



Бош илмий-методик
марказ

**FARG‘ONA DAVLAT
UNIVERSITETI HUZURIDAGI
PEDAGOG KADRLARNI QAYTA
TAYYORLASH VA ULARNING
MALAKASINI OSHIRISH
MINTAQAVIY MARKAZI**



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

O‘QUV –USLUBIY MAJMUUA

**D.Usmonova - FarDU katta
o‘qituvchisi**

2022

Modulning ishchi dasturi Oliy va oʻrta maxsus taʼlim vazirligining 2020 yil 7 dekabrda 648-sonli buyrugʻi bilan tasdiqlangan oʻquv dasturi va oʻquv rejasiga muvofiq ishlab chiqilgan va FarDU Ilmiy kengashining 2021 yil «30» dekabrda 5 -sonli qarori bilan tasdiqlangan.

Tuzuvchi:

D.Usmonova - FarDU katta oʻqituvchisi

Taqrizchi

N.Toirova – Ingliz tili oʻqitish metodikasi kafedra mudiri, f.f.d., PhD

MUNDARIJA

I.	ISHCHI DASTUR	6
II.	MODULNI O‘QITISHDA FOYDALANILADIGAN INTERFAOL TA‘LIM METODLARI	11
III.	NAZARIY MASHG‘ULOT MATERIALLARI	17
IV.	AMALIY MASHG‘ULOT MATERIALLARI	39
III.	KEYSLAR BANKI	90
IV.	O‘QUV MATERIALLARIVA ULARNI O‘TKAZISH BO‘YICHA TAVSIYALAR VA TOPSHIRIQLAR	110
V.	BITIRUV ISHLARI MAVZULARI	114
VI.	GLOSSARIY	115
VII.	ADABIYOTLAR RO‘YXATI	125

ISHCHI DASTUR

KIRISH

Dastur O‘zbekiston Respublikasining 2020 yil 23 sentabrda tasdiqlangan “Ta’lim to‘g‘risida”gi Qonuni, **O‘zbekiston Respublikasi Prezidentining 2017 yil 30-iyundagi “O‘zbekiston Respublikasi Fanlar akademiyasi huzurida O‘zbekistonning eng yangi tarixi bo‘yicha Jamoatchilik kengashi faoliyatini tashkil etish to‘g‘risida”gi PQ-3105-sonli Qarori**, 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 ta’lim muassasalari rahbar va pedagog kadrlarining uzluksiz malakasini oshirish tizimini joriy etish to‘g‘risida”gi PF-5789-son, 2019 yil 8 oktabrdagi “O‘zbekiston Respublikasi oliy ta’lim tizimini 2030 yilgacha rivojlantirish konsepsiyasini tasdiqlash to‘g‘risida”gi PF-5847-sonli Farmonlari hamda O‘zbekiston Respublikasi Vazirlar Mahkamasining 2019 yil 23 sentabrdagi “Oliy ta’lim muassasalari rahbar va pedagog kadrlarining malakasini oshirish tizimini yanada takomillashtirish bo‘yicha qo‘shimcha chora-tadbirlar to‘g‘risida”gi 797-sonli Qarorida belgilangan ustuvor vazifalar mazmunidan kelib chiqqan holda tuzilgan bo‘lib, u oliy ta’lim 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 ingliz tilini 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, yozish, gaplashish ko‘nikmalari bo‘yicha testlar tahlil qilish va yaratish ko‘nikmasini shakllantirish bo‘yicha bilim, ko‘nikma va malakalarini shakllantirishni nazarda tutadi.

Dastur doirasida berilayotgan mavzular ta’lim sohasi bo‘yicha pedagog kadrlarni qayta tayyorlash va malakasini oshirish mazmuni, sifati va ularning tayyorgarligiga qo‘yiladigan umumiy malaka talablari va o‘quv rejalari asosida shakllantirilgan bo‘lib, bu orqali oliy ta’lim muassasalari pedagog kadrlarining sohaga oid zamonaviy ta’lim va innovatsiya texnologiyalari, ilg‘or xorijiy tajribalardan samarali foydalanish, axborotkommunikatsiya texnologiyalarini o‘quv jarayoniga keng tatbiq etish, chet tillarini intensiv o‘zlashtirish darajasini

o'shish hisobiga ularning kasb mahoratini, ilmiy faoliyatini muntazam yuksaltirish, oliy ta'lim muassasalarida o'quv -tarbiya jarayonlarini tashkil etish va boshqarishni tizimli tahlil qilish, shuningdek, pedagogik vaziyatlarda optimal qarorlar qabul qilish bilan bog'liq kompetensiyalarga ega bo'lishlari ta'minlanadi.

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

“Til kompetensiyalarini baxolash mexanizmi” 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.

“Til kompetensiyalarini baxolash mexanizmi” 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'nikma va malakalariga qo'yiladigan talablar

“Til kompetensiyalarini baxolash mexanizmi”

- moduli bo'yicha tinglovchilarning bilim, ko'nikma va malakalariga qo'yiladigan talablar tegishli ta'lim sohasi bo'yicha pedagog kadrlarni qayta tayyorlash va malakasini oshirish mazmuni, sifati va ularning tayyorgarligi hamda kompetentligiga qo'yiladigan umumiy malaka talablari bilan belgilanadi.
- Maxsus fanlar bo'yicha tinglovchilar quyidagi yangi bilim, ko'nikma, malaka hamda kompetensiyalarga ega bo'lishlari talab etiladi:
- Tinglovchilar:

- xorijiy tilni o‘qitishdagi o‘quv jarayonini rejalashtirish;
- xorijiy tilni o‘qitishda sinf menejmenti;
- o‘quv jarayonini baholash mezonlari;
- til o‘rganish jarayonlari va baholash usullari haqida **bilishi** kerak;
- 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;
- tinglovchilarni o‘zlashtirishini tahlil qilish, baholash va turli amaliy topshiriqlarniyarata olish;
- o‘z kasbiy maxoratini oshirishda kasbga doir adabiyotlardan va shaxsiy amaliyotdan unumli foydalanishni yo‘lga qo‘yish **ko‘nikmalariga ega bo‘lishi** zarur.
- chet tili ta’limida bilim, ko‘nikma, malakalarni baholash;
- lingvistik va madaniyatlararo kompetentsiyalarni baholash;
- nutq kommunikatsiyasi elementlari, sotsiolingvistik elementlar pragmatik elementlar, baholashning asosiy turlarini o‘zlashtirish;
- mavzuga oid olimlar tomonidan bildirilgan fikrlarni tahlil qilish malakalarini egallashi lozim.

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

“**Texnologiyaga asoslangan til o‘qitish**” moduli mazmuni o‘quv rejadagi diskurs tahlili o‘quv moduli bilan uzviy bog‘langan holda pedagoglarning til ko‘nikmalarini talab darajasida qo‘llay olish malakasini orttirishga xizmat qiladi.

Modulning oliy ta’limdagi o‘rni

Modulni o‘zlashtirish orqali tinglovchilar ilg‘or xorijiy mamlakatlarda o‘qitishni tashkil qilishning xorijiy tajribalarni o‘rganish, amalda qo‘llash va baholashga doir kasbiy kompetentlikka ega bo‘ladilar. So‘nggi yillarda xorijiy tillar sohasidagi yutuqlar va istiqbollar oliy o‘quv yurtlaridagi ta’lim jarayonining mazmunini boyitishga xizmat qiladi.

“Til kompetensiyalarini baholash mexanizmlari” modulining soatlar bo‘yicha taqsimoti

№	Modul mavzulari	Tinglovchining o‘quv yuklamasi, soat				
		Hammasi	Auditoriya o‘quv yuklamasi			Ko‘chma mashg‘ulot
			Jami	Jumladan		
				Nazariy	Amaliy mashg‘ulot	
1.	Til bilish darajasini baholash tamoyillari. Baholash mezonlarini yaratish va ulardan samarali foydalanish.	4	4	2	2	
2.	Baholash va material to‘plash. Kommunikativ va vazifaga asoslangan til o‘rgatishda baholash mezonlarini qo‘llash. Alternativ baholash. Sinfni tadqiq qilish.	2	2		2	
3.	Fidbek berish. O‘z o‘zini baholash. Baholashning asosiy turlari. Mavzuga oid olimlar fikrlari tahlili. Portfolio tuzish. CEFR tizimi bo‘yicha baholash.	4	4		4	
4.	Testlar yaratish usullari. Test prinsiplari.	4	4	2	2	
5	Audio matnlarni tanlash va matn xaritasini yaratish.	4	4		4	
6	Tinglab tushunish, o‘qish, yozuv, gaplashish ko‘nikmasi bo‘yicha testlar taxlil qilish va yaratish ko‘nikmasini shakllantirish.	2	2		2	
Jami:		20	20	4	16	0

NAZARIY MASHG‘ULOT MATERIALLARI

1-Mavzu: Til bilish darajasini baholash tamoyillari. Baholash mezonlarini yaratish va ulardan samarali foydalanish.

Baholash va material to‘plash. Kommunikativ va vazifaga asoslangan til o‘rgatishda baholash mezonlarini qo‘llash. Alternativ baholash. Sinfni tadqiq qilish. Fidbek berish. O‘z o‘zini baholash. Baholashning asosiy turlari. Mavzuga oid olimlar fikrlari tahlili. Portfolio tuzish. CEFR tizimi bo‘yicha baholash.

2-Mavzu: Testlar yaratish usullari. Test prinsiplari.

Audio matnlarni tanlash va matn xaritasini yaratish. Tinglab tushunish, o‘qish, yozuv, gaplashish ko‘nikmasi bo‘yicha testlar taxlil qilish va yaratish ko‘nikmasini shakllantirish.

AMALIY MASHG‘ULOTLAR

1-Amaliy mashg‘ulot. Til bilish darajasini baholash tamoyillari. Baholash mezonlarini yaratish va ulardan samarali foydalanish.

2-Amaliy mashg‘ulot. Baholash va material to‘plash. Kommunikativ va vazifaga asoslangan til o‘rgatishda baholash mezonlarini qo‘llash. Alternativ baholash. Sinfni tadqiq qilish.

3-Amaliy mashg‘ulot. Fidbek berish. O‘z o‘zini baholash. Baholashning asosiy turlari. Mavzuga oid olimlar fikrlari tahlili. Portfolio tuzish. CEFR tizimi bo‘yicha baholash.

4-Amaliy mashg‘ulot. Testlar yaratish usullari. Test prinsiplari.

5-Amaliy mashg‘ulot. Audio matnlarni tanlash va matn xaritasini yaratish.

6-Amaliy mashg‘ulot. Tinglab tushunish, o‘qish, yozuv, gaplashish ko‘nikmasi bo‘yicha testlar taxlil qilish va yaratish ko‘nikmasini shakllantirish.

ADABIYOTLAR RO'YXATI

1. Бепесова, Ж. (2017). The impact of the CEFR on teaching and testing English in the local context. *Theory and Practice in Language Studies*, 7(11), 959-964.
2. Brown, J. D., & Abeywickrama, P. (2014). *Language assessment. Principles and classroom practices*. White Plains, NY: Pearson Education.
3. Coupland, N., & Jaworski, A. (2019). *The New Sociolinguistics Reader*. Palgrave: Macmillan.
4. Chomsky, N. (2014). *Aspects of the theory of Syntax*. Cambridge, MA: M.I.T. Press.
5. Croft, W. (2011). *Radical construction grammar*. Cambridge: Cambridge University Press.
6. Celce-Murcia, M. & Olshtain, E. (2015). *Discourse and context in language teaching: A guide for language teachers*. Cambridge: Cambridge University Press.
7. Cumming, A. and R. Berwick (Eds.). (2010) *Validation in Language Testing*. Clevedon, England: Multilingual Matters, Ltd.
8. David L. Chiyesa (Ed), U. Azizov, S. Xan, K. Nazamutdinova, K. Tangirova. *Re-conceptualizing Language Teaching: An-In-service Teacher education course in Uzbekistan*. Baktria Press 2019. ISBN 978-9943-5809-9-2. 205.
9. Diane Larsen-Freyman, Marti Anderson. *Techniques & Principles in Language Teaching* (3rd edition). Oxford University Press. 2011.
10. Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Sagari & J. Banerjee (Eds.), *Handbook of second language assessment: Volume 12 of handbooks on applied linguistics* (pp. 413-428). Berlin, Germany: Walter de Gruyter GmbH & Co KG. doi: <https://doi.org/10.1515/9781614513827-027>
11. Jennings, G. (2014). *A guide to language testing: Development, evaluation, research*. Newbury House Publishers.
12. Hughes, A. (2013). *Testing for Language Teachers*. 2nd ed. Cambridge: Cambridge University Press.
13. Langacker, R. W. (2011). *Foundations of Cognitive Grammar (Vol. 2): Descriptive Application*. Stanford: Stanford University Press
14. Gardner J. (2013) *Multiple Intelligences: The Theory and Practice*, New York: Basic Books.
15. Mattyews, P. J. (2014). *Syntax*. Cambridge: Cambridge University Press.
16. McMillan, J. (2014). *Fundamental Assessment Principles for Teachers and School Administrators. Practical Assessment, Research and evaluation*. Retrieved from <https://pareonline.net/getvn.asp?v=7&n=8> [Accessed July 17, 2018]

17. Madsen S. (2014). *Techniques in Testing*. England: Oxford University Press.
18. McNamara, T. (2014). *Language Testing*. UK. Oxford: Oxford University Press.
19. Knight, P., & Yorke, M. (2013). *Assessment, learning and employability*. Maidenhead: Society for Research into Higher Education & Open University Press.
20. Jack C. Richards, Willy A. Renandya. *Methodology in Language Teaching: An anthology of current practice*. Cambridge University Press. 2012.
21. Isaacs, T., Zara, C., Herbert, G., Coombs, S. J., & Smit, C. (2013). Assessment of learning. In *The SAGE Key Concepts Series: Key concepts in educational assessment* (pp. 12-17). London: SAGE Publications Ltd doi: 10.4135/9781473915077.n4
22. Richards, C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press. Stanley Kubrick. *A narrative and stylistic analysis*. An imprint of Greyenwood Publishing Group, Inc. Velikobritaniya (United Kingdom), 2011.
23. Stepien M. Alessi, Stanley R. Trollip. *Multimedia for Learning: Methods and Development* (3rd edition). Pearson; 3 edition (November 19, 2000), 580 pages, 2011.
24. Shamy, E. (2011). *The power of tests: A critical perspective on the uses of language tests*. Pearson Education Limited.
25. Taylor, L. (2019). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36.
26. Weir, C. (2015). *Language testing and validation: An evidence-based approach*. Palgrave
27. Wardaugh, R. (2016). *An Introduction to Sociolinguistics*. Sussex, U.K.: Wiley-Blackwell.

Интернет сайтлар

28. <http://edu.uz> – O‘zbekiston Respublikasi Oliy va o‘rta maxsus ta’lim vazirligi
29. <http://lex.uz> – O‘zbekiston Respublikasi Qonun hujjatlari ma’lumotlari milliy bazasi
30. <http://bimm.uz> – Oliy ta’lim tizimi pedagog va rahbar kadrlarini qayta tayyorlash va ularning malakasini oshirishni tashkil etish bosh ilmiy-metodik markazi
31. <http://ziyonet.uz> – Ta’lim portali Ziyonet
32. www.Britishcouncil.org -Британия кенгаши веб сайти
33. www.ziyonet.uz -Таълим портали
34. www.tefl.net- Тил ўрганиш веб сайти
busyteacher.org -Тил ўрганиш веб сайти

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.

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.NAZARIY MASHG‘ULOTLAR

NAZARIY MATERIALLAR

1. LECTURE: INTRODUCTION TO LANGUAGE ASSESSMENT AND TESTING (Тил билиш даражасини баҳолаш тамойиллари. Баҳолаш мезонларини яратиш ва улардан самарали фойдаланиш.)

Plan:

- 1. Introduction.**
- 2. History of assessment and testing**
- 3. Assessment of language competences**
- 4. Writing Effective Assessment Criteria**

Key words: Testing assessment; language skills; assessment for learning; assessment of learning, anxiety and self-doubt, confirmations of inadequacy, commercially administered standardized test, knowledge of lexical items, data analysis techniques

This lesson 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 twomain areas of language assessment and testing known as Assessment for Learning and Assessment of Learning.

INTRODUCTION

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 shrive
- d. a banker or money changer who evaluates coin

3. hadal

- a. relating to the deepest parts of the ocean below 20,000 feet
- b. one of seven stations in the Islamic hajj, or pilgrimage, to Mecca
- c. a traditional Romanian folk dance performed at spring festivals
- d. pertaining to Hades

4. chary

- a. discreetly cautious and vigilant about dangers and risks
- b. pertaining to damp, humid weather before a rainstorm
- c. optimistic, positive, looking on the bright side
- d. expensive beyond one's means

5. yabby

- a. overly talkative, obnoxiously loquacious

- b. any of various burrowing Australian crayfishes
- c. a small horse-drawn carriage used in Victorian England for transporting one or two persons
- d. in clockwork mechanisms, a small latch for calibrating the correct time

Now, how did that make you feel? Probably just the same as many learners 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.

ASSESSMENT OF LANGUAGE COMPETENCES

Language competence is a broad term which includes linguistic or grammatical competence, discourse competence, sociolinguistic competence, and strategic competence. The specific learning outcomes under the heading Language Competence deal with knowledge of the language and the ability to use that knowledge to interpret and produce meaningful texts appropriate to the situation in which they are used. Language competence is best developed in the context of learning activities or tasks where the language is used for real purposes, in other words, in practical applications. Grammatical competence was seen to encompass “knowledge of lexical items and of rules of morphology, syntax, sentence-grammar semantics, and phonology” (Canale and Swain, 1980, p. 29).

Discourse competence was defined as the ability to connect sentences in stretches of discourse and to form a meaningful whole out of a series of utterances. Sociolinguistic competence was defined as involving knowledge of the sociocultural rules of language and of discourse. Sociolinguistic component – assesses the use of linguistic forms in language performance. For example when a student bumps into a professor, spilling her coffee on the professor’s dress, “Sorry!” would probably constitute an inadequate apology. This category assesses the speakers’ control over the actual language forms used to realize the speech function, in this case referred to as a speech act (such as, “sorry,” “excuse me,” “very sorry,” “really sorry”), as well as their control over register or formality of the utterance

from most intimate to most formal language. Strategic competence was seen to refer to “the verbal and nonverbal communication strategies that may be called into action to compensate for breakdowns in communication due to performance variables or due to insufficient competence” (Canale and Swain, 1980, p. 30).

Writing Effective Assessment Criteria

Assessment criteria are descriptive statements that provide learners and instructors with information about the qualities, characteristics, and aspects of a given learning task. Assessment criteria make it clear to learners what they are expected to do to demonstrate achievement of the learning outcomes and factors instructors will take into account when making judgments about their performance. Sharing assessment criteria with students at the beginning of the course is an effective way to help students build confidence in their learning and improve their performance. Making assessment criteria explicit helps them recognize what is important and valued in the curriculum, focus their efforts on key learning outcomes and evaluate their own performance through self-assessment and reflection. Assessment criteria are a way to provide formative feedback throughout a course to support ongoing learning, as well as to provide end-of-term summative evaluation.

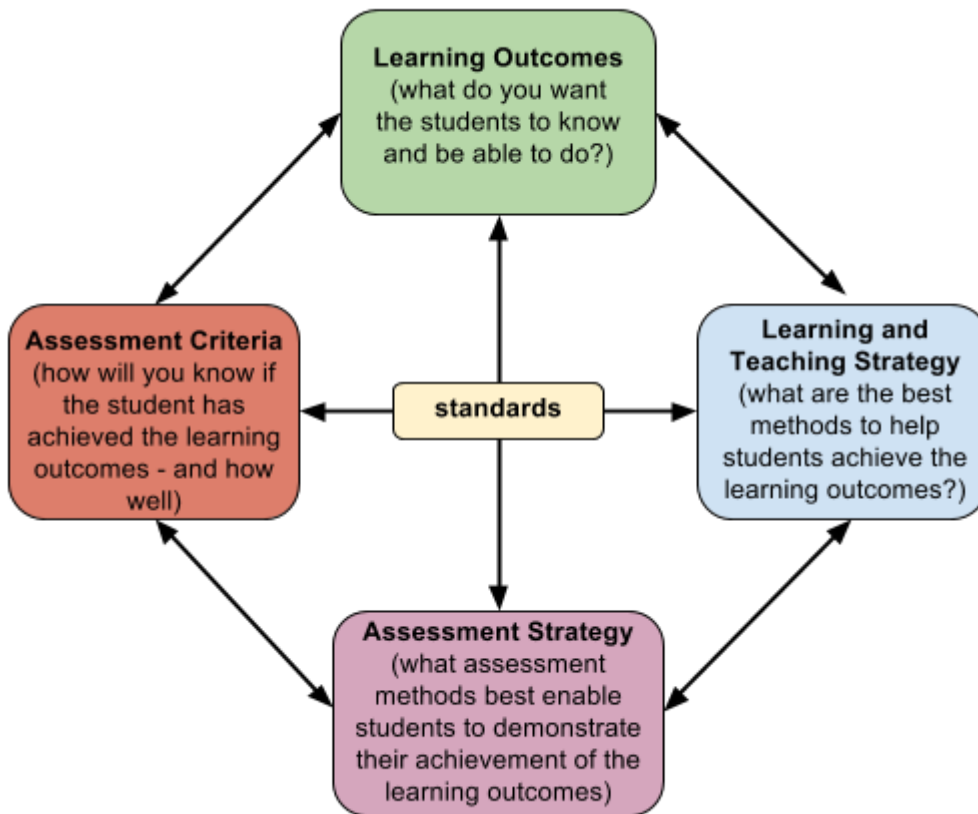
Assessment criteria take the “guess-work” out of grading for instructors and students. Well-defined assessment criteria allow instructors to evaluate learners’ work more openly, consistently and objectively. This increases a sense of fairness from the students’ perspective and makes it easy for the instructor to explain / justify how marks have been awarded.

EXAMPLES OF ASSESSMENT CRITERIA

- create a design brief that incorporates design process and principles
- analyze distributions using probability and data analysis techniques
- use tools and models to integrate quantitative and qualitative information in problem-based analysis using relevant tools and models
- adopt a variety of roles on a consulting team task force
- justify management's actions in selecting specific recommendations among alternative possibilities

RELATIONSHIP OF ASSESSMENT CRITERIA TO A PROGRAM OF STUDY

Before writing assessment criteria it is important to understand how assessment criteria relate to course design. The following diagram illustrates how assessment criteria both inform and are informed by learning outcomes.



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 were 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?

2 LECTURE: Methods of creating tests. Test principles.

(Тестлар яратиш усуллари. Тест принциплари.)

Plan:

1. Principles of assessment
2. Planning the Test
3. Preparing the Test
4. Try Out of the Test
5. Evaluating the Test

Key words:reliability, validity, authenticity, washback, entry performance, summative tests, an elaborate design, **weightage**, learning outcomes, **be free from ambiguity**, unintentional clues.

1.Principles of assessment serve as a guidelines to ensure that the test is useful, appropriate, effective, and plausible.^[1] This principles are crucial to be taken into consideration because assessment is an important aspect of educational process which determines the level of accomplishments of students.^[2]

There are five general principles of assessment:^{[1][3]}

1. practicality
2. reliability
3. validity
4. authenticity
5. washback

Practicality

This principle refers to the time and cost constraints during the construction and administration of an assessment instrument.^[1] Meaning that the test should be economical to provide. The format of the test should be simple to understand. Moreover, solving a test should remain within suitable time. It is generally simple to administer. Its assessment procedure should be particular and time-efficient.^[4]

Reliability

The principle of reliability refers to the stability of scores over time and different raters. There are four types of reliability: student-related which can be personal problems, sickness, or fatigue, rater-related which includes bias and subjectivity, test administration-related which is the conditions of test taking process, test-related which is basically related to the nature of a test.^{[5][1][4]}

Validity

Validity refers to the tests that measure what it claims to measure.^{[1][5]}

- content validity
- criterion validity
- construct validity
- consequential validity
- face validity

Authenticity

The assessment instrument is authentic when it is contextualized, contains natural language and meaningful, relevant, and interesting topic, and replicates real world experiences.^[1]

Washback

This principle refers to the consequence of an assessment on teaching and learning within classrooms.^[1] Washback can be positive and negative. Positive washback refers to the desired effects of a test, while negative washback refers to the negative consequences of a test. In order to have positive washback, instructional planning can be used.^[6]

2. Planning the Test:

Planning of the test is the first important step in the test construction. The main goal of evaluation process is to collect valid, reliable and useful data about the student.

Therefore before going to prepare any test we must keep in mind that:

- (1) What is to be measured?
- (2) What content areas should be included and
- (3) What types of test items are to be included.

Therefore the first step includes three major considerations.

1. Determining the objectives of testing.
2. Preparing test specifications.
3. Selecting appropriate item types.

Determining the Objectives of Testing:

A test can be used for different purposes in a teaching learning process. It can be used to measure the entry performance, the progress during the teaching learning process and to decide the mastery level achieved by the students. Tests serve as a good instrument to measure the entry performance of the students. It answers to the questions, whether the students have requisite skill to enter into the course or not, what previous knowledge does the pupil possess. Therefore it must be decided whether the test will be used to measure the entry performance or the previous knowledge acquired by the student on the subject.

Tests can also be used for formative evaluation. It helps to carry on the teaching learning process, to find out the immediate learning difficulties and to suggest its remedies. When the difficulties are still unsolved we may use diagnostic tests. Diagnostic tests should be prepared with high technique. So specific items to diagnose specific areas of difficulty should be included in the test.

Tests are used to assign grades or to determine the mastery level of the students. These summative tests should cover the whole instructional objectives and content areas of the course. Therefore attention must be given towards this aspect while preparing a test.

2. Preparing Test Specifications:

The second important step in the test construction is to prepare the test specifications. In order to be sure that the test will measure a representative sample of

the instructional objectives and content areas we must prepare test specifications. So that an elaborate design is necessary for test construction. One of the most commonly used devices for this purpose is ‘Table of Specification’ or ‘Blue Print.’

Preparation of table of specification is the most important task in the planning stage. It acts, as a guide for the test construction. Table of specification or ‘Blue Print’ is a three dimensional chart showing list of instructional objectives, content areas and types of items in its dimensions.

It includes four major steps:

- (i) Determining the weightage to different instructional objectives.
- (ii) Determining the weightage to different content areas.
- (iii) Determining the item types to be included.
- (iv) Preparation of the table of specification.

(i) Determining the weightage to different instructional objectives:

There are vast arrays of instructional objectives. We cannot include all in a single test. In a written test we cannot measure the psychomotor domain and affective domain. We can only measure the cognitive domain. It is also true that all the subjects do not contain different learning objectives like knowledge, understanding, application and skill in equal proportion. Therefore it must be planned how much weightage to be given to different instructional objectives. While deciding this we must keep in mind the importance of the particular objective for that subject or chapter.

For example if we have to prepare a test in General Science for Class—X we may give the weightage to different instructional objectives as following:

Table 3.1. Showing weightage given to different instructional objectives in a test of 100 marks:

Objectives	Weightage in %	No. of Questions (Marks)
Knowledge	20%	20
Understanding	40%	40
Application	40%	40

(ii) Determining the weightage to different content areas:

The second step in preparing the table of specification is to outline the content area. It indicates the area in which the students are expected to show their performance. It helps to obtain a representative sample of the whole content area.

It also prevents repetition or omission of any unit. Now question arises how much weightage should be given to which unit. Some experts say that, it should be decided by the concerned teacher keeping the importance of the chapter in mind. Others say that it should be decided according to the area covered by the topic in the text book. Generally it is decided on the basis of pages of the topic, total page in the book and number of items to be prepared. For example if a test of 100 marks is to be prepared then, the weightage to different topics will be given as following.

Weightage of a topic:

$$= \frac{\text{Total number of items/marks}}{\text{Total number of pages in the book}} \times \text{Number of pages in the topic.}$$

If a book contains 250 pages and 100 test/items (marks) are to be constructed then the weightage will be given as, following:

Table 3.2. Table showing weightage given to different content areas:

Sl. No.	Topic	No. of Items/Marks	% of Items/Marks
1.	Topic No—1 Page—1—24	10	10
2.	Topic No—2 Page—25—55	12	12
3.	Topic No—3 Page—56—114	24	24
4.	Topic No—4 Page—115—218	42	42
5.	Topic No—5 Page—219—250	12	12
	Total	100	100

(iii) Determining the item types:

The third important step in preparing table of specification is to decide appropriate item types. Items used in the test construction can broadly be divided into two types like objective type items and essay type items. For some instructional purposes, the objective type items are most efficient where as for others the essay questions prove satisfactory.

Appropriate item types should be selected according to the learning outcomes to be measured. For example when the outcome is writing, naming supply type items are useful. If the outcome is identifying a correct answer selection type or recognition type items are useful. So that the teacher must decide and select appropriate item types as per the learning outcomes.

(iv) Preparing the Three Way Chart:

Preparation of the three way chart is last step in preparing table of specification. This chart relates the instructional objectives to the content area and types of items. In a table of specification the instructional objectives are listed across the top of the table, content areas are listed down the left side of the table and under each objective the types of items are listed content-wise. Table 3.3 is a model table of specification for X class science.

Table 3.3. Table of Specification for Science (Biology) Class—X

Instructional Objectives Content Areas	Knowledge				Understanding				Application				TOTAL
	SU.	MC.	MT	T.F.	SU	M.C.	M.T.	T.F.	SU.	M.C.	MT.	T.F.	
1. Food	0	1	0	1	1	1	0	2	2	1	1	0	10
2. Food Deficiency Diseases	1	0	1	0	2	1	1	1	1	2	1	1	12
3. Diseases due to overtake of food	1	2	1	1	2	3	2	3	4	2	2	2	24
4. Food yields	2	3	1	2	6	2	7	2	6	3	6	2	42
5. Essentials for Good health	1	0	1	1	1	1	1	1	1	2	0	1	12
Total	5	6	4	5	12	8	11	9	14	10	10	6	100

Note : SU = Supply Type, M.C. = Multiple Choice type items, M.T. = Matching type items, T.F. = True false items.

3. Preparing the Test:

After planning preparation is the next important step in the test construction. In this step the test items are constructed in accordance with the table of specification. Each type of test item need special care for construction.

The preparation stage includes the following three functions:

- (i) Preparing test items.
- (ii) Preparing instruction for the test.
- (iii) Preparing the scoring key.

(i) Preparing the Test Items:

Preparation of test items is the most important task in the preparation step. Therefore care must be taken in preparing a test item. The following principles help in preparing relevant test items.

1. Test items must be appropriate for the learning outcome to be measured:

The test items should be so designed that it will measure the performance described in the specific learning outcomes. So that the test items must be in accordance with the performance described in the specific learning outcome.

For example:

Specific learning outcome—Knows basic terms

Test item—An individual is considered as obese when his weight is % more than the recommended weight.

2. Test items should measure all types of instructional objectives and the whole content area:

The items in the test should be so prepared that it will cover all the instructional objectives—Knowledge, understanding, thinking skills and match the specific learning outcomes and subject matter content being measured. When the items are constructed on the basis of table of specification the items became relevant.

3. The test items should be free from ambiguity:

The item should be clear. Inappropriate vocabulary and awkward sentence structure should be avoided. The items should be so worded that all pupils understand the task.

Example:

Poor item —Where did Gandhi born

Better —In which city did Gandhi born?

4. The test items should be of appropriate difficulty level:

The test items should be proper difficulty level, so that it can discriminate properly. If the item is meant for a criterion-referenced test its difficulty level should be as per the difficulty level indicated by the statement of specific

learning outcome. Therefore if the learning task is easy the test item must be easy and if the learning task is difficult then the test item must be difficult.

In a norm-referenced test the main purpose is to discriminate pupils according to achievement. So that the test should be so designed that there must be a wide spread of test scores. Therefore the items should not be so easy that everyone answers it correctly and also it should not be so difficult that everyone fails to answer it. The items should be of average difficulty level.

5. The test item must be free from technical errors and irrelevant clues:

Sometimes there are some unintentional clues in the statement of the item which helps the pupil to answer correctly. For example grammatical inconsistencies, verbal associations, extreme words (ever, seldom, always), and mechanical features (correct statement is longer than the incorrect). Therefore while constructing a test item careful step must be taken to avoid most of these clues.

6. Test items should be free from racial, ethnic and sexual biasness:

The items should be universal in nature. Care must be taken to make a culture fair item. While portraying a role all the facilities of the society should be given equal importance. The terms used in the test item should have an universal meaning to all members of group.

(ii) Preparing Instruction for the Test:

This is the most neglected aspect of the test construction. Generally everybody gives attention to the construction of test items. So the test makers do not attach directions with the test items.

But the validity and reliability of the test items to a great extent depends upon the instructions for the test. N.E. Gronlund has suggested that the test maker should provide clear-cut direction about;

- a. The purpose of testing.
- b. The time allowed for answering.
- c. The basis for answering.
- d. The procedure for recording answers.
- e. The methods to deal with guessing.

Direction about the Purpose of Testing:

A written statement about the purpose of the testing maintains the uniformity of the test. Therefore there must be a written instruction about the purpose of the test before the test items.

Instruction about the time allowed for answering:

Clear cut instruction must be supplied to the pupils about the time allowed for whole test. It is also better to indicate the approximate time required for answering each item, especially in case of essay type questions. So that the test maker should carefully judge the amount of time taking the types of items, age and ability of the students and the nature of the learning outcomes expected. Experts are of the opinion that it is better to allow more time than to deprive a slower student to answer the question.

Instructions about basis for answering:

Test maker should provide specific direction on the basis of which the students will answer the item. Direction must clearly state whether the students will select the answer or supply the answer. In matching items what is the basis of matching the premises and responses (states with capital or country with production) should be given. Special directions are necessary for interpretive items. In the essay type items clear direction must be given about the types of responses expected from the pupils.

Instruction about recording answer:

Students should be instructed where and how to record the answers. Answers may be recorded on the separate answer sheets or on the test paper itself. If they have to answer in the test paper itself then they must be directed, whether to write the correct answer or to indicate the correct answer from among the alternatives. In case of separate answer sheets used to answer the test direction may be given either in the test paper or in the answer sheet.

Instruction about guessing:

Direction must be provided to the students whether they should guess uncertain items or not in case of recognition type of test items. If nothing is stated about guessing, then the bold students will guess these items and others will answer only those items of which they are confident. So that the bold pupils by chance will

answer some items correctly and secure a higher score. Therefore a direction must be given ‘to guess but not wild guesses.’

(iii) Preparing the Scoring Key:

A scoring key increases the reliability of a test. So that the test maker should provide the procedure for scoring the answer scripts. Directions must be given whether the scoring will be made by a scoring key (when the answer is recorded on the test paper) or by a scoring stencil (when answer is recorded on separate answer sheet) and how marks will be awarded to the test items.

In case of essay type items it should be indicated whether to score with ‘point method’ or with the ‘rating’ method.’ In the ‘point method’ each answer is compared with a set of ideal answers in the scoring Key. Then a given number of points are assigned. In the rating method the answers are rated on the basis of degrees of quality and determine the credit assigned to each answer. Thus a scoring key helps to obtain a consistent data about the pupils’ performance. So the test maker should prepare a comprehensive scoring procedure along with the test items.

4. Try Out of the Test:

Once the test is prepared now it is time to be confirming the validity, reliability and usability of the test. Try out helps us to identify defective and ambiguous items, to determine the difficulty level of the test and to determine the discriminating power of the items.

Try out involves two important functions:

- (a) Administration of the test.
- (b) Scoring the test.

(a) Administration of the test:

Administration means administering the prepared test on a sample of pupils. So the effectiveness of the final form test depends upon a fair administration. Gronlund and Linn have stated that ‘the guiding principle in administering any class room test is that all pupils must be given a fair chance to demonstrate their achievement of learning outcomes being measured.’ It implies that the pupils must be provided congenial physical and psychological environment during the time of testing. Any other factor that may affect the testing procedure should be controlled.

Physical environment means proper sitting arrangement, proper light and ventilation and adequate space for invigilation, Psychological environment refers to these aspects which influence the mental condition of the pupil. Therefore steps should be taken to reduce the anxiety of the students. The test should not be administered just before or after a great occasion like annual sports on annual drama etc.

One should follow the following principles during the test administration:

1. The teacher should talk as less as possible.
2. The teacher should not interrupt the students at the time of testing.
3. The teacher should not give any hints to any student who has asked about any item.
4. The teacher should provide proper invigilation in order to prevent the students from cheating.

(b) Scoring the test:

Once the test is administered and the answer scripts are obtained the next step is to score the answer scripts. A scoring key may be provided for scoring when the answer is on the test paper itself Scoring key is a sample answer script on which the correct answers are recorded.

When answer is on a separate answer sheet at that time a scoring stencil may be used for answering the items. Scoring stencil is a sample answer sheet where the correct alternatives have been punched. By putting the scoring stencil on the pupils answer script correct answer can be marked. For essay type items separate instructions for scoring each learning objective may be provided.

Correction for guessing:

When the pupils do not have sufficient time to answer the test or the students are not ready to take the test at that time they guess the correct answer, in recognition type items.

In that case to eliminate the effect of guessing the following formula is used:

$$\text{Score} = R - \frac{W}{n-1}$$

Where **R** = No. of right responses
W = No. of wrong responses
n = No. of alternatives.

But there is a lack of agreement among psychometricians about the value of the correction formula so far as validity and reliability are concerned. In the words of Ebel “**neither the instruction nor penalties will remedy the problem of guessing.**” Guilford is of the opinion that “when the middle is excluded in item analysis the question of whether to correct or not correct the total scores becomes rather academic.” Little said “**correction may either under or over correct the pupils’ score.**” Keeping in view the above opinions, the test-maker should decide not to use the correction for guessing. To avoid this situation he should give enough time for answering the test item.

5. Evaluating the Test:

Evaluating the test is most important step in the test construction process. Evaluation is necessary to determine the quality of the test and the quality of the responses. Quality of the test implies that how good and dependable the test is? (Validity and reliability). Quality of the responses means which items are misfit in the test. It also enables us to evaluate the usability of the test in general class-room situation.

Evaluating the test involves following functions:

- (a) Item analysis.
- (b) Determining validity of the test.
- (c) Determining reliability of the test.
- (d) Determining usability of the test.

(a) Item analysis:

Item analysis is a procedure which helps us to find out the answers to the following questions:

- a. Whether the items functions as intended?
- b. Whether the test items have appropriate difficulty level?
- c. Whether the item is free from irrelevant clues and other defects?
- d. Whether the distracters in multiple choice type items are effective?

The item analysis data also helps us:

- a. To provide a basis for efficient class discussion of the test result
- b. To provide a basis for the remedial works
- c. To increase skill in test construction
- d. To improve class-room discussion.

Item Analysis Procedure:

Item analysis procedure gives special emphasis on item difficulty level and item discriminating power.

The item analysis procedure follows the following steps:

1. The test papers should be ranked from highest to lowest.
2. Select 27% test papers from highest and 27% from lowest end.

For example if the test is administered on 60 students then select 16 test papers from highest end and 16 test papers from lowest end.

3. Keep aside the other test papers as they are not required in the item analysis.
4. Tabulate the number of pupils in the upper and lower group who selected each alternative for each test item. This can be done on the back of the test paper or a separate test item card may be used (Fig. 3.1)

5. Calculate item difficulty for each item by using formula:

$$\text{Item difficulty} = \frac{R}{T} \times 100$$

Where R= Total number of students got the item correct.

T = Total number of students tried the item.

In our example (fig. 3.1) out of 32 students from both the groups 20 students have answered the item correctly and 30 students have tried the item.

The item difficulty is as following:

$$\begin{aligned} \text{Item difficulty} &= \frac{R}{T} \times 100 \\ &= \frac{20}{30} \times 100 = 66.67. \end{aligned}$$

It implies that the item has a proper difficulty level. Because it is customary to follow 25% to 75% rule to consider the item difficulty. It means if an item has a item difficulty more than 75% then is a too easy item if it is less than 25% then item is a too difficult item.

6. Calculate item discriminating power by using the following formula:

$$\frac{R_U - R_L}{T/2}$$

Item discriminating power =

Where R_U = Students from upper group who got the answer correct.

R_L = Students from lower group who got the answer correct.

$T/2$ = half of the total number of pupils included in the item analysis.

In our example (Fig. 3.1) 15 students from upper group responded the item correctly and 5 from lower group responded the item correctly.

$$\begin{aligned} \text{Discriminating power} &= \frac{R_U - R_L}{T/2} \\ &= \frac{15 - 5}{16} = \frac{10}{16} = .63 \end{aligned}$$

A high positive ratio indicates the high discriminating power. Here .63 indicates an average discriminating power. If all the 16 students from lower group and 16 students from upper group answers the item correctly then the discriminating power will be 0.00.

It indicates that the item has no discriminating power. If all the 16 students from upper group answer the item correctly and all the students from lower group answer the item in correctly then the item discriminating power will be 1.00 it indicates an item with maximum positive discriminating power.

7. Find out the effectiveness of the distracters. A distractor is considered to be a good distractor when it attracts more pupil from the lower group than the upper group. The distractors which are not selected at all or very rarely selected should be revised. In our example (fig. 3.1) distractor 'D' attracts more pupils from.

Fig.—3.1 The item analysis card.

Subject		Unit							
Class									
Item No.									
Group	No. of Pupils	Alternatives					Total No. of pupils responded	Item difficulty	Item Dis. power
		A	B	C*	D	E			
Upper	16	0	0	15	1	0	16	66.67	.63
Lower	16	5	4	5	0	0	14		

*Correct answer.

Upper group than the lower group. It indicates that distracter 'D' is not an effective distracter. 'E' is a distracter which is not responded by any one. Therefore it also needs revision. Distracter 'A' and 'B' prove to be effective as it attracts more pupils from lower group.

Preparing a test item file:

Once the item analysis process is over we can get a list of effective items. Now the task is to make a file of the effective items. It can be done with item analysis cards. The items should be arranged according to the order of difficulty. While filing the items the objectives and the content area that it measures must be kept in mind. This helps in the future use of the item.

(b) Determining Validity of the Test:

At the time of evaluation it is estimated that to what extent the test measures what the test maker intends to measure.

(c) Determining Reliability of the Test:

Evaluation process also estimates to what extent a test is consistent from one measurement to other. Otherwise the results of the test can not be dependable.

(d) Determining the Usability of the Test:

Try out and the evaluation process indicates to what extent a test is usable in general class-room condition. It implies that how far a test is usable from administration, scoring, time and economy point of view.

Questions for discussions;

1. How many are there general principles of assessment and what they are?
2. How many are there types of reliability?
3. What is the assessment instrument authentic?
4. What is washback?
5. Discuss the steps of constructing the test

References

1. *Brown, Douglas; Abeywickrama, Priyanvada (2010). Language Assessment, Principles and Classroom Practices. The United States of America: Pearson Longman. ISBN 978-0-13-814931-4.*
2. *Oxford Brookes University. "Purposes and principles of assessment". www.brookes.ac.uk. Retrieved 2018-10-09.*
3. *Oxford Brookes University. "Principles of assessment". www.brookes.ac.uk. Retrieved 2018-10-09.*
4. *"Reliability, Validity and Practicality | Teach English | Englishpost.org". Englishpost.org. 2012-06-26. Retrieved 2018-10-30.*
5. *Fawcett, Alison (2013). Principles of Assessment and Outcome Measurement for Occupational Therapists and Physiotherapists: Theory, Skills and Application. JohnWiley&Sons. ISBN 9781118709696.*
6. *"Understanding Assessment: Washback and Instructional Planning". www.cal.org. Retrieved 2018-10-29.*

https://en.wikipedia.org/wiki/General_principles_of_assessment

<https://www.yourarticlelibrary.com/education/test/top-4-steps-for-constructing-a-test/64781>

IV.AMALIY MASHG'ULOTLAR MATERIALLARI

1 PRACTICAL: ASSESSMENT FOR LEARNING: GATHERING DATA

Plan:

1. Differences of testing, assessment, and evaluation;
2. The Assessment for Learning construct;
3. 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 inaccurately – 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; Assessmentfor-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: _____

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?

1.) What are the positive qualities of the conclusion?

2.) What suggestions do you have for the author?

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).

2. PRACTICAL: ASSESSMENT FOR LEARNING: USING GATHERED DATA

Plan:

1. How to use gathered data to make changes to lesson plans, teaching materials, and syllabus;
2. How to incorporate dynamic assessment into teaching practice;
3. How to adjust teaching instructions based on the obtained data about students;
4. 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.

ACTION

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 on 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?

3 PRACTICAL: ASSESSMENT OF LEARNING: PRINCIPLES AND TYPES OF ASSESSMENT

Plan:

- 1. Distinction between Assessment for Learning and Assessment of Learning;**
- 2. Principles of language assessment;**
- 3. How to apply key principles of assessment to the quality of assessment tools;**
- 4. How to reflect on the challenges and identify ways to address them.**

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

INTRODUCTION AND OVERVIEW

Assessment is an inseparable part of teaching because language teachers have a dual role of teaching and assessing, which ultimately will have an impact on students' learning and motivation. They make decisions about who could pass or fail a quiz, test, or course of study; they determine whether the class is going well, and if the teaching they do is effective (Harding & Kremmel, 2016). Assessment of Learning measures 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 select 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 are three types of validity from Cumming & Berwick (1996), which are commonly referred to in the language testing literature:

- Construct validity: ‘a mutual verification of the measuring instrument and the theory of the construct it is meant to measure’, p.29; ‘we examine the psychological trait, or construct, presumed to be measured by the test and we cause a continuing, research interplay to take place between the scores earned on the test and the theory underlying the construct’ (p. 26)
- Content validity: ‘review of the test by subject-matter experts and a verification that its content represents a satisfactory sampling of the domain’ (p. 22)
- Face validity: ‘the appearance of validity...for example that the language and contexts of test items be expressed in ways that would look valid and be acceptable to the test taker and to the public generally’ (pp. 23 -24)

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

discrepancy in different raters' marking and the consistency of the same rater across students.

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

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

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

Table 12.Types of Assessments.

Assessment types	Main purpose
Placement test	places students at an appropriate level of instructions within a program
Diagnostic test	identifies students' strengths and weaknesses
Progress test	provides information about mastery or difficulty that learners have with course materials
Achievement test	Provides information about students' attainment of learning outcomes at the end of a course

Aptitude test measures a general ability to learn a foreign language

Proficiency test measures global competence in a language

REFLECTION

Which of the above are the most common type of tests in your teaching context? Which have you never used? Why? Test Specifications. Before designing any kind of assessment, a test developer must prepare test specifications which contain all the information about the test to be constructed. This blueprint is an informative document and also is also kept for preparing further variants of the same kinds of test in the future. Developed once, test 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 seen by the teacher and pay attention to how each principle of language assessment is reflected.

My teaching context is undergraduate students, which I taught the four skills of listening, reading, writing and speaking for Practical English. Apart from formative assessment that they have during the semester, they take one midterm and one final exam in each of the two terms in the academic year. Therefore, I usually deal with

achievement tests based on the syllabus content. It means that my tests need to reflect what students have presumably learnt and assess their knowledge of the skills that they have developed through the course. In this endeavor, my concern has always been how to develop language tests used for classroom assessment in a professional way. Having learnt the significance of different test qualities, I find it important to bear in mind that “a test cannot be valid for more than one purpose” (Henning, 1987, p. 89), which I used to ignore without realizing how inappropriate a certain test can be. I made no serious attempt to check the validity and reliability of the tests that I created or selected for assessment. All my decisions have been based on my own intuitions and the instructions from the administration and the head of my department. However, what I am inclined to do as of now is to make validity and reliability central in the development of my assessment tools. My aim from this is to make sure that students are tested for what they are intended to be tested and the tasks across all the variants of tests are of equal difficulty. Usually my department administers midterm and final tests in several slots within the same day by grouping students, as there are over 400 students to be tested. As a result, my task as a test developer is to create several variants for every slot. At the same time, I have to make sure that 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 inconsiderations with regard to main principles of an effective and useful test, it is impossible to know how good the test is unless it is piloted. My second aim is to work with other teachers in my department, particularly with those who are also

responsible for creating assessment tools. Usually when a test needs to be constructed, teachers search for ready tests that can be applicable for their purpose. This means that test selection is as important as test designing.

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

In this respect, Davies (1997) maintains that professional ethics and morality in language testing is important, which in my context is rarely discussed. It seems very necessary to “develop assessment literacy” (Taylor, 2009, p. 31) so that teachers could approach their task with a better understanding, not relying exclusively on their intuition. Only by consulting relevant theories and practices followed by collaboration and 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.

SUMMARY

In conclusion, language assessment in education is a great responsibility for the teacher because assessments can influence decisions about students and possibly their futures. Therefore, teachers should reflect each step in the development assessment instruments. Without understanding validity, reliability, practicality as well as impact (washback), it is difficult to come up with an effective language assessment. The adherence to these principles is vital. At the same time, there is no universal formula of creating an effective language assessment tool. The same test can be the best and the worst assessment instrument

depending on where and when it is used. Therefore, the main considerations are the purpose of a test and the context where it is used. Only by giving sufficient attention to these two factors can one design or choose an effective language assessment tool, which will be appropriate just for that occasion.

HOMEWORK TASK THIRTEEN

Imagine you have completed (i.e., taught) the lesson you chose for Homework Task One (A) and you want to make a short quiz (10 min) that will measure what you taught in the class. For this homework task, please write the answers to the seven questions 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 hotelworkers?)
- 6) What text types should be chosen as stimulus materials -- written and/or spoken?
- 7) What sort of tasks are required -- discrete point, integrative, simulated 'authentic', objectively assessable? (That is, what will the test-takers do?)

4 PRACTICAL: DESIGNING TESTS: TEST SPECIFICATIONS.

Plan:

- 1. How to start test designing process;**
- 2. Determining the purpose of a test;**
- 3. Designing clear, unambiguous objectives;**
- 4. Drawing up test specifications;**
- 5. 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. 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. 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 fifthgrade 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). 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.

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.

5. PRACTICAL: USING STATISTICS: OBJECTIVELY SCORED ITEMS

Plan:

1. Significance of assessment statistics for teaching and learning;
2. How to conduct basic statistical analysis of results from objectively scored items;
3. 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 fraught 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:

In other words, the mean or (\bar{X}) 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:

$14+18+19+20+21+21+21+24+26+27 = 213$ (the sum of all scores)

the sum of all scores number of students $213 : 10 = 21.3$ (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 213, 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:

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 \bar{X} 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

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

53.29 10.89 5.29 1.69 0.09 0.09 0.09 22.09 22.09 32.49

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?

6 PRACTICAL: USING STATISTICS: SUBJECTIVELY-SCORED ASSESSMENTS

Plan:

1. What subjectively-scored assessments are;

2. Difference between holistic and analytic assessments;
3. 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

Advantages	Disadvantages
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	

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?

V.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 does 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. O‘QUV MATERIALLARI VA ULARNI O‘TKAZISH BO‘YICHA
TAVSIYALAR VA TOPSHIRIQLAR**

Shaklan bir-biriga yaqin savol-javoblar

Савол	Жавоб
Tests are ...	scaring,unavoidable,measurement
Language competence is a broad term which includes linguistic or, discourse competence, sociolinguistic competence, and strategic competence.	grammatical competence
..... was defined as the ability to connect sentences in stretches of discourse and to form a meaningful whole out of a series of utterances.	Discourse competence
..... was defined as involving knowledge of the sociocultural rules of language and of discourse.	Sociolinguistic competence
..... was seen to refer to “the verbal and nonverbal communication strategies that may be called into action to compensate for breakdowns in communication due to performance variables or due to insufficient competence	Strategic competence
..... was seen to encompass “knowledge of lexical items and of rules of morphology, syntax, sentence-grammar semantics, and phonology”	Grammatical competence
.....(principle) refers to the time and cost constraints during the construction and administration of an assessment instrument	Practicality
..... refers to the	Reliability

stability of scores over time and different raters	
..... refers to the tests that measure what it claims to measure	Validity
The assessment instrument is when it is contextualized, contains natural language and meaningful, relevant, and interesting topic, and replicates real world experiences	authentic
..... refers to the consequence of an assessment on teaching and learning within classrooms	Wasback
..... of the test is the first important step in the test construction	Planning
After planning is the next important step in the test construction	preparation
..... helps us to identify defective and ambiguous items, to determine the difficulty level of the test and to determine the discriminating power of the items.	Try out
.....is an inseparable part of teaching, influencing the decisions teachers make, the actions they take, and the suggestions they offer for student learning.	Assessment
There are three popular terms in the literature that are sometimes misunderstood and used interchangeably. They are testing,, and evaluation.	assessment
..... or more widely known as formative assessment is used with the intention to form	Assessment for Learning

knowledge, skills and competencies to develop further and reach expected results.	
..... – 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 assessment
One of the ways to involve students in assessment is using	peer-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	Continuous Assessment
Teacher feedback plays a significant role in Assessment for Learning and has been identified in the literature as	feedforward
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.
The most useful feature of allows a teacher to observe changes happening during the process of assessment because of his/her influence on student performance.	dynamic assessment
..... – also known as summative assessment, is usually used at the end of a course of study.	Assessment of Learning
It is undeniable that any kind of has a certain impact.	assessment

It is undeniable that any kind of assessment has a certain	impact.
McNamara (2000) differentiates 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”	Washback
..... places students at an appropriate level of instructions within a program	Placement test
..... identifies students’ strengths and weaknesses	Diagnostic test
..... provides information about mastery or difficulty that learners have with course materials	Progress test
..... provides information about students’ attainment of learning outcomes at the end of a course	Achievement test
..... measures a general ability to learn a foreign language	Aptitude test

..... measures global competence in a language	Proficiency test
The best way to find out whether the created assessment tool works well is	piloting
Without understanding validity,, practicality as well as impact (washback), it is difficult to come up with an effective language assessment.	reliability
Without understanding, reliability, practicality as well as impact (washback), it is difficult to come up with an effective language assessment.	validity
Without understanding validity, reliability, as well as impact (washback), it is difficult to come up with an effective language assessment.	Practicality
The same test can be the best and the worst instrument depending on where and when it is used.	Assessment
..... are all receptive, or	Multiple-choice items

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.	
An Objectively-scored item is a question where there is only one fixed correct answer. It is also known as test.	closed-answer
.....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 assessment
...are descriptive statements that provide learners and instructors with information about the qualities, characteristics, and aspects of a given learning task.	Assessment criteria
..... can be positive and negative	Washback
..... competence was defined as involving knowledge sociocultural rules of language and of discourse.	Sociolinguistic

The roots of the standardizing testing back to ancient	China
In ancient Greece, Socrates had used a system where he tested his students through	dialogue.
In ancient Greece, had used a system where he tested his students through dialogue.	Socrates
... is a broader process involving various ways of collecting data including the use of tests	Assessment

Yopiq test savollari

Savol	Javob
Discourse competence	It was defined as the ability to connect sentences in stretches of discourse and to form a meaningful whole out of a series of utterances.
Sociolinguistic competence	It was defined as involving knowledge of the sociocultural rules of language and of discourse.
Strategic competence	It was seen to refer to “the verbal and nonverbal communication strategies that may be called into action to compensate for breakdowns in communication due to performance variables or due to insufficient competence
Grammatical competence	It was seen to encompass “knowledge of lexical items and of rules of morphology, syntax, sentence-grammar semantics, and phonology”
Practicality	This principle refers to the time and cost constraints during the construction and administration of an assessment instrument
Validity	This principle refers to the tests that measure what it claims to measure
Authenticity	The assessment instrument is authentic when it is contextualized, contains natural language and meaningful, relevant, and interesting topic, and replicates real world experiences

Washback	This principle refers to the consequence of an assessment on teaching and learning within classrooms
Planning	One of the first important step in the test construction
Try out	It helps us to identify defective and ambiguous items, to determine the difficulty level of the test and to determine the discriminating power of the items.
Assessment	It is an inseparable part of teaching, influencing the decisions teachers make, the actions they take, and the suggestions they offer for student learning.
Diagnostic assessment	It 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
Continuous Assessment	It involves a procedure of collecting evidence about students' knowledge and achievements throughout the course which then results in one final score at the end
Data-driven learning	A relatively new concept for the Uzbek EFL context is teaching and learning based on the data collected about students.
Dynamic assessment	The most useful feature of it allows a teacher to observe changes happening during the process of assessment because of his/her influence on student performance.
Assessment of Learning	It is also known as summative assessment, is usually used at the end of a course of study.
Placement test	It places students at an appropriate level of instructions within a program
Diagnostic test	It identifies students' strengths and weaknesses
Progress test	It provides information about mastery or difficulty that learners have with course materials
Achievement test	It provides information about students' attainment of learning outcomes at the end of a

	course
Aptitude test	It measures a general ability to learn a foreign language
Proficiency test	It measures global competence in a language
Multiple-choice items	They 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.
Assessment criteria	They are descriptive statements that provide learners and instructors with information about the qualities, characteristics, and aspects of a given learning task.

VII.Bitiruv ishi mavzulari

1. Tests - an integral part of a student's ongoing classroom development.
2. **Principles of assessment** in teaching foreign language.
3. **Types of assessment and their benefits**
4. **Designing tests for intermediate level students**
5. The role of tests in the development of four skills
6. The importance of assessment in teaching English as a foreign language
7. Types of assessment in Uzbekistan institutions.
8. Types of assessment in education and their main features.
9. Importance of assessments in teaching speaking
- 10.Importance of assessments in teaching writing
11. Importance of assessments in teaching listening
12. Importance of assessments in teaching reading
13. Effective methods of assessments in teaching EFL
- 14.CEFR in Uzbekistan education
- 15.Integrating four skills in assessment students' knowledge.

VII. Glossariy

Termin	O‘zbek tilidagi sharhi	Ingliz tilidagi sharhi
aptitude test	ishda muvaffaqiyatni bashorat qilish qobiliyatini yoki umumiy qobiliyatini (masalan, chet tili kursiga borishdan oldin) o‘lchash uchun mo‘ljallangan test...	a test designed to measure capacity or general ability a priori (e.g., before taking a foreign language course) to predict success in that undertaking...
Assessment	tizimli va mazmunli asoslangan protseduralarga muvofiq ma’lum bir qiziqish obyekti to‘g‘risida ma’lumot to‘plashning doimiy jarayoni	an ongoing process of collecting information about a given object of interest according to procedures that are systematic and substantively grounded
authenticity	berilgan til sinovi topshirig‘ining xususiyatlarining maqsadli til vazifasi xususiyatlariga muvofiqligi darajasi	the degree of correspondence of the characteristics of a given language test task to the features of a target language task
autonomy	tashqi maqsadlarni rag‘batlantirmasdan o‘z maqsadlarini belgilash va muvaffaqiyatni mustaqil ravishda nazorat qilish qobiliyati	the ability to set one’s own goals and independently monitor success without the presence of an external stimulus
biased for best	talabanning tilning "eng kichik" elementlarini (masalan, xarflar, pyecalar, so‘zlar) sinab ko‘rishda maqbul ishlashi uchun shart-sharoitlarni ta’minlash va keyinchalik ularni tobora kattaroq elementlarga birlashtirish	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	so‘zlar o‘chirilgan matn va test ishtirokchisi bo‘sh joyga mos keladigan so‘zni taqdim etishi	a text in which words are deleted and the test-taker must provide a word that fits the

	kerak	blank space
appropriate-word scoring	asl matndagi bo'sh joyga mos keladigan, grammatik va ritorik jixatdan maqbul so'zni qabul qiladigan skorklama usuli	a scoring method that accepts a suitable, grammatically and rhetorically acceptable word that fits the blank space in the original text
communicative test	test sinovida qatnashuvchining mazmunli va 'aqiqiy tildan foydalanish qobiliyatini yuzaga keltiradigan test	a test that elicits a test-taker's ability to use language that is meaningful and authentic
competence	tilni ijro etish qobiliyatiga asoslangan faraz qilingan (empirik ravishda kuzatib bo'lmaydigan) qobiliyat	one's hypothesized (empirically unobservable) underlying ability to perform language
computer-adaptive test (CAT)	test sinovlari talablariga javob beradigan va umuman ularning ishlash darajasiga mos keladigan savollar to'plamini oladigan asboblar	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)	tilni o'rganish va o'qitishda kompyuter texnologiyalarini qo'llash	the application of computer technology to language learning and teaching
C-test	boshqa xar qanday so'zning ikkinchi yarmi olib tashlangan va test topshiruvchi to'liq so'zni taqdim etishi kerak bo'lgan matn	a text in which the second half of every other word is eliminated and the testtaker must provide the whole word
diagnostic test	tilning aniq jixatlarini aniqlash uchun mo'ljallangan test	a test that is designed to diagnose specified aspects of a language
dialogue journal	o'quvchining o'qituvchi o'qigan va unga javob bergan fikrlari, xissiyotlari va reaksiyalarini yozadigan o'z-o'zidan yozish mashqlari	a self-writing exercise in which a student records thoughts, feelings, and reactions which a teacher reads and responds to

dichotomous scoring	faqat bitta to'g'ri javob berishga imkon beradigan, test oluvchining javobi to'g'ri yoki noto'g'ri bo'lishi mumkin bo'lgan ball olish usuli	a method of scoring that allows only one correct response so that a test-taker's response is either right or wrong
dictation	test topshiruvchilar matni tinglashi va eshitganlarini yozib olishlari orqali baholash usuli	a method of assessment in which test-takers listen to a text and write down what they hear
dicto-comp	imtixon topshiruvchilar nisbatan uzunroq matni (masalan, bir nechta jummalardan iborat abzatsni) tinglashi va tarkibini, ba'zi bir iboralarni va yoki asosiy leksik elementlarni ichki xolatga keltirishga urinib ko'rgan matni qayta tiklash uchun ishlatadigan diktantning bir varianti.	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	test topshiruvchi aniq maqsad vazifasini bajaradigan baxolash usuli; bilvosita sinovdan farqli o'laroq	an assessment method in which the test-taker actually performs the target task; as opposed to indirect testing
discrete point test	tilni tarkibiy qismlarga ajratish va ushbu qismlarni muvaffaqiyatli sinovdan o'tkazish mumkin degan taxmin asosida ishlab chiqilgan baxolashlar	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	grammatika, so'z boyligi yoki gap tuzishda o'z xoxish-irodasini ko'rsatish uchun, odatda, so'rovga javoban ishlab chiqarilgan yozuv; haqiqiy yozuvdan farqli o'laroq	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	testlar, boshqa baholashlar va yoki o'qituvchilarning xisobotlari natijalariga ko'ra qarorlar va yoki baxoli qarorlarni qabul qilish	making decisions and/or value judgments based on the results of tests, other assessments, and/or teachers' reports
focus on form	tilning tashkiliy tuzilishiga (grammatika, fonologiya, so'z boyligi va boshqalar) e'tibor berish	attention to the organizational structure (grammar, phonology, vocabulary, etc.) of a language
formal assessment	o'qituvchi va talabaga o'quvchilarning yutuqlarini baxolash uchun tuzilgan tizimli, rejalashtirilgan mashqlar yoki protseduralar	systematic, planned exercises or procedures constructed to give teacher and student an appraisal of student achievement
formative assessment	ularning o'sish jarayonini davom ettirishga yordam berish maqsadida ularning malakalari va ko'nikmalarini "shakllantirish" jarayonida talabalarni ba'olash	evaluating students in the process of "forming" their competencies and skills with the goal of helping them continue that growth process
form -focused assessment	tilning tashkiliy tarkibiy qismlariga (masalan, grammatika, so'z boyligi) yo'naltirilgan baxo	assessment that focuses on the organizational components (e.g., grammar, vocabulary) of a language
high-frequency word	yozma va og'zaki matnlarda tez-tez uchraydigan va tilning muhim foydalanuvchilari egallagan so'z boyligi bilimining bir qismi bo'lgan so'z; past chastotali so'zdan farqli o'laroq	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	test topshiruvchilar (masalan, kursga maktabga kirish) to'g'risida muhim qarorlar qabul qilinadigan ma'lumot beruvchi vosita.	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	test oluvchining til ishlab	an approach that uses a single

	chiqarishiga global baxo berish uchun yagona umumiy o'lhovdan foydalanadigan yondashuv; analitik ballardan farqli o'laroq	general scale to give a global rating for a test-taker's language production; as opposed to analytic scoring
idiom	ma'nosini so'zma-so'z tarifi bilan aniqlab bo'lmaydigan, lekin metafora ma'nosi umumiy foydalanish orqali ma'lum bo'lgan nutq figurasi	figure of speech whose meaning cannot be determined by the literal definition but whose metaphorical meaning is known through common use
impact	testdan foydalanishning individual test topshiruvchilar, muassasalar va jamiyatga ta'siri	the effect of the use of a test on individual test-takers, institutions, and society
indirect testing	test topshiruvchidan maqsadli vazifani bajarishi talab qilinmaydigan ba'olash usuli; aksincha, maqsadga muvofiq bo'lmagan vazifalarni bajarish natijasida xulosa qilinadi; to'ridan-to'g'ri sinovdan farqli o'laroq	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	tasodifiy, rejadan tashqari shar'lar va javoblar, shuningdek, o'quvchiga murabbiylik va boshqa tezkor javoblar	incidental, unplanned comments and responses, along with coaching and other impromptu feedback to the student
information transfer	bir malakadan qayta ishlangan maolomot (masalan, telefon xabarini tinglash) boshqa ko'nikmalarni bajarish uchun foydalaniladigan jarayon (masalan, telefon qo'niroini qaytarish uchun ism raqamni yozish)	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	shaklda aniqlangan va ijtimoiy taosir o'tkazish uchun ishlatiladigan uzoqroq	a longer utterance that is fixed in form and used for social interaction (e.g., "how do you

	soʻzlashuv (masalan, "qanday qilasiz?")	do?")
integrative test	til kompetensiyasini grammatika, soʻz boyligi, oʻqish, yozish, gapirish va tinglashning oʻzaro taosir qobiliyatlarining yagona toʻplami sifatida koʻrib chiqadigan test	a test that treats language competence as a unified set of interacting abilities of grammar, vocabulary, reading, writing, speaking, and listening
interactive (skills)	tildan foydalanishda bir nechta koʻnikmalardan foydalanishni (oʻqish, yozish, gapirish va tinglash) birlashtirish	combining the use of more than one skill (reading, writing, speaking, and listening) in using language
low-frequency word	yozma yoki ogʻzaki matnlarda kamdan-kam uchraydigan soʻz; yuqori chastotali soʻzdan farqli oʻlaroq	a word that seldom or rarely appears in written or spoken texts; as opposed to high-frequency word
macroskill	jumla darajasidan tashqaridagi til kompetensiyasini oʻz ichiga olgan lisoniy kompetensiyalar (nutq, pragmatik, ritorik vositalar); mikroskilldan farqli oʻlaroq	linguistic competencies that involve language competence beyond the sentence level (discourse, pragmatics, rhetorical devices); as opposed to microskill
measurement	aniq protseduralar yoki qoidalarga muvofiq sinovdan oʻtgan shaxsning ish faoliyatini miqdoriy aniqlash jarayoni	a process of quantifying a test-taker's performance according to explicit procedures or rules
mechanical task	test topshiruvchisi nimani ishlab chiqarishini oldindan belgilaydigan vazifa (masalan, ovoz chiqarib oʻqish yoki gapni takrorlash)	a task that determines in advance what the test-taker will produce (e.g., reading aloud or sentence repetition)
multiple-choice test	predmetlar ikki yoki undan ortiq roʻyxatlangan variantlar orasida tanlov beruvchiga tanlov taklif qiladigan baʼolash	an assessment instrument in which items offer the testtaker a choice among two or more listed options

	vositasi	
multiple intelligences	intellektning anoanaviy IQ asosidagi tushunchalardan tashqariga chiqadigan turlari, masalan, fazoviy, musiqiy, kinestetik, tabiatshunoslik, shaxslararo va shaxslararo intellekt	types of intelligence that extend beyond traditional IQbased concepts, such as spatial, musical, kinesthetic, naturalist, interpersonal, and intrapersonal intelligence
narrative evaluation	o'quvchining faoliyati to'risidagi individual yozma mulo'azalar shakli, baazida harflar baxosiga alternativa yoki qo'shimcha sifatida ishlatiladi	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	har bir test topshiruvchining ballari o'rtacha (o'rtacha ball), median (o'rtacha ball), standart o'ish (ballarning farqlanish darajasi) va yoki foizli darajaga qarab sharxlanadigan 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	oldindan aniqlangan javoblarga ega bo'lgan testlar	tests that have predetermined fixed responses
options	test topshiruvchisi tili o'quv dasturi yoki maktabda tanlashi mumkin bo'lgan turli xil javoblar	different responses from which a test-taker can choose in an item language curriculum or school
practicality	testni ishlab chiqishva boshqarish uchun mavjud bo'lgan resurslar va vaqt qay darajada boshqarilishi mumkinligi	the extent to which resources and time available to design, develop, and administer a test are manageable and feasible
prefabricated language	yangi jumalarni yaratish uchun modellarni taqdim etadigan tayyor jumla va butun jumlar yoki yodlangan jumlar	ready-made sentence stems and whole sentences or memorized chunks of sentences that provide models for the creation of new sentences

primary-trait scoring	yozuv testida, matnning asosiy maqsadiga erishishda samaradorligini ko'rsatadigan yagona ball	in a writing test, a single score indicating the effectiveness of the text in achieving its primary goal
process	tilni tushunish yoki ishlab chiqarish uchun foydalaniladigan protseduralarga (qadamlar, strategiyalar, vositalar, qobiliyatlar) qatnashish; ma'sulotdan farqli o'laroq	attending to the procedures (steps, strategies, tools, abilities) used to comprehend or produce language; as opposed to product
product	lingvistik xarakterning yakuniy natijalariga (masalan, yozma ravishda, "yakuniy" qog'ozga, qog'ozni tuzishda ishtirok etadigan turli bosqichlarga nisbatan) qatnashish; jarayondan farqli o'laroq	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	til bo'yicha biron bir kurs, o'quv dasturi yoki bitta ko'nikma bilan cheklanmagan test; aksincha, bu umumiy global qobiliyatni sinovdan o'tkazadi	a test that is not limited to any one course, curriculum, or single skill in the language; rather, it tests overall global ability
rubrics	talaba reyting shkalasida ma'lum bir nuqtada nimalarni amalga oshirishi mumkinligini tavsiflovchi bayonotlar; ba'zan tasma tavsiflovchi deb 'am ataladi	statements that describe what a student can perform at a particular point on a rating scale; sometimes also called band descriptors
subjective tests	oldindan aniqlangan yoki mutlaqo to'g'ri javoblarning yo'qligi to'g'ri va noto'g'ri javoblarni aniqlash uchun o'qituvchining xukmini talab qiladigan testlar	tests in which the absence of predetermined or absolutely correct responses require the judgment of the teacher to determine correct and incorrect answers
subtechnical	bir qator registrlar yoki predmet soxalarida	a word that occurs across a range of registers or subject

word	uchraydigan soʻz	areas
summative test	talaba nimani anglaganini oʻlchash yoki umumlashtirishga qaratilgan va odatda kurs yoki oʻqitish birligi oxirida sodir boʻladigan 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	test topshiruvchisi javoblar uchun tanlashi mumkin boʻlgan variantlar	options a test-taker can choose from for responses
task	xaqiqiy maqsad, muammo yoki talabni amalga oshiradigan lingvistik xarakterli majmui yoki topshiriq	a set or subset of linguistic actions that accomplish a real-world purpose, problem, or demand
task-based assessment	oʻquvchilarni oʻlchashni rejalashtirgan xatti-xarakterli amalga oshirishda ishtirok etadigan baxolashlar	assessments that involve learners in actually performing the behavior that one purports to measure
test	insonning maʼlum bir soxadagi qobiliyatini, bilimini yoki ishlashini oʻlchash usuli yoki tartibi	a method or procedure for measuring a person's ability, knowledge, or performance in a given domain
validity	baxolash natijalaridan kelib chiqadigan xulosalar darajasi, maqsadga muvofiq ravishda mazmunli va foydalidir	the extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the assessment
concurrent validity	test natijalarini testning oʻzidan tashqarida boʻlgan boshqa nisbatan soʻnggi koʻrsatkichlar bilan qoʻllab-quvvatlash darajasi	the extent to which results of a test are supported by other relatively recent performance beyond the test itself
consequential validity	testning taʼsiri, shu jumladan uning moʻljallangan mezonlarni oʻlchashdagi aniqligi, test sinovlarini	a test's impact, including such considerations as its accuracy in measuring intended criteria, its effect on the preparation of

	topshiruvchilarni tayyorlashga ta'siri va testni talqin qilish va ulardan foydalanishning (mo'ljallangan va kutilmagan) ijtimoiy oqibatlarini kabi fikrlar.	test-takers, and the (intended and unintended) social consequences of a test's interpretation and use
construct validity	inoikos olamidagi kuzatilgan xodisalarni tushuntirishga urinadigan xar qanday nazariya, gipoteza yoki model	any theory, hypothesis, or model that attempts to explain observed phenomena in one's universe of perceptions
content-related validity	test natijalari, natijada xulosalar chiqarilishi kerak bo'lgan mavzuni qanchalik aniq namunalar	the extent to which a test actually samples the subject matter about which conclusions are to be drawn
criterion-related validity	testning lingvistik mezonlari (masalan, belgilangan sinf maqsadlari)ning o'lganadigan darajasi va oldindan belgilangan darajadagi natijalarga erishilganligi	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	Sinov ishtirokchisi baxoni adolatli, dolzarb va o'rganishni takomillashtirish uchun foydali deb baxolash darajasi	The extent to which a test-taker views the assessment as fair, relevant, and useful for improving learning
predictive validity	kelajakda ishlashni baxolash uchun test natijalaridan qay darajada foydalaniladi	the extent to which results of a test are used to gauge future performance
washback	ba'olashning sinfda o'qitish va o'rganishga ta'siri	the effect of assessments on classroom teaching and learning

IX. ADABIYOTLAR RO‘YXATI

1. Beresova, J. (2017). The impact of the CEFR on teaching and testing English in the local context. *Theory and Practice in Language Studies*, 7(11), 959-964.
2. Brown, H. D., & Abeywickrama, P. (2014). *Language assessment. Principles and classroom practices*. White Plains, NY: Pearson Education.
3. Coupland, N., & Jaworski, A. (2019). *The New Sociolinguistics Reader*. Palgrave: Macmillan.
4. Chomsky, N. (2014). *Aspects of the theory of Syntax*. Cambridge, MA: M.I.T. Press.
5. Croft, W. (2011). *Radical construction grammar*. Cambridge: Cambridge University Press.
6. Celce-Murcia, M. & Olshtain, E. (2015). *Discourse and context in language teaching: A guide for language teachers*. Cambridge: Cambridge University Press.
7. Cumming, A. and R. Berwick (Eds.). (2010) *Validation in Language Testing*. Clevedon, England: Multilingual Matters, Ltd.
8. David L. Chiesa (Ed), U. Azizov, S. Xan, K. Nazamutdinova, K. Tangirova. *Re-conceptualizing Language Teaching: An-In-service Teacher Education course in Uzbekistan*. Baktaria Press 2019. ISBN 978-9943-5809-9-2. 205.
9. Diane Larsen-Freeman, Marti Anderson. *Techniques & Principles in Language Teaching* (3rd edition). Oxford University Press. 2011.
10. Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment: Volume 12 of handbooks on applied linguistics* (pp. 413-428). Berlin, Germany: Walter de Gruyter GmbH & Co KG. doi: <https://doi.org/10.1515/9781614513827-027>
11. Henning, G. (2014). *A guide to language testing: Development, evaluation, research*. Newbury House Publishers.

12. Hughes, A. (2013). *Testing for Language Teachers*. 2nd ed. Cambridge: Cambridge University Press.
13. Langacker, R. W. (2011). *Foundations of Cognitive Grammar (Vol. 2): Descriptive Application*. Stanford: Stanford University Press
14. Gardner H. (2013) *Multiple Intelligences: The Theory and Practice*, New York: Basic Books.
15. Matthews, P. H. (2014). *Syntax*. Cambridge: Cambridge University Press.
16. McMilan, J.H. (2014). *Fundamental Assessment Principles for Teachers and School Administrators. Practical Assessment, Research and Evaluation*. Retrieved from <https://pareonline.net/getvn.asp?v=7&n=8> [Accessed July 17, 2018]
17. Madsen H.S. (2014). *Techniques in Testing*. England: Oxford University Press.
18. McNamara, T. (2014). *Language Testing*. UK. Oxford: Oxford University Press.
19. Knight, P., & Yorke, M. (2013). *Assessment, learning and employability*. Maidenhead: Society for Research into Higher Education & Open University Press.
20. Jack C. Richards, Willy A. Renandya. *Methodology in Language Teaching: An anthology of current practice*. Cambridge university press. 2012.
21. Isaacs, T., Zara, C., Herbert, G., Coombs, S. J., & Smith, C. (2013). *Assessment of learning*. In *The SAGE key Concepts Series: Key concepts in educational assessment* (pp. 12-17). London: SAGE Publications Ltd doi: 10.4135/9781473915077.n4
22. Richards, C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press. Stanley Kubrick. *A narrative and stylistic analysis*. AN imprint of Greenwood Publishing Group, Inc. Великобритания (United Kingdom), 2011.
23. Stephen M. Alessi, Stanley R. Trollip. *Multimedia for Learning: Methods and Development (3rd Edition)*. Pearson; 3 Edition (November 19, 2000), 580 pages, 2011.
24. Shohamy, E. (2011). *The power of tests: A critical perspective on the uses of language tests*. Pearson Education Limited.

25. Taylor, L. (2019). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36.
26. Weir, C. (2015). *Language testing and validation: An evidence-based approach*. Palgrave
27. Wardhaugh, R. (2016). *An Introduction to Sociolinguistics*. Sussex, U.K.: Wiley-Blackwell.

IV. Internet saytlar

28. <http://edu.uz> – O‘zbekiston Respublikasi Oliy va o‘rta maxsus ta’lim vazirligi
29. <http://lex.uz> – O‘zbekiston Respublikasi Qonun hujjatlari ma’lumotlari milliy bazasi
30. <http://bimm.uz> – Oliy ta’lim tizimi pedagog va rahbar kadrlarini qayta tayyorlash va ularning malakasini oshirishni tashkil etish bosh ilmiy-metodik markazi
31. <http://ziyonet.uz> – Ta’lim portali Ziyonet
32. www.Britishcouncil.org -Britaniya kengashi veb sayti
33. www.ziyonet.uz -Ta’lim portali
34. www.tefl.net- Til o‘rganish veb sayti
35. busyteacher.org -Til o‘rganish veb sayti