

Karimov Rivojiddin Gulyamovich, Avilova Khalida Nabijanovna

2021

T

O'quv uslubiy-majmua

Il ko'nikmalarini baholash mexanizmlari

**O'ZDJTU HUZURIDAGI CHET TILLARINI
O'QITISHNING INNOVATSIYAVIY
METODIKALARINI RIVOJLANTIRISH RESPUBLIKA
ILMIY-AMALIY MARKAZI**

Filologiya va tillarni o'qitish (ingliz tili)

Toshkent-2021

**O'ZBEKISTON RESPUBLIKASI
OLY VA O'RTA MAXSUS TAHLIM VAZIRLIGI**

**OLY TA'LIM TIZIMI PEDAGOG VA RAHBAR KADRLARINI
QAYTA TAYYORLASH VA ULARNING MALAKASINI OSHIRISHNI
TASHKIL ETISH BOSH ILMIY - METODIK MARKAZI**

**O'ZDJTU HUZURIDAGI CHET TILLARINI O'QITISHNING
INNOVATSIYAVIY METODIKALARINI RIVOJLANTIRISH RESPUBLIKA
ILMIY-AMALIY MARKAZI**

FILOLOGIYA VA TILLARNI O'QITISH: INGLIZ TILI

**“TIL KOMPETENSIYALARINI BAHOLASH MEXANIZMLARI”
MODULI BO'YICHA**

O'Q U V – U S L U B I Y M A J M U A

Toshkent - 2021

Modulning o'quv- uslubiy majmuasi Oliy va o'rta maxsus, kasb-hunar ta'limi o'quv-metodik birlashmalari faoliyatini Muvofiqlashtiruvchi kengashining 2020 yil 7 dekabrda 648 -sonli buyrug'i bilan maqullangan o'quv dasturi va o'quv rejasiga muvofiq ishlab chiqilgan.

Tuzuvchi: f.f.n, R. Karimov, Kh. Avilova

Taqrizchilar: Prof. X. Paluanova, f.f.d. G. Ergasheva

O'quv-uslubiy majmua O'zRIAIM Ilmiy- metodik Kengashining 2020 yil 25 dekabrda 10 - sonli qarori bilan nashrga tavsiya qilingan

MUNDARIJA

I.	ISHCHI DASTUR	5
II.	MODULNI O'QITISHDA FOYDALANILADIGAN INTREFAOL TA'LIM METODLARI	12
III.	MA'RUZA MASHG'ULOTLARI	15
IV.	AMALIY MASHG'ULOTLAR	46
V.	KEYSLAR BANKI	92
VI.	GLOSSARIY	105
VII.	ADABIYOTLAR RO'YXATI	110

I. ISHCHI DASTUR

Kirish

Dastur O'zbekiston Respublikasi Prezidentining 2017 yil 7 fevraldagi "O'zbekiston Respublikasini yanada rivojlantirish bo'yicha Harakatlar strategiyasi to'g'risida"gi PF-4947-son, 2019 yil 27 avgustdagi "Oliy tahlim muassasalari rahbar va pedagog kadrlarining uzluksiz malakasini oshirish tizimini joriy etish to'g'risida"gi PF-5789-son, 2019 yil 8 oktyabrdagi "O'zbekiston Respublikasi oliy tahlim tizimini 2030 yilgacha rivojlantirish kontsepsiyasini tasdiqlash to'g'risida"gi PF-5847-son va 2020 yil 29 oktyabrdagi "Ilm-fanni 2030 yilgacha rivojlantirish kontsepsiyasini tasdiqlash to'g'risida"gi PF-6097-sonli Farmonlari hamda O'zbekiston Respublikasi Prezidentining 2012 yil 10 dekabrda "Chet tillarni o'rganish tizimini yanada takomillashtirish chora-tadbirlari to'g'risida"gi PQ-1875-son hamda O'zbekiston Respublikasi Vazirlar Mahkamasining 2019 yil 23 sentyabrdagi "Oliy tahlim muassasalari rahbar va pedagog kadrlarining malakasini oshirish tizimini yanada takomillashtirish bo'yicha qo'shimcha chora-tadbirlar to'g'risida"gi 797-sonli qarorlarida belgilangan ustuvor vazifalar mazmunidan kelib chiqqan holda tuzilgan bo'lib, u oliy tahlim muassasalari pedagog kadrlarining kasb mahorati hamda innovatsion kompetentligini rivojlantirish, sohaga oid ilg'or xorijiy tajribalar, yangi bilim va malakalarni o'zlashtirish, shuningdek amaliyotga joriy etish ko'nikmalarini takomillashtirishni maqsad qiladi.

Dastur mazmuni tilni bilish darajasini baholash tamoyillari, baholash mezonlarini yaratish va ulardan samarali foydalanish, baholash va material to'plash, muqobil baholash, sinfni tadqiq qilish, fidbek berish, o'z-o'zini baholash, test yaratish usullari va tamoyillari, tinglab tushunish, o'qish, yozuz, gaplashish ko'nikmalari bo'yicha testlar tahlil qilish va yaratish ko'nikmasini shakllantirish bo'yicha bilim, ko'nikma va malakalarini shakllantirishni nazarda tutadi.

Qayta tayyorlash va malaka oshirish yo'nalishining o'ziga xos xususiyatlari hamda dolzarb masalalaridan kelib chiqqan holda dasturda tinglovchilarning maxsus fanlar doirasidagi bilim, ko'nikma, malaka hamda kompetensiyalariga qo'yiladigan talablar o'zgartirilishi mumkin.

Modulning maqsadi va vazifalari

Modulning maqsadi til kompetensiyalarini baholash mexanizmlari moduli bo'yicha oliy ta'lim muassasalari professor-o'qituvchilarida amaliy xorijiy tilni o'qitish jarayonini tashkil etish, uni rejalashtirish va baholash metodlarini hozirgi kun talablari asosida tashkil etish metodikasini o'zlashtirish va amaliyotda qo'llay olish ko'nikmalarini rivojlantirishdan iboratdir.

Modulning vazifalari - chet tili ta'limida bilim, ko'nikma, malakalarni baholash, lingvistik va madaniyatlararo kompetensiyalarni baholash bilan birga nutq kommunikatsiyasi elementlari, sotsiolingvistik elementlar, pragmatik elementlar, baholashning asosiy turlarini o'zlashtirish bilan birga mavzuga oid olimlar tomonidan bildirilgan fikrlarni tahlil qilish ko'nikmalarini ham mustahkamlaydi. Modul portfolio tuzish, CEFR tizimi bo'yicha baholash, testlarni ishlab chiqish printsiplari bilan ham tanishtirish, o'quv adabiyotlarga bo'lgan ehtiyojni o'rganish, o'quv maqsadlari va kutilayotgan natijalarni to'g'ri belgilay olish, baholash, fikr-mulohazaga asoslangan baholash mexanizmini mustaqil holda tashkil etishni o'z ichiga oladi.

Modul bo'yicha tinglovchilarning bilimi, ko'nikmasi, malakasi va kompetensiyalariga qo'yiladigan talablar

Tinglovchi:

- Til bilish darajasini baholash tamoyillarini;
- Baholash mezonlarini yaratish va ulardan samarali foydalanishni;
- kommunikativ va vazifaga asoslangan til o'rgatishda baholash mezonlarini;
- testlar yaratish usullarini;
- xorijiy tilni o'qitishdagi o'quv jarayonini rejalashtirish;
- xorijiy tilni o'qitishda sinf menejmentini;
- o'quv jarayonini baholash mezonlarini;
- til o'rganish jarayonlari va baholash usullari bo'yicha **bilimga ega bo'lishi kerak.**

Tinglovchi:

- til o'qitilayotgan va o'rganilayotgan sharoitga qarab, o'quv jarayonini rejalashtirish uslublarini to'g'ri qo'llay olish;
- til o'rganuvchining ehtiyojlarini, bilim darajasini to'g'ri baholay olish va o'qitishning eng samarali usullarini qo'llay olish;
- talabalar o'zlashtirishini tahlil qilish, baholash va turli amaliy topshiriqlarni yarata olish;
- o'z kasbiy maxoratini oshirishda kasbga doir adabiyotlardan va shaxsiy amaliyotdan unumli foydalanishni yo'lga qo'yish **ko'nikmlarini egallashi lozim.**

Tinglovchi:

- chet tili ta'limida bilim, ko'nikma, malakalarni baholash;
- lingvistik va madaniyatlararo kompetensiyalarni baholash;
- nutq kommunikatsiyasi elementlari, sotsiolingvistik elementlar pragmatik elementlar, baholashning asosiy turlarini o'zlashtirish;
- mavzuga oid olimlar tomonidan bildirilgan fikrlarni tahlil qilish **malakalariga ega bo'ladi.**

Tinglovchi:

- Modul bo'yicha portfolio tuzish, CEFR tizimi bo'yicha baholash; testlarni ishlab chiqish printsiplari bilan ham tanishtirish;
- o'quv adabiyotlarga bo'lgan ehtiyojni o'rganish;
- o'quv maqsadlari va kutilayotgan natijalarni to'g'ri belgilay olish;
- baholash, fikr-mulohazaga asoslangan baholash mexanizmini mustaqil holda tashkil etish **kompetensiyalariga ega bo'lishi talab etiladi.**

Modulni tashkil etish va o'tkazish bo'yicha tavsiyalar

Modullarni o'qitishda darslik, o'quv qo'llanmalar, tarqatma materiallaridan, ma'ruza matnlaridan, xorijda sohaga oid chop etilgan so'nggi maqolalardan foydalaniladi. Mashg'ulotlar amaliy va vokshop ta'lim shaklida olib boriladi va kichik guruhlarda ishlash, amaliy o'yinlar, o'quv loyihalarini yaratish, "Keys-stadi" kabi interfaol metodlardan foydalangan holda amalga oshiriladi.

Kurs davomida tinglovchilar dastlab namunaviy darslarda “til o’rganuvchi” rovida qatnashadilar, so’ngra o’qituvchi sifatida tavsiya etilgan darslar va mashqlarni tahlil qilishadi, maxsus testlarni yechishadi, portfolio topshiriqlarini bajarishadi, namunaviy dars o’tib berish orqali o’zlarining til o’qitish ko’nikmalarini rivojlantirishadi. Kurs ishtirokchilaridan kasbiy malakalaridan kelib chiqqan holda chet tillarni o’qitishdagi dolzarb masalalarni belgilab olishlari shuningdek, guruh ahzolari bilan darsda hamkor holda ish yuritishlari, o’zaro fikr almashishlari, interfaol usullarining samarali shakllarini tahlim sohasiga tadbiq etishga erishishlari, mavjud ish tajribalarini hamkasblari bilan o’rtoqlashishlari, xususan, ilg’or tajribalarni seminar-treninglar, amaliy kurslar, didaktik materiallar, metodik qo’llanmalar ko’rinishida ommalashtirish va ta’lim tizimida qo’llashlari kutilmoqda. Shuningdek, tinglovchilar o’zlarining kasbiy mahoratlarini oshirishda muhim bo’lgan mavzular bilan tanishadilar va chet tilidagi muloqot ko’nikmalarini rivojlantirish ustida ish olib boradilar.

Kurs davomida tinglovchilarning ilg’or tajribalarini o’rganish va ommalashtirish nazarda tutilgan. Amaliy mashg’ulotlarda va mustaqil ta’limda qo’llaniladigan usullar:

- amaliy topshiriq;
- jarayon davomida o’rganish;
- muhokama;
- loyiha ishi;
- prezentatsiya;
- potrfolio qaydlari/tahlim olish kundaligini yuritishi;
- o’quv adabiyotlarni, video va audio materiallarni baxolash va tanlashga mo’ljallangan matn, topshiriq va vazifalar;
- so’rovnoma o’tkazish;
- Internet va boshqa manbalardan mavzularga oid mahlumotlar to’plash;
- turli situatsiyalar, rolli o’yinlar;

Modulning o'quv rejadagi boshqa modullar bilan bog'liqligi va uzviyligi

Til kompetensiyalarini baholash mexanizmlari moduli mazmuni o'quv rejadagi "Til va ta'limga integrativ yondashuv: Post-metod davri", "Kommunikativ tilshunoslik va til kompetensiyalari" o'quv modullari bilan uzviy bog'langan holda ingliz tili o'qituvchilarini xorijiy tillarni o'qitishdagi zamonaviy yondashuvlar, pedagogik texnologiyalar va interaktiv uslublar bilan tanishtiradi.

Modul bo'yicha soatlar taqsimoti

№	Modul mavzulari	Tinglovchining o'quv yuklamasi, soat			
		Auditoriya o'quv yuklamasi			Mustaqil ta'lim
		Jami	jumladan		
			Nazariy	Amaliy mashg'ulot	
1.	Mechanisms of the Language Competences Assessment. Assessment in TFL	2	2		
2	Language Competences Assessment. History of Language Testing	2	2		
3.	Assessment for Learning: Gathering Data	2		2	
4.	Assessment for Learning: Using Data	2		2	
5.	Assessment of Learning: Principles of Assessment	2		2	
6.	Assessment of Learning : Types of Assessment	2		2	
7.	Designing Tests	2		2	
8.	Test Specifications	2		2	

9.	Using Statistics: Objectively-scored Assessment	2		2	
10.	Using Statistics: Subjectively-scored Assessment	2		2	
	Жами	20	4	16	

O'quv materiallar mazmuni

1. **Mavzu:** Til kompetensiyalarini baholash mexanizmlari modulining maqsadi va vazifalari
2. **Mavzu:** Til kompetensiyalari. Baholash va testlar yaratilish tarixi
3. **Mavzu:** Til o'qitish uchun baholash. Test qilish, nazorat o'tkazish va baholashning farqi. O'z-o'zini, juftini va joriy nazorat baholash
4. **Mavzu:** Baholash natijasida olingan ma'lumotlarni tahlil qilish va dars jarayonoga tatbiq etish. Fidbek turlari va ularning xususiyatlari
5. **Mavzu:** Til o'qitish uchun baholash va til bilish darajasini baholashdagi farqlar
6. **Mavzu:** Baholash tamoyillari va ulardan foydalanish usullari
7. **Mavzu:** Test yaratish tamoyillari
8. **Mavzu:** Test tuzish hususiyatlari
9. **Mavzu:** Baholashda statistik ma'lumotlarning muhimligi. Ob'ektiv baholash. Statistik ma'lumotlarni tahlil qilish.
10. **Mavzu:** Subyektiv baholash. Umumiy va alohida baholashning farqi

Mashg'ulotlarni tashkil etish bo'yicha

ko'rsatma va tavsiyalar

Amaliy va nazariy mashg'ulotlarda tinglovchilar o'quv modullari doirasidagi ijodiy topshiriqlar, keyslar, o'quv loyihalari, texnologik jarayonlar bilan bog'liq vaziyatli masalalar bajaradilar. Amaliy mashg'ulotlar zamonaviy ta'lim uslublari va innovatsion texnologiyalarga asoslangan holda o'tkaziladi. Bundan tashqari, mustaqil holda o'quv va ilmiy adabiyotlardan, elektron resurslardan, tarqatma materiallardan foydalanish tavsiya etiladi.

Dasturning axborot-metodik ta'minoti

Modulni o'qitish jarayonida ishlab chiqilgan o'quv-metodik materiallar, tegishli soha bo'yicha ilmiy jurnallar, Internet resurslari, mul'timedia mahsulotlari va boshqa elektron va qog'oz variantdagi manbalardan foydalaniladi.

II. MODULNI O'QITISHDA FOYDALANILADIGAN INTERFAOL TA'LIM METODLARI

1- Method: Lexical approach.

It presupposes the analysis of contextual meanings, the correlation of denotative and connotative meanings, comparison of contextual and dictionary meanings. This method is most relevant to text semantics since it enables the reader to discern subtle shades of meanings the language units acquire in the text.

2- Method: Content-based approach.

It deals with, the semantic content of language units and is based on the linguistic postulate that lexical meaning can be segmented into minimal semantic components (semes). The method includes techniques of a multi-stage definitional analysis and that of associative field. Componential analysis was first applied to the study of the semantic word structure, word agreement, lexicographical descriptions, syntagmatic and paradigmatic links of the word. In text linguistics, this analysis may be helpful in revealing conceptual features constituting the cognitive structure of words.

3- Method: Cognitive approach.

It is aimed at revealing associative links of the analyzed unit and building its associative field. This method can be implemented in two ways: 1) on the basis of an associative experiment; 2) on the material of associative dictionaries and thesauruses. An associative experiment has been elaborated in cognitive thinking, and it is based on the assumption that a certain stimulus presupposes some reaction: $S \rightarrow R$. The second way to uncover a net of associations inherent in the analyzed unit is to use the material supplied by various dictionaries of associations.

4- Method: Brainstorming.

It 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 contributed by its members. The term was popularized by Alex Faickney Osborn in the 1953 book *Applied Imagination*.

5- Method: Interactive method.

There are three distinct reasons for interactive teaching. It is an attempt to see what actually exists in the brains of your students. This is the "summative" aspect. It is the easiest aspect to understand and it is well described in the literature. But, it is far from

being the only perspective! The second reason is "formative", where the teacher aims through the assigned task to direct students' mental processing along an appropriate path in "concept-space". The intent is that, as students think through the issues necessary in traversing the path, the resulting mental construction that is developed in the student's head will possess those properties that the teacher is trying to teach. As Socrates discovered, a good question can accomplish this result better than, just telling the answer. The third may be termed "motivational". Learning is hard work, and an injection of motivation at the right moment can make all the difference.

6- Method: Discourse analysis.

It is based on the situational interpretation of the discourse and aims a) to hypothesize about the author's pragmatic intention on the basis of verbal signals; b) to characterize the linguistic personality's social status, cultural background and psychological inner world; c) to substantiate the appropriateness and effectiveness of communication by the verbalized in the text direct or indirect indications of a pragmatic intention and the addressee's verbal or non-verbal reaction to it.

7- Method: Methods of parameterization.

It is the study of linguistic units including texts, based on a set of parameters and principles of categorization. It is used to define complex linguistic notions and hierarchy of distinctive features constituting them.

8- Method: Method of categorization.

It is a kind of taxonomical activity aimed at sorting out and combining somewhat similar notions and corresponding units into larger categories. This method is based on cognitive operations of comparing, contrasting, identifying, setting up similarities and differences, etc. In the domain of text linguistics, the method of categorization may be helpful in analyzing text as a conceptual system, and revealing hierarchical relationships between its constituents.

9- Method: Critical Thinking Skills development (Bloom's Taxonomy).

It is a comprehensive, multi-stage analysis of concepts, which presupposes description of a) the concept structure and its constituents (notional, figurative and evaluative spheres); b) hierarchical taxonomy of cognitive features inferred in the process of conceptualization; c) distribution of cognitive features according

to the "field" principle, i.e. their reference to either the nucleus or periphery of a concept.

10- Method: Problem solving strategies.

These are based on the theoretical conception that metaphor is not only a stylistic device, but also a means of conceptualization. It is one of the main mental operations, a means of cognition, categorization and evaluation (Lacoff, 1980). Metaphorical analysis aims to disclose the mechanism of metaphorization, which lies in the interaction of the two, source and target, spheres. The system of knowledge structures of a source sphere serves as a basis for cognitive modelling of a target sphere. Cognitive metaphorical analysis is of a special significance for fiction since imagery is considered an inalienable property of this text type.

11-Cognitive mapping.

It helps penetrate into "deep" semantics of the language units. It is presented as a step-by-step procedure including the study of:

- dictionary definitions and lexicographical interpretations of a word;
- properties conditioned by paradigmatic and syntagmatic links;
- word - building potential;
- phraseological units and paroemia affiliated to the analyzed word;
- all possible contexts of the word indicative of its conceptual senses.

12- Method: Cross-cultural analysis.

It is based on comparing and contrasting languages and cultures. It consists in cognitive interpretation of a) culturally and conceptually relevant language units in the text (linguoculturemes); b) universal and nationally-specific properties of language units including texts; c) cultural concepts manifesting particular domains.

III. MA'RUZA MASHG'ULOTLARI

1. LECTURE: MECHANISMS OF THE LANGUAGE COMPETENCES ASSESSMENT. ASSESSMENT IN TFL

PLAN:

1. PROFESSIONAL COMPETENCES OF THE FL TEACHERS

2. ASSESSMENT IN TFL

KEYWORDS: professional competence, linguistic competence, sociolinguistic competence, sociocultural competence, discourse competence, strategic competence

1. PROFESSIONAL COMPETENCES OF THE FL TEACHERS

1. Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.

Skills in choosing appropriate, useful, administratively convenient, technically adequate, and fair assessment methods are prerequisite to good use of information to support instructional decisions. Teachers need to be well-acquainted with the kinds of information provided by a broad range of assessment alternatives and their strengths and weaknesses. In particular, they should be familiar with criteria for evaluating and selecting assessment methods in light of instructional plans.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be able to use the concepts of assessment error and validity when developing or selecting their approaches to classroom assessment of students. They will understand how valid assessment data can support instructional activities such as providing appropriate feedback to students, diagnosing group and individual learning needs, planning for individualized educational programs, motivating students, and evaluating instructional procedures. They will understand how invalid information can affect instructional decisions about students. They will also be able to use and evaluate assessment options available to them, considering among other things, the cultural, social, economic, and language backgrounds of students. They will be aware that

different assessment approaches can be incompatible with certain instructional goals and may impact quite differently on their teaching. Teachers will know, for each assessment approach they use, its appropriateness for making decisions about their pupils. Moreover, teachers will know of where to find information about and/or reviews of various assessment methods.

Assessment options are diverse and include text- and curriculum-embedded questions and tests, standardized criterion-referenced and norm-referenced tests, oral questioning, spontaneous and structured performance assessments, portfolios, exhibitions, demonstrations, rating scales, writing samples, paper-and-pencil tests, seatwork and homework, peer- and self-assessments, student records, observations, questionnaires, interviews, projects, products, and others' opinions.

2. Teachers should be skilled in developing assessment methods appropriate for instructional decisions.

While teachers often use published or other external assessment tools, the bulk of the assessment information they use for decision-making comes from approaches they create and implement. Indeed, the assessment demands of the classroom go well beyond readily available instruments. Teachers who meet this standard will have the conceptual and application skills that follow. Teachers will be skilled in planning the collection of information that facilitates the decisions they will make. They will know and follow appropriate principles for developing and using assessment methods in their teaching, avoiding common pitfalls in student assessment. Such techniques may include several of the options listed at the end of the first standard. The teacher will select the techniques, which are appropriate to the intent of the teacher's instruction.

Teachers meeting this standard will also be skilled in using student data to analyze the quality of each assessment technique they use. Since most teachers do not have access to assessment specialists, they must be prepared to do these analyses themselves.

3. The teacher should be skilled in administering, scoring and interpreting the

results of both externally-produced and teacher-produced assessment methods.

It is not enough that teachers are able to select and develop good assessment methods; they must also be able to apply them properly. Teachers should be skilled in administering, scoring, and interpreting results from diverse assessment methods. Teachers who meet this standard will have the conceptual and application skills that follow. They will be skilled in interpreting informal and formal teacher-produced assessment results, including pupils' performances in class and on homework assignments. Teachers will be able to use guides for scoring essay questions and projects, stencils for scoring response-choice questions, and scales for rating performance assessments. They will be able to use these in ways that produce consistent results. Teachers will be able to administer standardized achievement tests and be able to interpret the commonly reported scores: percentile ranks, percentile band scores, standard scores, and grade equivalents.

They will have a conceptual understanding of the summary indexes commonly reported with assessment results: measures of central tendency, dispersion, relationships, reliability, and errors of measurement. Teachers will be able to apply these concepts of score and summary indices in ways that enhance their use of the assessments that they develop. They will be able to analyze assessment results to identify pupils' strengths and errors. If they get inconsistent results, they will seek other explanations for the discrepancy or other data to attempt to resolve the uncertainty before arriving at a decision. They will be able to use assessment methods in ways that encourage students' educational development and that do not inappropriately increase students' anxiety levels.

4. Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.

Assessment results are used to make educational decisions at several levels: in the classroom about students, in the community about a school and a school district, and in society, generally, about the purposes and outcomes of the educational enterprise. Teachers play a vital role when participating in decision-making at each of these levels and must be able to use assessment results effectively. Teachers who meet this standard will have the conceptual and application skills that follow. They will be able to use accumulated assessment information to organize a sound instructional plan for facilitating students' educational development.

When using assessment results to plan and/or evaluate instruction and curriculum, teachers will interpret the results correctly and avoid common misinterpretations, such as basing decisions on scores that lack curriculum validity. They will be informed about the results of local, regional, state, and national assessments and about their appropriate use for pupil, classroom, school, district, state, and national educational improvement.

5. Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.

Grading students is an important part of professional practice for teachers. Grading is defined as indicating both a student's level of performance and a teacher's valuing of that performance. The principles for using assessments to obtain valid grades are known and teachers should employ them.

Teachers who meet this standard will have the conceptual and application skills that follow. They will be able to devise, implement, and explain a procedure for developing grades composed of marks from various assignments, projects, inclass activities, quizzes, tests, and/or other assessments that they may use. Teachers will understand and be able to articulate why the grades they assign are rational, justified, and fair, acknowledging that such grades reflect their preferences and judgments. Teachers will be able to recognize and to avoid faulty grading procedures such as using grades as punishment. They will be able to evaluate and to modify their grading

procedures in order to improve the validity of the interpretations made from them about students' attainments.

6. Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.

Teachers must routinely report assessment results to students and to parents or guardians. In addition, they are frequently asked to report or to discuss assessment results with other educators and with diverse lay audiences. If the results are not communicated effectively, they may be misused or not used. To communicate effectively with others on matters of student assessment, teachers must be able to use assessment terminology appropriately and must be able to articulate the meaning, limitations, and implications of assessment results. Furthermore, teachers will sometimes be in a position that will require them to defend their own assessment procedures and their interpretations of them. At other times, teachers may need to help the public to interpret assessment results appropriately. Teachers who meet this standard will have the conceptual and application skills that follow.

Teachers will understand and be able to give appropriate explanations of how the interpretation of student assessments must be moderated by the student's socio-economic, cultural, language, and other background factors. Teachers will be able to explain that assessment results do not imply that such background factors limit a student's ultimate educational development. They will be able to communicate to students and to their parents or guardians how they may assess the student's educational progress. Teachers will understand and be able to explain the importance of taking measurement errors into account when using assessments to make decisions about individual students.

Teachers will be able to explain the limitations of different informal and formal assessment methods. They will be able to explain printed reports of the results of pupil assessments at the classroom, school district, state, and national levels.

7. Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

Fairness, the rights of all concerned, and professional ethical behavior must undergird all student assessment activities, from the initial planning for and gathering of information to the interpretation, use, and communication of the results. Teachers must be well-versed in their own ethical and legal responsibilities in assessment. In addition, they should also attempt to have the inappropriate assessment practices of others discontinued whenever they are encountered. Teachers should also participate with the wider educational community in defining the limits of appropriate professional behavior in assessment. Teachers who meet this standard will have the conceptual and application skills that follow.

They will know those laws and case decisions which affect their classroom, school district, and state assessment practices. Teachers will be aware that various assessment procedures can be misused or overused resulting in harmful consequences such as embarrassing students, violating a student's right to confidentiality, and inappropriately using students' standardized achievement test scores to measure teaching effectiveness.

2. ASSESSMENT IN TFL

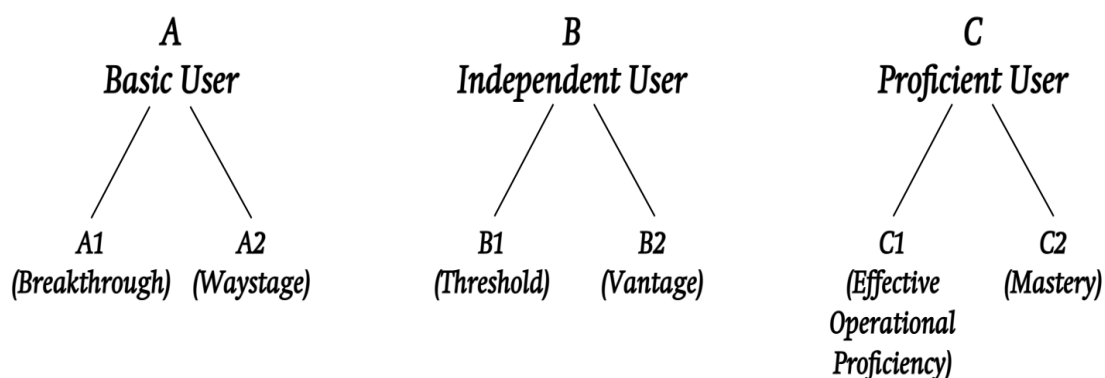
In 1991, the implementation of this framework was raised during an intergovernmental symposium in Switzerland with 3 objectives: establish a system of language levels to clearly and objectively describe proficiency, define a common framework and be able to apply it to any language. The guide, **CEFR: Learning, Teaching, Assessment**, published in 2001, lays out the fundamental principles and gives practical advice for applying the CEFR.

The importance of assessment One of the major areas impacted by the CEFR is assessment. What developed historically before the creation of the European

framework was a patchwork system of different scores and levels from one testing organization and training institute to another. The 1980s saw the proliferation of tests like the TOEIC, TOEFL and IELTS, each with their own scoring system. This made it hard to compare scores from one test to another and one candidate to another. The advantage of a framework like the CEFR is not only its ability to standardize language proficiency but also its objectivity and independence – it takes into consideration the theoretical concepts behind language learning and skill, including the 4 main competency areas: speak, listen, write and read. This means assessing speaking, listening, writing and reading skills.

A common framework in 6 levels

The CEFR consists of three broad divisions that can be divided into 6 levels:



Many European institutes of higher education require students to demonstrate a B2 level before graduating in order to ensure they will be able to find a job in the global talent market. This makes sense as global enterprises view B2 as the minimum level of business proficiency, particularly for jobs where communicating with international customers, colleagues and suppliers is par for the course. However, companies mainly based in one country with little or primarily written communication with other countries can get away with hiring employees with a B1 level. So what level should we expect candidates to be?

The Language(s) of Education (LE) is an overarching concept for language as a subject (LS), language across the curriculum (LAC) and foreign languages (FL). The aims of LE should be broadly described in terms of individual competences and of cultural competences required by society. The core purposes of LE teaching and learning in a social perspective are related to certain general aims of the Council of Europe: social cohesion, participatory democracy and respect for diversity within the member states. In this context language has an important role to play. It is a means of communication and personal development as well as a tool for learning and thinking, for artistic expression and for building identities. Mastering the LE is a prerequisite for being able to participate in society in all the different discourse communities a learner will meet. This includes a variety of skills and competences starting from the most basic reading and writing skills to being able to understand and reflect upon complicated texts and situations that will imply advanced linguistic and cultural competences. A definition of the competences covered by LE will therefore include what the learner knows, what the learner can do and what the learner is able to understand, think about and reflect upon. The aims for LE cover all of these.

The four domains of competences listed in the Common European Framework of Reference for Languages (CEFR): knowledge, skills/ know-how, existential competence, ability to learn (*savoir, savoir faire, savoir être and savoir apprendre*) might also be useful concepts for LS. The desirable outcome of the learning process is described in terms of competences and exemplified through assessment. The way in which we understand these aims is, however, deeply imbedded in culture and tradition. This again means that to a great extent we can describe and agree upon general competence and criteria for correctness. Criteria for quality, however, may have to be discussed bearing in mind cultural differences.

In general terms, we can specify aims for building up **knowledge** in LS in terms of:

- knowledge of the language: - grammar - sociolinguistics - tools for interpretation and analysis of language and text
- knowledge of texts: - text forms/ genres - text linguistics - composition - media and communication with texts
- knowledge of the

textual and linguistic tradition - history of language - history of literature - history of textual communication.

The aim for LS is to develop the three categories of **skills**: 8 • oral skills: - presenting knowledge and experiences - participating in discussion - presenting / interpreting literary texts - giving talks on various issues • reading skills: - understanding content of both continuous and discontinuous written texts - interpreting and reflecting upon literature of all genres - understanding and analysing a broad spectrum of textual forms - communicating reading experiences - learning from written texts - writing skills: - mastering linguistic conventions, orthography, grammar etc. - composing texts for presenting thoughts and meaning - composing texts for practical purposes - developing a spectrum of writing strategies for different genres and situations - mastering composition systems for different genres - mastering tools for gathering information, composing and presenting texts - being able to learn from text models and monitoring.

Aims for the development of **existential competence** can be expressed as follows: - consideration of “the other” - critical thinking and sound judgement - flexibility in thinking and argumentation - courage to express personal opinions - expressing understanding and experiences - exploring one’s own thinking and values. A learner might be able to write according to school standards for text construction, spelling and grammar. He/she might even master different genres and produce quite acceptable content. However, we expect more. Written and oral utterances are always made in a social context and will somehow reflect the author’s/speaker’s relationship to other people, other discourses and other texts. An awareness of these circumstances or a lack of such awareness will be imbedded in any utterance.

Aims for the development of **the ability** to learn in LS Traditionally, we assumed that learning to learn has been an ability acquired intuitively over years of schooling. We have, however, discovered that that this is not necessarily the case. Many learners will have to acquire this as a skill in addition to other skills at school. Not only do learners learn through different approaches and preferences, they often need help in

order to identify these approaches. Through reflection they can learn to be aware of what they do when they succeed, not just how they fail. Through self-assessment they can learn to see their own work from new perspectives and participate in their progress. The following abilities may be mentioned here as educational aims: - self-assessment - approaches and strategies for student work

Aims for teaching/learning : - developing subject specific concepts and genres - developing varied strategies for learning through language use - developing meta-perspectives on knowledge - learning to learn - developing mediation skills - developing cognitive academic language proficiency

Competence is a general ability based on knowledge, values, aptitudes, enabling to establish relationship between knowledge and situation, to reveal a procedure (knowledge and action), suitable for a problem.

Communicative competence is a term in linguistics which refers to a language user's grammatical knowledge of syntax, morphology, phonology and the like, as well as social knowledge about how and when to use utterances appropriately. The term was coined by Dell Hymes in 1966, reacting against the perceived inadequacy of Noam Chomsky's (1965) distinction between *linguistic competence* and *performance*.

Linguistic Competence refers to the mastery of knowledge of the language code itself. This involves controlling the formal organisation of the language for producing or recognising “correct” sentences and organising them to form texts. Linguistic Competence includes the rules of word formation and vocabulary (lexicon), pronunciation (phonology) and sentence formation (syntax). This knowledge of the language code is framed in terms of understanding the literal meaning of the utterance.

Discourse Fluency refers to the ability to use the rules and conventions of combining grammatical forms and meanings to achieve unified spoken texts in different genres. This unity of text is achieved through cohesion in form and coherence in meaning. Cohesive devices include pronouns, synonyms, conjunctions and parallel structures which help to link individual utterances and show the logical or

chronological relations among a series of utterances. Coherence refers to the logical sequencing of the ideas in a text. Discourse fluency is seen as an overall measure of spontaneous speech behaviour in peer-interactive situations. The observational categories pertaining to the appropriateness of language used in a natural way, combined with the functions implementing the communicative goal, can be characterised as a global measure of discourse fluency and include both linguistic and paralinguistic behaviour.

Sociolinguistic Competence refers to the mastery of cultural rules of use of the language and rules of discourse. With respect to cultural rules of use, the emphasis is on appropriateness of communicative acts and the naturalness of speech within given socio-cultural contexts (i.e. kinesics, proxemics and prosodics). With respect to the rules of discourse, the focus is on expressiveness using paralinguistic communication (gestures and mimical effects), and the rules of cohesion (the linking of utterances) and coherence (the logical sequencing of ideas). Illocutionary Force refers to the ability to use socially appropriate illocutionary acts in discourse. These include those acts (i.e. complaining, requesting, inviting, claiming etc) directed at achieving rhetorical effects, mimical effects and feedback.

Strategic Competence refers to the mastery of verbal and non-verbal strategies to compensate for breakdowns in communication and to enhance the effectiveness of communication, by paraphrases, avoiding, gestures, varying intonation, speed or rhythm, repeating, feedback, turn taking and topic switching. Some of these strategies are related to linguistic competence (e.g. paraphrasing, avoiding) and some to sociolinguistic competence (e.g. gesturing, varying intonation, speed or rhythm). Strategic Competence incorporates two communicative components: planning and execution. The planning component retrieves the necessary items from linguistic competence and formulates a plan for realising the communicative goal. The execution component draws on psychophysiological mechanisms to implement the plan in a modality (receptive/productive) and channel (audio/visual) that is appropriate to the

communication goal and context. It is also concerned with such performance factors as coping with the nuisance of background noise or using gap fillers.

State Standards on TFL in the Republic of Uzbekistan (08.05.2013) contain descriptors of the content of teaching and descriptors of the requirements to gradulators of each level. These descriptors must be key points in assessment.

—to track their progress toward meeting goals they have set for themselves in learning English This lecture addresses language assessment and testing and will provide information and resources on how to better support the relationship among teaching, learning, and assessment for EFL teachers in Uzbekistan. The guiding question for this lesson is the following: 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? The goal of this lesson is to familiarize teachers with the two main areas of language assessment and testing known as Assessment for Learning and Assessment of Learning.

Assessment and evaluation are essential components of teaching and learning in English language arts. Without an effective evaluation program, it is impossible to know whether students have learned, whether teaching has been effective, or how best to address student learning needs. The quality of the assessment and evaluation in the educational process has a profound and well-established link to student performance. Research consistently shows that regular monitoring and feedback are essential to improving student learning. What is assessed and evaluated, how it is assessed and evaluated, and how results are communicated results send clear messages to students and others about what is really valued—what is worth learning, how it should be learned, what elements of quality are most important, and how well students are expected to perform. Although the terms assessment and evaluation are often used interchangeably, in actuality they are two parts of the same process. Assessment is the process of gathering evidence of what the learner can do.

Evaluation is the process that follows this collection of data, including analysis and reflection, as well as decisions based on the data. The assessment and evaluation of literacy needs from the learner's perspective is an important part of an instructional program. Although they may say they just want to "learn English," they frequently have very specific learning goals and needs; for example, to be able to read to their children, speak with their learner's teachers, or to get a job. If their needs are not met, they are more likely to drop out than to voice their dissatisfaction. Therefore, using informal, self-assessment tools to gauge learner needs and goals is important. Also important, of course, is using formal assessment tools to gauge learner progress. The needs assessment and evaluation process can be used as the basis for developing curricula and classroom practice that are responsive to learners' needs. It encompasses both what learners know and can do and what they want to learn and be able to do. Learners also need opportunities to evaluate what they have learned.

Bloom's taxonomy is a set of three hierarchical models used to classify educational learning objectives into levels of complexity and specificity. The three lists cover the learning objectives in cognitive, affective and sensory domains. The cognitive domain list has been the primary focus of most traditional education and is frequently used to structure curriculum learning objectives, assessments and activities.

The models were named after Benjamin Bloom, who chaired the committee of educators that devised the taxonomy. He also edited the first volume of the standard text, *Taxonomy of Educational Objectives: The Classification of Educational Goals*.

Knowledge

Knowledge involves recognizing or remembering facts, terms, basic concepts, or answers without necessarily understanding what they mean. Its characteristics may include:

- Knowledge of specifics—terminology, specific facts

- Knowledge of ways and means of dealing with specifics—conventions, trends and sequences, classifications and categories, criteria, methodology
- Knowledge of the universals and abstractions in a field—principles and generalizations, theories and structures

Example: Name three common varieties of apple.

Comprehension

Comprehension involves demonstrating an understanding of facts and ideas by organizing, comparing, translating, interpreting, giving descriptions, and stating the main ideas.

Example: Compare the identifying characteristics of a Golden Delicious apple with a Granny Smith apple.

Application

Application involves using acquired knowledge—solving problems in new situations by applying acquired knowledge, facts, techniques and rules. Learners should be able to use prior knowledge to solve problems, identify connections and relationships and how they apply in new situations.

Example: Would apples prevent scurvy, a disease caused by a deficiency in vitamin C?

Analysis

Analysis involves examining and breaking information into component parts, determining how the parts relate to one another, identifying motives or causes, making inferences, and finding evidence to support generalizations. Its characteristics include:

- Analysis of elements
- Analysis of relationships
- Analysis of organization

Example: List four ways of serving foods made with apples and explain which ones have the highest health benefits. Provide references to support your statements.

Synthesis

Synthesis involves building a structure or pattern from diverse elements; it also refers to the act of putting parts together to form a whole. Its characteristics include:

- Production of a unique communication
- Production of a plan, or proposed set of operations
- Derivation of a set of abstract relations

Example: Convert an "unhealthy" recipe for apple pie to a "healthy" recipe by replacing your choice of ingredients. Explain the health benefits of using the ingredients you chose vs. the original ones.

Evaluation

Evaluation involves presenting and defending opinions by making judgments about information, the validity of ideas, or quality of work based on a set of criteria. Its characteristics include:

- Judgments in terms of internal evidence
- Judgments in terms of external criteria

Example: Which kinds of apples are best for baking a pie, and why?

Functions of assessment

Assessment of language learning serves one of two functions: either to measure learners' proficiency without reference to a language course, or to measure the extent to which they have achieved the goals of a particular repertoire) or recordings in audio or video which demonstrate the learner's oral capacities. It is often presented in a portfolio.

Alternative forms of assessment like portfolios have three advantages over tests: evidence may be collected under non-threatening conditions, which gives it

greater validity as evidence of a learner's true ability; evidence may be derived from the performance of real-world tasks that have been identified as particularly important for the learners in question; and there is greater potential to judge learners' performance holistically and thus to focus on their underlying ability to complete tasks successfully. Tests sometimes entail sanctions while alternative forms of assessment are often intended to enhance learners' motivation .

The European Language Portfolio as an assessment ular programme of learning. Within the latter function it is usual to distinguish between formative and summative assessment. Formative assessment takes place during the course of learning in order to provide learners with feedback on their progress and alert the teacher to any aspects of the course that may need adjustment; it is sometimes referred to as 'assessment for learning'. Summative assessment takes place at the end of the course and seeks to measure overall learning achievement; it is sometimes referred to as 'assessment of learning'.

Forms of assessment

Assessment usually takes one of two forms: either a test that generates a score which can be translated into a statement about the learner's proficiency/achievement, or a compilation of evidence that illustrates what the learner can do in his or her target language. The evidence may take the form of written text (essays, letters or other documents relevant to the learner's target

Instrument

The Council of Europe's European Language Portfolio (ELP) includes checklists of 'I can' descriptors arranged according to the activities and proficiency levels of the *Common European Framework of Reference for Languages* (CEFR). The checklists are used by the learner to identify learning targets and assess learning outcomes. When evidence of proficiency is systematically linked to checklist descriptors, the ELP can complement or replace a test that is linked to the CEFR. The use of the ELP as an assessment instrument requires continuous support from the

teacher, especially as self-assessment will not have played a role in the previous educational experience of many adult migrants.

QUESTIONS FOR DISCUSSING

1. What are the aims and objective of the module?
2. What requirements to TFL are introduced in CEFR?
3. What is the aim of TFL around the world?
4. What is competence?
5. What is communicative\language competence?
6. What requirements to assessment in TFL are presented in CEFR and State Standards on TFL in the Republic of Uzbekistan?

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 you feel yourself tense up when your teacher mentioned a test? The anticipation of the upcoming “moment of truth” provoked feelings of anxiety and self-doubt along with 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 negative emotions a student can experience, and the most common instrument inflicting such fear is the test. You are not likely to view a test as positive, pleasant, or affirming, and, like most ordinary mortals, you intensely wish for a miraculous exemption from the ordeal.

And yet, tests seem as unavoidable as tomorrow’s sunrise in virtually all educational settings around the world. Courses of study in every discipline are marked by these periodic milestones of progress (or sometimes, in the perception of the learner, confirmations of inadequacy) that have become conventional methods of measurement. The gate-keeping function of tests—from classroom achievement tests to large-scale standardized tests—has become an acceptable norm.

Now, just for fun, take the following quiz. All five of the words are found in standard 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 thrive
 - 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 feel when 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 did worse 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 an appropriate example of classroom-based achievement testing, nor is it intended to be. It was designed to be overly difficult, to offer you no opportunity to use contextual clues, and to give you little chance of deciphering the words from your knowledge of English. It's simply an illustration of how tests make us feel much of the 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? Can they become an integral part of a student's ongoing classroom development? Can they 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 motivating assessment procedures that are appropriate for their context and designed to offer constructive 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 how do 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 according to the scores of tests, mainly with questions about Confucian philosophy and 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.

Factors in the test:

- 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.

Factors related to students:

- 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:

- is not too expensive,

- 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.	Past tense	4
		Asking wh-questions	3
Sociolinguistic competence	The learner’s ability to express	Introducing themselves	2

	accurate meaning using specific functions.	Complaining	2
		Responding to a complaint	1
Discourse competence	The ability to produce coherent text.	Using pronouns	4
		Using conjunctions	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 Mr Murdstone?

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 for a 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 obvious. They are to be approached *with* caution!

4. Gap-filling and completion

This 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 difficulties should be taken into consideration while creating tests?

IV. AMALIY MASHG'ULOTLAR

PRACTICAL 1: ASSESSMENT FOR LEARNING: GATHERING DATA

Plan:

- Differences of testing, assessment, and evaluation;
- The Assessment for Learning construct;
- Ways to use self-, peer-, and continuous assessments

Keywords: formative and summative assessment; testing; assessment, evaluation; diagnostic assessment; self and peer-assessment.

Assessment is an inseparable part of teaching, influencing the decisions teachers make, the actions they take, and the suggestions they offer for student learning. Classroom assessment, specifically, refers to methods and procedures used by the language teacher for gathering, analysing, interpreting and using information about student's language abilities for decision making purposes. The terms often identified in the language assessment and testing literature to refer to assessment of students are formative and summative assessment. In this section, we focus on formative assessment or what is accurately – and more recently – identified as *Assessment for Learning (AfL)*. AfL is about assessments, both formal and informal, which provide information for teachers and students *to identify the next steps for learning*.

Think about the following:

- 1) How do student assessment results influence the way we teach, design syllabi, and/or approach the language classroom context?
- 2) Should we involve students in the language testing and assessment process; if so, how?

UZBEK VIGNETTES

An English language teacher at a prestigious 'foreign language university' discussed assessment at her university. She explained the following:

"I have observed that some teachers diagnose their student's language levels at the beginning of the course and some do not. One teacher explained the course she teaches does not need to be diagnosed as the content is new and therefore she simply begins by providing the planned input. My other colleague insists on checking students' language level (e.g., knowledge and skills) before providing any instructions. However, she struggles with choosing proper ways to learn about the student's abilities."

REFLECTION

Think about the case above. Are the issues raised in it familiar to you in your teaching context? How important is assessing students before a course begins? How do you think a teacher could gather information about students?

KEY CONCEPTS

There are eight key concepts in this section: testing; assessment; evaluation; Assessment-*for-Learning*; diagnostic assessment; peer-assessment; self-assessment; and continuous assessment. We will briefly explain each one below.

Testing, Assessment, Evaluation. Before learning about any concepts in the field of language assessment and testing, a teacher should be cognizant of the differences among testing, assessment, and evaluation in order to use each in the right place in the right time. There are three popular terms in the literature that are sometimes misunderstood and used interchangeably. They are *testing*, *assessment*, and *evaluation*. However, each of these terms has a different meaning, scope, and function. A **test** is a specific technique for gathering information about students' knowledge or abilities while **assessment** is a broader process involving various ways of collecting data including the use of tests. Therefore, a test is a part and a means of assessment; it serves as a tool in this process. **Evaluation**, in its turn, is the broadest concept out of the three, which encompasses both assessment and a test. It looks at the whole picture of teaching context and functions to make decisions based on the obtained evidence regarding the whole educational setting.

Assessment for learning: Why? How? When? Imagine a doctor and a patient situation: a doctor cannot start treating a patient until one does not learn about the background and the current health condition of the patient. Based on a diagnosis, the doctor then can decide what to start from and how to treat the patient. A doctor gathers information about his/her patient in order to make decisions for treating an illness. In the same manner teachers first need to gather information about their students before commencing the learning process. Those teachers who immediately provide instructions may not be aware of learners' background knowledge and the strengths and weaknesses of their students, which could have an effect on what is taught. This is called *Assessment for Learning*. From its name, we can see that *Assessment for Learning* or more widely known as *formative assessment* is used with the intention to form knowledge, skills and competencies to develop further (Brown & Abeywickrama, 2010) and reach expected results. Information gathered through any assessment, even summative can be used formatively, thus determining what has been achieved, what needs more attention and how to remediate a certain issue. The effective implementation of *Assessment for Learning* depends on teachers' knowledge of its

principles and techniques. It is important to consider the **aim** for a certain formative assessment instrument (what information is needed), the **way** in which the assessment is administered (what method is used) and finally **actions** to be taken afterwards (how the gathered information is then used to make changes in teaching and learning). Unfortunately, teachers tend to look at the students' results obtained at formal assessment period, however, every day of classroom communication is a source for teachers to make their decisions how to teach, where to remediate, and when to act on the issues identified.

Diagnostic assessment – is an act of gathering information which helps teachers identify learners' strengths and weaknesses with regards to course content (i.e., skills and knowledge) and is tied directly to the course syllabus. Diagnostic assessments are often used before teaching commences. Using diagnostic assessments can support a teacher's decisions about the content that will be taught as well as the approach of assessing that content. The focus of diagnosis is most important. For example, if you teach how to speak English during an interview, then, you should assess student's abilities during a *spoken* interview. If the student writes a paragraph about themselves, then, you are measuring a different construct than the oral interview – although the 'topic' might be similar. Diagnostic assessments need to match what is expected to be taught in the course syllabus. Thus, they can provide valuable data about what should be given more attention and what students already know, thus helping teachers to use class time efficiently.

Gathering information about students can help teachers to modify the way in which they teach, influence the decisions made and prevent from the wrong focus and waste of time. At the same time, it is important that teachers are aware of what they need to look for when diagnosing and most importantly what to do with the data collected. Teachers should be able to communicate the results to students, administration and other stakeholders in the appropriate way, targeting improvement.

Self and Peer-assessment. One of the ways to involve students in assessment is using peer-assessment. Students can develop their own understanding of how to measure one's or others' knowledge. In other words, they become active participants of an assessment process and learn what they need to take into account when assessing themselves or their peers. For example, you might provide them with a set of questions and by answering students can realize where they are in their learning; what they know and what they need to learn more about. Then, at the end of the course you can provide the same measurement tool and learners can compare what they knew before and after the course. For example, below is a traditional assessment tool that a teacher can use before and after a course. Dr. Kathleen M. Bailey from the Middlebury Institute of

International Studies at Monterey used this self-assessment mechanism for her course on Language Assessment and Testing:

LANGUAGE ASSESSMENT AND TESTING SELF-ASSESSMENT	
Name: _____	Date: _____
Directions: Complete each of the following items by marking a “T” for True or an “F” for False.	
_____ I understand and can explain the difference between direct, indirect, and semi-direct tests.	
_____ I understand and can explain the difference between norm-referenced and criterion-referenced tests	
_____ I understand and can explain the difference between analytic, holistic, primary trait and objective scoring of writing (and speaking)...	

She provided students with all the topics of the course and the learners marked T (true) or F (false) if they are (or not) able to understand and can explain the statements. Then, at the end of the course Dr. Bailey provided the same handout to students and they re-filled it in. She then distributed the student’s first comments with their last comments so that they can see if they grew in knowledge and skills during study.

In addition to self-assessments, you could also provide learners criteria to help them assess their colleague’s work (i.e., peer-assessments). In the same way, students can look at each other’ written works and/or observe them speaking and mark the performances by following the procedures set by a teacher. For example, if you teach a speaking class that focuses on giving speeches, you can provide a peer-assessment tool that could support the conclusion of a speech.

PEER-ASSESSMENT FOR SPEECHES, CONCLUSION

Speaker’s Name: _____	Yes	No
1.) Does the author (speaker) signal the end of the speech?		
2.) Is the central idea reinforced by one of the following: Summarizing the main points; ending with a quotation;		

making a dramatic statement; or referring to the introduction?		
3.) Is the conclusion limited to 5-10 percent of the entire speech?		
4.) Has the author worked on the language of the conclusion in detail?		
<p>1.) What are the positive qualities of the conclusion? 2.) What suggestions do you have for the author?</p>		

By including peer-assessments, teachers support inclusive education, which enhances students' involvement and role in classroom assessment. Therefore, the two (i.e., self and peer-assessments) are important constituents in *Assessment for Learning*.

Continuous Assessment (CA) – involves a procedure of collecting evidence about students' knowledge and achievements throughout the course which then results in one final score at the end. CA is different from a final achievement test. The final test reflects student performance only at the end of the course under certain constrained conditions. One important features of continuous assessment is a form of assessment is that it **allows teachers track the progress of students over a period**. Another strong quality is that CA results are generated based on various instances, tasks, and activities allowing both teachers and students **to check and recheck strengths and weaknesses**. There are two ways of approaching continuous assessment in Uzbekistan: accumulation of scores to reach the final top and deduction of average results based on all the obtained scores. Teachers have an opportunity to choose which of the assessment instances will be recorded and which will not be; this choice can even be decided upon mutually, among teachers and students. When students are involved in the process, assessment can become less intimidating, as they know that they have numerous chances to improve their scores.

ACTION

(1) Please discuss the approaches to gather information described above by answering the following question: How do you think the approaches can support teaching and learning in your class?

(2) Also, using your own teaching syllabus, think of the ways you can gather information about your students. Explain to a colleague how you can use this information?

SUMMARY

Assessment *for* Learning provides opportunities for both teachers and students to **form their further path in learning**. Used purposefully and proactively, Assessment *for* Learning can bring about significant changes to all stakeholders involved in the assessments. It is important to remember that assessment never ceases; it exists in everything a teacher does; in other words, anything that is done before and during class sections can provide valuable information to a teacher to act upon.

HOMEWORK TASK ELEVEN

Based on what you have learned in this section, develop or choose an available diagnostic assessment tool that will measure the concepts you want to teach in the lesson plan that you chose in Homework Task. Thus, explain the brief diagnostic assessment you will use to measure the constructs BEFORE the class. Finally, explain how you will use the information to make informed decisions about your lesson plan (2 pages total).

4. PRACTICAL 2: ASSESSMENT FOR LEARNING: USING GATHERED DATA

Plan:

- How to use gathered data to make changes to lesson plans, teaching materials, and syllabus;
- How to incorporate dynamic assessment into teaching practice;
- How to adjust teaching instructions based on the obtained data about students;
- Difference between the features of feedback and feedforward.

Keywords: Using collected Data; Dynamic Assessment; Feedback; and Feedforward.

INTRODUCTION AND OVERVIEW

Data about student's language abilities collected either prior to or during the teaching process can inform a teacher about further actions to be taken in his or her classes. How such data are used depends on the teacher's intention and focus. Sometimes, understanding student's language skills happens in interaction – either in the oral or written form – between a student and teacher. Thus, understanding how teacher's feedback plays a role in the student's learning process is also a very important

area for language teachers to consider. Teacher feedback plays a significant role in Assessment *for* Learning and has been identified in the literature as *feedforward* – *feedforward* is anticipatory in that it provides students with focused information that they can use in the next steps of the task, for future learning, or for specific outcomes (Hattie & Timperley, 2017). Bourke and Mentis (2013) argue that *feedforward* is an approach that “all students benefit from” (p. 857).

Think about the following:

- 1) What different types of data can you gather about student’s language abilities? (Reflect on Section 3.1) and what kinds of decisions can you make after you collect that information?
- 2) How can student assessment results influence the way we teach? Please be as specific as you can.
- 3) How should feedback be delivered by a teacher to help students progress further?

UZBEK VIGNETTE

Case 1

At the beginning of every academic year we receive a new cohort and the first thing we want to know is their level. Administrations often ask the departments to test the students. The students are tested and classes start. However, what happens after? Well, after this assessment, there is nothing. These test results are stored to serve no use. Both curriculum and syllabus are predetermined. The textbooks and teaching materials are prepared with accordance to set standards and regulations. All the decisions are made top-down.

Case 2

I teach writing skills. The main challenge in my practice is that my students do not seem to progress as they repeat the same mistakes. I feel frustrated every time I have to tell them the same rules and advice. I check their essays and nothing changes, they seem to keep making these mistakes. I always try to teach them using the syllabus and the teaching materials that have been approved by my department. But I feel there should be a way to address my problem of student recurring issues, there must be a solution...

REFLECTION

Think about the cases above. What are the ways the gathered data can inform a teacher about learners? How can assessment data be used to benefit teaching process? Have you ever had an experience of using assessment results for making changes in your syllabus, teaching materials, or the way you teach?

KEY CONCEPTS

There are four key concepts in this section: Using collected Data; Dynamic Assessment; Feedback; and Feedforward. We will briefly explain each one below.

Using Collected Data. A relatively new concept for the Uzbek EFL context is teaching and learning based on the data collected about students. It is called *data-driven learning*. Data-driven learning allows teachers to address students' immediate needs and remediate the issues they have. We frequently just rely on our assumptions in our choices of input. But can we really be sure what our students need without collecting the required information about their current state of knowledge.

Gathered data can inform language teachers about:

- the current state of students' knowledge and skills
- aspects that need to be given more attention to during class and/or after
- aspects that can be given less attention during class
- additional input that is required to remediate the identified lack of knowledge or skill
- sequence of the teaching input (i.e., you might want to change the teaching process)
- corrections to lesson plans
- changes should be made to the curriculum, syllabus and teaching materials (e.g., this information can be discussed with the members of your department.)

Reliable data about students are not the results of one test only; teachers are to collect their data continuously to check whether they are right about certain conclusions they have drawn. This approach in English as a Foreign Language pedagogy relies on the use of collection of computer readable texts which is called corpus. For example, many students' essays can be gathered to learn about the most common issues in learners' writing. This information can then be a foundation for you to create teaching materials, modify your teaching approaches, and make plans in the teaching content and focus.

Dynamic Assessment. One of popular forms of Assessment *for* Learning is Dynamic Assessment. It is quite new for the Uzbek context but very widespread around the world. The most useful feature of dynamic assessment allows a teacher to observe changes happening during the process of assessment because of his/her influence on student performance. Dynamic Assessment relies on interaction between an assessor and a student and is based on principles from Vygotsky's sociocultural theory. According to theory, students can perform at a certain level at a test; however, given support they can reach their highest potential within the zone of proximal

development (ZPD) – “the distance between the actual development level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers” (Vygotsky, 1978, p. 86). This quotation from Vygotsky emphasizes how an expert or someone more capable is needed to help someone learn. This happens as a result of interaction between a teacher and a student. Dynamic assessment is different from psychometric tests (i.e., tests that produce a numerical score at the end) for example, a multiple-choice test, which allows observing students’ performance in one instance only. In contrast, by the help of dynamic assessment, we cannot see whether this student can perform better if more support is provided or conditions are changed. Dynamic assessment in this regard is more flexible while psychometric is static. As the interaction continuous, the assessor can see some features and skills, which were not initially demonstrated, emerging.

As a teacher, you can use dynamic assessment to scaffold your students in the process of learning. For example, you can give them a speaking task and as each student responds, you try to notice what a student already knows and where he or she has difficulties. However, you do not stop at the point of identifying an issue you keep the conversation going to see how further you can take the student by clarifying what he or she has said and extending the opportunities to arrive at better performance than initially. It’s always useful to know how far your students can go. This can influence planning the further lessons and teaching materials, as you already learn about their potential.

Feedback and Feedforward. One of the key constituents in data collection is providing valuable feedback. Both *written* and *oral* feedback can be useful input for students to work on their weaknesses. It is important that feedback is timely (so that students can act on it to make changes during the learning process), clear (so that students understand and feel confident about actions to be taken) and future-oriented (so that students feel that the efforts that they are going to make are meaningful for and influential on their academic success) Thus, teacher’s role in this process is not only indicating the problem but also suggesting ways to overcome and address the identified problem. In the literature this is called *feedforward* i.e. advice and instructions from a teacher to students for further improvement. *Feedforward* might be dependent on not only how comments are written but also on what students do with them.

Think how you can use the results of assessments (classroom, midterm, etc.) to make changes in your teaching. Describe the challenges that you might face? Then write an action plan focusing on the followings:

- describe an assessment tool used to gather data about students (what the task or activity involves)
- explain how can the gathered information be used by you to make changes
- speculate on possible issues that might arise in this process based on your experience in your teaching context
- provide some plausible ways of addressing these issues and making the implementation of data effective and useful

SUMMARY

Using assessment data is central to evidence-based discovery of issues. If teachers want to see improvements in the way they teach as well as the way students learn, they need to be conscious of immediate data that they have at hand. One of the primary measures taken in response to identified issues is feedback, which should aim *forward* to make an impact on students' skills and knowledge enhancement. Therefore, to make this impact even stronger, assessment forms such as dynamic assessment are critical at both identifying issues, digging for more evidence as the assessment process goes and finally remediating them.

HOMEWORK TASK TWELVE

Please refer to the Diagnostic Assessment Tool you chose for Homework Task Eleven. In a page, please explain how you will use the information you learn to make informed decisions about your lesson plan DURING the lesson. For instance, are there key areas you are interested in that might cause some confusion for your students?

5. PRACTICAL 3: ASSESSMENT OF LEARNING: PRINCIPLES OF ASSESSMENT

Plan:

- Distinction between Assessment *for* Learning and Assessment *of* Learning;
- Principles of language assessment;
- How to apply key principles of assessment to the quality of assessment tools;
- 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 these concerns. 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 often that what I teach during the semester is not 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 and concerned 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; 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. 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-referenced and 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 you will 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 choice assessments, 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. This quality 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.

HOMETASK

REFLECTION
Which of the above test qualities are the most difficult to apply? Which have you never used? Why? Create sample test due to the subject you teach which can meet the quality requirements.

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

Table 12. Types of Assessments.	
<u>Assessment types</u>	<u>Main purpose</u>
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

TYPES of TESTING

1. Direct testing:

Testing is said to be direct when the student is required to perform directly the skill which we wish to measure. E.g. we ask students to write compositions if we want to know how well they can write compositions. We ask them to speak if we want to know how well they can pronounce a language.

2. Indirect testing:

Indirect testing attempts to measure the abilities which underlie the skills in which we are interested. E.g. we test pronunciation ability by asking students to identify pairs of words which rhyme with each other.

3. Objective testing:

It doesn't require judgement on the part of the scorer because scoring here is objective. It won't change even if the scorer has been changed. Multiple choice test is an example of this kind of tests.

4. Subjective testing:

It requires judgement on the part of the scorer because scoring here is subjective. The grades in subjective testing depend on the impressions of the scorer. These impressions are not the same among different scorers. Scoring of a composition is an example of this kind of testing.

5. Discrete point testing:

It refers to the testing of one element at a time, item by item. This kind of testing is always indirect. Each testing involves a particular item. Testing particular grammatical structures is an example of this kind of testing.

6. Integrative testing:

It includes many language elements in the completion of a task. It might involve writing a composition, taking notes while listening to a text and completing a cloze passage.

7. Norm-referenced testing:

This kind of testing relates one student's performance to that of other students. We don't say that student is capable of doing well in the language but we say the student gained a score that placed him/her in the top five students who have taken the same test.

8. Criterion-referenced testing:

The purpose of this kind of testing is to classify students according whether they are able to perform some tasks satisfactorily. Who perform the tasks satisfactorily 'pass', those who don't, 'fail'. We measure students' progress in relation to meaningful criteria.

HOMETASK

REFLECTION

Which of the above are the most common type of tests in your teaching context?

Which have you never used? Why?

7. PRACTICAL 5: DESIGNING TESTS

Plan:

- How to start test designing process;
- Determining the purpose of a test;
- Designing clear, unambiguous objectives;
- Drawing up test specifications;
- Designing multiple-choice items.

Key words: test objectives; test specifications; grading; scoring; feedback; multiple-choice questions; distractor.

In this lesson, you will draw on those foundations and tools to begin the process of designing tests or revising existing tests. So for now, for classroom purposes, let's start the process by asking some critical questions.

1. What is the purpose of the test? Why are you creating this test, or why was it created by, say, a textbook writer? What is its significance relative to your course (for example, to evaluate overall proficiency or place a student in a course)? How important is the test compared to other student performance? What will its impact be on you and your students before and after the assessment? Once you have established the major purpose of a test, it then becomes easier to specify its objectives.

2. What are the objectives of the test? What exactly are you trying to find out? Establishing appropriate objectives involves a number of issues, from relatively simple ones about forms and functions covered in a course unit to much more complex ones about constructs to be represented on the test. Included here are decisions about what language abilities are to be assessed.

3. How will the test specifications reflect both the purpose and the objectives? To design or evaluate a test, you must make sure that the test has a structure that logically follows from the unit or lesson it is testing. The class objectives should be present in the test through appropriate task types and weights, a logical sequence, and a variety of tasks.

4. How will the test item types (tasks) be selected and the separate items arranged? The tasks need to be practical. To have content validity, they should also mirror tasks of the course, lesson, or segment. They should also be authentic, with a progression biased for best performance. Finally, the tasks must be ones that can be evaluated reliably by the teacher or scorer.

5. In administering the test, what details should I attend to in order to help students achieve optimal performance? Once the test has been created and is ready to administer, students need to feel well prepared for their performance. An otherwise

effective, valid test might fail to reach its goal if the conditions for test taking are inadequately established. How will you reduce unnecessary anxiety in students, raise their confidence, and help them view the test as an opportunity to learn?

6. What kind of scoring, grading, and/or feedback is expected? The appropriate form of feedback on tests will vary, depending on their purpose. For every test, the way results are reported is an important consideration. Under some circumstances a letter grade or a holistic score may be appropriate; other circumstances may require that a teacher offer substantive washback to the learner.

These six questions should form the basis of your approach to designing, administering, and making maximum use of tests in your classroom.

ASSESSMENT SCENARIOS

For the purposes of making practical applications in this lesson, we will consider a scenario as we proceed through the six steps for designing an assessment. These common classroom contexts should enable you to identify with real-world assessment situations.

Scenario 1: Reading Quiz

The first context is an intermediate-level class for secondary school students in an English class in Uzbekistan. The students have been assigned a two-page short story to read on their own for homework, and the teacher has decided to begin class the next day with a brief “pop quiz”: 10 short-answer written comprehension questions.

The quiz will (a) give students a sense of how well they understood the story and (b) act as a starting point for a teacher-led discussion on each of the items. Results of the quiz will not be recorded in the teacher’s record book.

Scenario 2: Midterm Essay

In a writing course in a university in Uzbekistan, students at the advanced level have been working for half a semester on writing essays, mostly narrative and description essays. In the second half of the course, students will move on to cause/effect, argument, and opinion essays.

The midterm essay is an opportunity for students to demonstrate their ability to write a coherent essay with relatively few grammatical and rhetorical errors. The essay will be given in class (in a 80-minute class period). The students do not know the topic ahead of time but are allowed to use a bilingual dictionary to look up words or spelling. The curriculum specifies quality of writing over quantity. The teacher will read essays over the weekend and make comments but not give a grade or a score. During the next

week, there will be peer conferences with the goal of each student to revise his or her essay, followed by a student-teacher conference after a revision has been turned in.

DETERMINING THE PURPOSE OF A TEST

You may think that every test you devise must be a wonderfully innovative instrument that impresses your colleagues and students alike. Not so. First, new and innovative testing formats take a lot of effort to design and a long time to refine through trial and error. Second, traditional testing techniques can, with a little creativity, conform to the spirit of an interactive, communicative language curriculum. Your best course of action as a new teacher is to work within the guidelines of accepted, known, traditional testing techniques. Slowly, with experience, you can attempt bolder designs. In that spirit, let's consider some practical steps in constructing classroom tests.

The first and perhaps most important step in designing any sort of classroom assessment (or in determining the appropriateness of an existing test) is to step back and consider the overall purpose of the exercise that your students are about to perform. The purpose of an assessment is what Bachman and Palmer (1996, pp. 17-19) refer to as test usefulness or, very simply put, to what use will you put an assessment? Consider the checklist on the next page for determining purpose and usefulness of an assessment:

PURPOSE AND USEFULNESS CHECKLIST

- 1. Do I need to administer a test at this point in my course? If so, what purpose will it serve the students and/or me?
- 2. What is its significance relative to my course?
- 3. Is it simply an expected way to mark the end of a lesson, unit, or period of time?
- 4. How important is it compared to other student performance?
- 5. Do I want to use results to determine if my students have met certain predetermined curricular standards?
- 6. Do I genuinely want students to be recipients of beneficial washback?
- 7. Will I use the results as a means to allocate my own pedagogical efforts in the days or weeks to follow?
- 8. What will its impact be on what I do, and what students do, before and the test?

Now look back at each of the two assessment scenarios and think about the purpose of each. Before reading on, do some personal brainstorming on just how the eight questions in the checklist will be answered for each scenario.

Reading Quiz. To start your thinking process, let's look at the purpose of the first scenario—the reading quiz. The quiz is designed to be an instructional tool to guide classroom discussion for one classroom period. Its significance is minor but not trivial when viewed against the backdrop of the whole course. Because it is a surprise test and a tool for teaching and self-assessment, the results will justifiably not be recorded, and so one student's performance compared to others is irrelevant. It is entirely formative in nature, with the almost exclusive purpose of providing beneficial washback. Forcing students to think independently about the reading passage allows them to see areas of strength and weakness in their comprehension skills.

Can you now consider the second scenario and think about the overall purpose of it given the context described and the information given? Your understanding of the purpose of an assessment procedure governs, to a great extent, the next steps you take in identifying clear objectives, designing test specifications, constructing tasks, and determining scoring and reporting criteria.

DESIGNING CLEAR, UNAMBIGUOUS OBJECTIVES

In addition to knowing the purpose of the test you're creating, you need to know as specifically as possible what it is you want to test.

Remember that every curriculum should have appropriately framed, assessable objectives, that is, objectives that are stated in terms of overt performance by students.

Thus an objective that states "Students will learn tag questions" or simply names the grammatical focus of "tag questions" is not testable. You don't know whether students should be able to understand them in spoken or written language, or whether they should be able to produce them orally or in writing. Nor do you know in what context (a conversation? an essay? an academic lecture?) those linguistic forms should be used. Your first task in designing a test, then, is to determine appropriate objectives, stated as explicitly as possible.

DRAWING UP TEST SPECIFICATIONS

Test specifications (specs) for classroom use can be an outline of your test—what it will "look like." Think of your test specs as a blueprint of the test that include the following:

- a description of its content

- item types (methods, such as multiple-choice, cloze, etc.)
- tasks (e.g., written essay, reading a short passage, etc.)
- skills to be included
- how the test will be scored
- how it will be reported to students

For classroom purposes (Davidson & Lynch, 2002), the specs are your guiding plan for designing an instrument that effectively fulfills your desired principles, especially validity. It's important to note here that for large-scale standardized tests that are intended to be widely distributed and therefore are broadly generalized, test specifications are much more formal and detailed (Spaan, 2006). They are also usually confidential so that the institution that is designing the test can ensure the validity of subsequent forms of a test. Such secrecy is not a part of classroom assessment; in fact, one facet of effectively preparing students for a test is giving them a clear picture of the type of items and tasks they will encounter.

The more meticulous you are in specifying details of an assessment procedure, the better off you will ultimately be in providing your students with appropriate opportunities to perform well. Other test specs may look more complex. Suppose you have two or more skills that students will perform. In that case, as described below, test specs will involve several elicitation techniques and a number of categories of student responses. In all cases, specifications are not the actual test items or tasks but rather a description of limitations, boundaries, directions, and other details that you will adhere to. The next step is to design tasks and items that fit the specs.

DEVISING TEST ITEMS

At this point it is important to note that test development is not always a clear, linear process. In reality, test design usually involves a number of “loops” as you discover problems and other shortcomings. With that fair warning, let's look at the midterm essay scenario again and this time more specifically in terms of item design.

Midterm Essay. The single test task described in Scenario 2 includes the prompt, directions, and evaluation criteria. This could be one of the easiest kinds of test tasks to create, because only one “item” is involved, student responses are open ended, and evaluation criteria have already been covered well in previous instruction. Let's see what the prompt might look like:

Choose one of the following topics. Write an essay of about three paragraphs on the topic you have chosen. Assume that you are writing this to share with your classmates.

A. Based on the changes and future developments we have read about and discussed in class, invent a possible job of the future. Use your imagination.

Write an essay that describes this job.

B. Describe your present job (or profession) or the job of a parent or a friend that you know quite well.

You have 90 minutes to complete your work. You may want to begin with a very quick first draft, an outline, or some freewriting and then write a final draft.

For your final draft, do your best to write a legible, neat essay. However, you will have an opportunity to revise, rewrite, and correct this essay, so don't worry about a few words or phrases or sentences that may be crossed out.

The criteria for evaluation will be

X Content

X Organization

X Rhetorical discourse (coherence, cohesion, appropriateness, etc.)

X Grammar/mechanics

Given the constraints of the curriculum and the context described for Scenario 2, do you feel that this prompt is effective? Does it adhere to the five principles of practicality, reliability, validity (in various forms), authenticity, and washback potential?

DESIGNING MULTIPLE-CHOICE ITEMS

Multiple-choice items, which may on the surface appear to be simple items to construct, are actually very difficult to design correctly. Hughes (2003, pp- 76-78) cautions against a number of weaknesses of multiple-choice items:

- The technique tests only recognition knowledge.
- Guessing may have a considerable effect on test scores.
- The technique severely restricts what can be tested.
- It is very difficult to write successful items.
- Beneficial washback may be minimal.
- Cheating may be facilitated.

The two principles that stand out in support of multiple-choice formats are, of course, practicality and reliability. With their predetermined correct responses and

time-saving scoring procedures, multiple-choice items offer overworked teachers the tempting possibility of an easy and consistent process of scoring and grading. But is the preparation phase worth the effort? Sometimes it is, but you might spend even more time designing such items than you save in grading the test. Of course, if your objective is to design a large-scale standardized test for repeated administrations, then a multiple-choice format does indeed become viable.

As you face the task of designing the listening comprehension section of the fifth-grade English exam, let's first consider some important terminology.

1. Multiple-choice items are all receptive, or selective response, items in that the test-taker chooses from a set of responses (commonly called a supply type of response) rather than creating a response. Other receptive item types include true/false questions and matching lists. (In the discussion here, the guidelines apply primarily to multiple-choice item types and not necessarily to other receptive types.)
2. Every multiple-choice item has a stem (the "body" of the item that presents a stimulus) and several (usually between three and five) options or alternatives to choose from.
3. One of those options, the key, is the correct response, whereas the others serve as distractors.

Because there will be occasions when multiple-choice items are appropriate, consider the following four guidelines for designing multiple-choice items for both classroom-based and large-scale situations (adapted from Gronlund, 1998, pp. 60-75, and J. D. Brown, 2005, pp. 48-50).

1. Design each item to measure a single objective.

Consider the following item from a secondary school class in English at the intermediate level. The objective is wh- questions:

Test-takers hear: Where did George go after the party last night?

Test-takers read:

- A. Yes, he did.*
- B. because he was tired*
- C. to Elaine's place for another party*
- D. around eleven o'clock*

Distractor A is designed to ascertain that the student knows the difference between an answer to a wh- question and a yes/no question. Distractors B and D, as well as the key item, C, test comprehension of the meaning of where as opposed to why and when. The objective has been directly addressed.

On the other hand, here is an item that was designed to test recognition of the correct word order of indirect questions:

Excuse me, do you know _____ ?

A. where is the post office

B. where the post office is

C. where post office is

Distractor A is designed to lure students who don't know how to frame indirect questions and therefore serves as an efficient distractor. But what does distractor C actually measure? In fact, the missing definite article {the} is what J. D. Brown (2005) calls an "unintentional clue"(p. 48)—a flaw that could cause the test taker to eliminate C automatically. In the process, no assessment has been made of indirect questions in this distractor. Can you think of a better distractor for C that would focus more clearly on the objective?

2. State both stem and options as simply and directly as possible.

We're sometimes tempted to make multiple-choice items too wordy. A good rule of thumb is to get directly to the point. Here's a negative example:

My eyesight has really been deteriorating lately. I wonder if I need glasses. I think I'd better go to the _____ to have my eyes checked.

A. pediatrician

B. dermatologist

C. optometrist

You might argue that the first two sentences of this item give it some authenticity and accomplish a bit of schema setting. But if you simply want a student to identify the type of medical professional that deals with eyesight issues, those sentences are superfluous. Moreover, by lengthening the stem, you have introduced a potentially confounding lexical item, deteriorate, that could distract the student unnecessarily.

Another rule of succinctness is to remove needless redundancy from your options. In the following item, "which were" is repeated in all three options. It should be placed in the stem to keep the item as succinct as possible.

We went to visit the temples, _____ fascinating.

A. which were beautiful

B. which were especially

C. which were holy

3. Make certain that the intended answer is clearly the only correct one.

In the test item described earlier, which turned out to be suitable, a draft of the item appeared as follows:

Test-takers hear: Where did George go after the party last night?

Test-takers read: A. Yes, he did.

B. because he was tired

C. to Elaine's place for another party

D. He went home around eleven o'clock.

A quick consideration of distractor D reveals that it is a plausible answer (because of the mention of “home”), along with the intended key, C. Eliminating unintended possible answers is often the most difficult problem of designing multiple-choice items. With only a minimum of context in each stem, a wide variety of responses may be perceived as correct.

As you look over the items, are there some that need to be revised before you finalize them? In revising your draft, ask yourself the following questions: Suggestions for revising your test:

1. Are the directions to each section absolutely clear?
2. Is there an example item for each section? If not, are the directions and format so familiar to students that they will clearly understand the tasks they are being asked to perform?
3. Does each item measure a specified objective?
4. Is there a single correct answer for each question?
5. Is each item stated in clear, simple language?
6. Does each multiple-choice item have appropriate distractors; that is, are the wrong items clearly wrong and yet sufficiently “alluring” that they aren’t ridiculously easy?
7. Is the difficulty of each item appropriate for your students?
8. Is the language of each item sufficiently authentic?
9. Is there a balance between easy and difficult items?

10. Do the sum of the items and the test as a whole adequately reflect the learning objectives?

ADMINISTERING THE TEST

The moment has arrived. You have designed your test based on your carefully considered purposes, objectives, and specs. Could anything now go awry in these best laid plans? Of course, you know the answer is yes. So consider some of the measures you can take to ensure that the actual administration of the test accomplishes everything you want it to. Here's a list of pointers:

Pre-test considerations (the day before the in-class essay):

1. Provide appropriate pre-test information on
 - a. the conditions for the test (time limits, no portable electronics, breaks, etc.)
 - b. materials that students should bring with them
 - c. the kinds of items (item types) that will be on the test
 - d. suggestions of strategies for optimal performance
 - e. evaluation criteria (rubrics, show benchmark samples)
2. Offer a review of components of narrative and description essays
3. Give students a chance to ask any questions, and provide responses

In this lesson, guidelines and tools were provided to enable you to address the five questions posed at the outset: (a) how to determine the purpose of the test, (b) how to state objectives, (c) how to design test specifications, (d) how to design or select test tasks, including evaluating those tasks with item indices, and (e) how to begin to address scoring and grading. This five-part template, shown in Figure 3.2, can serve as a pattern as you design classroom tests.

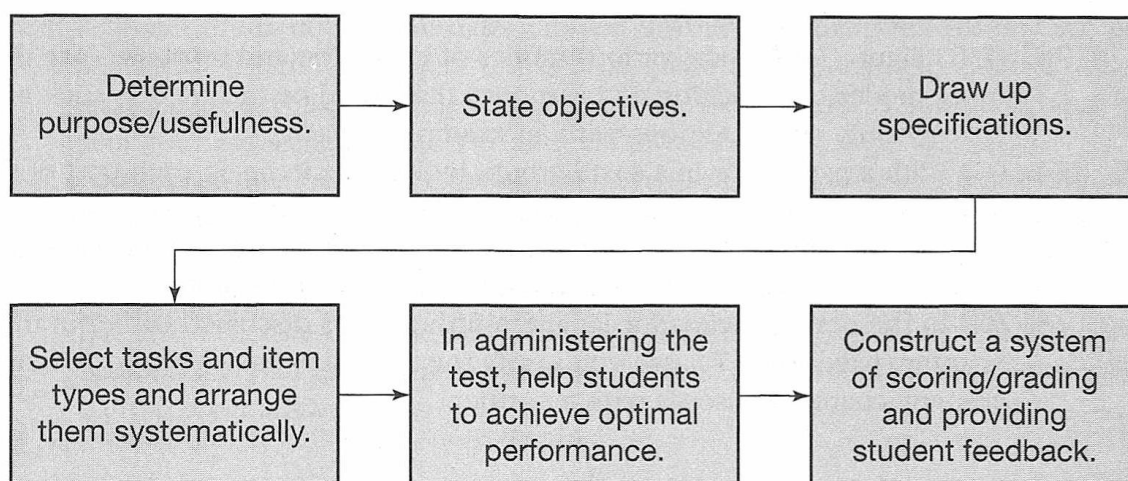


Figure 3.2. Steps to designing an effective test

HOMETASK

- 1) The first issue discussed in this lesson was determining the purpose of a proposed test, and a checklist was offered. In small groups, have one or two members share an experience they had either taking or giving a test, then systematically discuss the probable answers to each item on the checklist. The group may be able to help solve certain problems or dilemmas that came up. Report back to the class any notable surprises or questions that were unresolved.
- 2) Select a language class in your immediate environment for the following project: In small groups, design a multiple-choice test for an authentic reading material. Follow the guidelines in this lesson for developing an assessment procedure. When it is completed, present your assessment project to the rest of the class.

8. PRACTICAL 6: 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 specifications can serve for a long time and be revised when needed (Similar to the design and implementation 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) 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 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 issues and the features of the test that the test developers otherwise might never be able to spot themselves. 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 set 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 every 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 testing instruments, which means that no matter how much effort is made in considerations 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 consultation 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 comprehensive 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.

HOMEWORK TASK

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 that will 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 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?)

9. PRACTICAL 7: USING STATISTICS: OBJECTIVELY SCORED ITEMS

Plan:

- Significance of assessment statistics for teaching and learning;
- How to conduct basic statistical analysis of results from objectively scored items;
- Interpretation of the results of conducted statistics.

Key words: Objectively-scored assessments; Measures of Central Tendency; Mean; Median; Mode; Measures of Dispersion; Standard Deviation.

INTRODUCTION AND OVERVIEW

Most teachers try to steer clear of seemingly complicated mathematical calculations of statistical analysis. However, this process is not as daunting as it might be assumed. Below, you will learn the most basic procedures that every teacher should be able to apply while considering the validity of the objectively-scored assessment results.

Think about the following:

- 1) What do you know about Measures of Central Tendency and Measures of Dispersion?
- 2) Why do you think it is important to conduct statistical analysis of assessment results?

UZBEK VIGNETTE

Teacher: During an academic year, we have two midterms and two final tests. The assessment tools are designed based on what is taught in the class. Teachers collect different tasks from textbooks and we also take stuff from the Internet on reading, listening, as well as grammar and vocabulary.

Interviewer: Why do you choose these skills to assess?

Teacher: They are easy to check; we do not have to spend much time on scoring.

Interviewer: What happens after the tests?

Teacher: We check the tests without any criteria and announce scores. Then, these scores are used to provide a final score for the student.

Interviewer: OK, what happens after?

Teacher: That's it. We finalize. Students receive their scores.

Interviewer: Do you ever analyse the students' scores?

Teacher: No, we do not have time. We need to start the new term.

REFLECTION

Think about the case above. What issues come up? Is the teacher's assessment procedure similar to others at your university? Having been introduced to *Assessment for Learning* and *Assessment of Learning* what can you suggest a teacher can and should do with the test results?

KEY CONCEPTS

There are eight seven concepts in this section: Objectively-scored assessments; Measures of Central Tendency; Mean; Median; Mode; Measures of Dispersion; Standard Deviation. We will briefly explain each one below.

Objectively-Scored Assessment. An Objectively-scored item is a question where there is only one fixed correct answer. It is also known as closed-answer test. One of the strongest advantages of using this type of assessment is its high reliability and accuracy in generating a total score. In the Uzbekistan context, using objective

scoring to measuring language has been the main approach in assessment. The most popular is being multiple-choice tests with four variants of responses. Even though it is a highly reliable method of testing, it does have certain dangers. Multiple-choice items are notoriously hard to design. Other closed-response item test method types might be easier to develop but even they are fraud with issues, like guessing. Therefore, statistics can reveal not only the quality of a test but also the preparedness of students.

Measures of Central Tendency. After students have taken an objectively-scored test you might want to know how your strong students did as opposed to the weak ones? Or, you might be interested in knowing how well one class did in comparison to another group. When a teacher obtains students' test results, this becomes informative data. Usually in statistics we look for an average result, which is also referred to as **central tendency**. Central tendency can be informed by *mean, median, and mode*.

Mean is the average of all the available scores from a test. The formula can be represented mathematically as:

$$\bar{X} = \frac{\sum X}{n}$$

In other words, the mean or (X bar) is the sum (addition) of all scores in a set divided by the number of test takers. Here is an example: A class of 10 students were assessed in reading with a progress test consisting of 30 closed-item questions, in which the maximum score was 30. The procedure for obtaining the mean is as follows:

- 1) Present the Distribution of Scores

Table 13. Distribution of Scores.

Student Number	1	2	3	4	5	6	7	8	9	10
Score	14	18	19	20	21	21	21	26	26	27

- 2) All the scores are added up and divided by the number of students:

$$14+18+19+20+21+21+21+24+26+27 = \mathbf{213 \text{ (the sum of all scores)}}$$

the sum of all scores

number of students

$$213 : 10 = \mathbf{21.3} \text{ (this is the average score and it is also called mean)}$$

We need to know **mean** to see how well our students did on average. And here, with total score of 30, the mean is **21.3**.

- 3) Interpreting the mean score. To interpret the mean, you need to think about what type of test you used (e.g., progress test, proficiency, achievement, etc.). For example, the mathematical distribution above was for a progress test. In a progress test a teacher hopes for higher scores, which means the students have learned the knowledge or skills. 21.3 is a low average and informs the teacher that the students did not understand the materials as best as they could. However, to more fully understand the central point of understanding, we will need to also look at the median and mode.

Median is derived by means of, firstly, setting scores in ascending order (see Table 13) and then identifying the score that appears in the middle of the list. Thus, the median is the point at which 50% of the scores are higher and 50% of the scores are lower. Because there are an even number of students (i.e.,10) we will take Student 5 and Student 6 scores, which are both respectively 21 and 21. Then, we find the average of these scores. Median in our case is **21**.

Mode is the most commonly occurring score. To find the mode, you find the score that is used most often in the data set. In our case, it is **21** (if you look at Table 13 above, 21 is the score of three students).

Interpreting overall results of the Measures of Central Tendency. We have identified that the mean is 21.3, the median is 21, the mode is 21. Because this test is a progress test and most students were not successful – as the total score is 30 – we will need to revisit some topics that students did not understand

Measures of Dispersion. Apart from the Measures of Central Tendency indicators (i.e., mean, median, and mode), we are also interested in *how* spread out the scores are from the mean. These mathematical procedures are called Measures of Dispersion (i.e., standard deviation).

Standard Deviation is the average distance of scores from the mean. The lower number you receive for standard deviation to 0, the more the students in the class are similar. The larger number you obtain for standard deviation, the less similar (i.e., more different) the students are in the class. The standard deviation formula is mathematically represented as follows:

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

In other words, there are five steps we need to take to complete the standard deviation (if we calculate the standard deviation by hand). Let's refer to our data set from the reading quiz:

Student Number	1	2	3	4	5	6	7	8	9	10
Score	14	18	19	20	21	21	21	26	26	27

1) Find the mean. The mean is represented by X bar in the formula. We found the mean to be 21.3.

2) For each data point, find the square of its distance to the mean: Here is an example for the first data point: 14, from Student 1:

a. $(14 - 21.3) = -7.3$

b. $(-7.3)^2 = 53.29$

Student Number	1	2	3	4	5	6	7	8	9	10
Score	14	18	19	20	21	21	21	26	26	27
Square of the distance from the mean	53.29	10.89	5.29	1.69	0.09	0.09	0.09	22.09	22.09	32.49

$$(x - \bar{x})^2$$

3) Sum the values:

a. $53.29 + 10.89 + 5.29 + 1.69 + 0.09 + 0.09 + 0.09 + 22.09 + 22.09 + 32.49$

b. Sum = 148.10

4) Divide by the number of data points minus one.

a. 10 students took the class; $10 - 1 = 9$

b. 148.10 divided by 9 equals 16.45

5) Take the square root.

a. $\sqrt{16.45}$

b. 4.06

Interpreting the standard deviation: The closer the number is to 0, the more similar the class is; the farther away from 0 the number is, the more different the students in the class are. Usually, for language teachers, you would like your class standard deviation to be between 0.00 and 1.00. However, the standard deviation for the groups of students here is 4.06, which means the students are very spread out and you have various ranges of levels of students in your class.

Interpretation of assessment scores. To fully interpret your results you will need to combine the results from the Measures of Central Tendency (i.e., mean, median, and mode) with the standard deviation.

ACTION

You have learned the main ways of statistical test results analysis, now, in groups, consider the following case and compare the results of two classes.

Class One:

Student Number	1	2	3	4	5	6	7	8	9	10
Score	14	18	19	20	21	21	21	26	26	27

Class Two

Student Number	1	2	3	4	5	6	7	8	9	10
Score	10	12	17	18	21	21	27	28	29	30

You already know the mean in the Group 1, find out the mean the Group 2.

Derive the mean, median as well as the standard deviation. What can you realize when making class comparisons using basic statistical procedures? Which class did better (i.e., which class is stronger). If you were the teacher of these classes what actions would you take next?

**HOMETASK
MATCHING**

STANDARD DEVIATION	is the average distance of scores from the mean.
MEASURES OF CENTRAL TENDENCY	Are mean, median, and mode
MEASURES OF DISPERSION	<i>how</i> spread out the scores are from the mean
MODE	is the most commonly occurring score
THE MEDIAN	is the point at which 50% of the scores are higher and 50% of the scores are lower.
MEAN	is the average of all the available scores from a test
INFORMATIVE DATA	when a teacher obtains students' test results
CENTRAL TENDENCY	an average result
OBJECTIVELY-SCORED ITEM	is a question where there is only one fixed correct answer
MULTIPLE-CHOICE TESTS	the most popular objectively-scored test in Uzbekistan
OBJECTIVELY-SCORED ASSESSMENT	is popular because of its high reliability and accuracy in generating a total score.

10. PRACTICAL 8: USING STATISTICS: SUBJECTIVELY-SCORED ASSESSMENTS

Plan:

- What subjectively-scored assessments are;
- Difference between holistic and analytic assessments;
- How to distinguish and apply holistic and analytic scoring.

Key words: Assessment criteria; holistic scoring; analytic scoring; inter/intra-rater reliability; benchmarking; informing and explaining a score.

INTRODUCTION AND OVERVIEW

Recently, teachers in Uzbekistan are highly encouraged to use subjectively scored methods of assessment based on performance (e.g. role-play, presentation) and product (e.g. essay, report, portfolio). This approach to assessment involves judgement from one or more assessors who will need to use either holistic and/or analytic criteria (see below).

Think about the following:

- 1) What do you think are the main challenges in using subjectively-scored assessments?
- 2) How difficult is it to agree on a score with a colleague? Please think about examples from your own experience during your time as a pre-service or current in-service teacher life.
- 3) Have you ever had any trouble understanding assessment criteria? If you had, why do you think that happened?

UZBEK VIGNETTE

At the end of a course, students are assessed based on the presentation of their project work. The process involves 20-minute speech delivered by students, and a question-answer part from members of a specially assigned assessing committee. Then, the presenter is to leave the room and the committee members discuss and agree on the score to be announced. Sometimes, students disagree with the granted scores and consider them unfair. For example, the other day, one of the students presented on a topic that he liked a lot during the course and he received instruction on it during classes. However, when he presented, one of the committee members, was very critical of the selected approaches in the project work and he marked the student down claiming the absence of the required

procedures. The student was shocked as, in his project work; he followed the procedures taught by his teacher. Therefore, the mark seemed to be very unfair.

REFLECTION

Discuss the case above. What do you think is a possible violation – of the four traditional criteria of language assessment (i.e., validity, reliability, practicality, and washback) – that the assessor made? How do you think assessors can be sure that their assessment is fair? How should the score be explained to a student, when such a situation happens?

KEY CONCEPTS

There are six key concepts in this section: Assessment criteria; holistic scoring; analytic scoring; inter/intra-rater reliability; benchmarking; informing and explaining a score. We will briefly explain each one below.

Assessment criteria. Marking criteria or rubric is central to making sure that subjectively-scored assessments are valid and fair for subjectively scored assessments. We believe a teacher cannot and should not assess a student unless there are clear criteria from which he or she is to be assessed. A piece of work (i.e., essay) or oral response (i.e., interview) by a student that is assessed without using any criteria is regarded unacceptable. A common issue in Uzbekistan is that criteria for one assessment is used for multiple situations and does not address the appropriate content to be measured. (Thus, many subjectively-scored tests are not valid.) It should be noted that adapting criteria should be approached with specific attention, as the quality and clearness of scale descriptors can affect scoring and its validity. Another common issue is interpreting scale descriptors for assessment. Therefore, creators of criteria must make sure that there is no ambiguity in descriptors and teacher, in their turn, need to notify the issues observed. Responsibility of fair assessment lies on teachers. The proper use of rating criteria is also crucial, which means that teachers need to be attentive in making their decisions.

Holistic and Analytic Scoring. There are two main types of scoring: holistic and analytic. Holistic scoring looks at the whole picture of the student's work. Thus, you will need to provide one general score. Analytic scoring, however, is concerned with separate constituents when assessing. Your task is to look at various aspects or scoring facets to generate a score. Here is an example of holistic criteria that was used by Pat Blogger (as cited in Bailey, 1999) for his test, which was an attempt to measure

English language ability with respect to understanding written passages first and then to work with that information in order to construct arguments:

Holistic Scale for a Speaker’s Effectiveness of Argumentation

7	Relevant arguments are presented in an interesting way, with main ideas prominently and clearly stated, with completely effective supporting material; arguments are effectively related to the speaker’s view.
6	Relevant arguments are presented in an interesting way; main ideas are highlighted with effective supporting material, and are well related to the speaker’s own views.
5	Arguments are well presented with relevant supporting material and an attempt to relate them to the speaker’s views.
4	Arguments are presented but it may be difficult for the rater to distinguish main ideas from supporting material; main ideas may not be supported; their relevance may be dubious; arguments may not be related to the speaker’s views.
3	Arguments are presented, but may lack relevance, clarity, consistency or support; they may not be related to the speaker’s views.
2	Arguments are inadequately presented and supported; they may be irrelevant; if the speaker’s views are presented, their relevance maybe difficult to see.
1	Some elements of information are present but the rater is not provided with an argument, or the argument is mainly irrelevant.
0	A meaning comes through occasionally but it is not relevant.

This rubric is holistic because the assessor will provide one score, from 0 to 7, to the student based on the descriptors for each level. There are both advantages and disadvantages to using holistic scoring methods.

Table 14. Advantages and Disadvantages for Holistic Scoring	
<u>Advantages</u>	<u>Disadvantages</u>
Higher rater reliability can be achieved	May mask differences across individual assignments

Scoring scale can be understood by students and teachers.	Does not provide much useful diagnostic feedback.
Applicable to many different topics	Fails to capture important differences
Emphasizes strengths rather than weakness	
Great potential for positive washback	

Here is an example of analytic criteria that was used by Dr. David Chiesa when assessing students discourse analysis paper after analyzing dialogue from an American Movie. The analytic criteria were adapted from Ferris and Hedgcock (2014).

Analytical Writing Scoring Criteria			
Names:		Movie Title:	
Facilitator: David Chiesa		Date:	
Record Score	Grade	Score Range	
Content			
	A	36-40	Superior understanding of topic and writing context; all questions answered and supported with sound generalizations and substantial, specific, and relevant details; rich distinctive content that is original, perceptive, and/or persuasive; strong reader interest.
	B	32-35	Accurate grasp of topic and writing context; many questions answered and supported with sound generalizations and relevant details; substantial reader interest.
	C	28-31	Acceptable but cursory understanding of topic and writing context; few questions answered and supported with adequate generalizations and relevant details; suitable but predictable content that is somewhat sketchy and overly general; occasional irrelevant detail or unsound generalizations; average reader interest

	D/F	1-27	Little or no grasp of the topic or writing context; little to no questions answered; inadequate supporting points or details; irrelevant details, numerous unsound generalizations; insufficient, unsuitable, unclear, vague, or weak content; minimal or no reader interest.
Academic Writing: Rhetorical Structure			
	A	36-40	Exceptionally clear plan connected to purpose; plan developed with consistent attention to proportion, emphasis, logical order, flow, and synthesis of ideas; paragraphs coherent, unified, and effectively developed; striking introduction, body paragraphs, and conclusion.
	B	32-35	Clear plan connected to purpose; plan developed with proportion, emphasis, logical order, and synthesis of ideas; paragraphs coherent, unified and adequately developed; smooth transitions between paragraphs; effective introduction, body paragraphs, and conclusion.
	C	28-31	Conventional plan apparent but routinely presented; paragraphs adequately unified and coherent, but minimally effective in development; weak topic sentences; transition between paragraphs apparent but abrupt, mechanical, or monotonous; routine introduction, body paragraphs, and conclusion.
	D/F	1-27	Plan not apparent, inappropriate, underdeveloped, or developed with irrelevance, redundancy, inconsistency, or inattention to logical progression; paragraphs incoherent, underdeveloped, or not unified; transitions between paragraphs unclear, ineffective, or nonexistent; weak introduction, body paragraphs, and conclusion.
Grammatical Form: Including Mechanics			
	A	18-20	Sentences skillfully constructed, unified, coherent, forceful, effectively varied; deftness in coordinating,

			subordinating, and emphasizing ideas; harmonious agreement of content and sentence design; impressive use of grammatical structure; clarity and effectiveness of expression enhanced by consistent use of conventional punctuation, capitalization, and spelling.
	B	16-17	Sentences accurately and coherently constructed with some variety; evident and varied coordination, subordination, and emphasis of ideas; some errors in complex patterns; effective and clear use of grammatical structures. Flow of communication only occasionally diverted by errors in conventional punctuation.
	C	14-15	Sentences constructed accurately but lacking in distinction; minimal skill in coordinating and subordinating ideas; little variety in sentence structure; clarity weakened by occasional awkward, incomplete, fused and/or improperly predicated clauses and complex sentences; marginal to adequate use of grammatical structures. Adequate clarity and effectiveness of expression, though diminished by punctuation, capitalization, and/or spelling errors.
	D/F	1-13	Sentences marred frequently enough to distract or frustrate the reader; numerous sentences incoherent, fused, incomplete, and/or improperly predicated; monotonous, very simple sentence structure; unacceptable use of grammatical structures. Communication hindered or obscured by frequent violations of punctuation, capitalization, and/or spelling conventions.
Total Score (out of 100 total	Grade		Comments:

point s)			
-------------	--	--	--

Both approaches to scoring subjectively (holistic and analytic) have advantages and disadvantages. The decision regarding which approach to choose – holistic or analytic – depends on the purpose of assessment.

Inter(Intra)-rater reliability. One of the significant challenges in performance and direct assessments is inter-rater reliability (between two people)/ intra-rater reliability (amongst yourself as an assessor). It is very hard to achieve consistency in marking not only between one rater with the other, but also among various instances of assessment made by the same rater. It appears subjectivity cannot be avoided and we find quite frequently teachers who are quite generous and kind in marking, while others are quite harsh and demanding. To mitigate this endless problem, raters are usually asked to have benchmarking before they conduct assessment. **Benchmarking** is when two raters are both assigned to check the same written work or conduct a spoken examination together. In this case, both specialists must agree on a certain score. A third rater might be invited in cases when the judgement of the two differ significantly. In McNamara (2000, p.58) minimum acceptable inter-rater agreement range from 50%, which other 50% are the cases when raters disagree. Much acceptable is the case of 80% agreement to 20% disagreement, which can be the result of accurate wording or criteria or proper training of the raters. For information on the mathematical formulas on how to conduct inter-rater and intra-rater reliability, we recommend referring to Bailey (1998).

Informing and Explaining a Score. Very often, in the Uzbekistan context, students tend to ask for the explanation of a score. Arguments over a score can be avoided if transparent scoring is in place, which means that students need to know how are they assessed and why they are granted a certain score. Issues come out from the absence and ambiguity of marking criteria. Therefore, the primary concern of a rater

must be whether the granted score is thoroughly considered and whether the necessary procedures have been used.

SUMMARY

In conclusion, subjectively-scored assessment can be very challenging for a teacher and it is hard to abstain from impressionistic marking. We recommend sticking closely to using either a holistic or analytic scoring method. However, a teacher must be conscious of ethics in assessment, as whatever mark we-teachers settle on might change the course of events in the lives of our students. Therefore, assessment is great responsibility.

HOMEWORK TASK FOURTEEN

You now have gone through five sections in the chapter and learnt the salient aspects in language assessment and testing. By this time, you should have developed ideas regarding what changes you can make in your teaching and assessment practices. Based on what you have learnt in this chapter and discussions at classes, write an action plan (minimum 1 page) describing the problematic areas which you are planning to address after you complete this in-service education course, provide details of:

- Why you think it is a significant issue
- How you are going to address it; and,
- What is the expected result?

VI. KEYSLAR BANKI

Case 1: Over Their Heads

Designing and adapting materials and teaching English for specific purposes.

by Richard Watson Todd

Kim had been working at a private language school in a small town in Spain for nearly 18 months now, and planned to continue for another six months before she went back to the UK to look for something more permanent. The school, like the town, was small, and nearly all the courses were for adolescents. Although Kim still enjoyed the classes full of vibrant teenagers, recently she had been growing a bit bored with the lack of variety and was on the lookout for a new challenge to revitalise her teaching. When David came to talk to her about a small hi-tech company which had contacted the school about in-company classes, she felt that this might be the challenge she needed.

David was the Director of Studies at the school and a long-term resident in the town. He was central to the management of the school and the owner relied on him a lot. He had set up all the courses, produced the materials and even written out suggested plans for all the lessons. Since his suggestions generally worked well, this made life easy for the teachers at the school.

David explained to Kim that the company wanted its white-collar staff to be able to use English in their work, and that this would involve teaching such skills as reading instruction manuals, understanding and writing business correspondence, and speaking on the telephone. Kim realised that content like this would be the refreshing change she was looking for, but was worried that she wouldn't be capable of preparing lessons on these topics. She was reassured when David went on to explain that, although he would be back in the UK when the course started, he would talk to the company about the course, design the syllabus and prepare the materials for her to use before he went. Kim readily accepted the responsibility of teaching the course.

Over the next three weeks, David was busy preparing the course which was to cover forty hours in two lessons a week over ten weeks. Occasionally, he would call Kim to his office and explain to her how the course and the materials would work, so that by the Saturday he left for his holiday, Kim felt well-prepared and confident about the course.

The next Tuesday was the first day of the course and the company sent a van to pick Kim up. Arriving at the factory, she was treated well and taken to a well-appointed room that was to be her classroom. The staff taking the course all trooped into the room

in the next five minutes which suggested that the company was taking the course seriously. There were twelve learners in the class, evenly split into men and women, and all aged in their late twenties and early thirties. Kim felt happy and thought the situation looked promising.

For the first lesson, David had suggested an easy-going getting-to-know-you introductory lesson with no overt business focus to help Kim establish relationships with the learners before the 'real work' began. He had suggested an adaptation of the warm-up lesson for upper-intermediate learners at the school, which Kim was familiar with. This started with a small New Name activity which the learners at the school usually found easy and interesting. They would try to translate their name into English and find the nearest English equivalent to their name. Following this, there was a "Find a person who ..." questionnaire (with questions such as "Find a person whose favourite childhood toy was a teddy bear") which learners usually found amusing.

The lesson didn't go as well as Kim had expected. The learners showed little interest in the task, but instead dourly went through the procedure of asking their classmates for information. In addition, they had some surprisingly large gaps in their vocabulary and some of them had difficulty formulating questions to ask their classmates. Two of the men, Manuel and Juan, in particular, seemed to be having big problems coping with even the simplest English; one of the women, Sophia, resolutely refused to have anything to do with the task; and the rest of the learners appeared reluctant and unsure of what they were doing. Kim hoped that these were just teething problems for the course, perhaps because the company staff had not been in the role of learners for several years. She felt that the second lesson with more technical, business-oriented language probably familiar to the learners would be more of a success.

The lesson on the next Friday, however, was a disaster. Focusing on instruction manuals, the prepared lesson aimed to help the learners understand the organisation of instruction manuals and analyse the language used. From the beginning of the lesson, Kim found herself doing all the talking while the learners watched her with blank faces. When, after ten minutes of the lesson, Juan put his hand up and asked "What mean 'manual'?", Kim realised that she had been talking completely over their heads. The rest of the lesson was a nightmare. Kim vainly tried to follow the lesson plan that David had prepared, but it was all way beyond the learners' level. The lesson had changed from the joint exploration of the language of instruction manuals that David had intended into a desperate succession of teacher explanations of unknown vocabulary by Kim.

When the two hours were up, Kim felt released. The lesson had been her worst ever teaching experience. In the van home, however, it struck her that she would have to teach the learners again the next Tuesday. The materials she had available were obviously completely inappropriate. This weekend she would have to come up with something different. David wouldn't be back from his holiday for another two weeks, and the other teachers at the school had no experience of preparing materials or of business English. She was on her own. How on earth could she come up with any useful, appropriate, business-oriented materials by next Tuesday?

Questions

1. Kim's immediate preoccupation is that the materials are too difficult for the students. In a situation like this, there are three choices facing the teacher: to discard the existing materials and design new materials from scratch, to adapt the existing materials to make them more appropriate for the learners, and to find other ready-made materials as a replacement. Which of these choices do you think would be the most suitable for Kim? Why?
2. If Kim decides to design new materials, how should she do this? What are the stages in materials design? Do you think that these stages represent an idealised design process or can they be followed in practice?
3. If Kim decides to adapt the existing materials, on what bases should she decide what aspects of the materials to retain and what aspects to discard? Is there anything she should be particularly wary about when she adapts the materials?
4. If Kim decides to replace the existing materials, how can she find other ready-made materials? If she finds several possible alternative sets of materials, how can she decide between them? What criteria can guide her decision?
5. Another aspect of the case study that we can consider is why Kim sees the problems with the materials as so serious. This can be viewed as a longer-term problem of staff development. David, although presumably usually efficient, does not see staff development as part of his duties as Director of Studies. Rather than helping the teachers at the school become competent and independent at non-classroom aspects of teaching, such as materials design, David takes the whole of this work himself. How could David help the staff at the school develop?
6. The materials described in the case study are specifically aimed at business, focusing on the language used in business communication and revolving around business topics. David, in designing such materials, would presumably argue that business learners

need English for business and that the best way to prepare such learners is to provide them with English used in business situations. An alternative viewpoint is that, by providing the learners with business English only, they are being limited to a specific area of English rather than being exposed to English in all its wide uses and so are missing out on a broad English education. Are there any other viable viewpoints? Which do you agree with? Why?

7. The case study talks about materials "way beyond the learners' level". As teachers we often use the word level as a term of convenience to describe learners' competence. Although convenient, the term level has many problems. First, it is difficult to define. Does it refer to the students' ability in grammar, the size of their vocabulary, their fluency, some other aspect of language, or a combination of these? Second, comparing the levels of two learners is problematic, since all learners' interlanguage is idiosyncratic. Third, it is very difficult to describe a given level, so that we are usually at a loss if a colleague asks us, "What level is your class?" Do you think that the convenience of level outweighs these problems? If you believe that the term level is useful, how can you define it and how would you describe the level of one of your classes?

Further reading

In deciding whether to design, adapt or replace, Block (1991) argues the case for materials design, Nunan (1991) argues against adaptation, and Robinson (1991) looks at the pros and cons of design and replacement. Nunan (1991) also includes an 8-stage model for designing materials, which it is interesting to compare with the checklist in Dubin and Olshtain (1986) and the process of materials design given by Jolly and Bolitho (1998). Hutchinson and Waters (1987) suggest a different approach particularly applicable to ESP materials design such as the business English materials in the case study. Brown (1995) and Maley (1998) describe the process of materials adaptation with detailed examples. For evaluating materials as possible replacements, a whole book (Cunningsworth, 1995) is devoted to the subject, and Brown (1994), Ellis (1997) and Jordan (1997) also provide useful guidelines. Impey and Underhill (1994) and White et al. (1993) look at the responsibilities of people in management positions in ELT, such as David, including their responsibilities for staff development. The arguments for and against using materials for specific purposes are humorously presented in Widdowson (1984). Lastly, the problems in defining students' communicative competence are accessibly examined in Omaggio (1986).

Case 2: Working Together

Giving instructions and observing other teachers to help their development.

by Richard Watson Todd

Twenty years in the same secondary school had turned Paolo into an automaton. For most of the time he had stuck to the tried and trusted methods of drilling, long grammar explanations and rote learning of word lists.

Six months ago, however, a new teacher, Sophia, had arrived at the school with bright innovative ideas that were a breath of fresh air to Paolo. Sophia talked about students discovering language for themselves, interacting in groups, and becoming confident with using English.

Hearing these ideas, Paolo had realized that he had been treating his classes like a production line, something he had sworn he would never do when he had started teaching twenty years earlier. Quickly becoming friends, Sophia and Paolo had decided to work together to help each other develop as teachers. They decided that they would talk English outside the classroom to improve their language, they would subscribe to a couple of journals for teachers, and they would encourage each other to innovate and try out new ideas in the classroom.

They had worked together like this for over four months now. Paolo felt that his interest in teaching had been rekindled and that his students enjoyed themselves more and seemed to look forward to learning English with him. Most of Paolo's lessons worked well, and now he sometimes found himself trying to persuade other colleagues to try out new techniques.

Today's lesson, however, had severely shaken his confidence. In his plan, the lesson had appeared straightforward and effective, but in the classroom it had ended in shambles. The focus of the lesson was reading, and Paolo had decided to do something different from the old 'Read the passage and answer the questions' approach.

In one of the journals he and Sophia were subscribing to he had found a technique called Jigsaw Reading. Cutting a long passage into pieces, he would give each of the pieces to a different group of students. The groups would read and try to understand their section of the passage. Then new groups consisting of one student from each of the previous groups would be formed. The new groups would try to reconstruct the whole passage. Paolo thought that the students would pay a lot more attention to the reading if he used a Jigsaw Reading technique. In addition, the technique would generate a lot of beneficial student-student interaction.

Paolo had been looking forward to trying out Jigsaw Reading with his fourth-year students. In the classroom, however, the new technique had been fraught with problems. The seemingly simple procedures of Jigsaw Reading turned out to be almost

impossible to convey to the students, even when Paolo resorted to Italian. In giving the instructions before the activity, he found himself using longer and longer sentences with all sorts of convoluted phrasing to explain whether he was talking about the original grouping of students or the regrouping halfway through the activity.

Because they had been unclear about the purpose and organization of the activity, the students had been uncertain of what to do while reading. Regrouping the students had taken a full ten minutes of class time, and once they had been regrouped, the students just sat there not knowing what to do next. Paolo had had to explain all the stages of the activity over and over again, until he was relieved to hear the bell ring at the end of the class.

Exhausted and dispirited after the lesson, Paolo sought out Sophia. He explained all that had gone wrong while Sophia listened attentively and made sympathetic noises. When he finished, they discussed the possible causes of the problem. They decided that everything came down to the clarity of his instructions. The next problem, then, was how Paolo could improve his instructions. Paolo, still discouraged by the lesson, did not feel capable of improving his instructions by himself. He wanted Sophia to help him.

Together, Paolo and Sophia brainstormed ways of helping Paolo overcome his problems with instructions. Obviously, he could pay more attention to his instructions during the planning stage, and Sophia could help him here by working through the plan with him. But Paolo was more concerned about what would happen once he was in the actual classroom. He knew that Sophia was free when he taught his third-year class and he wanted her to come into his classroom and watch him give instructions. While Sophia felt flattered that Paolo trusted her so much, she was worried about observing his teaching. First, what would the students think? Paolo was a far more senior colleague, so it would look strange if Sophia went into his classroom and took notes on his teaching.

What's more, Sophia wasn't sure about how much help she could be to Paolo by observing him. She didn't know what sort of things she should look for when Paolo gave instructions, and she didn't think she would be able to identify what his problems were. Another thing that worried her was what she should say to Paolo after the lesson. Although they had built up a close relationship, Sophia knew she wouldn't feel comfortable criticizing Paolo's teaching and wasn't sure what his reactions would be. With these misgivings in her mind, Sophia was undecided about whether to accept Paolo's invitation to observe his teaching.

Questions

1. Paolo's instructions for the Jigsaw Reading activity are at the root of his problems. It is suggested that he could pay more attention to his instructions during the planning stage. One of the skills of planning is predicting what problems will arise during teaching and pre-empting these problems.

How do you think Paolo can improve his problem-predicting skills while planning?

2. Instructions are vital to the success of an activity. What content should be included in instructions? Should instructions be explained or demonstrated, or both? How can a teacher tell how effective any set of instructions is?

3. The regrouping of the students in the Jigsaw reading activity causes most of Paolo's problems.

Should all of the instructions be given at the start of the activity or should the instructions be broken down into two sets, one given at the start of the activity and one given in the middle before regrouping the students?

If you were going to use a Jigsaw reading activity in your teaching, what instructions would you give? Can you predict any problems which might arise from your instructions?

4. Regrouping the students creates a transition in the middle of the Jigsaw reading activity. Transitions, if not well-organized, may waste valuable time and possibly lead to chaos. How can transitions be managed to reduce the chances of time-wasting and chaos occurring?

5. Paolo and Sophia decide that Sophia should help Paolo while planning. How should Sophia help Paolo?

Should she be actively involved all through the planning process or should she only comment on the final plan that Paolo produces?

6. Paolo also wants Sophia to observe his teaching and give him feedback. Sophia, however, is unsure of what to look for and how to observe.

If Sophia observes Paolo, do you think she should just write down any comments she thinks might be useful as she observes? Alternatively, Sophia could use an observation sheet, in the form of, say, a table which would help her to categorise certain aspects of Paolo's teaching.

If she chooses to use an observation sheet, what aspects of Paolo's teaching should Sophia include on the sheet? What form should the observation sheet take?

If you were going to observe a colleague focusing on his/her instructions, how would you record your observations? If you decided to use an observation sheet, what would it look like?

7. Sophia is worried about giving feedback to Paolo on his teaching. Feedback can often seem critical to the person receiving it, leading to defensiveness and relationship problems. If Sophia needs to give Paolo feedback, how can she give feedback so that such problems do not occur?

Further reading

A full description of Jigsaw Reading, including a detailed rationale, can be found in de Berkeley-Wykes (1983). Teacher instructions are considered in depth in Watson Todd (1997), while Ur (1996) gives some basic guidelines for giving instructions and Campbell and Kryszewska (1995) contains training activities for improving teachers' instructions. Froyen (1993) gives examples and guidelines for managing transitions effectively. Regarding working with colleagues, Knezevic and Scholl (1996) is a case study of collaborative planning; Day (1990) provides an excellent background to observation; Richards and Lockhart (1994) give guidelines for observing colleagues; Wajnryb (1992) is a comprehensive collection of useful observation sheets, including one for instructions; and Wallace (1998) discusses different ways in which data can be collected through observation. Finally, Gebhard (1990), Wallace (1991) and Woodward (1992) contain a variety of ways of giving feedback with the probable effects of each feedback technique.

Case 3. Planning, being flexible in the classroom, and dealing with the unexpected.

by Richard Watson Todd

Mustafa was proud of his BEd. Now in his first job as a teacher, he had great plans for helping his students learn English, and he knew that what he had learnt from his BEd would help him reach this goal. All through his years as a secondary school student, he had thought that teaching was easy, but his degree had made him realise that teaching was far more complicated when seen from the teacher's perspective than from the student's seat. The most important thing that Mustafa had learnt from his degree was the importance of planning. His tutors had constantly emphasised and re-emphasised the need to think before teaching. Planning, he had been told, was often more important for the success of a lesson than the teaching. Having been asked to

teach an impromptu lesson and then compare it with a planned lesson, Mustafa firmly believed his tutors.

Mustafa had been teaching at a technical college in Cairo for two weeks now. His next lesson was on Saturday with an evening class of older students. He picked up the textbook assigned for the class and started planning. On his degree, he had been told to follow a given sequence for planning and to write his plan according to a model format. Mustafa didn't need to remind himself of the sequence or the format since he had used them so often already. Starting from the unit in the textbook, he identified the objectives to be covered in the lesson, used a grammar book to check on his knowledge of these objectives, looked through the reading passage, prepared quick explanations of unknown words, checked the answers to the comprehension questions, decided how to present the grammar points, and wrote up instructions for the pair work activity. As a final flourish, he decided to devote three minutes at the start of the lesson to chatting to the students.

Looking over his lesson plan, Mustafa was pleased. It looked perfect. He could easily imagine his old tutor giving him an A grade for the plan. With a plan like this, he felt sure that he could help his students understand the grammar easily and that they would enjoy learning.

At six o'clock on the Saturday, Mustafa went into the classroom to find all of his students waiting for him. He checked the register and let the students calm down.

"OK, what did you do in the last week?", he asked. "Yes, Fatima?"

"I went to the cinema."

"You went to the cinema. Very good. OK, Ahmed what did you do?"

"I went to see my uncle near from Alexandria."

"Near Alexandria. No 'from'. OK. Hafiz?"

"I got married."

Mustafa smiled. "You got married. That's interesting." The three minutes he had set aside for chatting were up. "Now turn to page 17 in your books."

Mustafa asked the students to read the passage and to identify unknown words. After the students had finished reading, he asked, "Right, what words didn't you know?"

"Trapped."

"Trapped, right." Mustafa looked at his lesson plan. "Trap means to catch. So the boy was trapped means the boy was caught. OK?"

The students were silent.

"Any other words?"

"Pick."

"OK, pick means to select."

"But I don't understand. Here the book has that pick the lock."

"Yes, lock means the thing that you open with a key."

"But I don't understand."

Mustafa wondered what was wrong with Hafiz who was usually a good student.

Maybe it was his marriage affecting him. "What do you mean?"

"Pick a lock means select a lock. I don't understand."

"Never mind. Any more words? Yes, Abdullah?"

"Freezer."

"I taught you freezer last week. You already know the word. Yes, Miriam?"

"Jog."

Mustafa looked a bit put out. He had noticed the word when he had prepared his lesson, but he had assumed that the students would know it. He knew that in the short time he had been teaching them, the students hadn't come across jog, but it was such a simple word he had thought they must know it. "Um, jog means run."

Ahmed looked up brightly. "Run. Like Said Aouita. Yeah, good runner."

Mustafa was flustered. "No, not like Said Aouita. He runs very fast, but jog is running slowly."

It was now Ahmed's turn to look puzzled. "But if you run, you want to win. Why people run slow?" He then switched to Arabic and used the slang expression for 'They must be cheats'.

Mustafa felt that he was starting to lose control. This wasn't in his lesson plan. "No. You don't jog when you run in a race. Jog is run slowly for exercise. If you want to get fit, you can run but you only need to run slowly. So people jog for exercise or to get fit."

Now it was Miriam who looked confused. "What mean exercise and fit?"

Mustafa felt himself in danger of falling into a never-ending circle of definitions. He decided that he had to avoid this at all costs. So he quickly wrote the three problem words in English on the board with their Arabic equivalents. All of the students looked satisfied and dutifully copied these down into their exercise books. Mustafa was still worried, however. First, he had broken the climate of English which he had tried so hard to establish in the classroom. Second, he realized that he was already five minutes behind his lesson plan. He would have to rush through everything to get the lesson finished on time.

The rest of the lesson consisted of a mad rush on Mustafa's part to catch up with the times written in his lesson plan. In this he was frustrated by several unexpected questions and incorrect answers from the students which he felt duty-bound to deal with. The lesson turned into a race between Mustafa and the clock. He didn't give the students enough time to answer the comprehension questions; his grammar explanation was so rushed that he then had to spend a lot of time dealing with students' misunderstandings; he skimmed through the pairwork instructions at such a rate that the students had little idea of what they were supposed to do; and by the time the bell rang at the end of the lesson none of the pairs were anywhere close to finishing the activity.

After the lesson ended, the students left the room and Mustafa collapsed at his desk. What had gone wrong? His lesson plan had been so good. He looked back over it. The only problem he could see was that he should have predicted the need to teach jog. But surely such a little mistake couldn't have made his lesson go so awry. Nevertheless, it was the only problem he could find. He resolved to be more careful in his lesson planning in future. He would need to check every word in the reading passages, and prepare explanations for most of them. Although he didn't look forward to this, he knew that good lesson plans were vital, and the more he prepared the better his lesson plans and his teaching would be. With a sigh, he started reading the passage for the next lesson he would teach.

Questions

1. The BEd that Mustafa took placed a heavy emphasis on lesson planning. How important do you think lesson planning is to the success of lessons? How much emphasis should be given to lesson planning on teacher training programmes?
2. Mustafa had been taught to follow a certain sequence and format when planning. What do you think this sequence and format consisted of? How helpful do you think such a model is to beginning teachers? While models of planning can help give security to beginning teachers, they are also restrictive. Do you think the benefits of such models outweigh the extent to which they restrict teachers?
3. The lesson started with "chatting". What are the purposes of chatting to students? Why is it used so often as a way of beginning lessons?
4. Chatting to students can be considered an attempt to bring the characteristics of natural conversation (such as unpredictability and the need to constantly negotiate topics) into the classroom. However, the way in which Mustafa runs the chatting session does not reflect interaction in the real world. In what ways does the classroom

chatting in the extract differ from interaction in the real world? How do you think Mustafa can change the way he runs the chatting session in order to make it reflect real world interaction more closely?

5. To help students understand the reading passage, Mustafa asks them to identify unknown words. Do you think unknown words should be taught before the students read or after they have finished reading? Do all unknown words need to be explained? In explaining the meaning of the unknown words to the students, Mustafa seems to regard the context in which the words appear as not being very important. What problems do this lead to and how can Mustafa overcome these problems?

6. Mustafa's main problems occur when he is required to teach something he has not planned for. Improvisation is a vital teaching skill. Do you think that improvisation is teachable? If so, how can improvisation be taught? All through the lesson Mustafa is being controlled by the plan rather than controlling it. To what extent should plans be followed? When and why should teachers deviate from their prepared plans?

VI. GLOSSARY

Termin	Ingliz tilidagi sharhi
aptitude test	a test designed to measure capacity or general ability <i>a priori</i> (e.g., before taking a foreign language course) to predict success in that undertaking
Assessment	an ongoing process of collecting information about a given object of interest according to procedures that are systematic and substantively grounded
authenticity	the degree of correspondence of the characteristics of a given language test task to the features of a target language task
autonomy	the ability to set one's own goals and independently monitor success without the presence of an external stimulus
biased for best	providing conditions for a student's optimal performance on a test "smallest" elements (e.g., letters, syllables, words) of language and then combining them into increasingly larger elements
cloze	a text in which words are deleted and the test-taker must provide a word that fits the blank space
appropriate-word scoring	a scoring method that accepts a suitable, grammatically and rhetorically acceptable word that fits the blank space in the original text
communicative test	a test that elicits a test-taker's ability to use language that is meaningful and authentic
competence	one's hypothesized (empirically unobservable) underlying ability to perform language
computer-adaptive test (CAT)	instruments in which test-takers receive a set of questions that meet test specifications and that are generally appropriate for their performance level
computer-assisted language learning (CALL)	the application of computer technology to language learning and teaching

C-test	a text in which the second half of every other word is eliminated and the testtaker must provide the whole word
diagnostic test	a test that is designed to diagnose specified aspects of a language
dialogue journal	a self-writing exercise in which a student records thoughts, feelings, and reactions which a teacher reads and responds to
dichotomous scoring	a method of scoring that allows only one correct response so that a test-taker's response is either right or wrong
dictation	a method of assessment in which test-takers listen to a text and write down what they hear
dicto-comp	a variant of dictation whereby test-takers listen to a relatively long text (e.g., a paragraph of several sentences or more) and try to internalize the content, some phrases, and/or key lexical items and then use them to recreate the text
direct testing	an assessment method in which the test-taker actually performs the target task; as opposed to indirect testing
discrete point test	assessments designed on the assumption that language can be broken down into its component parts and that those parts can be tested successfully
display writing	writing that is produced, usually in response to a prompt, to show competence in grammar, vocabulary, or sentence formation; as opposed to real writing
evaluation	making decisions and/or value judgments based on the results of tests, other assessments, and/or teachers' reports
focus on form	attention to the organizational structure (grammar, phonology, vocabulary, etc.) of a language
formal assessment	systematic, planned exercises or procedures constructed to give teacher and student an appraisal of student achievement

formative assessment	evaluating students in the process of “forming” their competencies and skills with the goal of helping them continue that growth process
form -focused assessment	assessment that focuses on the organizational components (e.g., grammar, vocabulary) of a language
high-frequency word	a word that appears most often in written and oral texts and is part of the foundation of vocabulary knowledge that proficient users of the language have acquired; as opposed to low-frequency word
high-stakes test	an instrument that provides information on the basis of which significant decisions are made about test-takers (e.g., admission to a course/school)
holistic scoring	an approach that uses a single general scale to give a global rating for a test-taker’s language production; as opposed to analytic scoring
idiom	figure of speech whose meaning cannot be determined by the literal definition but whose metaphorical meaning is known through common use
impact	the effect of the use of a test on individual test-takers, institutions, and society
indirect testing	an assessment method in which the test-taker is not required to perform the target task; rather, inference is made from performance on non-target tasks; as opposed to direct testing
informal assessment	incidental, unplanned comments and responses, along with coaching and other impromptu feedback to the student
information transfer	a process in which information processed from one skill (e.g., listening to a telephone message) is used to perform another skill (e.g., writing down name/number to return a phone call)
institutionalized expression	a longer utterance that is fixed in form and used for social interaction (e.g., “how do you do?”)

integrative test	a test that treats language competence as a unified set of interacting abilities of grammar, vocabulary, reading, writing, speaking, and listening
interactive (skills)	combining the use of more than one skill (reading, writing, speaking, and listening) in using language
low-frequency word	a word that seldom or rarely appears in written or spoken texts; as opposed to high-frequency word
macroskill	linguistic competencies that involve language competence beyond the sentence level (discourse, pragmatics, rhetorical devices); as opposed to microskill
measurement	a process of quantifying a test-taker's performance according to explicit procedures or rules
mechanical task	a task that determines in advance what the test-taker will produce (e.g., reading aloud or sentence repetition)
microskill	detailed specific linguistic competencies that involve processing up to and including the sentence-level (phonology, morphology, grammar, lexicon); as opposed to macroskill
multiple-choice test	an assessment instrument in which items offer the test-taker a choice among two or more listed options
multiple intelligences	types of intelligence that extend beyond traditional IQ-based concepts, such as spatial, musical, kinesthetic, naturalist, interpersonal, and intrapersonal intelligence
narrative evaluation	a form of individualized written feedback about a student's performance, sometimes used as an alternative or supplement to a letter grade
norm -referenced test	a test in which each test-taker's score is interpreted in relation to a mean (average score), median (middle score), standard deviation (extent of variance in scores), and/or percentile rank
objective tests	tests that have predetermined fixed responses

options	different responses from which a test-taker can choose in an item language curriculum or school
practicality	the extent to which resources and time available to design, develop, and administer a test are manageable and feasible
prefabricated language	ready-made sentence stems and whole sentences or memorized chunks of sentences that provide models for the creation of new sentences
primary-trait scoring	in a writing test, a single score indicating the effectiveness of the text in achieving its primary goal
process	attending to the procedures (steps, strategies, tools, abilities) used to comprehend or produce language; as opposed to product
product	attending to the end result of a linguistic action (e.g., in writing, the “final” paper, versus the various steps involved in composing the paper); as opposed to process
proficiency test	a test that is not limited to any one course, curriculum, or single skill in the language; rather, it tests overall global ability
rubrics	statements that describe what a student can perform at a particular point on a rating scale; sometimes also called <i>band descriptors</i>
subjective tests	tests in which the absence of predetermined or absolutely correct responses require the judgment of the teacher to determine correct and incorrect answers
subtechnical word	a word that occurs across a range of registers or subject areas
summative test	a test that aims to measure, or summarize, what a student has grasped and typically occurs at the end of a course or unit of instruction
supply items	options a test-taker can choose from for responses

task	a set or subset of linguistic actions that accomplish a real-world purpose, problem, or demand
task-based assessment	assessments that involve learners in actually performing the behavior that one purports to measure
test	a method or procedure for measuring a person's ability, knowledge, or performance in a given domain
validity	the extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the assessment
concurrent validity	the extent to which results of a test are supported by other relatively recent performance beyond the test itself
consequential validity	a test's impact, including such considerations as its accuracy in measuring intended criteria, its effect on the preparation of test-takers, and the (intended and unintended) social consequences of a test's interpretation and use
construct validity	any theory, hypothesis, or model that attempts to explain observed phenomena in one's universe of perceptions
content-related validity	the extent to which a test actually samples the subject matter about which conclusions are to be drawn
criterion-related validity	the extent to which the linguistic criteria of the test (e.g., specified classroom objectives) are measured and implied predetermined levels of performance are actually reached
face validity	The extent to which a test-taker views the assessment as fair, relevant, and useful for improving learning
predictive validity	the extent to which results of a test are used to gauge future performance
washback	the effect of assessments on classroom teaching and learning

VII. FOYDALANILGAN ADABIYOTLAR RO'YXATI

O'zbekiston Respublikasi Prezidentining asarlari:

1. Karimov I.A. O'zbekiston mustaqillikka erishish ostonasida. - T.:“O'zbekiston”, 2011.
2. Mirziyoev SH.M. Buyuk kelajagimizni mard va olijanob halqimiz bilan birga quramiz. – T.: “O'zbekiston”. 2017. – 488 b.
3. Mirziyoev SH.M. Milliy taraqqiyot yo'limizni qatiyat bilan davom ettirib, yangi bosqichga ko'taramiz – T.: “O'zbekiston”. 2017. – 592 b.

Normativ-huquqiy hujjatlar

1. O'zbekiston Respublikasining Konstitutsiyasi. – T.: O'zbekiston, 2019.
2. O'zbekiston Respublikasi Prezidentining 2017 yil 7 fevraldagi “O'zbekiston Respublikasini yanada rivojlantirish bo'yicha Harakatlar strategiyasi to'g'risida”gi 4947-sonli Farmoni.
3. O'zbekiston Respublikasining “Tahlim to'g'risida”gi Qonuni.
4. O'zbekiston Respublikasi Prezidentining "Oliy tahlim tizimini yanada rivojlantirish chora-tadbirlari to'g'risida”gi 2017 yil 20 apreldagi PQ-2909-sonli qarori.
5. O'zbekiston Respublikasi Prezidentining “Oliy mahlumotli mutaxassislar tayyorlash sifatini oshirishda iqtisodiyot sohalari va tarmoqlarining ishtirokini yanada kengaytirish chora-tadbirlari to'g'risida”gi 2017 yil 27 iyuldagi PQ-3151-sonli qarori.
6. O'zbekiston Respublikasi Prezidentining “Nodavlat tahlim xizmatlari ko'rsatish faoliyatini yanada rivojlantirish chora-tadbirlari to'g'risida”gi 2017 yil 15 sentyabrdagi PQ-3276-sonli qarori.
7. O'zbekiston Respublikasi Prezidentining “2019-2021 yillarda O'zbekiston Respublikasini innovatsion rivojlantirish strategiyasini tasdiqlash to'g'risida”gi 2018 yil 21 sentyabrdagi PF-5544-sonli Farmoni.
8. O'zbekiston Respublikasi Prezidentining “Oliy tahlim muassasalarida tahlim sifatini oshirish va ularning mamlakatda amalga oshirilayotgan keng qamrovli islohotlarda faol ishtirokini tahminlash bo'yicha qo'shimcha chora-tadbirlar

to'g'risida"gi 2018 yil 5 iyundagi PQ-3775-sonli qarori.

9. O'zbekiston Respublikasi Prezidentining 2015 yil 12 iyundagi "Oliy tahlim muasasalarining rahbar va pedagog kadrlarini qayta tayyorlash va malakasini oshirish tizimini yanada takomillashtirish chora-tadbirlari to'g'risida" gi PF-4732-sonli Farmoni.

10. O'zbekiston Respublikasi Vazirlar Mahkamasining 2012 yil 26 sentyabrdagi "Oliy tahlim muassasalari pedagog kadrlarini qayta tayyorlash va ularning malakasini oshirish tizimini yanada takomillashtirish chora-tadbirlari to'g'risida"gi 278-sonli Qarori.

Maxsus adabiyotlar

1. Chiesa D., Azizov U. and others (2019) Reconceptualizing Language Teaching: an in-service teacher education course in Uzbekistan. Tashkent. Baktria press
2. Ferris, D., & Hedgcock, J. (2014). Teaching L2 Composition: Purpose, Process, and Practice. (3rd ed.). New York/London: Routledge.
3. McNamara, T. (2000). Language Testing. UK. Oxford: Oxford University Press.
4. Brown, J. D. (2006). Testing in language programs: A Comprehensive guide to English language assessment. McGraw-Hill Companies: New York.
5. Mary J. Pitoniak, John W. Young (2009). Guidelines for the Assessment of English Language Learners. Educational Testing Service.

Elektron ta'lim resurslari

O'zbekiston Respublikasi Oliy va O'rta maxsus ta'lim vazirligi:

www.edu.uz.

- Toshkent axborot texnologiyalari universiteti: www.tuit.uz.
- O'zbekiston Respublikasi Davlat Hokimiyati portali: www.gov.uz
- O'zbek internet resurslarining katalogi: www.uz
- <http://www.Britishcouncil.org>
- <http://www.ziyonet.uz/>
- <http://www.edu.uz/>
- <http://www.pedagog.uz/>
- <http://www.tefl.net>
- busyteacher.org, onestopenglish.com

- file:///C:/Users/User/Desktop/3.3/assessment_matters_self-assessment_peer_assessment.pdf
- file:///C:/Users/User/Desktop/3.3/CER_6168_V1_APR19_Cambridge_English_First_Handbook_WEB_v3.PDF
- file:///C:/Users/User/Desktop/3.3/Assessment_of_EFL_learners_Sociolinguistic_and_Pr.pdf
- <file:///C:/Users/User/Desktop/3.3/compensatory%20competence.pdf>
- <file:///C:/Users/User/Desktop/3.3/disc%20competence.pdf>
- file:///C:/Users/User/Desktop/3.3/ielts_task_2_writing_band_descriptors.pdf
- file:///C:/Users/User/Desktop/3.3/ManualLanguageTest-Alte2011_EN.pdf.pdf
- file:///C:/Users/User/Desktop/3.3/ManualLanguageTest-Alte2011_EN.pdf.pdf
- [file:///C:/Users/User/Desktop/3.3/Reading%20Asst%20Criteria%20Revised%20new%20headings_sample%20\(5\).pdf](file:///C:/Users/User/Desktop/3.3/Reading%20Asst%20Criteria%20Revised%20new%20headings_sample%20(5).pdf)
- <file:///C:/Users/User/Desktop/3.3/speaking-assessment-criteria-updated-2018.pdf>
- <https://www.ed.ac.uk/reflection/facilitators-toolkit/assessment/criteria>