop only multiple-choice questions.

E) Students will know about different test types.

F) Portfolios, logs, and project works are considered as an alternative ways of assessment

G) Students will only discuss how language skills are assessed

PROJECT WORK

Lesson 4. Activity 1 Project work as one of the vehicles for assessing students' performance

Objective: to give participants an opportunity to analyse a piece of project work done by students

Time: 15 min

Materials: 4 booklets, 4 posters, board, markers, flipchart paper

Preparation: write the questions from the first bullet point on the board

►Procedure:

☺☺☺(5-7 min) Put participants in 4 groups. Explain that students of one university were given an assessed task to produce booklets and posters. Distribute one booklet and one poster to each group. Ask groups to look through the materials and answer the following questions (written on the board):

~ What do you think the task of the project was?

~ How many students do you think worked on the task?

~ What do you think the steps in students' work were?

~ How much time do you think the project took?

☺(7-8 min) Invite groups to share the results of their discussion.

Summarise the activity by saying that project work can be quite an effective way of learning, teaching and assessment and in answering the questions above participants have thought about certain things that should be taken into consideration when setting such tasks.

GROUP WORK

Lesson 2. Activity 5. <u>Specifying topics and writing objectives with outcomes</u>

Time: 25 min

Materials: handout 4

Procedure:

- Distribute handout 4, ask participants to work in groups and read the topics written in the first column

- Ask participants by referring to the example fill in the table with possible subtopics, objectives and outcomes for each topic.
- Say that answers of each group can vary.

Once they have finished, elicit random answers in the plenary

Lesson 4. Activity 2 Skills that can be developed through alternative ways of assessment

Invite groups to share the marks they have agreed on and ask each group representative to comment on how they came up to the decision.

NB Be sure to lead the group to the concept of assessment criteria through this discussion. While group representatives are presenting their ideas, listen in and make a list (on the board) of assessment criteria mentioned by them. When all the representatives have spoken, draw the whole group's attention to the list of criteria they were using to decide on a mark and ask if anything should be added. (Make sure participants understand what assessment criteria are, refer to the session on Giving Feedback on Writing in module 1)

Make the point that if a task aims at developing the ability to work in groups, or decision making skills, these skills also have to be recognized and credited in the assessment process.

Lesson 4. Activity 2. Task for students

1. You are required to work in small groups of no more than 4 people. The task for you is to write a newspaper article about a local zoo. You will have to:

1. think about the specific things in the zoo you would like to concentrate on;

2. visit the zoo and collect all the necessary information;

3. analyse the information you have collected and think how you can use it to write an

Lesson 4. Activity 3 More ideas

Ask participants to work in 4 groups. Assign one type of assessment to each group. Tell groups to imagine that they will have to use one of the assessment types in their teaching. Ask them to think how they can incorporate this type of assessment into their system of assessment (they do not need to think of a task instruction in great details, only a brief idea). Ask groups to fill in the following proforma in handout 3:

Type of assessment	
Progress or achievement	
What skills it will test	
The percentage weighting (from the overall 100 percent for the semester)	
Task description (very briefly, just an idea would be fine here)	
Time allocated for students	
Individual or group	
Assessment criteria	

Lesson 6, Activity 1 Assessing speaking

Put participants in groups. Tell them that they are going to listen to a sample of students' speech. Ask participants to assess the student's answer according to the scale on the board. Play the tape œ.

☺ ☺ ☺ (10 min) Ask participants to discuss their results in groups. Tell them to think what aspects of the spoken language they took into account and ask them to

come up with the criteria for assessing spoken language. Ask the following question:

Lesson 6. Activity 2 Rating scales for assessing speaking

Put participants in 5 groups and tell them that each group will receive a scrambled description of one of the criteria. Tell them that they will need to put the descriptors according to the levels of proficiency from A1 (beginner) to C2 (advanced) paying attention to the key words that help define the level of proficiency.

- Give out the handout 1 to participants.

☺(10 min) Ask groups to report on their work. Tell them to speak particularly about the language in the chart that helped them define the levels.

Lesson 7. Activity 4 Principles of testing

Objective: to work out the main principles of a good achievement test

Time: 15 min

Materials: board, markers

►Procedure:

☺☺☺(5 min) Tell participants to work in groups and based on their discussions of good and bad items draw up some principles of a good test.

☺(10 min) Collect participants' ideas on the board.

Lesson 8. Activity 1 Testing reading skills

Put participants in groups of 5. Ask groups to study the tests and answer the following questions:

1. What reading sub-skills does each item test?

2. Do you actually have to read the text in order to answer the questions?

3. What kinds of texts (authentic or non-authentic) are used in the test?

4. What did you like and dislike about this test?

☺(5 min) Elicit random answers to each question one at a time and expand on them.

REFLECTION

Lesson 4. Hometask

You are also required to write a brief personal reflection (150 words maximum) about how you worked together and your individual contribution to the project. Submit both papers (an article and a report) in 4 weeks time

MATCHING

Lesson 5. Lead-in activity

Time: 15 min

Materials: Handout 1

Procedure:

Ask participants to match classroom research procedures to their definitions.

Possible answers:

Teaching journals	Written or recorded accounts of teaching experiences.
Lesson reports	Written accounts of lessons which describe the main features of the lessons.
Surveys and questionnaires.	Activities such as administering a questionnaire or completing a survey, designed to collect information on a particular aspect of teaching or

	learning.
Audio and video recordings	Recordings of a lesson, or part of a lesson.
Observation.	Tasks completed by a student teacher observing a cooperating teacher's class, or peer observation (i.e., tasks completed by a teacher visiting a colleague's class).

Theme 1. LECTURE: INTRODUCTION TO LANGUAGE ASSESSMENT AND TESTING (2 hours)

Plan:

1. **Introduction. Assessment as an ongoing processs**

2. **Summative assessment and formative assessment**

3.   **Fundamental principles of assessment**

4. **Scholars points about assessment**

Key words: formative assessment, summative assessment, formal and informal assessment, final and continuous assessment, product-driven assessment, process-based assessment, convergent, divergent, validity, reliability, cohort referencing.

1. Introduction. Assessment as an ongoing process

Assessment is an ongoing process aimed at understanding and improving student learning. It involves:

•making our expectations explicit and public;

•setting appropriate criteria and high standards for learning quality;

• systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and standards;

•and using the resulting information to document, explain, and improve performance.

We accept here that assessment: … is an ongoing process aimed at understanding and improving student learning. It involves making our expectations explicit and public; setting appropriate criteria and high standards for learning quality; systematically gathering, analysing, and interpreting evidence to determine how well performance matches those expectations and standards; and using the

resulting information to document, explain, and improve performance. When it is embedded effectively within larger institutional systems, assessment can help us focus our collective attention, examine our assumptions, and create a shared academic culture dedicated to assuring and improving the quality of higher education. (Angelo, 1995, pp. 7-9)

2. Formative and summative assessment

Assessment is usually subdivided into two categories, often known as summative assessment and formative assessment.

➢ Formative assessment has a developmental purpose and is designed to help learners learn more effectively by giving them feedback on their performance and on how it can be improved and/or maintained. Reflective practice by students sometimes contributes to formative assessment. Formative assessment is designed to assist the learning process by providing feedback to the learner, which can be used to highlight areas for further study and hence improve future performance. Self and diagnostic assessment are types of formative assessment with specific purposes. (Computer based test, which provides feedback on areas of weakness, an essay which is annotated with the lecturer's comments, but no overall mark) Formative assessment on the other hand allows the students and other interested parties to form a more detailed opinion of their abilities, which can then be used to inform further study, concentrating students' efforts on the more appropriate areas and hence improving overall performance.

➢ Summative assessment is used to indicate the extent of a learner's success in meeting the assessment criteria used to gauge the intended learning outcomes of a module or programme. Summative assessment is for progression and/or external purposes, given at the end of a course and designed to judge the students' overall performance. (IQ tests, driving tests, traditional examinatins)

Formal or informal

➢ Formal assessments are where the students are aware that the task that they are doing is for assessment purposes. Formal assessments are perceived to be

'fairer'. Criteria tend to be more explicit and have less room for bias. Students know they are to be assessed and behave accordingly. However, such assessments can induce stress sometimes causing students to perform less well; others may cram and perform well, but without deep understanding.

➤ With informal assessment the judgements are integrated with other tasks. Informal assessments can reduce stress, and give a more valid view of students' abilities, however some students may feel 'cheated' out of their chance to shine. There can also be problems with hidden prejudices and stereotypes influencing the judgement of the assessor when informal assessments are used.

Final or continuous

➤ Final (terminal) assessment is that which takes place only at the end of a course while continuous assessment is scattered throughout the course.

➤ The primary advantage of final assessment is that it is simple to organise and condenses the assessment process into a short space of time. This means, however, the timing of the examination becomes of great importance. Illness at an unfortunate time can unduly influence the result. Furthermore final assessment cannot be used for formative purposes.

➤ The main advantages of continuous assessment are that both students and lecturers obtain feedback from the process which can then be used to improve teaching and learning, and the final result is based on evidence gathered over the span of the learning period. Disadvantages include the increased workload inherent with this mode of assessment, and difficulties associated with students from different backgrounds tackling the same material and being assessed in exactly the same way.

➤ Final assessment may be appropriate where there is evidence that learning each new field of study contributes to the understanding of every other, and hence learning can only be assessed as a complete whole rather than as constituent parts. Continuous assessment is more appropriate where student feedback is required and a series of pieces of information are required across the course to build up a picture

of students' abilities. Computer-assisted assessment (CAA) can provide a powerful means of continuous assessment, providing rapid and detailed feedback to students and academics about the learning process

Process or product

➢ With the rapidly changing nature of modern society, increased emphasis is being placed on skills and abilities rather than knowledge. It is therefore important to consider whether you wish to assess the product of student learning, or the process undertaken.

➢ Product-driven assessments are usually easier to create, as the assessment criteria tend to be more tangible. They can also be more easily summarised. Process-based assessments however can give more useful information about skills, and can highlight to students the importance of learning generalised techniques rather than specific knowledge. Some students do see process-based assessments as unfair 'How are we supposed to write an essay on Quarks when you haven't taught us about them?'. Therefore, the reasons for assessing in this manner, the criteria that will be applied, and what will be considered evidence must be explained carefully to students.

➢ Process-based assessments are best where the learning is skill or ability-based, while product- based assessments are best where the knowledge content of the course is fundamental. Most assessments are mixtures of the two forms. The balance is critical in ensuring that the assessment is fit for the purpose

Convergent or divergent

Convergent assessments are those which have one correct answer that the student is trying to reach. Divergent assessments appreciate a range of answers based on informed opinion and analysis. Convergent assessments are generally easier to mark - both by automated and human means. They tend to be quicker to deliver and give more specific and directed feedback to individuals and can also provide greater curricular coverage. However, they can be limited in scope and can

occasionally degenerate into a 'quiz' of facts about the area of study. There is also a temptation to only test that which is easily translated into a convergent form.

CAA is an increasingly common form of convergent assessment. Computers offer particular advantages in extending the scope and authenticity of convergent assessments, however good questions and tests require skilled design and construction. Divergent assessments can be more authentic, and make it easier to assess higher cognitive skills. However, they can be time consuming to set and mark. They also require greater marking skill than convergent assessments, this can involve training markers and/or detailing criteria.

➢ Where knowledge is the primary issue, convergent assessments can be very useful. Because of their wide curricular coverage, they can be very important in formative assessment to quickly and effectively highlight areas of weakness and gaps in students' knowledge. Where there is a core of knowledge that is a fundamental base for the study of the subject, convergent assessment can be an effective way of ensuring that it is in place. CAA is often used to provide broad and rapid assessment of students' knowledge, it can effectively identify gaps in students' knowledge using statistical analysis and reports.

➢ Divergent assessments by contrast are best suited when there may be a difference of opinion based on interpretation. This is most obvious in an area such as literary criticism, but can also be important in a medical diagnosis for example. A divergent assessment - requiring students to explain the basis for their diagnosis - can check students' reasoning, and uncover misapprehensions that they might be under. It also allows for valid diagnoses that may not have occurred to the question setter.

Purposes of assessment

Assessment serves a number of purposes. The main purpose of summative assessment is to measure student learning in a way that recognises it through the award of credits or equivalent (the combination of which can then lead to a named qualification). However, of equal importance is the recognition that assessment

should also be an integral part of learning, or that summative as well as formative assessment can, and does, facilitate student learning.

To ensure that the communication is as honest as possible it is crucial to make sure that the assessment is both valid - in that it tests a relevant skill or ability, and that it is reliable - in that the same result would be achieved if you repeated the assessment

For the student, individual pieces of assessment provide a source of motivation for study; they promote learning by providing feedback on performance and help students to identify their strengths and weaknesses.

For the lecturer, assessment provides an opportunity to evaluate the knowledge, understanding, ability and skills attained by different students. The overall profile of student performance offers useful information for assessing the effectiveness of course content and teaching methods, thereby facilitating improvement.

For the institution, assessment provides information upon which decisions as to students' progression and the receipt of awards may be based. The assessment process enables the institution to ensure that appropriate standards are being met, in accordance with nationally agreed frameworks, such as subject benchmark statements and the frameworks for higher education qualifications. Information generated by assessment, such as mark or grade distributions, forms a valuable tool for quality assurance and enhancement.

3. Fundamental principles of assessment
Validity and reliability.

Although validity and reliability are separate entities it makes sense to consider them together because jointly they define the overall quality of assessment. Conventional wisdom suggests that a valid test must always be reliable, although this is being challenged in some quarters. A valid assessment is one which measures that which it purports to measure, while a reliable assessment is one where the same results are gained time after time.

A valid assessment is one which measures that which it is supposed to measure. For example, a Physics assessment which insisted that answers had to be written in German would not be a valid assessment as there is a good chance that you would be testing students' knowledge of German rather than their abilities in Physics. It is important when designing an assessment that you consider whether it does actually assess what you intend it to. There are several different types of validity and it is worth considering each of these in turn.

Curricular (content) validity

The first overarching form is curricular validity - ensuring that the aims of the curriculum are in keeping with what the students need to know. Without curricular validity, not only is the assessment quality doubtful, but calls into question the quality of the whole course. Ensuring curricular validity means ensuring that the learning objectives for the course are closely related to the desirable outcomes of a successful student.

struct validity

Construct validity is essentially how closely the assessment relates to the domain that you wish to assess. Most assessments require broadly based skills beyond the subject domain (e.g. the ability to read questions involving technical terminology, to construct an essay, even the ability to turn up to the exam hall on time). Some of these skills can be validly included as part of the assessment as they could be considered to be implicit criteria within the learning objectives, while other skills may not be. For example, a CAA which required a high level of information technology skills would be inappropriate if you were testing students' ability to read geological maps. Ensuring construct validity means ensuring that the assessment content is closely related to the learning objectives of the course.

Predictive validity

Predictive validity suggests that predictions made on the basis of the assessment results will be valid. For example you might predict that someone who

scored an A in Biology (at A-level) might perform better in a degree course in Biology than someone who failed. If that is the case, then the assessment can be considered to have predictive validity. This type of validity is most important when the primary purpose of the assessment is selective. Ensuring predictive validity means ensuring that the performance of a student on the assessment is closely related to their future performance on the predicted measure.

Reliability

A reliable assessment consistently gives the same results under identical circumstances. A physics assessment which gave the same candidate three different grades on three consecutive occasions, without any substantive change in the candidate's abilities in-between, would not be a reliable assessment. It is important when designing an assessment that you consider whether the results achieved will be consistent. There are several different ways of measuring reliability.

Test-retest reliability

Test-retest reliability is the correlation between candidates' attempts at the same test. Where there is little test-retest reliability, the people who did well first time round may not do well second time round. Obviously this is an important consideration as it suggests that some element of the measure may be due to chance rather than actual skills, ability and knowledge.

Parallel forms reliability

Parallel forms reliability is the correlation between candidates' attempts at two tests which are supposed to be identical. Where this type of reliability is lacking, there is evidence that the tests are testing different things; suggesting that one or both are not testing the pre-defined knowledge and skills - or domain- intended.

Internal consistency

The internal consistency of a test is essentially a pseudo-measure of reliability. Most of the time we do not have the luxury of constructing two separate tests, or testing the students twice. Internal consistency is designed to measure what would have happened had we done that. It is essentially the correlations between the test items. It can be thought of as an estimate of the correlation between the test that was actually delivered, and all of the other possible tests that might have been constructed using those items.

Referencing

The referencing of an assessment is the basis of the judgement. There are three main ways of referencing: against peers (norm-related referencing), whereby the judgement is essentially a comparison between the student and other people; against objective criteria (criterion referencing) where the judgement is a comparison between the student's abilities and the contents of a pre-defined domain; and against the student her/himself (ipsotive referencing) where the judgement is a comparison of the student's performance on one area as against prior performance, or performance on other areas.

Norm-related referencing

Norm-related referencing is the comparison of individuals with their peers. This was popular through the mid-20th century, but has become rather unfashionable in modern testing. It can be useful for selective purposes (e.g. for the distribution of a scholarship to the 5 best students, or extra tuition to the 5 which are struggling most), but gives little information about the actual abilities of the candidates.

Norm referencing

Classic norm referencing involves delivering a test to a representative sample of the type of students that you wish to assess, and developing norms based on the results. These norms are then used to grade subsequent groups of students. This can lead to anomalies where the group on which the norm was based becomes

very different from the group that is currently taking the examination. This type of referencing is normally credited with maintaining standards across time however, as the curriculum and intake changes, these will not be reflected in the assessment leading to unreliable results.

Cohort referencing

Cohort referencing is similar to norm referencing, however, it takes the subgroup of candidates attempting the assessment as its base-line. Under this type of referencing, the highest results are given to students who attain the best marks relative to their peers who also took the assessment at the time. Unless you can be confident that the intake remains unchanged, this makes for unreliable comparisons across student groups, particularly where the cohort is small. Attainment of a high grade can be as dependent on the performance of the other students taking the assessment as on your own performance.

Criterion referencing

Criterion referencing is a comparison of an individual with pre-defined criteria. It can be used for both formative and summative purposes, both highlighting areas of weakness and determining whether candidates have achieved an acceptable level in the areas they are expected to know about. Results can often be misinterpreted, particularly by those who are more familiar with the older, norm (related) referencing. It must be made clear to users of the assessment data that the criteria for success is performance against learning objectives, rather than performance against other students.

Ipsotive referencing

Ipsotive referencing is a comparison of an individual against him/herself. Although generally unsuitable for selective purposes, Ipsotive referencing can be extremely useful for diagnostic or formative purposes.

Difficulty (facility)

The difficulty of a question (or mark point) can be thought of as the proportion of students who get the question correct. In order that students are separated out as

much as possible it is desirable for assessments overall to have a difficulty level of about 0.5 - so that the mean mark is roughly half of the marks available.

Where one question in an assessment carries a high proportion of the marks (e.g. a 25 mark essay question on a paper worth 40 marks), it is desirable for the difficulty level of that question to be close to 0.5. In contrast where an individual question is worth a lower proportion of the marks, it is quite acceptable for it to have a higher or lower facility value.

Where a test is comprised of many questions, each worth a low proportion of the total marks available, it is desirable to have questions which vary in difficulty, so that candidates at all points of the ability stratum may be fully tested . It is, however, undesirable for questions to have facility values above 0.85 or below 0.15.[1], because at this level they are contributing little to overall measurement. The closer the questions come to having a facility value of 0.5, the more they are contributing to the measurement of the candidates.

Discrimination

Discrimination is a measure of how well the question distinguishes between students - and thus how much information the question is providing. There are several methods used to calculate the discrimination of a question, the most common being the Pearson product- moment correlation between the question and total score. This measure assumes unidimensionality. Where this is not the case and the test is designed to examine more than one content area or skill, it may be better to use the correlation between the question and the total of other questions within the same domain as a measure of discrimination.

Being essentially a correlation, question discrimination can vary from +1.0 (where there is a perfect relationship between those who score high marks on the question and those who score high marks on the test) to -1.0 (where there is a perfect *inverse* relationship between those scoring high marks on the question and on the test overall).

In general question discrimination should be positive, unless there is good reason to suppose that the assumption of unidimensionality has been violated. In such a case, question discrimination should be positive within the sub-domain that the question tests, or (if it is the only question representing the sub-domain) with another more representative indicator of performance.

Negative question discrimination with a valid criterion should always be regarded as suspect, however, there is no upper limit for this statistic: the higher the correlation, the better the question discrimination, the better the question. In general values below 0.2 are weak, and values above 0.4 are desirable. It should be noted that questions with lower maximum marks, and those with extreme difficulty levels have less potential for variance than those with higher maximum marks, and hence are likely to have a lower discrimination

GRADING

Grading involves comparing a student's performance with a pre-defined set of standards. The two types of grading most commonly in use are norm referenced grading, where the candidate's performance is compared to other people who are considered to have 'set the standard', and mastery learning where the candidate's performance is compared with a set of learning objectives. In practice most types of grading involve combining the two types.

4. Scholars points about assessment

There is no generally agreed definition of assessment. Various definitions of assessment and the role it plays in teaching and learning:

1. Assessment involves the use of empirical data on student learning to refine programs and improve student learning. (Assessing Academic Programs in Higher Education by Allen 2004)
2. Assessment is the process of gathering and discussing information from multiple and diverse sources in order to develop a deep understanding of what students know, understand, and can do with their knowledge as a result

of their educational experiences; the process culminates when assessment results are used to improve subsequent learning. (Learner-Centered Assessment on College Campuses: shifting the focus from teaching to learning by Huba and Freed 2000)

3. Assessment is the systematic basis for making inferences about the learning and development of students. It is the process of defining, selecting, designing, collecting, analyzing, interpreting, and using information to increase students' learning and development. (Assessing Student Learning and Development: A Guide to the Principles, Goals, and Methods of Determining College Outcomes by Erwin 1991)

4. Assessment is the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving student learning and development. (Assessment Essentials: planning, implementing, and improving assessment in higher education by Palomba and Banta 1999)

In all academic settings, assessment is viewed as closely related to instruction. Assessment is needed to help teachers and administrators make decisions about students' linguistic abilities, their placement in appropriate levels, and their achievement. The success of any assessment depends on the effective selection and use of appropriate tools and procedures as well as on the proper interpretation of students' performance. Assessment tools and procedures, in addition to being essential for evaluating students' progress and achievement, also help in evaluating the suitability and effectiveness of the curriculum, the teaching methodology, and the instructional materials.

Educational assessment can only truly reap benefits for education if it is conceptualized as having the ultimate purpose of ensuring quality education. However, the educational assessment water is muddied by different classifications of assessment that focus on various levels of interrogation. The classification of assessment as either summative or formative leaves something to be desired, as the grouping of summative provides little indication of the use of the data (Newton,

35

2007). This has led to the interrogation of the purposes of assessment on a much more detailed level, such as assessment for: genuine improvement actions; instructional purposes; supporting conversations; professional development; encouraging self-directed learning; policy development and planning; meeting accountability demands; legitimizing actions; motivating students and staff; personnel decisions; student monitoring; placement; diagnosis; and resource allocation (Newton, 2007; Schildkamp and Kuiper, 2010). The intended use and audience/s impact the design and implementation of any assessment and causes complications when data are used for anything other than its original purpose and consumers. This approach of differentiating very specific uses, while valuable when working in a particular topic area, is too detailed to help us conceptualize the role of assessment in establishing overall educational quality on a systemic level. Newton (2007) deals extensively with the importance of establishing clear, meaningful and distinct groupings of assessment purposes. In line with Newton's (Newton, 2007) work and with adaptation from the arguments of Brookhart (2001) and Black et al. (2003, 2010), this article employs a category grouping approach based on assessment purposes.

Assessment to support learning is often referred to as formative assessment. In line with the arguments of Brookhart (2001) and Black et al. (2003, 2010), who emphasize the pedagogical role that summative assessment plays in supporting learning. Summative assessment has been judged by both students (Brookhart, 2001) and educators (Black et al., 2010) to serve not merely as a tool for reporting learners' progress but to support learning. Assessment to support learning refers to the interaction between learning and assessment that is forward going. This means employing assessment data in a diagnostic approach to determine competence, gaps, and progress so learners may adapt their learning strategies and teachers their teaching strategies (Black and Wiliam, 1988; Black, 1998). This role is usually, but not solely associated with formative assessment. This type of assessment—be it formative or summative—may be a distinct event or integrated into the teaching practice. It is employed to determine the degree of mastery attained to that point and to inform the learning required to move toward mastery.

Formative assessment in particular usually has a high frequency and focuses on smaller units of instruction (Bloom et al., 1971; Newton, 2007).

There is no doubt that concerted efforts are directed toward improving the classroom assessment environment and progress is being made (Black et al., 2010). Timely, formative data are essential for continuing such informed improvements in education. Unfortunately though, such data are often lacking.

The classroom environment is where the majority of regular assessments take place to support learning. This mostly represents a low stakes environment where cheating, inadequate student and educator effort, classroom dynamics, and school complexities make for a less than optimum testing environment (Dorans, 2012; Wise and Smith, 2016; Brown, 2017). The value and importance of valid, reliable assessment data to inform decision-making and to plan monitoring is uncontested (William et al., 2004; DeLuca and Bellara, 2013), however, highlights the inconvenient truth that our assumption that assessment data are always trustworthy and subject to appropriate scrutiny is a fallacy. This basic assumption often remains unchallenged and leads to decisions based on data which has no more value than the gut instinct of educators or teacher experience. This is a particular risk in the non-standardized, formatively orientated assessment environment of the individual classroom, which is mostly employed in support of learning.

An over-emphasis of assessment to support learning may lead to an over-dependence on norm referencing. As both formative and summative assessment usually takes place in the classroom environment, the educator's judgment is easily modulated to focus on the skill levels of the group. Learner performance is thus compared to that of the other learners in the class or level. In such a case, it may well happen that the learners in a particular school have commenced at a disadvantage, having not attained the quality standards required of the particular level of study. In the absence of comparison to these standards and performance of learners in other institutions throughout the system, the learners' attainment may be over-estimated, leaving them ill-equipped to compete with learners from other institutions.

2.LECTURE: LANGUAGE COMPETENCES ASSESSMENT.
HISTORY OF LANGUAGE TESTING
PLAN:

1. **History of testing**
3. **Testing design**
4. **Testing specifications**

Keywords: Testing

TESTING

Tests have a way of scaring students. How many times in your school days did youfeel yourself tense up when your teacher mentioned a test? The anticipation of theupcoming "moment of truth" provoked feelings of anxiety and self-doubt alongwith a fervent hope that you would come out on the other end with at least a sense of worthiness. The fear of failure is perhaps one of the strongest negativeemotions a student can experience, and the most common instrument inflictingsuch fear is the test. You are not likely to view a test as positive, pleasant, oraffirming, and, like most ordinary mortals, you intensely wish for a miraculousexemption from the ordeal.

And yet, tests seem as unavoidable as tomorrow's sunrise in virtually alleducational settings around the world. Courses of study in every disciplineare marked by these periodic milestones of progress (or sometimes, in the perceptionof the learner, confirmations of inadequacy) that have become conventionalmethods of measurement. The gate-keeping function of tests—from classroom achievement tests to large-scale standardized tests—has become anacceptable norm.

Now, just for fun, take the following quiz. All five of the words are found instandard English dictionaries, so you should be able to answer all five items easily,right?

Directions: In each of the five items below, select the definition that correctly defines the word. You have two minutes to complete this test!

1 .onager

a. a large specialized bit used in the final stages of oil well drilling

b. in cultural anthropology, an adolescent approaching puberty

c. an Asian wild ass with a broad dorsal stripe

d. a phrase or word that quantifies a noun

2. shroff

a. (Yiddish) a prayer shawl worn by Hassidic Jews

b. a fragment of an ancient manuscript

c. (Archaic) past tense form of the verb to shrive

d. a banker or money changer who evaluates coin

3. hadal

a. relating to the deepest parts of the ocean below 20,000 feet

b. one of seven stations in the Islamic hajj, or pilgrimage, to Mecca

c. a traditional Romanian folk dance performed at spring festivals

d. pertaining to Hades

4. chary

a. discreetly cautious and vigilant about dangers and risks

b. pertaining to damp, humid weather before a rainstorm

c. optimistic, positive, looking on the bright side

d. expensive beyond one's means

5. yabby

a. overly talkative, obnoxiously loquacious

b. any of various burrowing Australian crayfishes

c. a small horse-drawn carriage used in Victorian England for transporting
one or two persons

d. in clockwork mechanisms, a small latch for calibrating the correct time

Now, how did that make you feel? Probably just the same as many learners feelwhen they take multiple-choice (or shall we say multiple-guess?), timed,

"tricky"tests. To add to the torment, if this were a commercially administered standardized

test, you would probably get a score that, in your mind, demonstrates that you didworse than hundreds of people! If you're curious about how you did on the quiz,check your answers by looking at end of this lesson.

Of course, this little quiz on obscure, infrequently used English words is not anappropriate example of classroom-based achievement testing, nor is it intended tobe. It was designed to be overly difficult, to offer you no opportunity to use contextualclues, and to give you little chance of deciphering the words from yourknowledge of English. It's simply an illustration of how tests make us feel much ofthe time.

Here's the bottom line: Tests need not be degrading or threatening to your students.

Can they build a person's confidence and become learning experiences? Canthey become an integral part of a student's ongoing classroom development? Canthey bring out the best in students? The answer is yes. That's mostly what this module is about: helping you as a teacher create more authentic, intrinsically motivatingassessment procedures that are appropriate for their context and designed to offerconstructive feedback to your students.

To reach this goal, it's important to understand some basic concepts: What do we mean by assessment! What is the difference between assessment and a test? And howdo various categories of assessments and tests fit into the teaching-learning process?

Answers to the analogies quiz: 1. c, 2. d, 3. a, 4. a, 5. B

BRIEF HISTORY OF STANDARDIZED TESTING

We can trace the roots of the standardizing testing back to ancient China where, government jobs were assigned accordingto thescores oftests, mainly withquestions aboutConfucian philosophyand poetry.

This Imperial examination system, formally started during the Sui Dynasty in 605 B.C., consisted of military strategy, civil law, revenue and taxation, agriculture, geography, and the philosophical works of Confucius and his disciples. They designed tests for various levels of assignments within the government. Each degree required a more specific content knowledge.   In Ancient Greece, Socrates had used a system where he tested his students through dialogue. There were not necessarily right or wrong answers but creation of more dialogue and a higher knowledge was the purpose. Essay type tests were favored over multiple choice methods.

1838 American educators begin articulating ideas that would soon be translated into the formal assessment of student achievement.1840 to 1875 establishes several main currents in the history of American educational testing including formal written testing begins to replace oral examinations administered by teachers and schools at roughly the same time as schools changed their mission from servicing the elite to educating the masses. In 1900 College Entrance Examination Board is established and in 1901, the first examinations were administered around the country in nine subjects. By 1930 multiple-choice tests are firmly entrenched in the schools. Not surprisingly, the rapid spread of multiple choice tests kindled debate about their drawbacks. Critics accused them of encouraging memorization and guessing, of representing "reactionary ideals" of instruction, but to no avail. Efficiency and ''objectivity' won out.

With the Industrial Revolution, children were led to a more formalized schooling system where they were forced to study a set curriculum. Increasing number of students in schools created a need for a more practical systems

and standardized testing became a standard practice. This system further led to creation of SAT in 1926 as the Scholastic Aptitude Test, and then later ACT in 1956 (American College Testing) as alternatives and competitors.

A good language test should have a positive effect on learning and teaching. Such a test should aim at specifying areas of difficulties experienced by the class or the individual students so that additional practice and corrective exercises can be given.

A good language test should also measure students' performance without setting "traps" for them. It should be developed well to provide an opportunity for students to show their ability to perform certain language tasks.

On the other side, the test should enable the teachers to find out which parts of the language program cause difficulty for the class. In this way, the teachers can evaluate the effectiveness of the syllabus as well as the methods and materials they are using.

Although language testing is a fundamental part of learning and teaching in school today, it is at any level a highly complex process because it must be based on theory as well as practice.

A written language test is a classic example of formal assessment where students should be aware of being tested for a reason. There are some various characteristics that should be taken into account when constructing and using a language test. These qualities should be addressed with high consideration in order for a language test to have the positive effect.

A good language TEST SHOULD BE:

1. RELIABLE:

Reliability is consistency, dependence and trust. This means that the results of a reliable test should be dependable. They should be consistent (remain stable,

should not be different when the test is used in different days). A test that is called reliable yield similar results with similar group of students took the same test under identical conditions. Thus reliability has three aspects: reliability of the test itself, reliability of the way in which it has been marked, and reliability of the way in which it has been administered. The three aspects of reliability are named: equivalence, stability and internal consistency (homogeneity).

The first aspect, equivalence, refers to the amount of agreement between two or more tests that are administered at nearly the same point in time. Equivalence is measured through administering two parallel forms of the same test to the same group. This administration of the parallel forms occurs at the same time or following some time delay. The second aspect of reliability, stability, is said to occur when similar scores are obtained with repeated testing with the same group of respondents. In other words, the scores are consistent from one time to the next. Stability is assessed through administering the same test to the same individuals under the same conditions after some period of time.The third and last aspect of reliability is internal consistency (or homogeneity). Internal consistency concerns the extent to which items on the test are measuring the same thing. There are three factors affect test reliability:

1.   The length of the test. Longer tests produce more reliable results than very brief quizzes. In general, the more items on a test, the more reliable it is considered to be.

2.   The administration of the test which include the classroom setting (lighting, seating arrangements, acoustics, lack of intrusive noise etc.) and how the teacher manages the test administration.

3.   Affective status of students. Test anxiety can affect students' test results.

2. VALID:

The term validity refers to whether or not the test measures what it claims to measure. On a test with high validity the items will be closely linked to the test's

intended focus. Unless a test is valid it serves no useful function.One of the most important types of validity for teachers is content validity which means that the test assesses the course content and the outcomes using formats familiar to the students. Content validity is the extent to which the selection of tasks in a test is representative of the larger set of tasks of which the test is assumed to be a sample. A test needs to be a representative sample of the teaching contents as defined and covered in the curriculum. Like reliability there are also some factors which affect the validity of test scores.

Factorsinthetest:

- Unclear directions to students to respond the test.

- Difficulty of the reading vocabulary and sentence structure.

- Too easy or too difficult test items.

- Ambiguous statements in the test items.

- Inappropriate test items for measuring a particular outcome.

- Inadequate time provided to take the test.

- Length of the test is too short.

- Test items not arranged in order of difficulty.

Factors in test administration and scoring:

- Unfair aid to individual students, who ask for help,

- Cheating by students during testing.

- Unreliable scoring of essay type answers.

- Insufficient time to complete the test.

- Adverse physical and psychological condition at the time of testing.

Factorsrelatedtostudents:

- Test anxiety of the students.

- Physical and Psychological state of the student.

3. PRACTICAL:

Practical test is the test that is developed and administered within the available time and with available resources. Based on this definition, practicality can be measured by the availability of the resources required to develop and conduct the test. Practicality refers to the economy of time, effort and money in testing. Practical test should be easy to. design, easy to administer, easy to mark and easy to interpret its results. Traditionally, test practicality has referred to whether we have the resources to deliver the test that we design.

A test is practical when it:

- isnottooexpensive,

- stays with appropriate time constraints,

- is relatively easy to administer, and

- has a scoring/evaluation procedure that is specific and time efficient…

4. Discriminate:

All assessment is based on comparison, either between one student and another, or between students as they are now and as they were earlier. An important feature of a good test is its capacity to discriminate among the performance of different students or the same student in different points in time. The extent of the need to Discrimination varies according to the purpose of the test.

5. Authentic:

Authenticity means that the language response that students give in the test is appropriate to the language of communication. The test items should be related to the usage of the target language. Other definitions of authenticity are rather similar. The Dictionary of language testing, for instance, states that "a language test is said to be authentic when it mirrors as exactly as possible the content and skills under test". It defines authenticity as "the degree to which test materials and test conditions succeed in replicating those in the target situation". Authentic tests are

an attempt to duplicate as closely as possible the circumstances of real-life situations. A growing commitment to a proficiency-based view of language learning and teaching makes authenticity in language assessment necessary.

TABLE 1. THE SAMPLE OF THE TEST'S CONSTRUCTS

The construct	Definition	Sub-constructs	Number of questions
Grammatical competence	The learner's ability to use grammatical rules accurately in relation to their needs.	Pasttense	4
		Askingwh-questions	3
Sociolinguistic competence	The learner's ability to express accurate meaning using specific functions.	Introducingthemselves	2
		Complaining	2
		Respondingto a complaint	**1**
Discourse competence	The ability to produce coherent text.	Usingpronouns	4
		Usingconjunctions	4

Types of test techniques

Formal and informal testing:

http://www2.vobs.at/ludescher/testing/types_of_test_techniques.htm

Formal testing:

v   testees are told in advance what they need to know,

v   what the criteria are for success,

v   and so on.

Informal testing

v   a homework assignment

v   questions asked during the routine give-and-take of classroom interaction

v   textbook exercises.

ELICITATION TECHNIQUES

1.   Questions and answers.

Simple questions, very often following reading, or as part of an interview; may require short or long answers:

What is the (family) relationship between David Copperfield and MrMurdstone?

2.   **True/false**.

A statement is given which is to be marked true or false. This may also be given as a question, in which case the answer is yes or no.

Addis Ababa is the capital of Egypt. Is Addis Ababa the capital of Egypt?

3.   Multiple-choice.

The question consists of a stem and a number of options (usually four), from which the testee has to select the right one.

A person who writes books is called

a) a booker.   b) an editor.   c) an author.   d) a publisher.

4.   **Gap-filling and completion.**

The testee has to complete a sentence by filling a gap or adding something. A gap may or may not be signalled by a blank or dash; the word to be inserted may or may not be given or hinted at.

They (go) to Australia in 1980. Or

They _____ to Australia in 1980. (go) Or

A _____ is someone who writes books. Or

I've seen that film, (never)

5.   Matching.

The testee is faced with two groups of words, phrases or sentences;

each item in the first group has to be linked to a different item in the second.

large   small

unhappy   many

a lot   big

little   sad

6.   Dictation.

The tester dictates a passage or set of words; the testee writes them down.

7.   Cloze.

Words are omitted from a passage at regular intervals (for example, every seventh word). Usually the first two or three lines are given with no gaps.

The family are all fine, though Leo had a bad bout of flu last week. He spent most of it lying on the sofa watching _____ when he wasn't sleeping!

His exams _____ in two weeks, so he is _____ about missing school, but has managed to _____ quite a lot in spite _____ feeling ill.

8.   Transformation.

A sentence is given; the testee has to change it according to some given instruction.

Put into the past tense: I go to school by bus.

9.   Rewriting.

A sentence is given; the testee rewrites it, incorporating a given change of expression, but preserving the basic meaning.

He came to the meeting in spite of his illness. Although ....

10. Translation.

The testee is asked to translate expressions, sentences or entire passages to or from the target language.

11. Essay.

The testee is given a topic, such as 'Childhood memories', and asked to write an essay of a specific length.

12. **Monologue.**

The testee is given a topic or question and asked to speak about it fora minute or two.

Which you will choose to use for a certain testing purpose will probably depend mainly on the following considerations:

1)   What will it tell me about the testee's knowledge? In other words, for what type of knowledge might it be a valid test?

2)   How easy is it to compose?

3)   How easy is it to administer?

4)   How easy is it to mark?

Critical study of elicitation techniques

1. Questions and answers

These can be used to test almost anything. The more 'closed' the question is (that is, the fewer the possible options for correct answers), the easier the item will be to mark. It is fairly easy to compose and grade closed-ended questions; more open, thought-provoking ones are more difficult, but may actually test better.

2. True/false

This does not directly test writing or speaking abilities: only listening or reading. It may be used to test aspects of language such as vocabulary, grammar, content of a reading or listening passage. It is fairly easy to design; it is also easy to administer, whether orally or in writing, and to mark.

3. Multiple-choice

This may be used for the same testing purposes as true/false items; it does test rather more thoroughly since it offers more optional answers and is obviously *very* easy to mark. It is administered more conveniently through writing; but note that since the reading of the question-and-options is fairly time-consuming, the process of comprehension of the actual question items may take more time and effort than the point ostensibly tested, which raises problems of validity. Another important problem is that good multiple-choice questions are surprisingly difficult to design: they often come out ambiguous, or with no clear

right answer, or with their solutions over-obvious. They are to be approached *with* caution!

4. **Gap-filling and completion**

"his usually tests grammar or vocabulary, as in the examples. It is tedious to compose, though not so difficult as multiple-choice; it is more easily administered in writing than in speech; the marking is usually simple. You may need to be aware that there is more than one possible right answer.

5. Matching

This usually tests vocabulary, and is rather awkward to administer orally: thus it is best presented written on the board or on paper, though responses may be either oral or in writing. Items can be time-consuming and difficult to compose, and again, there may be alternative 'right' answers to any particular item. Answers are fairly easily checked.

6. Dictation

This mainly tests spelling, perhaps punctuation, and, perhaps surprisingly on the face of it, listening comprehension: people can only usually write words down accurately from dictation if they understand them. It does not, however, test other writing skills or speech, and involves very little reading. It may supply some information on testees' passive knowledge of pronunciation, grammar and vocabulary. It is very easy to prepare and administer; it is relatively easy to mark, though there may be a problem deciding how much weight to attribute to different mistakes.

7. Cloze

This tests (intensive) reading, spelling, and to some extent knowledge of vocabulary and grammar. It can be adapted to 'target' specific language items, by, for example, omitting all the verbs (in which case it is not, strictly speaking, 'cloze', but rather 'gap-filling'). It is fairly easy to prepare and administer. Marking can be tricky: you may find it difficult sometimes to decide if a specific item is 'acceptable' or not.

8. Transformation

This item is relatively easy to design, administer and mark, but its validity may be suspect. It tests the ability of the testee to transform grammatical structures, which is not the same as testing grammar: a testee may perform well on transformation items without knowing the meaning of the target structure or how to use it in context. Marking is fairly straightforward.

9. Rewriting

This tests the same sort of thing as transformation, but is likely to reflect more thorough knowledge of the target items, since it involves paraphrasing the entire meaning of a sentence rather than transforming a particular item. It is, however, more difficult to compose, and the marking may be more subjective. It is, as its name suggests, usually done in writing.

10. Translation

A technique which, at the time of writing, is for various reasons rather unpopular, but in my opinion undeservedly so. In a monolingual class whose teacher also speaks the learners' mother tongue, the translation of a 'bit' of language to or from the target language can give very quick and reliable information on what the testee does or does not know, particularly when it involves entire units of meaning (phrases, sentences) within a known context. Translation items are also relatively easy to compose - even improvise, in an informal test - and administer, in either

speech or writing. Marking may sometimes be more difficult, but not prohibitively so.

11. Essay

This is a good test of general writing abilities. It is relatively easy to provide a topic and tell the class to write an essay about it but marking is extremely difficult and time-consuming. It must be clear in advance, both to you and to the students, how much emphasis you are going to lay on language forms, such as spelling, grammar, punctuation, and how much on aspects of content, such as interest and originality of ideas, effectiveness of expression, organization

12. Monologue

This tests oral fluency in 'long turns' - something not everyone can do in their mother tongue! It also tests overall knowledge of pronunciation, grammar and vocabulary. To choose a topic and allot it is not so difficult; to assess is very difficult indeed, demanding concentration and a very clear set of criteria and weighting system.

Discussion questions:

1. How can language teachers use information about their student's knowledge and skills of language before, during, and after a lesson to better support their language development?
2. Can tests become an integral part of a student's ongoing classroom development?
3. When was the first standardized tests used?
4. What skills were tested in ancient China?
5. What was the difference between Chinese and Greece testing methods?
6. What competences does language competence include?
7. What does each competence assess in language learners?

TASK.Read the text *Critical study of elicitation techniques*and find proper answers.

What difficultiesshould be taken into consideration while creating tests?

IV. АМАЛИЙ МАШҒУЛОТ МАТЕРИАЛЛАРИ

1-Амалий машғулот: **Тил билиш даражасини баҳолаш тамойиллари. Баҳолаш мезонларини яратиш ва улардан самарали фойдаланиш.**

(2 соат)
Режа:

1. Factors to consider when developing assessments
2. Practicing designing assessment criteria

Language Testing and Assessment

Aim: to familiarize participants with course description and its assessment specifications

Objectives:
- to explore participants' experience/understanding on language testing and assessment;
- to familiarize participants with Course Description, Assessment Profile;
- to explore different assessment modes, assessment criteria and their importance in language teaching.

Time: 80 minutes

Handouts: Handout 1. Course description/assessment specifications, Handout 2. True/False statements, Handout 3. Assessment types, Handout 4. Assessment criteria

Activity 1 (5 min) Ask participants the following questions (if possible write key words on the white board):

- What is *testing*?
- What is *assessment*?
- What is evaluation?

Elicit answers from the Ps and sum up:

EVALUATION – thinking carefully about something before making

judgment about its value, importance or quality

ASSESSMENT – the evaluation of a student's achievement or

proficiency; a process which involves gathering and

documenting information about the achievement,

knowledge, skills and abilities of an individual.

TESTING -   an examination of somebody's knowledge or ability, consisting

of questions for them to answer or activities for them to carry

out. (Macmillan English Dictionary for Advanced Learners)

Tell Ps <u>aim</u> and <u>objective</u> of the session.

Invite groups to plenary discussion.

Activity 2. *(15 min)* Handout 1a. The course description

- Give the Language Testing and Assessment course description to participants (Handout 1a) and 2 minutes to read it (individually);
- Ask the Ps to read the course description and do the task (Handout 1b)

Give Handout 1b and ask Ps to do True /false statements

(key: A) F; B) T; C) F; D) F; E) T; F) T; G) F)

Activity 3. (10 min.) Handout 2. Assessment profile and specifications

Draw Ps' attention to assessment profile and specification, and ask them to discuss the following questions:
- How many types of assessment are there?

- What is Analytical Report? How many words do analytical report should contain? Do students develop new tests to analyze? Please comment your answer.

- What are the assessment criteria of this task?

- What do students do in project work?

Activity 4 (25 min) An extract from a lesson

Tell participants to take the role students for 20 minutes. Tell them that you are going to present an extract from a lesson. Ask them be active during the lesson.

Sample lesson

- Tell that Spolsky(1984) has referred to three generations in the evolution of language testing. Ask participants to match three generations to the brief definitions.
- Ask them to explain their choice in the plenary.

Generations	Definitions
a) Traditional	3. Tests linked to the grammar-translation method of teaching; items test language that is unrelated to any social context or use.
b) Modern	1. Tests linked to the behaviourist-structural method of teaching; stimulus-response items test isolated pieces of language
c) Post modern	2. Tests linked to the communicative method of teaching; test items seek to replicate the psycho-social nature of authentic language use..

- Ask participants look at the test items and say whether they are first, second or third generation items.

1. The oral interview	1st
2. The written composition, eg 'a School Speech Day'	1st
3.a multiple-choice item, eg the opposing of strong is: a)short b)poor c)weak d)good	2nd
4.a blank-completion item, eg: The man who flies an aeroplane is a _____.	2nd
5.picture description, two pictures are given and a student is asked to describe two pictures and answer the given question:	3rd

Tell that for each skill, three sample items are given. Tell participant to work in groups of three to arrange the sample chronologically by stating whether they are 1st, 2nd, or 3rd generation.

Reading	
A passage from an encyclopedia follow by series of multiple-choice questions.	2
A passage from a literary source followed by open-ended questions.	1
An authentic text from a financial; newspaper accompanied by a blank pie-chart which must be labeled with data from the text.	3
Writing	
Replying to an input letter	3
Writing 200 words about a visit to a zoo.	1
Completing a passage from which words have been deleted.	2
Listening	
An extract from a written text is read aloud; students then reproduce the text in their own words.	1
A specially-written dialogue is heard on tape; students then answer multiple-choice questions.	2
A tape of a railway station announcement is played; students complete a table with details of the train they want to catch.	3
Speaking	
Two students discuss a practical problem and then report their solution to the assessor.	3
The student reads a series of sentences aloud; the examiner gives marks for each minimal pair (ship/sheep) which is pronounced correctly and distinctly.	2
The student is interviewed by the examiner, who asks a series of unrelated questions.	1

- Tell that this is the end of a sample lesson. Ask participants to answer the following questions:

What is the topic of the lesson? (evaluation of language testing)

What objective is put on the lesson? Make students familiar with the evaluation of language testing.

Which of the following learning outcomes are expected?

- an ability to critically evaluate assessment tools (+)
- an ability to design and use limited range of assessment tools

Which of the following teaching approaches are used during the lesson?.

- Mini-lectures
- Workshop sessions (+)
- Self study
- Case studies

Analysis of assessment tools

Activity 5   **<u>Specifying topics and writing objectives with outcomes</u>**

Time: 25 min

Materials: handout 4

Procedure:

- Distribute handout 4, ask participants to work in groups and read the topics written in the first column
- Ask participants by referring to the example fill in the table with possible subtopics, objectives and outcomes for each topic.
- Say that answers of each group can vary.
- Once they have finished, elicit random answers in the plenary.

Suggested answers:

Topics	Possible Subtopics	Objectives	Outcomes
Assessment	eg. Summative assessment	eg. By the end of the module students will be aware of test types which take place at the end of language course	eg. Students will be able to design and use objective tests to test learners progress at the end of the course
Testing methods	eg. Objective testing (testing reading and listening) Subjective testing (testing writing and speaking)	eg. By the end of the lesson students will demonstrate the ability to work with the functions and applications of subjective and objective test methods	eg Students will be able to design and apply objective and subjective tests
How to assess language skills and aspects	eg. Testing reading	eg. By the end of the lesson students will be familiarized with the principles of testing reading	eg Students will be able to construct and administer reading tests appropriately.
Types of tests	eg. Test types (proficiency, progress, diagnostic, placement, achievement)	eg. By the end of the lesson students will be familiarized with the aims and principles of different types of language tests	eg Students will be able to design , evaluate and use different test types

Alternative ways of assessment	eg. Project works, portfolio assessment, Self-assessment and ect.	eg. By the end of the lesson students will be able to analyse the appropriateness of alternative ways of assessment to certain teaching contexts	eg Students will be able to creatively design different alternative tasks and critically evaluate them.

HANDOUTS

Activity 2, Handout 1a

Language Testing and Assessment

Introduction

Compulsory, Year 4, 44 hours, Semester 7 and 8.

Aims:

- to acquaint students with the theory and practice of assessment and testing
- to develop students' ability to differentiate between the purpose and focus of assessment

Objectives

By the end of the course students will:

- be aware of basic principles of language assessment
- be aware of different types of language tests
- be able to analyse the appropriateness of alternative ways of assessment to certain teaching contexts
- be aware of different types of continuous assessment
- be able to design assessment tools and plan assessment procedures

Indicative content

- Assessment:
    - summative; formative;
- Testing:
    - Evaluation of language testing
    - validity, reliability, practicality
    - Criterion referencing; norm referencing
    - Specification
    - Test types (proficiency, achievement, progress, etc.)
- CEFR (Common European Framework of Reference)
- Test methods (mcq., multiple matching, etc.); Basics of test design (instruction, distracters, etc.)
- How to test/ assess
    - Writing
    - Reading
    - Listening
    - Speaking
    - Integrated skills
    - Vocabulary
    - Grammar
- Alternative ways of assessment (portfolio assessment, self-assessment, project work, logs, reflection tools, etc.)
- Planning assessment
- Washback (influence of assessment on teaching)

Approaches to teaching and learning

- Mini-lectures
- Workshop sessions
- Self study

- Case studies
- Analysis of assessment tools

Learning outcomes

Students should have developed:

- an ability to critically evaluate assessment tools
- an ability to design and use limited range of assessment tools

Semester 7

Continuous Assessment	40%
• Participation	10%
• Test analysis (Analytical Report see specification)	30%
Mid course assessment	
• Test analysis	**30%**
Final assessment	
• Project work	**30%**

Activity 2, Handout 1b. True/False statements

A) This course is taught in semesters 5-6.

B) Students will be able to analyze, develop and plan assessment procedure.

C) Only theory of testing is introduced in this course.

D) Students will analyse and develop only multiple-choice questions.

E) Students will know about different test types.

F) Portfolios, logs, and project works are considered as an alternative ways of assessment

G) Students will only discuss how language skills are assessed. **Activity 3,**

Handout 2

<p align="center">**Language Testing and Assessment**</p>

<p align="center">**Assessment Specification**</p>

Test analysis

Analytical Report: students analyze language tests from previous language courses (Year1, Year 2) and write analytical report (max 350-550 words). The language test analysis should cover at least two of the following aspects: focus, instructions, face validity, construct validity, reliability, practicality, test methods.

Assessment criteria: 40%

- Task fulfillment   10%
- Awareness of principles for analysing the given aspects 10%
- Awareness of basics of test development   10%
- Participation   10%

Test analysis

Students develop assessment criteria for evaluating one of the alternative ways of assessment such as portfolios, logs, essays, reflective writing, diaries, project work, etc. students can use first or second year students' portfolios, written works, etc for completing this task.

Assessment criteria:30%

- Task fulfilment 10%
- Appropriateness of criteria   10%
- Appropriateness of assessment tools (does he/she choose appropriate tool to create criteria)   10%

Project work specification

Students, in groups of 4 people, develop a progress test with assessment criteriato assess(including at least 2 test methods: multiple choice, banked/gap filling, multiple matching):

- Group 1 – grammar + vocabulary
- Group 2 – reading
- Group 3 – listening
- Group 4 –writing
- Group 5 – speaking

Test specification also should be provided by each group.
(Note: Ss may use textbooks that are available at the secondary schools/colleges, and choose a unit)

Assessment criteria 30%

1. task fulfilment   6%
2. coverage of the focus (language skill/language area) 6%
3. appropriate use of different test methods   6%
4. design, layout, and instruction of the test   6%
5. appropriateness of assessment criteria and specification 6%

Activity 4, Handout 3a

Sample lesson

Match three generation of testing to the brief definitions. Explain your choice in the plenary.

Generations	Definitions

a) Traditional	1. Tests linked to the behaviourist-structural method of teaching; stimulus-response items test isolated pieces of language.
b) Modern	2. Tests linked to the communicative method of teaching; test items seek to replicate the psycho-social nature of authentic language use.
c) Post modern	3. Tests linked to the grammar-translation method of teaching; items test language that is unrelated to any social context or use.

Activity 4, Handout 3b

Look at the test items and say whether they are first, second or third generation items.

3. The oral interview	
4. The written composition, eg 'a School Speech Day'	
3.a multiple-choice item, eg the opposing of strong is: a)*short* b)*poor* c)*weak* d)*good*	
4.a blank-completion item, eg: The man who flies an aeroplane is a _____.	
5.picture description, two pictures are given and a student is asked to describe two pictures and answer the given question:	

Activity 4, Handout 3c

For each skill, three sample items are given. Work in groups of three to arrange the sample chronologically by stating whether they are 1st, 2nd, or 3rd generation.

Reading	
A passage from an encyclopedia follow by series of multiple-choice questions.	
A passage from a literary source followed by open-ended questions.	
An authentic text from a financial; newspaper accompanied by a blank pie-chart which must be labeled with data from the text.	

Writing	
Replying to an input letter	
Writing 200 words about a visit to a zoo.	
Completing a passage from which words have been deleted.	

Listening	
An extract from a written text is read aloud; students then reproduce the text in their own words.	
A specially-written dialogue is heard on tape; students then answer multiple-choice questions.	
A tape of a railway station announcement is played; students complete a table with details of the train they want to catch.	

Speaking	
Two students discuss a practical problem and then report their solution to the assessor.	
The student reads a series of sentences aloud; the examiner gives marks for each minimal pair (ship/sheep) which is pronounced correctly and distinctly.	
The student is interviewed by the examiner, who asks a series of unrelated questions.	

Activity 4, Handout 3d

Answer the following questions:

What is the topic of the lesson?

What objective is put on the lesson?

Which of the following learning outcomes are expected:

- an ability to critically evaluate assessment tools
- an ability to design and use limited range of assessment tools?

Which of the following teaching approaches are used during the lesson?.

- Mini-lectures

- Workshop sessions

- Self study

- Case studies

- Analysis of assessment tools

Activity 5, Handout 4

Topics	Subtopics	Objectives	Outcomes
Assessment	eg. Summative assessment	eg. By the end of the lesson students will be aware of test types which take place at the end of language course	eg. Students will be able to design and use objective tests to test learners progress at the end of the course
Testing methods	eg.	eg.	eg
How to assess language **skills and aspects**	eg	eg.	eg

Test types	eg.	eg.	eg
Alternative ways of assessment	eg.	eg.	eg

Режа:

1. Analysing what is task based language assessment
2. Four main features characterize TBLA.

WHAT IS TASK-BASED LANGUAGE ASSESSMENT,OR TBLA?

TBLA is a framework for language testing / assessment that takes the task as the fundamentalunit for assessment and testing. It is based on the same underlying principles as TBLT, but extends them from the learning-and-teaching domain to the testing domain. Specifically, as in TBLT methodology, testing / assessment in TBLA is also organized around tasks rather than in terms of grammar or vocabulary. For instance, Long and Norris (2000: 600) state that "genuinely task-based language assessment takes the task itself as the fundamental unit of analysis, motivating item selection, test instrument construction and the rating of task performance." As in TBLT methodology too, the main goal and validity of TBLA is measured against the extent to which it can successfully achieve a close link between the testee's performance during the test and his performance in the real world. For instance, Ellis (2003: 279) states that "task-based testing is seen as a way of achieving a close correlation between the test performance, i.e. what the testee does during the test, and the criterion of performance, i.e. what the testee has to do in the real world." Assessment tasks are thus viewed as "devices for eliciting

and evaluating communicative performances from learners in the context of language use that is meaning-focused and directed towards some specific goal" (ibid.).

Four main features characterize TBLA.

First, it is a formative assessment; that is, it is an assessment undertaken as part of an instructional program for the purpose of improving learning and teaching. Second, it is a performance-referenced assessment; that is, it is an assessment that seeks to provide information about learners' abilities to use the language inspecific contexts, that is directed at assessing a particular performance of learners, and thatseeks to ascertain whether learners can use the L2 to accomplish real target tasks. Third,itis a direct assessment; that is, it is an assessment that involves a measurement of language abilities that involves tasks where the measure of the testee's performance is incorporated into the task itself, like information-transfer test tasks such as information-gap, opinion-gap,and reasoning-gap tasks. It must be noted, however, that direct assessment still involvessome level of inferring because it is necessary to observe performance and then infer abilityfrom that performance. Put differently, you can measure outcomes, but you are still left withinferring the ability that produced the outcome. Fourth, it is an authentic assessment; thatis, it is an assessment that involves either real-world language use (or as close as possibleto this), or the kinds of language processing found in real-world language use, that is, thetest task's characteristics must match those of the target-language task (Ellis, 2003: 285).

COMPONENTS OF TASK-BASED LANGUAGE ASSESSMENT. Task-based language assessments consist of three basic components: a test task, an imple-mentation procedure, and a performance measure: TEST TASK DESIGN AND SELECTION. It is possible to identify two approaches to test task design and selection within the frame-work of TBLA: The first is called the construct-centered approach or the direct system-referenced tests. This approach entails the identification of a theory of language learning and language use as a basis for the

test task design and selection. This approach is used with tasks that seek "to establish the general nature of the testees' language proficiency"(Ellis, 2003: 286). The second approach is called the work-sample approach or the direct performance-referenced tests. This approach involves analysis of the target situation todetermine what tasks the testee will need to perform in the real world. This approach isused with tasks that seek "to find out what a learner can do in a particular situation" (Ellis,2003: 286). It should be noted, however, that the two approaches must be seen as com-plementary rather than mutually exclusive because any successful TBLA ideally requiresboth approaches.

IMPLEMENTATION PROCEDURES

There are two implementation procedures: The first is planning time. Planning time must be adopted as a key implementation procedure because it can improve the test taker'sperformance. The second procedure is the interlocutor (on oral test tasks). The role of the addressee (familiar or unfamiliar, native speaker or nonnative speaker) plays a significantrole in the performance of the testee on an assessment task. For instance, some researchers have shown that if you want to elicit the "best performance" from the testee, it may be preferable to set up a testing situation where the candidates interact with another nonnativespeaker than with a native speaker (e.g., Wigglesworth, 2001). Along the same lines, onewonders whether it is also possible to have real recipients for writing tasks, analogous to the interlocutor in oral tasks.

MEASURING PERFORMANCE

How do we assess the performance elicited from the testee in TBLA tests? Two principal methods for measuring learner performance are used in TBLA contexts. The first is the direct assessment of task outcomes. This method involves either the assessor observing a performance of a task and making a judgment, or no judgment on the part of the assessor(i.e., judgment / measurement is objective – the testee did or did not succeed in performingthe task) because it results in solutions that are either right or wrong. A good exampleof direct assessment of task

outcomes is a closed task that results in a solution that iseither right or wrong, like locating a journal article relating to a particular topic in a library(Robinson and Ross, 1996: 474–475). The main advantages of this method are that "it affords an objective measurement, involving no judgment on the part of the assessor, and it is easy and quick" (Ellis, 2003: 296).The second performance measure is external ratings. This method involves external judgment which is more subjective (but efforts are being made to make it more reliable).This measure also involves a holistic measure (scale) or an analytic measure (scale) of linguistic ability (e.g., paragraph or essay evaluation; overall linguistic ability versus the four language skills identified for rating separately). Examples of this assessment include judging that the learner's speaking ability is at the "expert" level in the oral interviewcomponent of the old ELTS test which specifies that the learner "can speak with authorityon a variety of topics; can initiate, expand, and develop a theme" (Ellis, 2003: 300).The main advantage of this method is that it enables the assessor to specify the learner's.

Task-Based Language Assessment competencies to be measured in more functional terms. Competencies in external ratings are determined in terms of learner performance levels or checklists. Along the same lines, Long and Norris, bearing in mind language professionals and classroom teachers, specify six practical steps for developing and implementing task-based language tests as follows:

1. The intended use(s) for task-based assessment within the language programme must be specified, minimally addressing the following four issues:

   - who uses information from the assessment?

   - What information is the assessment supposed to provide?

   - What are the purposes for the assessment?

   - who or what is affected, and what are the consequences of the assessment?

2. Target tasks or task-types emerging from the needs analysis are analyzed and classified according to a variety of task features. Analysis is undertakenin order to understand exactly what real-world conditions are associated withtarget tasks and should therefore be replicated under assessment conditions.

3. Based on information from the analysis of task features, test and itemspecifications are developed. Specifications delineate the formats tests should take, procedures involved, tasks or task-types to be sampled, format for test tasks (items), and how performance on the task-based test should be evaluated.

4. [Carrying out] identification and specification of rating criteria, which form the basis for interpretations of examinee performance and task accomplishment...Real-world criterial elements(aspects of task performance that will be evaluated) and levels (descriptions of what success looks like on these aspects of task performance) should be identified within initial needs analysis, with a view toward providing students and teachers with clear learningobjectives.

5. Task items, test instruments and procedures and rating criteria need to be evaluated (involving pilot-testing and revision) according to their efficiency,appropriacy and effectiveness with respect to the intended assessment uses.

6. Finally, task-based language assessment should incorporate procedures for systematic and ongoing validation of its intended use within the language programme.

3-4 АМАЛИЙ МАШҒУЛОТ.

Фидбек бериш. Ўз ўзини баҳолаш. Баҳолашнинг асосий турлари. Мавзуга оид олимлар фикрлари таҳлили. Портфолио тузиш. CEFR тизими бўйича баҳолаш. (4 COAT)

Plan:

1. **Distinction between Assessment for Learning and Assessment of Learning;**

2. **Principles of language assessment;**

3. **How to apply key principles of assessment to the quality of assessment tools;**

4. **How to reflect on the challenges and identify ways to address them.**

Key words: Assessment of Learning; norm-referenced tests; criterion-referenced tests; key principles of assessment; validity; reliability; practicality; washback; construct validity; content validity; face validity; types of assessment; placement tests; diagnostic tests; progress tests; achievement tests; aptitude tests; proficiency tests; testing specifications; and, piloting.

INTRODUCTION AND OVERVIEW

Assessment is an inseparable part of teaching because language teachers have a dual role of teaching and assessing, which ultimately will have an impact on students' learning and motivation. They make decisions about who could pass or fail a quiz, test, or course of study; they determine whether the class is going well, and if the teaching they do is effective (Harding &Kremmel, 2016). Assessment of Learning measures theseconcerns. However, it is often the case that assessment tools used by language teachers fail to provide the measurement of what is targeted. Such misuse causes unfair judgement of students' knowledge and preparedness.

Think about the following:

1) How do you think Assessment for Learning differs from Assessment of Learning?

2) What do you know about assessment principles?

3) How can teachers make sure the way they assess students' knowledge is fair?

4) How can you know whether your assessment is measuring what you want to measure?

UZBEK VIGNETTE

Case 1

I designed a test for the required level by state standards. Students showed very low results. I was blamed for designing poor quality assessment tools but what if the students are simply not at the expected level. How can I know that? Another challenge I have been facing again and again is the connection between teaching and assessment. It is quite o ften that what I teach during the semester isnot assessed at tests. The problem is that every teacher uses their own approach in teaching but in the end the final test is the same for all the students across the groups. As a result, students learn one thing but are assessed for another.

Case 2

"I have been responsible for designing tests in my department for several years. Every time I find it hard to make decisions and choices regarding what tasks to set in an assessment tool. I feel that I need to rely on some principles, however, what are they? Everything I do is usually based on my own intuitions as a teacher and the instructions by the head of the department. I collect tests from various sources, including internet and textbooks. I put all the parts together and the test is ready! But when I and my colleagues administer the test, it appears to be too easy or too hard, sometimes the designed test takes longer time than it was planned to take".

Case 3

A teacher in my department believes that test should not be designed by teachers themselves as they do not possess sufficient skills to do so. She notes that there should be a special person responsible for this task in each department. Developing a test requires much preparation and time, therefore, teachers who are already overloaded cannot spend sufficient time for this. However, my other

colleague, thinks that is a responsibility of every teacher to be involved in andconcerned with their students' assessment. This process is part of teaching.

REFLECTION

Think about the cases above. Are the issues raised in it familiar to you in your own teaching context? What do you think can be possible solutions for each case?

KEY CONCEPTS

There are eighteen key concepts in this section: Assessment of Learning; normreferenced tests; criterion-referenced tests; key principles of assessment; validity; reliability; practicality; washback; construct validity; content validity; face validity; types of assessment; placement tests; diagnostic tests; progress tests; achievement tests; aptitude tests; proficiency tests; testing specifications; and, piloting. We will briefly explain each one below.

Assessment of Learning – also known as summative assessment, is usually used at the end of a course of study. The intention is to summarise, to see if achievements are at the expected level. In this case, further learning is usually not considered. Knight and Yorke (2003) note that "because summative assessments tend to come at the end of learning sequence, it can be difficult to use them to give feedback to help learners to do better next time" (p. 16). In this case, the challenge for teachers can be developing an understanding of students that summative assessments such as midterms or final tests are not the end of the road but just the accomplishments of a certain phase. Students should be able to look at this assessment as a learning opportunity as well. Norm-referencedand Criteria-referenced assessment. There are two approaches in language assessment. Norm-referenced assessment is based on the comparison of all test-takers within a continuum. In other words, each test -taker is compared to how well (or poor) other test-takers did on the same exam. Oftentimes youwill be

notified you're your score is in the "X" percentile, or, you have done better than "X" number of test-takers who took the same test. The National Entrance Examination in Uzbekistan uses this approach for admission to higher educational institutions. The other approach – criterion-referenced assessment – looks at whether each candidate meets the requirements for a certain level or position based on characteristics, i.e. descriptions provided in criteria. An example for this type of scoring in Uzbekistan is the PRESETT

Exit Test. Another example is the more widely-known tests such CEFR level certification, which is created and implemented by the National Testing Center. Key Principles of Assessment: The notion a "good test" (Madsen, 1993, p. 178), implies that an assessment tool provides accurate and reliable information about learners' knowledge and skills. To make sure the selected or developed assessment tool is fair and successful, it should meet four criteria (we have provided guiding questions to define each term):

Validity – does a test measure what it is trying to measure?

Reliability – does the test produce consistent results?

Practicality – does the test take an adequate amount of time, money, energy and resources to make?

Impact – is there a positive effect on the teaching and learning with the distribution of this test?

These four principles work together and are sometimes contradictory. For instance, multiple choiceassessments, which are very practical to distribute and score, are often the most notoriously difficult to write. Thus, the closer you become to establishing all four principles in your assessment practices, the more you learn about your students' abilities, and will thus, be able to make the necessary choices about how to design your assessment and the appropriate feedback to give them. Knowing and understanding these principles are not only

important for those who design their own tests, but also for those who selects an assessment tool from existing ones. Inappropriate selection from well - constructed commercially available tests for a certain purpose of assessment can cause as much harm as developing one from a scratch. We discuss the four areas in detail below.Selecting or developing an assessment tool is a very responsible task for a teacher, as the information obtained through it will be used for making decisions. The administration of the test is as important as the test construction itself. A teacher should consider a fair conduction of assessment of any kind, whether it is a mere classroom check or high-stake official examination. Unfavourable assessment condition can influence the results.

Validity is critical for assessment. Thisquality signifies whether the intended purpose of assessment has been met and is reflected in the created assessment tool. In other words, by checking the validity of a certain test or task, we learn whether it provides the evidence about test-takers that we intended to obtain. Validity also influences how the assessment results are interpreted by teachers and delivered to stakeholders. It is, therefore, vital that we know what the purpose of a certain assessment tool is before checking whether it is valid or not. There three types of validity from Cumming&. Berwick (1996), which are commonly referred to in the language testing literature:

  ➢ Construct validity: 'a mutual verification of the measuring instrument and the theory of the construct it is meant to measure', p.29; 'we examine the psychological trait, or construct, presumed to be measured by the test and we cause a continuing, research interplay to take place between the scores earned on the test and the theory underlying the construct' (p. 26)
  ➢ Content validity: 'review of the test by subject-matter experts and a verification
  ➢ that its content represents a satisfactory sampling of the domain' (p. 22)

➢ Face validity: 'the appearance of validity…for example that the language and contexts of test items be expressed in ways that would look valid and be acceptable to the test taker and to the public generally' (pp. 23 - 24)

Reliability. A reliable assessment tool generates consistent results even if administered on several occasions. However, several factors must be taken into consideration when judging the reliability of a certain test. Brown and Abeywickrama (2010) highlight the main four: students, the scoring, the test administration and the assessment tool. It is possible that the same student can perform differently from a previous time due to physical conditions (e.g. fatigue, illness) or psychological (e.g.stress) (p. 27-28). There are also issues such as inter-rater and intra-rater reliability (these will be discussed in more detail in Section 3.4). Sufficient attention should be paid to develop reliability in terms of discrepancy in different raters' marking and the consistency of the same rater across students.

Impact. It is undeniable that any kind of assessment has a certain impact. This is also referred as washback or, sometimes in the literature as backwash. Such impact can be intentional or unintentional. McNamara (2000) differentiates washback from impact by referring to the formers as "the influence that testing has on teaching" and the latter as "the effect of a test beyond the classroom, the ripples or waves it makes in the wider educational and social world" (p. 72). Overall, washback can be positive or negative, narrow or broad, intended or unintended. Washback happens more with high stakes -tests than low-stakes tests.

Practicality. An assessment tool can be considered practical if it uses fits an available budget, does not exceed the optimal time constraints and easy in administration as well as scoring. In other words, a practical test is well-considered logistically. Therefore, departments responsible for hundreds of **students should make thorough considerations in planning such assessments.**

Types of Assessment. Below are types of assessment and the description of their main purposes:

Table 12.Types of Assessments.

Assessment types Main purpose

Placement test places students at an appropriate level of instructions within a program

Diagnostic test identifies students' strengths and weaknesses

Progress test provides information about mastery or difficulty that learners have with course materials

Achievement test Provides information about students' attainment of learning outcomes at the end of a course

Aptitude test measures a general ability to learn a foreign language

Proficiency test measures global competence in a language

REFLECTION

Which of the above are the most common type of tests in your teaching context? Which have you never used? Why?Test Specifications. Before designing any kind of assessment, a test developer must prepare test specifications which contain all the information about the test to be constructed. This blueprint is an informative document and also is also kept for preparing further variants of the same kinds of test in the future. Developed once, test specificationscan serve for a long time and be revised when needed (Similar to the design andimplementation of classroom lesson plans.) Another important feature of test specifications is they provide instructions to other specialists who will use them. It means they should be as clear and detailed as possible so that anyone other than the developer can use it for test creation. In

essence, test specifications can be different in content and form, as different developers prefer to focus on different aspects, however, this is not an issue as long as the document provides sufficient information need to know about the test.

There are seven specific questions all test specifications should answer:

1) What is the purpose of the test?

2) What sort of learners will be taking the test?

3) Whatlanguage skills should be tested (reading, writing, speaking and/or listening)?

4) What language elements should be tested (grammar, vocabulary, pronunciation, speech acts, etc.)?

5) What target language situation is envisaged for the test, and is this to be simulated in some way in the test content and method? (For instance, is this a test of academic French? Of English for international TAs? Of Japanese for hotel workers?)

6) What text types should be chosen as stimulus materials -- written and/or spoken?

7) What sort of tasks are required -- discrete point, integrative, simulated 'authentic', objectively assessable? (That is, what will the test-takers do?)

Piloting. The best way to find out whether the created assessment tool works well is piloting. A test should not be administered without prior piloting. Such trialing can inform us about weaknesses the test might possess. Using a test without piloting may bring about serious consequences such as providing the wrong evidence about learners' knowledge. Therefore, it is the department's and particularly the test developers' responsibility to strictly comply with the right procedures of the creation and use of assessments. The characteristics of

the group which will be used for piloting a test should be similar to those of the target population for the test itself. A teacher that piloted even one assessment tool will realize how useful this can be. it can reveal the issue s and the features of the test that the test developers otherwise might never be able to spotthemselves. Thus, the evidence gathered from piloting is used to make the necessary changes and the whole process recommences.

ACTION

Read the following reflection by an Uzbek language teacher who has learnt the main principles of language assessment and read extensively the literature. Please highlight the decisions and aims seen by the teacher and pay attention to how each principle of language assessment is reflected.

My teaching context is undergraduate students, which I taught the four skills of listening, reading, writing and speaking for Practical English. Apart from formative assessment that they have during the semester, they take one midterm and one final exam in each of the two terms in the academic year. Therefore, I usually deal with achievement tests based on the syllabus content. It means that my tests need to reflect what students have presumably learnt and assess their knowledge of the skills that they have developed through the course. In this endeavor, my concern has always been how to develop language tests used for classroom assessment in a professional way. Having learnt the significance of different test qualities, I find it important to bear in mind that "a test cannot be valid for more than one purpose" (Henning, 1987, p. 89), which I used to ignore without realizing how inappropriate a certain test can be. I made no serious attempt to check the validity and reliability of the tests that I created or selected for assessment. All my decisions have been based on my own intuitions and the instructions from the administration and the head of my department. However, what I am inclined to do as of now is to make validity and reliability central in the development of my assessment tools. My aim from this is to make sure that

students are tested for what they are intended to be tested and the tasks across all the variants of tests are of equal difficulty. Usually my department administers midterm and final tests in several slots within the same day by grouping students, as there are over 400 students to be tested. As a result, my task as a test developer is to create several variants for every slot. At the same time, I have to make sure that ever y variant is no different from another in terms of difficulty and other characteristics. Here my responsibility is to construct equivalent tests, or what Weir calls "parallel forms" (2005, p. 250) so no matter what variant is given to a certain student, he/she should be able to perform with the same results, which is the concern of reliability. An important consideration is checking a testing instrument by piloting. I have been through experiences when my test turned out to be too easy or too difficult. I was usually given feedback from other teachers in my department after the administration of the test. I realize here that what I lacked at that point is piloting my testinginstruments, which means that no matter how much effort is made inconsiderations with regard to main principles of an effective and useful test, it is impossible to know how good the test is unless it is piloted.My second aim is to work with other teachers in my department, particularly with those who are also responsible for creating assessment tools. Usually when a test needs to be constructed, teachers search for ready tests that can be applicable for their purpose. This means that test selection is as important as test designing.

Nevertheless, even when selecting a test instrument from a variety of sources that are available on the internet or in printed materials, it is important to ask whether the purpose of the chosen test is the same with that we want to use it for. "Too often the content of tests is determined by what is easy to test rather than what is important to test" (Hughes, 2003, p. 23). It is a serious disadvantage if all the testing instruments that are created or selected from available sources can be random and without considerations concerning whether they are valid, reliable, and practical or what impact they might have on test takers.

In this respect, Davies (1997) maintains that professional ethics and morality in language testing is important, which in my context is rarely discussed. It seems very necessary to "develop assessment literacy" (Taylor, 2009, p. 31) so that teachers could approach their task with a better understanding, not relying exclusively on their intuition. Only by consulting relevant theories and practices followed by collaboration and consultatio n amongst each other can teachers learn how to approach test development professionally and avoid serious mistakes.

It is essential to continuously research and collect data in one's own context to provide better testing conditions for test takers. A comp rehensive discussion of

the responsibilities of testers and the rights of test takers in Shohamy (2001) makes me realize how complex is the process of language assessment and how much thought every decision should be given to avoid any unfairness that can take place. Most students are usually scared of tests and this experience and causes then much stress. The impact of assessment is very important in my context, as this often has a great influence of decisions made concerning every student. The results of tests and, assessment on the whole, influences the students not only academically but also financially, as depending on the marks received, they become eligible for a certain amount of monthly student allowance paid by the government. Understandably, every student strives to be successful at the tests he/she sits. Such circumstances make test developers particularly attentive to the impact from the tests they create.

I also aim to balance required with available. In my department, practicality is the main concern when it comes to testing. The decisions made concerning test constructs are usually influenced by trying to decrease teacher workload. Therefore, oral and written tasks are used much less compared to those that can be checked faster and easier, such as multiple choice items in reading and listening. My position here is that even though it is important to take into account practicality aspect, students must not be deprived of being tested for what they are taught,

which, from my observation, quite often happens in my department. When developing achievement tests I find it important to align the assessment with the syllabus.

TASK

Having read the reflection and an action plan given above, consider what changes would you like in the way that you have been approaching assessment in your teaching practice. Prepare a short vignette focusing on a certain issue in your context, followed by a plausible action plan in which you need to suggest improvements and support your arguments with the literature (e.g. books, articles, websites, etc.) that you read.

SUMMARY

In conclusion, language assessment in education is a great responsibility for the teacher because assessments can influence decisions about students and possibly their futures. Therefore, teachers should reflect each step in the development assessment instruments. Without understanding validity, reliability, practicality as well as impact (washback), it is difficult to come up with an effective language assessment. The adherence to these principles is vital. At the same time, there is no universal formula of creating an effective language assessment tool. The same test can be the best and the worst assessment instrument depending on where and when it is used. Therefore, the main considerations are the purpose of a test and the context where it is used. Only by giving sufficient attention to these two factors can one design or choose an effective languageassessment tool, which will be appropriate just for that occasion.

HOMEWORK TASK THIRTEEN

Imagine you have completed (i.e., taught) the lesson you chose for Homework Task

One (A) and you want to make a short quiz (10 min) that will measure what you taught in the class. For this homework task, please write the answers to the seven questions thatwill help you make test specifications for the test. We have pasted the questions again here:

1) What is the purpose of the test?

2) What sort of learners will be taking the test?

3) What language skills should be tested (reading, writing, speaking and/or listening)?

4) What language elements should be tested (grammar, vocabulary, pronunciation, speech acts, etc.)?

5) What target language situation is envisaged for the test, and is this to be simulated in some way in the test content and method? (For instance, is this a test of academic French? Of English for international TAs?Of Japanese for hotelworkers?)

6) What text types should be chosen as stimulus materials -- written and/or spoken?

7) What sort of tasks are required -- discrete point, integrative, simulated 'authentic', objectively assessable? (That is, what will the test-takers do?)

5-АМАЛИЙ МАШҒУЛОТ. Тестлар яратиш усуллари.
Тест принциплари
Режа:

1. The purpose and principles of a test
2. Types of tests
3. Practicing developing tests

Principles of test designing

Objectives:

- to review current practices of testing and assessment
- to raise awareness about different purposes of testing
- to explore the advantages and disadvantages of multiple choice items
- to explore some principles of testing

Activity 1 Different types of tests

Objective: to discuss current practices of testing, different test types and their purposes

Time: 20 min

Materials: flipchart, markers

►Procedure:

☺(5 min) Ask participants the following questions. Take answers to each question

one at a time and comment and expand as necessary:

~ Do you test your students' knowledge? If you do, how do you do it?

~ Why do you have tests with your students?

~ Are there any obligatory tests at your universities? If yes, what kinds of test are these?

☺(10 min) Ask participants when each of the tests is given and why. Draw a timeline to illustrate when different tests are given taking into account different purposes of the tests. Use the following questions:

~ When are the tests conducted ?

~ What is the purpose of each of the tests?

1)Placement test / proficiency test

2)Progress test

3) Achievement test

4)Proficiency test

~ What is the difference between a progress test and an achievement test?

Possible answers:

~Answers may vary. Note that some teachers prefer to use other means of assessment rather than testing. Some teachers may tell you that they assess their students on a regular basis without conducting any tests, others may say that they have tests. Note that a test is usually an event when students have to complete some tasks under timed conditions in order to be assessed.

~To check where the students are in their learning; tests can help a teacher understand needs of their learners and decide on his/her teaching plan; tests give students a feeling of accomplishment; tests can help students understand their strengths and weaknesses.

~Participants may not know the exact terminology and the trainer will need to elicit ideas from participants and supply the correct terminology. Teachers may tell you about the existing system of assessment in their universities i.e. joriy nazorat (it is usually some kind of assessment of students' progress or a progress test 1), oralik nazorat (progress test 2) and yakuniy nazorat (final test).

A progress test serves to assess the progress of students in the course; an achievement test is similar to a progress test and is used to assess the overall progress of students at the end of the course; it aims to define whether students have acquired the knowledge and skills required by the course objectives and learning outcomes; a proficiency test doesn't take into account the syllabus but checks the overall language abilities of students (IELTS, TOEFL) against external standards; a placement test helps to group students according to their level so that there are students of the same language level in a group.

~A placement test is done before the beginning of the course; a progress test can be done at any time during a course; an achievement test is usually done at the end of the course; a proficiency test is usually external i.e. it is usually conducted outside the school or university and students have to register for it. It is done irrespective of the course because it is not based on the course materials.

☺(5 min) Establish that we often test our students' knowledge to assess their language skills. Say that there are different purposes of testing and therefore there are different types of tests. Some of these can define the students' level of English as in a proficiency test, others help teachers find out how much their students have acquired in the course as in an achievement test and progress test.

Activity 2 Analysis of existing tests

Objective: to analyse sample test items

Time: 25 min

Materials: handout 1, flipchart, markers

►Procedure:

Ask participants the following question:

~ What do you usually test?

Possible answers:

~ Answers may vary e.g. grammar, vocabulary, skills.

☺☺☺(15 min) Tell participants that they will now have a chance to analyse some grammar and vocabulary test items taken from an achievement test. Ask participants to work in groups. Distribute handout 1and ask participants to do the test and answer the questions. Write the questions on the board.

~ What kind of test methods are these?

~ What does each question test?

~ Can you see any problems with the tasks?

~ How can the instructions be improved?

~

Suggested answers:

I. Multiple choice items (1, 2, 3, 4, 5); multiple matching (10); open gap fill (11)

II. 1. articles; 2-4 vocabulary; 5 knowledge of political systems; 6-11 vocabulary.

III. 1. In this item two answers are possible (aand c); there is not enough context to answer the question.

2. There is no context in this item and therefore it is difficult to do the task. Moreover, the underlined word is an adjective and therefore the synonym should also be an adjective but the correct answer is a noun.

3. In this item any answer is possible, there is not enough context.

4. The sentence in the given item is unethical and may send the wrong message to learners; there is not enough context to do the task and more than one answer (b, c or d)is possible.

5. The item doesn't test English.

6. The sentence in the item is incorrect. Criminals usually break the law, rather than escape from it. The word combination 'escape from law' is impossible.

7. The second part of the sentence is impossible. The word combination 'repeat the book' is incorrect.

8. The sentence contains incorrect English. The word 'agree' requires a preposition afterwards.

10. There is no context after the gap in this stem and therefore any answer is possible.

IV. The instructions are different and thus can confuse students. Therefore, all the instructions in the test should be standardised e.g. Choose the correct answer.

☺(10 min) Collect participants' ideas, commenting on their answers when necessary.

- Summarise the activity by saying that we have looked at the problematic grammar and vocabulary items where there was more than one correct answer, not enough context to do the task and sometimes the test items didn't test knowledge of English at all. Tell participants that it is also common to refer to a grammar and vocabulary test as a Use of English test.

Activity 3 Context based multiple choice items (Use of English)

Objective: to give participants an opportunity to do and analyse multiple choice items based on the texts

Time: 20 min

Materials: handout 2

►Procedure:

☺☺☺(15 min) Tell participants that this time they will receive good items to discuss.

Tell participants that these tests were taken from a proficiency test that is used as a University entrance test in Uzbekistan. Write the questions on the board, then distribute handout 2 and ask participants to do the test in groups and answer the questions.

~ What test method is used in this test?

~ How is this test different from the test in activity 2?

~ What do these tasks test?

~ Are the instructions helpful?

Suggested answers:

~ Multiple choice questions; open gap fill

~ All items are given in context. There is only one correct answer in each item.

Tutankhamen

0 We need a word here which has the same meaning as 'due to the fact' or 'because' in order to provide a logical link between the two parts of the sentence. Of the four options available, only 'since' can fulfil this function in this context because here it is used to state the reason. Therefore a) is the correct answer.

Q1 In order to answer this item correctly, you have to know how to use the correct form of the verb 'to have' in the present perfect tense.

Option a)is not correct because the word 'tomb' is in the singular and therefore 'have' cannot be used.

Options b)and d)are also incorrect because they are non-finite forms of the verb, and cannot be used to form the present perfect tense.

Option c)is correct because the word 'tomb' requires the third person singular form of the present perfect tense, that is, 'has'.

Q2 In order to answer this item correctly, you have to know the meaning of a range of prepositions.

Option a)is not correct because 'over' is normally used with a period of time rather than a specific point in time and therefore cannot be used with the phrase '20thcentury' [text] in this context.

Option b)is also incorrect because the preposition 'on' is used with dates such as 'on the 20thof January' and days of the week, for example, 'on Wednesday'. It

cannot be used with the word 'century'.

Option c)is incorrect because the preposition 'at' is used in certain expressions of time, for example, 'at Christmas', 'at night'. It cannot be used with the word 'century'.

Option d)is correct because it combines the two nouns ('finds' and '20th

century') and the second noun gives more information about the first noun making it more definite: not just any 'finds' but the finds of a particular time.

Q3 In order to answer this item correctly, you have to know the correct form of the verb in the past perfect passive.

Option b) is correct because the word 'forgotten', which is the past participle of the verb 'forget', is necessary to form the past perfect passive.

Option a) is incorrect because it is in the present simple tense.

Option c) is incorrect because it is in the simple past.

Option d)is incorrect because it is an infinitive.

Q4 In order to answer this item correctly, you have to know the meaning of a range of prepositions.

Option c) is correct because the word 'by' is used after a passive construction to introduce the doer of the action i.e. 'a later dynasty' [text].

Options a)and d)do not go together with the phrase 'a later dynasty' [text] and are therefore incorrect.

Option b)does not fit on a semantic level and it does not go together with the verb 'to erase'.

Q5 In order to answer this item correctly, you have to know the meaning of a range of prepositions.

Option a) is correct because the word 'under' goes together with 'was hidden'[text] and 'ruins' [text] whereas the other options go together with only one of these words/phrases from the text and not both as is required to make the text read correctly. Therefore options b),c) and d) are incorrect.

Reality TV

0 The text requires a preposition before the word 'Sweden'. To show that the first reality TV programme was shown there we need a preposition of place. The

preposition 'in' indicates the location in this context i.e. 'Sweden'. Therefore 'in' would be the correct answer.

Q1 In the first line of the text we learn that "Expedition Robinson" is a television programme. The second line says that the country watched the final part of the programme. Therefore in this gap, 'a new kind of TV' [text] requires a word which goes together with TV e.g. 'programme', 'show' or 'series'. Any of these answers would be acceptable as would the American spelling of the word 'programme' [program].

Q2 The verb in this sentence is in the passive which requires the auxiliary verb 'to be' in order to complete the gap before 'filmed'. 'Two years later' [text] refers to an action which happened in the past, and 'the first part of Big Brother' [text] is in the third person singular. Therefore the correct answer is 'was'.

Q3 This sentence is about the 'Big Brother' programme mentioned in the text. To refer to the fact that there is only one final programme, we need the definite article. Therefore 'the' is the correct answer.

Q4 This sentence states that the number of countries showing 'Big Brother or Survivor' programmes is over 20. So the preposition 'than' is needed to link the two parts of the comparison 'more ….. 20 countries'.

Q5 'Take part' is a phrasal verb which requires the preposition 'in' and implies some kind of participation in something. Therefore the correct answer would be 'in'.

Q6 When speaking about large numbers e.g. hundreds, thousands, millions etc. we need the preposition 'of' to link the number with the noun. Thus the word missing here is 'of'.

☺ (5 min) Summarise the activity by saying that there are certain things that a good language test should consider. Tell participants that based on the discussions they had in their groups they will now think about the principles of a good Use of English test.

Activity 4 Principles of testing

Objective: to work out the main principles of a good achievement test

Time: 15 min

Materials: board, markers

►Procedure:

☺ ☺ ☺ (5 min) Tell participants to work in groups and based on their discussions of good and bad items draw up some principles of a good test.

☺ (10 min) Collect participants' ideas on the board.

A good Use of English (vocabulary and grammar) achievement test

should…

Possible answers:

~ not test what has not been taught

~ provide enough context to complete the task

~ have clear instructions

~ have examples

~ have distractors of approximately the same length/structure (in multiple choice items)

~ use more than one test method

~ use authentic/correct English

• Distribute handout 3to participants so that at home they can compare their explanations and answers with the ones suggested by the authors.

Summary

Establish that it's better to test English in a context which is sufficient and enables the test taker to decide on the correct answer. In order to provide test takers with more context the first and last sentences of a text are left intact. The instructions should be clear and consistent. If it is a progress or achievement test it should test what has been taught. A proficiency test has a different purpose and doesn't need to contain familiar topics or tasks.

TESTING AND ASSESSMENT

Activity 2, Handout 1

Do the test and answer the questions below.

1. What kind of test methods are illustrated in each of these tasks?

2. What does each question test?

3. What is the problem with each of the questions?

4. How can the instructions be improved?

1. Find the best answer.

Open … window please.

a) the b) all c) a d) –

2. Choose the synonym.

potential

a) refusal b) desire c) favourite d) possibility

3. There are four answers. Choose the correct answer.

A policeman stopped the driver and asked whether he had a ….

a) driving license b) passport c) bottle of alcohol d) screwdriver

4. Choose the best antonym for the underlined word.

"Death is beautiful", she said

a) imperfect b) unattractive c) disastrous d) ugly

5. Choose the correct answer.

Which country doesn't have its own Constitution?

a) Brazil b) the UK c) Italy d) the USA

6. Use a synonym:

Criminals try to escape from the law.

7. Use an antonym instead of underlined words.

She has agreedher daughter's marriage.

8. Complete the following sentence:

The boy is punished for the slightest …

TESTING AND ASSESSMENT

Activity 3, Handout 2

~ What test method is used in this test?

~ How is this test different from the test in activity 2?

~ What do these tasks measure/assess?

~ Are the instructions helpful?

Read the text below. Then circle the appropriate letter (a, b, c or d) for each gap in the text. There is an example at the beginning (0).

Tutankhamen was king of Egypt from 1361 to 1352 BC. As a ruler he was not remarkable, (0) __a___ he was very young when he died. But his tomb (Q1) _____ proved extremely valuable. It was one of the great historical finds (Q2) _____ the 20thcentury. Tutankhamen's burial site remained intact until the English scientist Howard Carter entered it in 1922. The tomb had been long (Q3) _____ because Tutankhamen's name was erased from a list of kings (Q4) _____ a later dynasty. By the 20thcentury the tomb was well hidden (Q5) _____ ruins. Only a systematic exploration of the Valley of the Kings revealed it.

Example

0. a) since   b) so   c) then   d) if

Q1 a) have   b) having c) has   d) to have

Q2 a) over   b) on   c) at   d) of

Q3 a) forget   b) forgotten   c) forgot   d) to forget

Q4 a) on   b) to   c) by   d) out

Q5 a) under   b) at   c) for   d) from

Read the text below. Some words are missing from the text. Use only one word to complete each gap. There is an example at the beginning (0).

Reality TV

The first reality TV show in the world was called Expedition Robinson. It was shown (0) __in___ Sweden in 1997. Half the population of the country watched the final part and a new kind of TV (Q1) _____ was born. Two years later in Holland, the first part of Big Brothers (Q2) _____ filmed. Again, it was a fantastic success and (Q3) _____ final programme was watched by 15 million people. There are now more (Q4) _____ 20 countries around the world which have Big Brothers or Survivor (Expedition Robinson) on their screens. The ordinary people take part (Q5) _____ the programmes are known by millions (Q6) _____ people in their own countries. Reality TV has become big, big business.

6-7-АМАЛИЙ МАШҒУЛОТ

Аудио матнларни танлаш ва матн харитасини яратиш. (4 соат)

Режа:

1. Listening assessment tools
2. Reading strategies and assessing reading
3. Requirements for assessing speaking.
4. Assessing writing.

Assessing receptive skills (LISTENING & READING)

Objectives:

- to explore current practices of testing and assessing receptive skills

- to explore other ways of testing listening and reading skills

- to identify good practice in testing reading and listening

Lead-in

Objective: to uncover the current practice of testing and assessing receptive skills

Time: 5 min

►Procedure:

☺(5 min) Refer participants to the sessions on Teaching Reading and Authenticity.

Ask them what kind of reading subskills they remember. Elicit scanning, skimming, reading intensively. Ask participants the following question:

~ How do you test or otherwise assess your students' reading and listening

skills?

- Elicit random answers.

- Establish that in our context reading skills are regarded as a tool for testing other skills like writing, speaking, pronunciation, etc. through such tasks as 'read and retell the text','read aloud', 'read and learn by heart'. Reading skills such as scanning, skimming and reading intensively are not tested. Testing listening skills is not very popular because of the lack of resources. Tell participants that this session focuses on testing listening and reading skills.

Activity 1 Testing reading skills

Objective: to let participants explore characteristics of a good reading test

Materials: handout 1

Time: 30 min

►Procedure:

☺ (5 min) Distribute handout 1and ask participants to do the reading test.

☺ (5 min) Check the answers.

Suggested answers:

1. Q1 The phrases from the second paragraph 'after eating sugary food', and 'some people feel tired' are the key phrases which help to answer this question.

Therefore, the correct answer is f).

Q2 The correct answer to this question comes from the sentences 'Alcohol might make you feel good … but long term is likely to lead to irritability.'So the correct answer is e).

Q3 The last paragraph contains the correct answer to this question. It says,

'vegetables like ….. can help improve intelligence' [text], so the correct answer is a).

2. Q1 The answer can be found in the second paragraph of the text. The key phrase is: 'the correct address for the headquarters was No. 4 Whitehall Place'. The correct answer would therefore be: 'No 4 Whitehall Place'.

Q2 This question asks you about the 'size' of the police force which refers to the number of people working for the police. Therefore you should look for numbers in the text. In this example, the correct answer is in paragraph 3: '1,000'.

Q3 This question also focuses on numbers. The number which relates to the year 1870, is 10,000. Therefore the correct answer is '10,000'. 3. b

☺ ☺ ☺ (5 min) Put participants in groups of 5. Ask groups to study the tests and answer the following questions:

1. What reading sub-skills does each item test?

2. Do you actually have to read the text in order to answer the questions?

3. What kinds of texts (authentic or non-authentic) are used in the test?

4. What did you like and dislike about this test?

☺ (5 min) Elicit random answers to each question one at a time and expand on them.

Suggested answers:

1. Task 1 tests skimming skills, task 2 tests scanning skills, task 3 tests knowledge of vocabulary rather than reading.

2. You have to read and use specific reading skills (scanning or skimming) in order to do the test; to do task 3 one does not have to read the whole text in order to find the antonym.

3. Tasks 1 and 2 exploit authentic (or semi-authentic) texts, while item 3 seems to be based on a non-authentic text.

4. Answers may vary.

☺☺☺(5 min) Ask participants to work in the same groups and make a list of characteristics of a good reading test.

Possible answer:

A good reading test:

~ focuses on reading sub-skills (scanning, skimming and reading intensively);

~ tests what it should test i.e. a reading test should test reading, not speaking or anything else;

~ uses authentic texts;

~ leaves little room for guessing;

~ contains only one correct answer per item etc.

☺(5 min) Elicit answers and write them on the board.

Establish that a good reading test focuses on reading sub-skills (scanning, skimming and reading intensively); it tests what it should test; uses authentic texts, tests reading in context etc.

Activity 2 Testing listening skills

Objective: to let participants explore characteristics of a good listening test

Materials: handout 2 (Jakeman V.& C.McDowell (1996). Cambridge Practice Tests for IELTS 1. Cambridge University Press. Listening Test 1. pp. 12-13.) accompanied by the cassette (Cassette 1, Side 1)

Time: 35 min

Preparation: Prepare the cassette so that it can be started from the right point

Procedure:

☺ (15 min) Tell participants that now they will try to look into what a good listening test is. Distribute handout 2 and ask them to listen to the recording and do the test.

 ☐ Play the recording.

 ☺ (5 min) Check the answers.

Answer key:

1. A

2. C

3. D

4. D

5. C

6. Prescott (must be correct spelling with capital 'P')

7. 41

8. Fountain (must have capital 'F')

9. 752239

10. £65

☺ (5 min) Ask participants the following questions and elicit random answers:

1. What listening subskills does each test item test (What do test-takers have to do in order to get it right)?

2. Do you actually have to listen to the recording in order to answer the questions?

3. What kind of texts (authentic or non-authentic) are used in the test?

4. Is there any visual support to help listening?

5. What did you like and didn't like about this test?

Possible answers:

1.Most of the test items test listening for specific information.

Questions 6-10 test spelling skills as well.

2.Yes.

3.Semi-authentic or non-authentic.

4.Yes.

Even if the texts may seem to be non-authentic, tasks are taken from

real life (e.g. calling Lost & Found office, describing an object, filling in

personal details forms etc. Visuals help listening and make the test not

boring and less frightening.)

☺ ☺ ☺ (5 min) Ask participants to work in the same groups as above and make a list of the characteristics of a good listening test.

☺ (5 min) Elicit answers and write them on the board.

Possible answers:

A good listening test:

~ focuses on listening sub-skills (listening for the main idea(s), listening for

> specific information);
>
> ~ tests what it should test;
>
> ~ uses authentic texts and tasks;
>
> ~ can be supported with visuals;
>
> ~ should require minimum output (e.g. writing) etc.

Establish that a good listening test focuses on listening sub-skills (listening for the main idea(s), listening for specific information etc.); it tests what it should test; uses authentic listening texts; is supported with visuals etc.

Summary

Summarise the session and establish that good reading and listening tests should:

~ test reading and listening subskills (scanning, skimming, listening for specific information etc.);

~ use authentic or semi-authentic texts and real life tasks;

~ include visual support;

~ require minimum output (e.g. writing)

TESTING RECEPTIVE SKILLS

Activity 1, Handout 1, Reading test samples

1. Read the text and match the questions 1-3 with the appropriate answers a-f. There is an example (0) at the beginning. There are more answers than you need.

Food for Mood

What you eat can affect how you feel, think and act. Let's look first at bad mood food. Keep caffeine down to one or two cups of coffee a day. More could make you anxious and depressed. A sugary snack will raise levels of brain serotonin almost instantly. This can give a short term high, but beware of let down. An hour or so after eating sugary foods, some people feel tired, unable to concentrate, even depressed.

Alcohol might make you feel good in the short-term but long-term is likely to lead to irritability, lethargy and depression. Damage to brain tissue may bring on forgetfulness, bad moods, confusion, inability to concentrate and sleep disturbance.

So what foods are feel-good enemies? Well, obviously, junk food, alcohol, caffeine, sugary cereals, snacks, sweets and drinks, and highly processed foods.

OK, so now let's take look at good mood food

Fruit and vegetables studies show people who eat more of these tend to be happier. Green leafy vegetables like asparagus, endive and spinach contain an amino acid, which can help improve intelligence, ease mental tiredness and fight depression.

Type of food and drinks	Effect produced
0 coffee	a) improved intelligence
Q1 sugary snacks	b) high temperature
Q2 alcohol	c) anxiety
Q3 vegetables	d) better memory
	e) irritability and depression
	f) feeling of tiredness

Answers

0 a b c d e f

 1 a b c d e f 2 a b c d e f 3 a b c d e f

2. Read the text and complete the sentences below using 1-5 words. There is an example at the beginning (0).

Sir Robert Peel founded the Metropolitan Police in 1829. It was headquartered in a district known as Scotland Yard. Scotland Yard originally denoted an area of Whitehall that contained many government buildings.

When Sir Robert Peel located the new Metropolitan Police in Whitehall, the correct address for the headquarters was No. 4 Whitehall Place. The police station there had an entrance from Scotland Yard proper. The press always referred to the police as Scotland Yard, until the term became synonymous with the force itself.

Scotland Yard grew rapidly in its early years. In 1829 the total force was 1,000, but a decade later it was 3,350. By 1850 it was more than 6,000, and would increase to 10,000 by 1870.

The task of the Yard was extraordinary: it was to police crime in an area with a population of two and a half million people.

0The Metropolitan Police, known as Scotland Yard, was established in __1829__ .

Q1 The address of the police headquarters was _____ .

Q2 In 1829 the size of Scotland Yard's total force was _____ .

Q3 By 1870 the police force had grown to _____ .

3. Read the text and find the antonym to the underlined word.

Theodore Dreiser, the great American progressive writer, was born in a poorfamily in 1871. He began to work for his living when he was sixteen. He had a number of jobs, and at one time was a newspaper reporter. As a reporter he gained a wide experience of life, which was a great help to him when he took up novel-writing.

a. happy   b. rich   c. deprived   d. friendly

8-Мавзу: Тинглаб тушуниш, ўқиш, ёзув, гаплашиш кўникмаси бўйича тестлар тахлил қилиш ва яратиш кўникмасини шакллантириш. (2 соат)

PLAN

5. Listening assessment tools
6. Reading strategies and assessing reading
7. Requirements for assessing speaking.
8. Assessing writing.

Assessing receptive skills (LISTENING & READING)

Objectives:

- to explore current practices of testing and assessing receptive skills
- to explore other ways of testing listening and reading skills
- to identify good practice in testing reading and listening

Lead-in

Objective: to uncover the current practice of testing and assessing receptive skills

Time: 5 min

►Procedure:

☺(5 min) Refer participants to the sessions on Teaching Reading and Authenticity.

Ask them what kind of reading subskills they remember. Elicit scanning, skimming, reading intensively. Ask participants the following question:

~ How do you test or otherwise assess your students' reading and listening

skills?

- Elicit random answers.
- Establish that in our context reading skills are regarded as a tool for testing other skills like writing, speaking, pronunciation, etc. through such tasks as 'read and retell the text','read aloud', 'read and learn by heart'. Reading skills such as scanning, skimming and reading intensively are not tested. Testing listening skills is not very popular because of the lack of resources. Tell participants that this session focuses on testing listening and reading skills.

Activity 1 Testing reading skills

Objective: to let participants explore characteristics of a good reading test

Materials: handout 1

Time: 30 min

►Procedure:

☺(5 min) Distribute handout 1and ask participants to do the reading test.

☺(5 min) Check the answers.

Suggested answers:

1.

Q1 The phrases from the second paragraph 'after eating sugary food', and 'some people feel tired' are the key phrases which help to answer this question.

Therefore, the correct answer is f).

Q2 The correct answer to this question comes from the sentences 'Alcohol might make you feel good … but long term is likely to lead to irritability.'So the correct answer is e).

Q3 The last paragraph contains the correct answer to this question. It says,

'vegetables like ….. can help improve intelligence' [text], so the correct answer is a).

2.

Q1 The answer can be found in the second paragraph of the text. The key phrase is: 'the correct address for the headquarters was No. 4 Whitehall Place'. The correct answer would therefore be: 'No 4 Whitehall Place'.

Q2 This question asks you about the 'size' of the police force which refers to the number of people working for the police. Therefore you should look for numbers in the text. In this example, the correct answer is in paragraph 3: '1,000'.

Q3 This question also focuses on numbers. The number which relates to the year 1870, is 10,000. Therefore the correct answer is '10,000'. 3. b

☺ ☺ ☺ (5 min) Put participants in groups of 5. Ask groups to study the tests and answer the following questions:

1. What reading sub-skills does each item test?

2. Do you actually have to read the text in order to answer the questions?

3. What kinds of texts (authentic or non-authentic) are used in the test?

4. What did you like and dislike about this test?

☺ (5 min) Elicit random answers to each question one at a time and expand on them.

Suggested answers:

1. Task 1 tests skimming skills, task 2 tests scanning skills, task 3 tests knowledge of vocabulary rather than reading.

2. You have to read and use specific reading skills (scanning or skimming) in order to do the test; to do task 3 one does not have to read the whole text in order to find the antonym.

3. Tasks 1 and 2 exploit authentic (or semi-authentic) texts, while item 3 seems to be based on a non-authentic text.

4. Answers may vary.

☺ ☺ ☺ (5 min) Ask participants to work in the same groups and make a list of characteristics of a good reading test.

Possible answer:

A good reading test:

~ focuses on reading sub-skills (scanning, skimming and reading intensively);

~ tests what it should test i.e. a reading test should test reading, not speaking or anything else;

~ uses authentic texts;

> ~ leaves little room for guessing;
>
> ~ contains only one correct answer per item etc.

☺ (5 min) Elicit answers and write them on the board.

Establish that a good reading test focuses on reading sub-skills (scanning, skimming and reading intensively); it tests what it should test; uses authentic texts, tests reading in context etc.

Activity 2 Testing listening skills

Objective: to let participants explore characteristics of a good listening test

Materials: handout 2 (Jakeman V.& C.McDowell (1996). Cambridge Practice Tests for IELTS 1. Cambridge University Press. Listening Test 1. pp. 12-13.) accompanied by the cassette (Cassette 1, Side 1)

Time: 35 min

Preparation: Prepare the cassette so that it can be started from the right point

Procedure:

☺ (15 min) Tell participants that now they will try to look into what a good listening test is. Distribute handout 2 and ask them to listen to the recording and do the test.

 Play the recording.

 ☺ (5 min) Check the answers.

Answer key:

1. A

2. C

3. D

4. D

5. C

6. Prescott (must be correct spelling with capital 'P')

7. 41

8. Fountain (must have capital 'F')

9. 752239

10. £65

☺ (5 min) Ask participants the following questions and elicit random answers:

1. What listening subskills does each test item test (What do test-takers have to do in order to get it right)?

2. Do you actually have to listen to the recording in order to answer the questions?

3. What kind of texts (authentic or non-authentic) are used in the test?

4. Is there any visual support to help listening?

5. What did you like and didn't like about this test?

Possible answers:

1. Most of the test items test listening for specific information.

Questions 6-10 test spelling skills as well.

2. Yes.

3. Semi-authentic or non-authentic.

4. Yes.

Even if the texts may seem to be non-authentic, tasks are taken from

real life (e.g. calling Lost & Found office, describing an object, filling in

personal details forms etc. Visuals help listening and make the test not

boring and less frightening.)

☺ ☺ ☺ (5 min) Ask participants to work in the same groups as above and make a list of the characteristics of a good listening test.

☺ (5 min) Elicit answers and write them on the board.

Possible answers:

A good listening test:

~ focuses on listening sub-skills (listening for the main idea(s), listening for specific information);

~ tests what it should test;

~ uses authentic texts and tasks;

~ can be supported with visuals;

~ should require minimum output (e.g. writing) etc.

Establish that a good listening test focuses on listening sub-skills (listening for the main idea(s), listening for specific information etc.); it tests what it should test; uses authentic listening texts; is supported with visuals etc.

Summary

Summarise the session and establish that good reading and listening tests should:

~ test reading and listening subskills (scanning, skimming, listening for specific information etc.);

~ use authentic or semi-authentic texts and real life tasks;

~ include visual support;

~ require minimum output (e.g. writing)

TESTING RECEPTIVE SKILLS

Activity 1, Handout 1, Reading test samples

1. Read the text and match the questions 1-3 with the appropriate answers a-f. There is an example (0) at the beginning. There are more answers than you need.

Food for Mood

What you eat can affect how you feel, think and act. Let's look first at bad mood food. Keep caffeine down to one or two cups of coffee a day. More could make you anxious and depressed. A sugary snack will raise levels of brain serotonin almost instantly. This can give a short term high, but beware of let down. An hour or so after eating sugary foods, some people feel tired, unable to concentrate, even depressed.

Alcohol might make you feel good in the short-term but long-term is likely to lead to irritability, lethargy and depression. Damage to brain tissue may bring on forgetfulness, bad moods, confusion, inability to concentrate and sleep disturbance.

So what foods are feel-good enemies? Well, obviously, junk food, alcohol, caffeine, sugary cereals, snacks, sweets and drinks, and highly processed foods.

OK, so now let's take look at good mood food

Fruit and vegetables studies show people who eat more of these tend to be happier. Green leafy vegetables like asparagus, endive and spinach contain an amino acid, which can help improve intelligence, ease mental tiredness and fight depression.

Type of food and drinks	Effect produced
0 coffee	a) improved intelligence
Q1 sugary snacks	b) high temperature
Q2 alcohol	c) anxiety
Q3 vegetables	d) better memory
	e) irritability and depression
	f) feeling of tiredness

Answers

0 a b c d e f

 1 a b c d e f 2 a b c d e f 3 a b c d e f

2. Read the text and complete the sentences below using 1-5 words. There is an example at the beginning (0).

Sir Robert Peel founded the Metropolitan Police in 1829. It was headquartered in a district known as Scotland Yard. Scotland Yard originally denoted an area of Whitehall that contained many government buildings.

When Sir Robert Peel located the new Metropolitan Police in Whitehall, the correct address for the headquarters was No. 4 Whitehall Place. The police station there had an entrance from Scotland Yard proper. The press always referred to the police as Scotland Yard, until the term became synonymous with the force itself.

Scotland Yard grew rapidly in its early years. In 1829 the total force was 1,000, but a decade later it was 3,350. By 1850 it was more than 6,000, and would increase to 10,000 by 1870.

The task of the Yard was extraordinary: it was to police crime in an area with a population of two and a half million people.

0The Metropolitan Police, known as Scotland Yard, was established in __1829__ .

Q1 The address of the police headquarters was _____ .

Q2 In 1829 the size of Scotland Yard's total force was _____ .

Q3 By 1870 the police force had grown to _____ .

3. Read the text and find the antonym to the underlined word.

Theodore Dreiser, the great American progressive writer, was born in a poorfamily in 1871. He began to work for his living when he was sixteen. He had a number of jobs, and at one time was a newspaper reporter. As a reporter he gained a wide experience of life, which was a great help to him when he took up novel-writing.

a. happy   b. rich c. deprived   d. friendly

V.   КЕЙСЛАР

Lesson 2. Activity 4 (25 min) An extract from a lesson

Tell participants to take the role students for 20 minutes. Tell them that you are going to present an extract from a lesson. Ask them be active during the lesson.

Lesson 5. Activity 3

Time: 15

Materials: Handout 3

Procedure:

- o The table below gives three paradigms of classroom investigation.

- o Ask participants to put each of the headings *{mixed; quantitative; qualitative)* next to the paradigm where they think it belongs.

- o Possible answers:

Lesson 5. Activity 4

paradigm 1: *Qualitative*	**paradigm 2:** *Quantitative*	**paradigm 3:** *Mixed*
1. non-experimental design a researcher kept a diary of her experiences as a language learner	*1. experimental or quasi- experimental design* the discussions of 10 groups of different sizes were recorded	*1. experimental or quasi-experimental design* a researcher gave a group of students two sheets and a questionnaire to fill in. They used these to record their observations of a lesson.
2. *qualitative data* the diarist recorded her feelings about different aspects of the course	2. *quantitative data* the data from the recorded discussions were coded according	2. *qualitative data* the researcher analysed their impressions and opinions of the lesson

Time: 20 min

Materials: Handout 4

Procedure:

- o Ask participants to work in pairs and fill in the table (handout 4) with the questions/problems and disscuss:
- • Is any qualitative data necessary to research any of the written research questions? If yes, how this data can be got? (students' tests, questionnaires, etc.)
- • What are the reasons of collecting quantitative data to research the question/problem?
- o Ask participants to share filled table and report about the discussed answers to the stated questions.

    Possible answers: participants' answers

List of research questions	Possible ways of collecting a quantitative data, if necessary
1. To what extent do my students progress in reading? What are the reasons?	- Examining students' tests, assignments - Students questionnaire
2. What are the ways of dealing with challenging students?	- Not necessary
3.	
4.	
5.	
6.	
7.	

Lesson 6.Activity 2 Rating scales for assessing speaking

Objectives:

●to raise awareness of the CEFR for assessing speaking;

●to explore different types of activities to assess students' spoken
language

Time: 20 min

Materials: handout 1 (scrambled band descriptors i.e. group 1 receives
scrambled

Range descriptors, group 2- Accuracy, etc)

►Procedure:

☺(5 min) Tell participants that there are different rating scales developed by
different educationalists for different purposes. Tell them that you are going to give
them the Common European Framework of Reference that was developed to assess
overall oral production. Tell participants that there are 6 levels in the scales that
correspond to 6 levels of proficiency.

A1 corresponds to a beginner level (basic user);

A2 – elementary;

B1 – pre-intermediate;

B2 – intermediate;

C1 – upper intermediate (independent user);

C2 – advanced level (proficient user).

Write the corresponding letters against the levels on the board if appropriate.
Tell participants that in the speaking scales there are 5 criteria in the CEFR i.e.
Range, Accuracy, Fluency, Interaction and Coherence.

VI. ГЛОССАРИЙ

Термин	Ўзбек тилидаги шарҳи	Инглиз тилидаги шарҳи
Assessment	Biror kishining qobiliyatini yoki o'qitish kursining sifati yoki muvaffaqiyatini o'lchash va boshqalar	The measurement of the ability of a person or the quality or success of a teaching course, etc
Evaluation	qaror qabul qilish uchun muntazam ravishda ma'lumot to'plash	the systematic gathering of information for purposes of decision making.
Feedback	O'qituvchilar yoki o'qituvchidan yoki boshqa o'quvchilardan o'rganish vazifalarining muvaffaqiyatli bajarilishi to'g'risida sharhlar yoki ma'lumotlar.	Comments or information learners receive on the success of a learning task, either from the teacher or from other learners.
Fluency	yozma yoki og'zaki nutq so'zlash qobiliyatini va jiddiy ikkilanmasdan ishlab chiqarish qobiliyatini o'z ichiga olgan muloqotni bilish darajasi	a level of proficiency in communication, which includes the ability to produce written or spoken language withease and without significant hesitation
Grid	O'quvchilar yoki o'qituvchi-ishtirokchilar tomonidan to'ldiriladigan jadval, ko'pincha fikrlarni umumlashtirish yoki aks ettirish uchun ishlatiladi	A chart to be filled in by learners or teacher-participants, often used to summarise ideas or to focus reflection
TESTING -	birovning bilishi yoki qobiliyatini tekshirish, ular javob beradigan savollar yoki ular amalga oshirishi kerak bo'lgan ishlardan iborat	an examination of somebody's knowledge or ability, consisting of questions for them to answer or activities for them to carry out
Assessment Profile	loyihalarni yaratish uchun asos bo'lishi mumkin bo'lgan baholashlar va hisobotlar to'plami.	a collection of assessments and reports that can be used as a basis for creating projects.

117

Assessment criteria	o'rganish baholanadigan standartlar.	the standards by which learning is judged.
Assessment Mode	Tarkibni baholashda talabalarning boshqa yorliqlarni yoki oynalarni ko'rish imkoniyatlarini kamaytirishga yordam beradigan Chrome kengaytmasi.	a Chrome extension that helps reduce the chance of students viewing other tabs or windows while taking Content Assessments.
a framework	foydali yoki konstruktsiyani kengaytiradigan narsalarni qurish uchun qo'llab-quvvatlash yoki ko'rsatma bo'lib xizmat qilish uchun mo'ljallangan haqiqiy yoki kontseptual tuzilish.	a real or conceptual structure intended to serve as a support or guide for the building of something that expands the structure into something useful.
Task	qilinadigan yoki bajariladigan ishning bir qismi	a piece of work to be done or undertaken.
Testee	imtihon topshiradigan kishi.	one who takes an examination.
performance-referenced assessment	o'quvchilarning o'quv bo'linmasi yoki bo'linmalaridan olingan ko'nikma va bilimlarni qo'llash qobiliyatini o'lchaydi	measures students' ability to apply the skills and knowledge learned from a unit or units of study.
Portfolio	dars ishlarining sifati, o'qish taraqqiyoti va o'quv yutuqlarini baholash maqsadida yig'ilgan o'quv ishlari va boshqa ma'lumotlarga oid ma'lumotlar to'plami	a compilation of academic work and other forms of educational evidence assembled for the purpose of evaluating coursework quality, learning progress, and academic achievement;
Self evaluation	maqsadlaringizga erishish uchun o'zingizning noyob kuchli tomonlaringizni qanday ishlatganingizni va shu bilan birga yaxshilashingiz mumkin bo'lgan sohalarda halol bo'lganingizni ta'kidlash	an opportunity to highlight how you used your unique strengths to accomplish your goals while also being honest about areas you could

	imkoniyati.	improve.
Project work	vazifani bajarishga qaratilgan ish.	work which focuses on completing a task.
Peer assessment	o'quvchilar yoki ularning tengdoshlari tomonidan o'qituvchining mezonlari asosida berilgan topshiriqlar yoki testlar.	a process whereby students or their peers grade assignments or tests based on a teacher's benchmarks.
Observation report	topilmalarni tasdiqlovchi hujjatlashtirilgan dalillar	the documented evidence to support findings.
Feedback	harakat, hodisa yoki jarayon haqida baholash yoki tuzatish ma'lumotlarini asl yoki boshqaruvchi manbaga etkazish	the transmission of evaluative or corrective information about an action, event, or process to the original or controlling source
Survey	odamlarga savollar berish orqali qilingan fikrlar, xulq-atvor va boshqalarni tekshirish.	an examination of opinions, behaviour, etc., made by asking people questions
Classroom observation	o'qitishni sinfda yoki boshqa o'quv muhitida amalga oshirilayotgan paytda rasmiy yoki norasmiy kuzatuv.	a formal or informal observation of teaching while it is taking place in a classroom or other learning environment.
Questionnaire	respondentlardan ma'lumot to'plash maqsadida bir qator savollardan iborat tadqiqot vositasi..	a research instrument consisting of a series of questions for the purpose of gathering information from respondents.
Common European Framework of Reference	til qobiliyatini tavsiflovchi xalqaro standart.	international standard for describing language ability.
level descriptors	ma'lum bir darajadagi o'qishga mos keladigan ta'limning keng	statements that provide a broad indication of learning

	ko'rsatkichlarini ta'minlaydigan, ushbu darajada kutilayotgan ta'limning xususiyatlari va kontekstini tavsiflovchi bayonotlar	appropriate to attainment at a particular *level*, describing the characteristics and context of learning expected at that *level*.
rating scales	respondentlarning fikr-mulohazalarini solishtirma shaklda muayyan o'ziga xos xususiyatlar / mahsulotlar / xizmatlar uchun ifodalash uchun foydalaniladigan yopiq so'rovnoma savoli	a closed-ended survey question used to represent respondent feedback in a comparative form for specific particular features/products/services.
Range,	bir xil umumiy tipdagi turli xil narsalar to'plami.	a set of different things of the same general type.
Accuracy,	to'g'ri yoki aniq bo'lgan sifat yoki holat.	the quality or state of being correct or precise.
Fluency,	o'zini oson va aniq ifodalash qobiliyati.	the ability to express oneself easily and articulately.
Coherence	qismlarning mantiqiy va tartibli va izchil aloqasi	logical and orderly and consistent relation of parts
Placement test	odatda ma'lum bir bilim yoki malakani aniqlash uchun ta'lim muassasasiga kiradigan talabaga beriladigan test.	a test usually given to a student entering an educational institution to determine specific knowledge or proficiency .
Progress test	kognitiv bilimlarning rivojlanishi va barqarorligini baholash uchun uzunlamasına, teskari yo'naltirilgan ta'limni baholash vositalari	longitudinal, feedback oriented educational assessment tools for the evaluation of development and sustainability of cognitive knowledge
Achievement test	odatda ta'lim sharoitida yuzaga keladigan rejalashtirilgan ko'rsatma orqali shaxs olgan bilimlari va akademik mahorat darajasini o'lchash	to measure the amount of knowledge and level of academic skills an individual has acquired through the

		planned instruction that typically occurs in educational settings
Proficiency test	talabalarning umumiy til qobiliyatini (IELTS, TOEFL) tashqi standartlarga nisbatan tekshiradi.	checks the overall language abilities of students (IELTS, TOEFL) against external standards;
Multiple choice test	respondentlardan ro'yxat sifatida berilgan tanlovlardan faqat to'g'ri javoblarni tanlashlari so'raladigan ob'ektiv baholash shakli.	a form of an objective assessment in which respondents are asked to select only correct answers from the choices offered as a list.
Receptive skills	Axborot olish uchun ishlatiladigan til qobiliyatlari (o'qish va tinglash).	Language skills used to receive information (reading and listening).
Productive skills	o'quvchilarga tilni yozma yoki og'zaki shaklda ishlab chiqarishga imkon beradigan ko'nikmalarga murojaat qiling	refer to the skills that enable the learners to produce language in written or spoken forms
Scanning	ma'lum bir ma'lumotni tezda topmoqchi bo'lganingizda foydalaniladigan o'qish texnikasi.	a reading technique to be used when you want to find specific information quickly.
Skimming	matnning asosiy g'oyalariga e'tiboringizni qaratadigan strategik, tanlab o'qish usuli.	a strategic, selective reading method in which you focus on the main ideas of a text.
Eliciting	o'qituvchiga ma'lumot berish o'rniga ma'lumot berish uchun o'qituvchiga yordam beradigan bir qator metodlarni tavsiflovchi atama.	term which describes a range of techniques which enable the teacher to get learners to provide information rather than giving it to them.

VII. ФОЙДАЛАНИЛГАН АДАБИЁТЛАР РЎЙХАТИ

I. Ўзбекистон Республикаси Президентининг асарлари

1. Мирзиёев Ш.М. Буюк келажагимизни мард ва олижаноб халқимиз билан бирга қурамиз. – Т.: "Ўзбекистон", 2017. – 488 б.

2. Мирзиёев Ш.М. Миллий тараққиёт йўлимизни қатъият билан давом эттириб, янги босқичга кўтарамиз. 1-жилд. – Т.: "Ўзбекистон", 2017. – 592 б.

3.Мирзиёев Ш.М. Халқимизнинг розилиги бизнинг фаолиятимизга берилган энг олий баҳодир. 2-жилд. Т.: "Ўзбекистон", 2018. – 507 б.

4.Мирзиёев Ш.М. Нияти улуғ халқнинг иши ҳам улуғ, ҳаёти ёруғ ва келажаги фаровон бўлади. 3-жилд.– Т.: "Ўзбекистон", 2019. – 400 б.

5.Мирзиёев Ш.М. Миллий тикланишдан – миллий юксалиш сари. 4-жилд.– Т.: "Ўзбекистон", 2020. – 400 б.

II. Норматив-ҳуқуқий ҳужжатлар

1.Ўзбекистон Республикасининг Конституцияси. – Т.: Ўзбекистон, 2018.

2. Ўзбекистон Республикасининг "Таълим тўғрисида"ги Қонуни.

3. Ўзбекистон Республикаси Президентининг 2015 йил 12 июндаги "Олий таълим муасасаларининг раҳбар ва педагог кадрларини қайта тайёрлаш ва малакасини ошириш тизимини янада такомиллаштириш чора-тадбирлари тўғрисида" ги ПФ-4732-сонли Фармони.

4. Ўзбекистон Республикаси Президентининг 2017 йил 7 февралдаги "Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида"ги 4947-сонли Фармони.

5. Ўзбекистон Республикаси Президентининг "Олий таълим тизимини янада ривожлантириш чора-тадбирлари тўғрисида"ги 2017 йил 20 апрелдаги ПҚ-2909-сонли Қарори.

6. Ўзбекистон Республикаси Президентининг "2019-2021 йилларда Ўзбекистон Республикасини инновацион ривожлантириш стратегиясини

тасдиқлаш тўғрисида”ги 2018 йил 21 сентябрдаги ПФ-5544-сонли Фармони.

7. Ўзбекистон Республикаси Президентининг 2019 йил 27 майдаги “Ўзбекистон Республикасида коррупцияга қарши курашиш тизимини янада такомиллаштириш чора-тадбирлари тўғрисида”ги ПФ-5729-сон Фармони.

8. Ўзбекистон Республикаси Президентининг 2019 йил 17 июндаги “2019-2023 йилларда Мирзо Улуғбек номидаги Ўзбекистон Миллий университетида талаб юқори бўлган малакали кадрлар тайёрлаш тизимини тубдан такомиллаштириш ва илмий салоҳиятини ривожлантири чора-тадбирлари тўғрисида”ги ПҚ-4358-сонли Қарори.

9. Ўзбекистон Республикаси Президентининг 2019 йил 27 августдаги “Олий таълим муассасалари раҳбар ва педагог кадрларининг узлуксиз малакасини ошириш тизимини жорий этиш тўғрисида”ги ПФ-5789-сонли фармони.

III. Махсус адабиётлар

1. Beresova, J. (2017). The impact of the CEFR on teaching and testing English in the local context. Theory and Practice in Language Studies, 7(11), 959-964.

2.Brown, H. D., & Abeywickrama, P. (2014). Language assessment. Principles and classroom practices. White Plains, NY: Pearson Education.

3.Cumming, A. and R. Berwick (Eds.). (2010) Validation in Language Testing. Clevedon, England: Multilingual Matters, Ltd.

4.Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), Handbook of second language assessment: Volume 12 of handbooks on applied linguistics (pp. 413-428). Berlin, Germany: Walter de Gruyter GmbH & Co KG. doi: https://doi.org/10.1515/9781614513827-027

5.Henning, G. (2014). A guide to language testing: Development, evaluation, research. Newbury House Publishers.

6.Hughes, A. (2013). Testing for Language Teachers. 2nd ed. Cambridge: Cambridge University Press.

7.Isaacs, T., Zara, C., Herbert, G., Coombs, S. J., & Smith, C. (2013). Assessment of learning. In The SAGE key Concepts Series: Key concepts in educational assessment (pp. 12-17). London: SAGE Publications Ltd doi: 10.4135/9781473915077.n4

8.Knight, P., & Yorke, M. (2013). Assessment, learning and employability. Maidenhead: Society for Research into Higher Education & Open University Press.

9.Madsen H.S. (2014). Techniques in Testing. England: Oxford University Press.

10.McMilan, J.H. (2014). Fundamental Assessment Principles for Teachers and School Administrators. Practical Assessment, Research and Evaluation. Retrieved from https://pareonline.net/getvn.asp?v=7&n=8 [Accessed July 17, 2018]

11.McNamara, T. (2014). Language Testing. UK. Oxford: Oxford University Press.

12.Taylor, L. (2019). Developing assessment literacy. Annual Review of Applied Linguistics, 29, 21-36.

13.Weir, C. (2015). Language testing and validation: An evidence-based approach. Palgrave

IV. Интернет сайтлар

1.Ўзбекистон Республикаси Олий ва ўрта махсус таълим вазирлиги: www.edu.uz.

2. Бош илмий-методик марказ: www.bimm.uz

3.www. Ziyonet. uz

4.Ўзбекистон Республикаси Президентининг расмий сайти: www.gov.uz

5.Ўзбекистон Республикаси Олий ва ўрта махсус таълим вазирлигининг расмий сайти: http://www.edu.uz/modules/ wfchannel

6.karsu.uz-Қорақалпоқ давлат университети сайти

Фарғона давлат университети ҳузуридаги педагог кадрларни қайта тайёрлаш ва уларнинг малакасини ошириш минтақавий маркази ўқитувчиси, Усмонова Донохоннинг "Тил компетенцияларини баҳолаш механизмлари" модули бўйича яратилган ўқув-услубий мажмуасига

ТАҚРИЗ

"Тил компетенцияларини баҳолаш механизмлари" модули инглиз тили йўналиши ўқитувчиларининг ҳалқаро эътироф этилган стандартларга мувофиқ хорижий тилларни билиш даражасини баҳолаш доирасидаги билим, кўникма, малака ҳамда компетенцияларни ривожлантиришга қаратилган.

Модулни ўзлаштириш жараёнида тингловчилар тил компетенцияларини баҳолашнинг асосий тамойиллари бўйича тасаввурга эга бўлишида , альтернатив баҳолаш турларини билиш ва улардан фойдаланиш ва тил кўникмаларини (ўқиш, гапириш, тинглаб-тушуниш, ёзув) баҳолаш мезонларини яратиш кўникмаларига эга бўлади.

"Тил компетенцияларини баҳолаш механизмлари" модули бўйича ўқув услубий мажмуа ўз ичига 6 та мавзуни олган. Улар тингловчиларга тил билиш даражасини баҳолаш тамойиллари ва усуллари бўйича билим бериш, мавзуга оид олимлар фикрларини таҳлил қилишни ўргатиш, баҳолаш мезонларини яратиш ва улардан самарали фойдаланишни амалиётга қўллашга ўргатиш, альтернатив баҳолаш, синфни тадқиқ қилиш, фидбек бериш, портфолио тузиш, ўз-ўзини баҳолаш учун материаллар яратишни ўргатиш, CEFR тизими бўйича баҳолашни ўргатиш каби энг долзарб ва керакли маълумотларни беради ва тингловчиларда баҳолаш тизими бўйича кўникмаларни шакллантиради. Ҳозирги кунда керакли тестлар яратиш усуллари ва принципларини билиш, тил кўникмалари бўйича тестлар яратиш кўникмасини шакллантиришни мажмуа ўз ичига олади.

Ўқув услубий мажмуада яратилган топшириқлар, маъруза матнлари уларнинг касбий фаолиятига қаратилган бўлиб, улар тингловчиларда катта қизиқишлар уйғотиши мумкин. Юқорида келтирилган ижобий фикрларни инобатга олиб "Тил компетенцияларини баҳолаш механизмлари" модули бўйича яратилган ўқув услубий-мажмуа педагог кадрларни қайта тайёрлаш ва уларнинг малакасини ошириш минтақавий маркази инглиз тили йўналиши тингловчилари учун қўлланма сифатига фойдаланишга тавсия этаман.

Тақризчи   ФарДУ "Инглиз тили ўқитиш методикаси" кафедраси кат.ўқитувчи ф.ф,б.ф.д.(PhD) Хошимова Н.