

## TIL KO'NIKMALARINI BAHOLASH: KLASSIK TEST NAZARIYASI, RASH VA KO'P QIRRALI RASH MODELLARI

**A.B. Normurodov, M.Dj. Ermamatov, A.R. Sattiyev, Q.A. Amonov**

*Bilim va malakalarni baholash agentligi huzuridagi Ilmiy-o'quv amaliy markazi, 100084, Toshkent sh., Bog'ishamol k., 12, [a.normurodov@uzbmb.uz](mailto:a.normurodov@uzbmb.uz)*

**Qisqacha mazmuni.** Ushbu maqolada klassik test nazariyasi, Rash va ko'p qirrali Rash modellari asosida til ko'nikmalarini baholash natijalarini tadqiq qilish imkoniyati ko'rib chiqildi. Tadqiqot uchun o'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash bo'yicha o'tkazilgan aprobatsiya test sinovi natijalari olindi. Dastlab test natijalari an'anaviy klassik test nazariyasi bilan test topshiriqlarining tavsif statistikasi, qiyinlik darajalari va umumiy ball bilan alohida test topshiriqlarining korelyatsiyasi o'rganildi. Test topshiriqlarining 3 ta shartli qiyinlik darajalari bo'yicha taqsimotining normal taqsimotga nisbatan farqlari, umumiy ball bilan korelyatsiya qiymatlarining test sifatiga ta'siri ko'rsatib berildi. Rash modeli asosida hisoblashlar bilan qiyinlik darajalari, Rayt xaritgasi, Rash modeli bilan moslik va test xarakteristikasi chiziqlari o'rganildi. Yuqori va quyi qobiliyat darajalari uchun test topshiriqlarining kam ekanligi, variantning qiyinlik darajasi va test topshiriqlariga berilgan javoblarning qobiliyat bilan bog'liqligi qanchalik Rash modeli bilan qanchalik mos kelishi aniqlandi. Yozish va gapirish ko'nikmalari ko'p qirrali Rash modeli bilan tadqiq qilindi. Yozish ko'nikmasi bo'yicha tadqiqot natijalariga ko'ra, element (savol)ning farqlanishi 1-esseda o'rtacha, 2-esseda esa kuchliroq, baholovchining qat'iylik qiymati farqi 1-esseda juda kichik, 2-esseda esa kattaroq ekanligi ko'rsatib berildi. Gapirish ko'nikmasi bo'yicha tahlil natijalariga ko'ra, elementlarning qiyinlik darajasi barqaror va muammosiz ekanligi, reyting shkalasida qadamlar (4-5-qadamlar) tartibsizligining mavjudligi, baholovchilarning umumiy qat'iyliги yaxshi baholanganligi va baholovchilar tomonidan ayrim qadamlar keskin farqli qo'llanilayotganligi aniqlandi.

**Kalit so'zlar:** qiyinlik darajasi, Rayt xaritasi, test ma'lumoti chizig'i, test xarakteristikasi chizig'i, klassik test nazariyasi, Rash modeli, ko'p qirrali Rash modeli.

### 1. Kirish

Baholash va pedagogik o'lchovlarning nazariy asoslariga ko'ra har bir test bo'yicha tuzilgan test topshiriqlari mazmuni ekspert tekshiruvidan o'tkazilgandan so'ng, keyingi bosqichda test sinovlari o'tkaziladi va

uning natijalari asosida testlarning xususiyatlari sifatini tashxislash uchun ularning statistik xarakteristikalari aniqlanadi. Statistik xarakteristikalar yordamida test topshiriqlariga qo'yilgan asosiy talablarning ko'rsatkichlari aniqlanadi.

O'zbekiston Respublikasi Prezidentining "2020–2030-yillarda o'zbek tilini rivojlantirish va til siyosatini takomillashtirish konsepsiyasini tasdiqlash to'g'risida" 2020-yil 20-oktabrdagi PF-6084-son Farmonida belgilangan vazifalardan biri o'zbek tilining xalqaro miqyosdagi o'rni va nufuzini oshirish, uni xorijiy mamlakatlarda keng targ'ib qilish hamda chet ellik fuqarolarga o'zbek tilini bilish darajasi bo'yicha malaka sertifikatini taqdim etish imkoniyatini yaratishdir. Shu maqsadda O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash (o'qish, tinglab tushunish, yozish va gapirish ko'nikmalari bo'yicha) uchun aprobatsiya test sinovi Qoraqalpog'iston Respublikasi, Toshkent shahri va viloyatlarda ro'yxatdan

o'tgan va Respublikaning turli hududlarida istiqomat qiluvchi hamda Oliy ta'lim tashkilotlarida ta'lim oluvchi xorijiy fuqarolar o'rtasida kompyuter orqali o'tkazildi.

O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash (o'qish va tinglab tushunish ko'nikmasi bo'yicha) uchun o'tkazilgan aprobatsiya test sinovida har bir variant 20 ta test topshirig'idan iborat bo'lib, mos ravishda ajratilgan vaqt 30 va 60 daqiqani tashkil qildi. Yozish ko'nikmasi bo'yicha talabgorlarga 2 ta esse yozish talabi qo'yildi va ajratilgan vaqt 60 daqiqani tashkil etdi. Gapirish ko'nikmasi bo'yicha 1 ta topshiriq berildi va ajratilgan vaqt 10 daqiqani tashkil etdi. Test sinovida jami 422 nafar talabgorlar ishtirok etdi.

Ushbu maqolada O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash bo'yicha o'tkazilgan aprobatsiya test sinovi natijalari klassik test nazariyasi, Rash va ko'p qirrali Rash modellari bilan tadqiq qilindi.

## 2. Natija va muhokamalar

### 2.1. Klassik test nazariyasi bo'yicha natijalar

Test topshiriqlariga qo'yiladigan asosiy talablar – topshiriqning qiyinligi, test ballarining dispersiyasi (o'zgaruvchanligi, farqlanishi), topshiriqning boshqa topshiriqlar bilan, shuningdek, umumiy ballar yig'indisi bilan korrelyatsiyasi (bog'liqligi)dan iborat. Topshiriqning qiyinlik dara-

jasini aniqlash usullaridan biri topshiriqni empirik sinovdan o'tkazib, to'g'ri javoblar salmog'ini aniqlash hisoblanadi. Test ballari (yoki to'g'ri javoblar)ning dispersiyasi test topshiruvchilarning tayyorgarlik darajasini aniqlashga, bilim darajalari bo'yicha ajratishga imkon beradi.

Shuningdek, test variantlari va test topshiriqlarining asosiy statistik xarakteristikalarini qatoriga o'rtacha qiymat, gistogrammani qurish, moda va mediana kabi ko'rsatkichlarni hisoblash hamda test ballarining umumiy dispersiyasi (standart og'ish) ham kiradi [1-6]. Test ballari (yoki to'g'ri javoblar)ning o'rtacha arifmetik qiymati fanlar, oliy ta'lim muassasalari va boshqa muhim belgilar kesimida aniqlanadi. Bu ko'rsatkich test ballari o'rtasidagi tafovutlarni umumlashtiradi, ularga xos bo'lgan yo'nalishni, qonuniyatni ochib beradi. Test sinovlari natijalari asosida aniqlangan test ballari taqsimoti gistogrammasi quriladi va uning normal taqsimotga yaqinligi baholanadi. Gistogrammaning normal taqsimotga yaqinligi testning sifatini, test sinovlarining ob'ektiv o'tkazilganligini bildiradi. Test ballarining eng ko'p takrorlanadigan qiymati statistikada moda, o'sish tartibida joylashtirilgan test ballari qatorining o'rtasida joylashgan qiymati esa mediana deyiladi. O'rtacha arifmetik qiymat, moda va mediana qiymatlari o'zaro teng bo'lganda test ballari taqsimoti simmetrik bo'ladi. Ushbu statistik ko'rsatkichlar biri-biridan qanchalik ko'p farq qilsa, ballar taqsimoti normal taqsimotdan shunchalik uzoq bo'ladi.

O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va ser-

tifikatlash (tinglab tushunish ko'nikmasi bo'yicha) uchun o'tkazilgan aprotatsiya test sinovi natijalarining tavsif statistikalarini 1-jadvalda keltirilgan.

1-jadvaldagi gistogrammadan o'rtacha qiymat 11,06 ga, moda va mediananing qiymatlari mos ravishda 17 va 12 ga teng ekanligi ko'rish mumkin. Asimmetriyaning absolyut qiymati -1 va 1 oralig'ida, ekstsessning qiymati esa -1 dan kichik. Ushbu ma'lumotlar taqsimotning normal taqsimotga nisbatan kenroq taqsimlanganligini bildiradi. Gistogrammaning 2 ta cho'qqiga ega ekanligi, ushbu aprotatsiya test sinovida qobiliyat darajalari bo'yicha 2 xil qobiliyat taqsimotiga ega guruh ishtirok etganligini ko'rsatadi.

O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash (o'qish ko'nikmasi bo'yicha) uchun o'tkazilgan aprotatsiya test sinovi natijalarining tavsif statistikalarini esa 2-jadvalda keltirilgan.

2-jadvaldagi gistogrammadan o'rtacha qiymat 12,5 ga, moda va mediananing qiymatlari mos ravishda 17 va 14 ga teng ekanligi ko'rish mumkin. Asimmetriya va ekstsessning absolyut qiymatlari esa -1 va 1 oralig'ida. Ushbu ma'lumotlar taqsimotning normal taqsimotga yaqinroq ekanligini bildiradi.

## 1-jadval

## Tinglab tushunish ko'nikmasi bo'yicha test sinovi natijalarining tavsif statistikalari

Test topshiriqlari soni	20	
Test topshiruvchilar soni	408	
Kronbax alfa koeffitsiyenti	0,87	
O'rta qiymat	11,16	
Moda	17	
Mediana	12	
Dispersiya	23,91	
Standart tafovut	4,89	
Asimmetriya	-0,283	
Ekstsess	-1,226	
O'lchashning standart xatoligi	1,76	

## 2-jadval

## O'qish ko'nikmasi bo'yicha test sinovi natijalarining tavsif statistikalari

Test topshiriqlari soni	20	
Test topshiruvchilar soni	423	
Kronbax alfa koeffitsiyenti	0,88	
O'rta qiymat	12,5	
Moda	17	
Mediana	14	
Dispersiya	23,11	
Standart tafovut	4,81	
Asimmetriya	-0,691	
Ekstsess	-0,792	
O'lchashning standart xatoligi	1,67	

Test natijalarining statistik tahliliga ko'ra, (1- va 2-jadvallar), O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash (tinglab tushunish va o'qish ko'nikmasi) uchun o'tkazilgan aprobatiya test sinovi natijalari bo'yicha test variantlarining ishonchlilik koeffitsiyenti (Kronbax alfa koeffitsiyenti) mos ravishda 0,87 va 0,88 ga teng ekanligi aniqlandi. Kronbax alfa koeffitsiyentining bu qiymati ushbu test variantlarining ichki muvofiqligi "yaxshi" ekanligini ko'rsatadi. Olib borilgan o'lchashning standart xatoligi tinglab tushunish ko'nikmasi uchun 1,76 ga va o'qib tushunish ko'nikmasi uchun 1,67 teng bo'ldi.

Test topshiriqlarining ichki muvofiqligi har bitta test topshirig'iga berilgan to'g'ri javoblarning umumiy

ball bilan korrelyatsiyasiga, talabgorlar olgan umumiy ballarning standart og'ishiga, har bitta test topshirig'iga berilgan javoblarning standart og'ishlari yig'indisiga hamda test topshiriqlari va test topshiruvchilar soniga bog'liq bo'ladi. Bundan tashqari test topshiriqlarining ichki muvofiqligi nafaqat test topshiriqlarining sifatiga, balki talabgorlarning tayyorgarlik darajasining past yoki yuqoriligiga ham bog'liqdir.

Quyida tinglab tushunish bo'yicha test topshiriqlari natijalari asosida test topshiriqlarining - ID raqami, to'g'ri javob berganlar foizi - Ans (foizda) va test natijalari bo'yicha aniqlangan qiyinlik darajalari V (1-, 2- va 3-qiyinlik darajalari) jadval shaklida keltirildi (3-jadval).

**3-jadval**

Tinglab tushunish ko'nikmasi bo'yicha test topshiriqlarining aniqlangan qiyinlik darajalari

No	ID	Ans (foizda)	V
1	q78	81,62	1
2	q74	80,39	1
3	q76	73,77	2
4	q87	71,57	2
5	q88	71,32	2
6	q82	70,83	2
7	q75	70,10	2
8	q73	68,38	2
9	q71	64,46	2

10	q72	62,50	2
11	q79	60,54	2
12	q80	60,05	2
13	q89	53,19	2
14	q83	49,02	2
15	q85	49,02	2
16	q86	44,36	2
17	q84	37,50	2
18	q81	23,77	3
19	q91	22,79	3
20	q90	0,74	3

O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash (tinglab tushunish ko'nikmasi bo'yicha) uchun o'tkazilgan aprobatsiya test sinovi natijalarining statistik tahliliga ko'ra (3-jadval), 20 ta test topshiriqlaridan 2 tasi (10,00 foiz) 1-qiyinlik darajasidagi test topshiriqlaridan, 15 tasi (75,00 foiz) 2-qiyinlik darajasidagi test topshiriqlaridan va 3 tasi (15,00 foiz) 3-qiyinlik darajasidagi test topshiriqlaridan iborat ekanligi aniqlandi. Bu esa ushbu variantda 1-qiyinlik darajasidagi test topshiriqlari soni normadan biroz kam, 2-qiyinlik darajasidagi test topshiriqlari soni esa normadan ko'p va 3-qiyinlik darajasidagi test topshiriqlari soni esa normada taqsimlanganligi bildiradi.

Variantdagi (3-jadval) 1-qiyinlik darajasidagi test topshiriqlari normadan biroz kam, 2-qiyinlik dara-

jasidagi test topshiriqlari soni esa normadan ko'p taqsimlanganligini hisobga olgan holda ushbu variantlardagi test topshiriqlari qiyinlik darajalari taqsimotini o'rganib chiqish kerak, chunki normal taqsimotni ta'minlash uchun, 1- va 3-qiyinlik darajali test topshiriqlarining soni test variantidagi test topshiriqlari sonining 16-25 (3-5 ta) foizi, 2-qiyinlik darajali test topshiriqlari soni esa 50-68 (10-14 ta) foiz bo'lishi talab qilinadi.

Har bir test topshirig'iga berilgan javoblarning umumiy test bali bilan korrelyatsiyasi (1, 2, 3, ... 20-test topshiriqlari va umumiy ball orasidagi korrelyatsiya) har bitta test topshirig'i bilan umumiy ball korrelyatsiya koeffitsiyentiga teng bo'ladi.

Umuman olganda, har bitta test topshirig'i bilan umumiy ball korrelyatsiya koeffitsiyenti qiymati

2-qiyinlik darajasidagi test topshiriqlari uchun 0,5 va undan katta bo'lsa, 1- va 3-qiyinlik darajasidagi test topshiriqlari uchun esa 0,25 va undan katta bo'lsa valid hisoblanadi. Har bitta test topshirig'i bilan umumiy ball korrelyatsiya koeffitsiyenti qiymati manfiy bo'lgan test topshiriqlari esa variantdan chiqariladi. Aks holda bilim darajalari past bo'lgan talabgorlar g'olib bo'lib, bilim darajalari yuqori

bo'lgan talabgorlar test topshiriqlarini yechishda noto'g'ri javobni tanlaydilar yoki ularni o'tkazib yuboradilar.

4-jadvalda o'qish ko'nikmasi bo'yicha test natijalari asosida test topshiriqlarining - ID raqami, to'g'ri javob berganlar foizi – Ans (foizda) va test natijalari bo'yicha aniqlangan qiyinlik darajalari V (1-, 2- va 3-qiyinlik darajalari) keltirilgan.

**4-jadval**

O'qib tushunish ko'nikmasi bo'yicha test topshiriqlarining aniqlangan qiyinlik darajalari

<b>No</b>	<b>ID</b>	<b>Ans (foizda)</b>	<b>V</b>
1	q72	82,74	1
2	q71	80,85	1
3	q82	77,07	1
4	q80	76,12	1
5	q85	76,12	1
6	q83	75,41	1
7	q75	75,18	1
8	q91	74,00	2
9	q79	71,63	2
10	q76	71,16	2
11	q84	68,79	2
12	q74	67,85	2
13	q86	65,96	2
14	q73	63,83	2
15	q78	60,05	2
16	q89	55,79	2
17	q88	42,08	2

18	q81	40,90	2
19	q87	12,77	3
20	q90	12,06	3

O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash (o'qib tushunish ko'nikmasi bo'yicha) uchun o'tkazilgan aprobatsiya test sinovi natijalarining statistik tahliliga ko'ra (4-jadval), 20 ta test topshiriqlaridan 7 tasi (35,00 foiz) 1-qiyinlik darajasidagi test topshiriqlaridan, 11 tasi (55,00 foiz) 2-qiyinlik darajasidagi test topshiriqlaridan va 2 tasi (10,00 foiz) 3-qiyinlik darajasidagi test topshiriqlaridan iborat ekanligi aniqlandi. Bu esa ushbu variantda 1-qiyinlik darajasidagi test topshiriqlari soni normadan biroz ko'p, 2-qiyinlik darajasidagi test topshiriqlari soni esa normada va 3-qiyinlik darajasidagi test topshiriqlari soni esa normadan biroz kam taqsimlanganligini bildiradi.

Variantdagi (4-jadval) 1-qiyinlik darajasidagi test topshiriqlari normadan biroz ko'p, 3-qiyinlik darajasidagi test topshiriqlari soni esa normadan biroz kam taqsimlanganligini hisobga olgan holda ushbu variantlardagi test topshiriqlari qiyinlik darajalari taqsimotini o'rganib chiqish kerak, chunki bu yerda ham normal taqsimotni ta'minlash uchun, 1- va 3-qiyinlik darajali test topshiriqlarining soni test variantidagi test topshiriqlari sonining 16-25 (3-5

ta) foizi, 2-qiyinlik darajali test topshiriqlari soni esa 50-68 (10-14 ta) foiz bo'lishi talab qilinadi.

5-jadvalda tinglab tushunish ko'nikmasi bo'yicha o'tkazilgan aprobatsiya test sinovi natijalari tahlili bo'yicha olingan test topshiriqlarining har bitta test topshirig'i bilan umumiy ball korrelyatsiya koeffitsiyenti qiymatlari keltirilgan.

O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash (tinglab tushunish ko'nikmasi bo'yicha) uchun o'tkazilgan aprobatsiya test sinovi natijalari bo'yicha har bitta test topshirig'i bilan umumiy ball korrelyatsiya koeffitsiyenti qiymatlari 0,25 dan kichik (5-jadvalda ajratib ko'rsatilgan ID raqamlari – q81 va q90) bo'lgan test topshiriqlarini o'rganish, muammo aniqlansa ularning kamchiliklarini tuzatish, aks holda bunday test topshiriqlarini bazaga qo'shmaslik maqsadga muvofiqdir.

6-jadvalda O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash (o'qish ko'nikmasi) uchun o'tkazilgan aprobatsiya test sinovi natijalari tahlili bo'yicha olingan test topshiriqlarining har bitta test topshirig'i bilan umumiy ball korrelyatsiya koeffitsiyenti qiymatlari keltirilgan.

Tinglab tushunish ko'nikmasi bo'yicha test topshiriqlarining har bitta test topshirig'i bilan umumiy ball korrelyatsiya koeffitsiyenti qiymatlari

No	ID	Har bitta test topshirig'i bilan umumiy ball korrelyatsiya koeffitsiyenti
1	q78	0,576
2	q74	0,527
3	q76	0,631
4	q87	0,721
5	q88	0,723
6	q82	0,614
7	q75	0,686
8	q73	0,610
9	q71	0,467
10	q72	0,600
11	q79	0,450
12	q80	0,571
13	q89	0,677
14	q83	0,486
15	q85	0,617
16	q86	0,489
17	q84	0,488
18	<b>q81</b>	0,168
19	q91	0,453
20	<b>q90</b>	0,085

O'qib tushunish ko'nikmasi uchun o'tkazilgan aprobatsiya test sinovi natijalari bo'yicha har bitta test topshirig'i bilan umumiy ball korrelyatsiya koeffitsiyenti qiymati manfiy bo'lgan ID raqamlari q87 va q90

bo'lgan test topshiriqlarini o'rganish, muammo aniqlansa, test topshiriqlarining kamchiliklarini tuzatish lozim, aks holda bunday test topshiriqlarini bazaga qo'shmaslik kerak.

## 6-jadval

O'qib tushunish ko'nikmasi bo'yicha test topshiriqlarining har bitta test topshirig'i bilan umumiy ball korrelyatsiya koeffitsiyenti qiymatlari

№	ID	Har bitta test topshirig'i bilan umumiy ball korrelyatsiya koeffitsiyenti
1	q72	0,622
2	q71	0,597
3	q82	0,722
4	q80	0,654
5	q85	0,693
6	q83	0,657
7	q75	0,489
8	q91	0,669
9	q79	0,591
10	q76	0,653
11	q84	0,659
12	q74	0,504
13	q86	0,653
14	q73	0,653
15	q78	0,548
16	q89	0,742
17	q88	0,374
18	q81	0,483
19	<b>q87</b>	<b>-0,193</b>
20	<b>q90</b>	<b>-0,148</b>

O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash (o'qish ko'nikmasi bo'yicha) uchun o'tkazilgan ap-robatsiya test sinovi natijalari bo'yicha ham har bitta test topshirig'i bilan umumiy ball korrelyatsiya koef-

fitsiyenti qiymatlari 0,5 dan kichik bo'lgan (2-qiyinlik darajasidagi test topshiriqlari uchun) test topshiriqlari va har bitta test topshirig'i bilan umumiy ball bilan korrelyatsiya koeffitsiyenti qiymati 0,25 dan kichik bo'lgan (1- va 3-qiyinlik darajasidagi

test topshiriqlari uchun) test topshiriqlari o'rganib chiqilishi va

qiyinlik darajalarini hisobga olgan holda kerakli o'zgarishlar qilish lozim.

## 2.2. Rash modeli bo'yicha natijalar

Zamonaviy test nazariyasida (IRT) parametrlar tajriba va modellarni moslashtirish orqali topiladi. Buni amalga oshirish uchun bir necha xil usullar ishlab chiqilgan. Zamonaviy test nazariyasi modellarida model parametrlarini aniqlash muhim rol o'ynaydi. Rash modeli bo'yicha bu parametrlar – yashirin qobiliyat va elementlar qiyinligidir [6-10].

Rash modeliga ko'ra, dixotomik elementlarga individual javoblar ehtimoli shaxsning qobiliyat darajasi va element qiyinligi bilan aniqlanadi. Buni quyidagi matematik formula orqali ifodalanadi:

$$P(X_{is} = 1 | \theta_s, b_i) = \frac{e^{\theta_s - b_i}}{1 + e^{\theta_s - b_i}}$$

bu yerda,  $X_{is} = 1$  s- o'quvchining i elementga to'g'ri javob berish ehtimolligi,  $\theta_s$ - qobiliyat o'zgaruvchisi,  $b_i$ - topshiriq qiyinlik darajasi, e- natural logarifm asosi ( $e = 2,7182818...$ ).

Rash modeli doirasida hisoblashlarda R dasturining dexter to'plamidan foydalanildi [11].

Rash modeli asosida aniqlangan qiyinlik darajalarini talabgorlar qobiliyatlariga qanchalik mosligini Rayt xaritasi yordamida tahlil qilish mumkin [12]. Rayt xaritasi – test topshiriqlarining qiyinlik darajalari va talabgorlarning qobiliyat darajalarini

o'zaro mos kelishini aniqlovchi diagrammadir. Bu test topshiriqlari qiyinlik darajasining talabgorlarning qobiliyat darajalariga qanchalik mos kelishini ko'rsatadi. Agar xarita elementning qiyinligi va talabgor qobiliyati o'rtasidagi yaxshi moslashuvni ko'rsatsa, u baholashning konstrukt validligini asosligini qo'llab-quvvatlaydi. Turli qobiliyat darajalarining qanchalik yaxshi ajralishini Rayt xaritasi bilan ko'rsatish mumkin. Agar test topshiriqlari qobiliyat darajalari bo'yicha yaxshi taqsimlangan bo'lsa, test ularning izchil o'lchovini ta'minlashi mumkin. Yaxshi tuzilgan Rayt xaritasi yuqori va quyi qobiliyat darajalarini samarali farqlay olishini ko'rsatadi, bu esa baholash natijalarining ishonchliligiga hissa qo'shadi.

Test topshiriqlarining qiyinlik darajasi va uning taqsimotini tasavvur qilish orqali baholashning mo'l-jallangan kontent sohalarini har tomonlama qamrab olishini ko'rish mumkin.

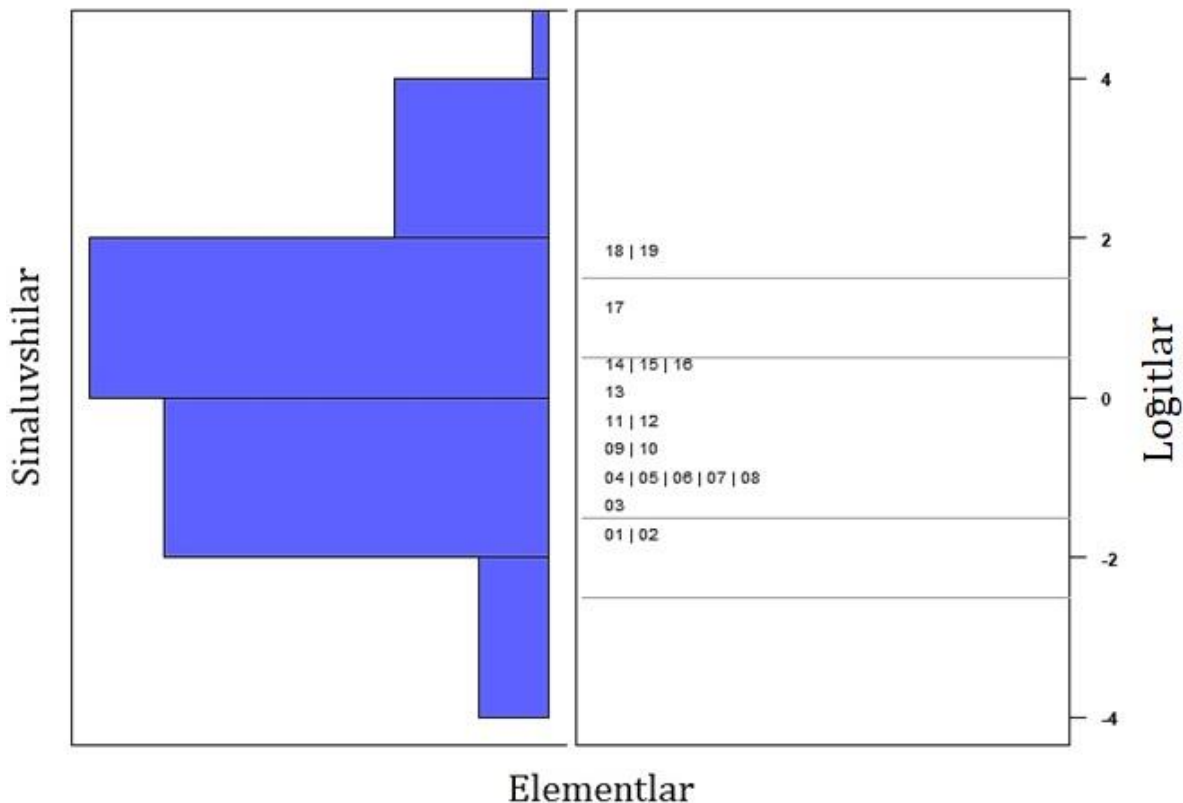
O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash bo'yicha (tinglab tushunish ko'nikmasi bo'yicha) o'tkazilgan aprobatsiya test sinovi natijasi asosida chizilgan Rayt xaritasi 1-rasmda keltirilgan.

Test natijalariga ko'ra qobiliyat darajalari –3,73 va 4,43 logit birligi

oralig'ida, test topshiriqlari qiyinlik darajalari  $-1,83$  va  $5,84$  oraliqda taqsimlangan.

Rasmdan test sinovlarining tinglab tushunish ko'nikmasi bo'yicha 20-test topshirig'ining (ID raqami - q90 bo'lgan test topshirig'i) qiyinlik darajasi ( $-3:3$ ) logit birligi oralig'idan

tashqarida ekanligi ko'rinadi. Test variantida qiyinlik darajasi yuqori bo'lgan test topshiriqlari ulushi kichik (ID raqami - q90-test topshirig'i bundan mustasno), bu esa test variantining qiyinlik darajasi pastroq ekanligini bildiradi.



1-rasm. Tinglab tushunish ko'nikmasi bo'yicha test natijalari asosida aniqlangan qobiliyat va qiyinlik darajalarining mosligi (Rayt xaritasi)

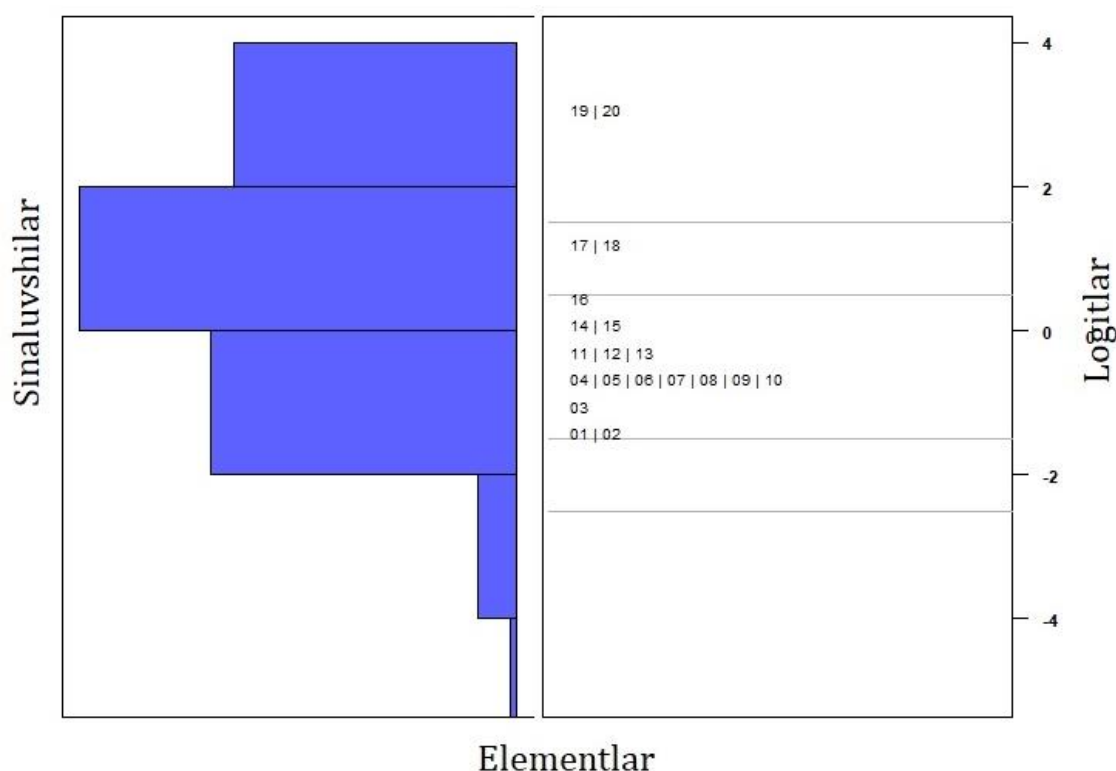
O'qib tushunish ko'nikmasi bo'yicha test sinovining natijasi asosida chizilgan Rayt xaritasi 2-rasmda keltirilgan.

Test natijalariga ko'ra qobiliyat darajalari  $-5,7$  va  $3,9$  logit birligi oralig'ida, test topshiriqlari qiyinlik darajalari  $-1,42$  va  $3,16$  oraliqda taqsimlangan. Rasmdan test si-

novlarining o'qib tushunish ko'nikmasi bo'yicha 17 va 20-test topshiriqlarining (ID raqamlari - q87 va q90 bo'lgan test topshiriqlari) qiyinlik darajalari ( $-3:3$ ) logit birligi oralig'idan tashqarida ekanligi ko'rinadi. Test variantida qiyinlik darajasi yuqori bo'lgan test topshiriqlari ulushi kichik (ID raqamlari - q87 va q90 test

topshiriqlari bundan mustasno), bu esa test variantining qiyinlik darajasi pastroq ekanligini bildiradi. Bunday hollarda yuqori qobiliyat darajalari uchun test topshiriqlari sonining taqsimotini to'g'rilash maqsadga muvofiq.

Barcha test topshiriqlariga to'g'ri va noto'g'ri javob bergan talabgorlar qobiliyatlarining xatoliklari juda katta bo'lishini inobatga olgan holda, bunday holatni oldini olish maqsadga muvofiq.



2-rasm. O'qib tushunish ko'nikmasi bo'yicha test natijalari asosida aniqlangan qobiliyat va qiyinlik darajalarining mosligi (Rayt xaritasi)

7-jadvalda O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash bo'yicha (tinglab tushunish ko'nikmasi) o'tkazilgan aprotatsiya test sinovida foydalanilgan test topshiriqlarining Rash modeli bilan aniqlangan qiyinlik darajalari keltirilgan.

Jadvaldan, Rash modeli bo'yicha q78 ID raqamli test topshirig'i eng oson, q90 ID raqamli test topshirig'i esa eng qiyin ekanligini ko'rish mumkin. Qolgan test topshiriqlarining qiyinlik darajalari shu oraliqda taqsimlangan.

Tinglab tushunish ko'nikmasi bo'yicha test topshiriqlarining Rash modeli bilan aniqlangan qiyinlik darajalari

No	ID	b
1	q78	-1,84
2	q74	-1,74
3	q76	-1,24
4	q87	-1,09
5	q88	-1,08
6	q82	-1,04
7	q75	-0,99
8	q73	-0,88
9	q71	-0,63
10	q72	-0,51
11	q79	-0,39
12	q80	-0,36
13	q89	0,05
14	q83	0,30
15	q85	0,30
16	q86	0,57
17	q84	0,97
18	q81	1,85
19	q91	1,92
20	q90	5,85

O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash bo'yicha (o'qib tushunish ko'nikmasi bo'yicha) o'tkazilgan aprobatsiya test sinovida foydalanilgan variantdagi test topshiriqlarining Rash modeli bilan aniqlangan qiyinlik darajalari 8-jadvalda keltirilgan.

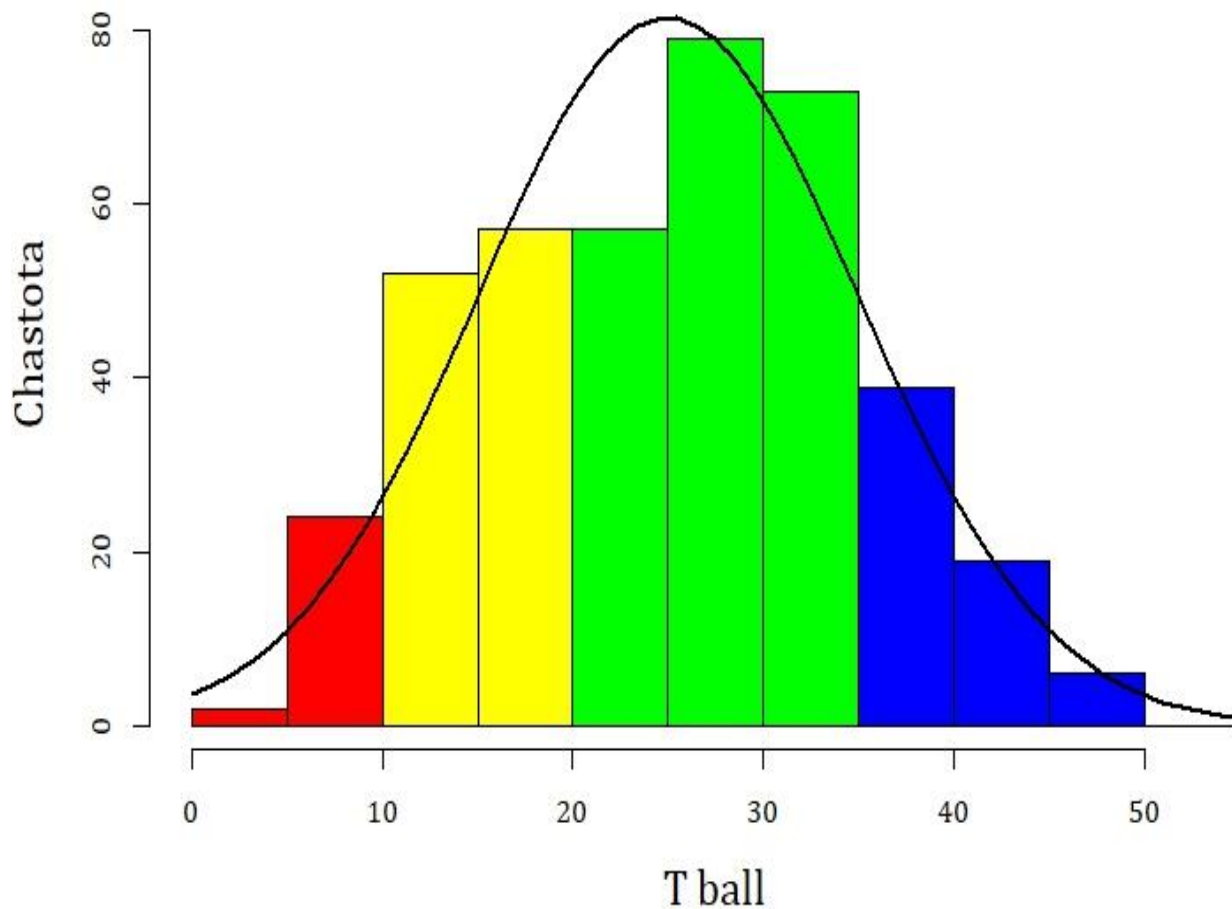
8-jadvaldagi ma'lumotlardan, Rash modeli bo'yicha q72ID raqamli test topshirig'i eng oson, q90 ID raqamli test topshirig'i esa eng qiyin ekanligini ko'rish mumkin. Qolgan test topshiriqlarining qiyinlik darajalari esa shu oraliqda taqsimlangan.

O'qib tushunish ko'nikmasi bo'yicha test topshiriqlarining Rash modeli bilan aniqlangan qiyinlik darajalari

<b>№</b>	<b>ID</b>	<b>b</b>
1	q72	-1,43
2	q71	-1,26
3	q82	-0,94
4	q80	-0,87
5	q85	-0,87
6	q83	-0,82
7	q75	-0,80
8	q91	-0,71
9	q79	-0,54
10	q76	-0,51
11	q84	-0,34
12	q74	-0,28
13	q86	-0,15
14	q73	-0,02
15	q78	0,22
16	q89	0,48
17	q88	1,25
18	q81	1,31
19	q87	3,10
20	q90	3,16

3-rasmda tinglab tushunish ko'nikmasi bo'yicha Rash modeli bilan aniqlangan haqiqiy ball (T-ball) taqsimoti gistogrammasi keltirilgan. Umumiy holda Rash modeli bilan aniqlangan haqiqiy ballar (T-ball)

taqsimoti gistogrammasi normal taqsimotga yaqin. Gistogrammalarning normal taqsimotga yaqinligi testning sifatini, test sinovining ob'ektiv o'tkazilganligini bildiradi.

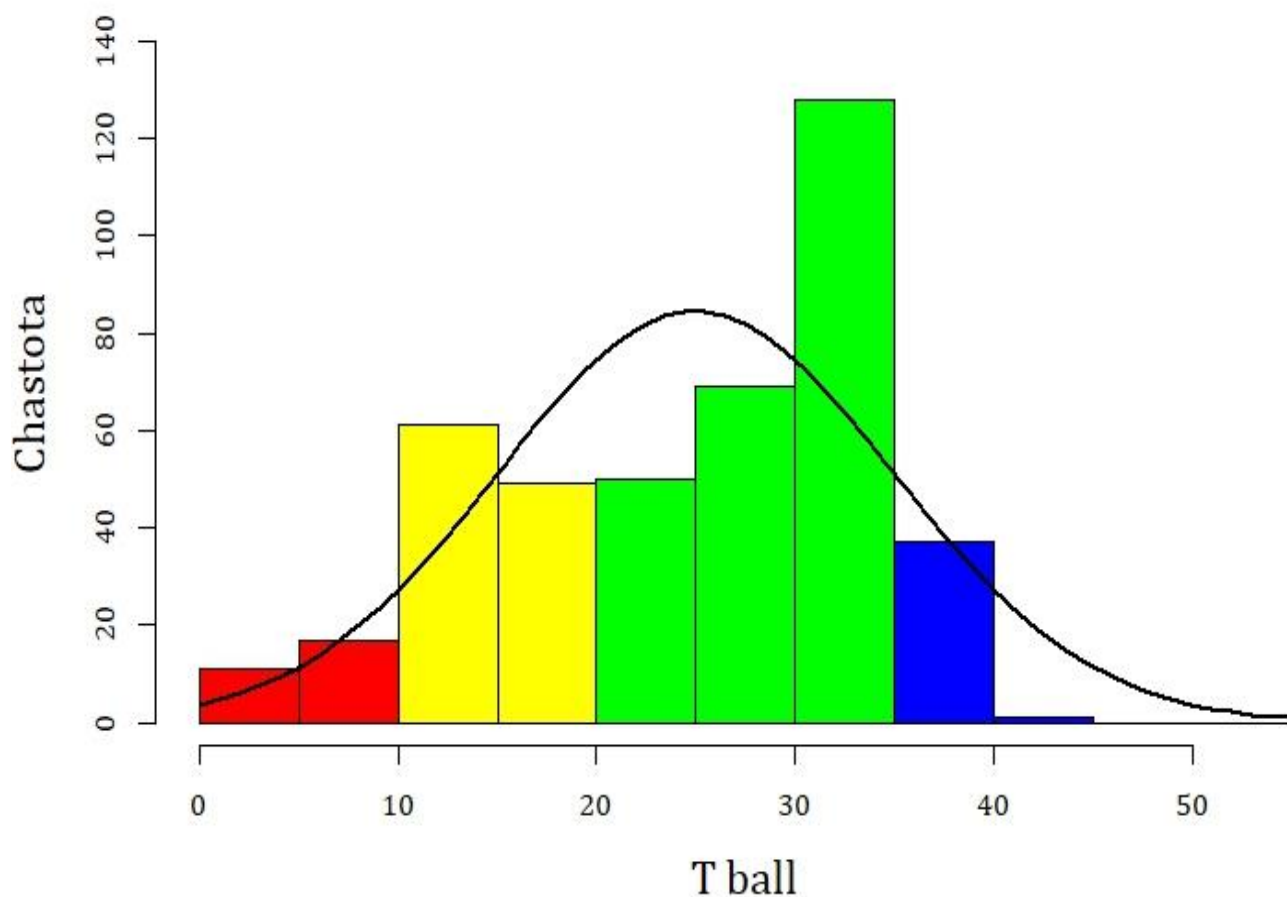


3-rasm. Tinglab tushunish ko'nikmasi bo'yicha Rash modeli bilan aniqlangan haqiqiy ball (T-ball) taqsimoti

4-rasmda o'qib tushunish ko'nikmasi bo'yicha Rash modeli bilan aniqlangan haqiqiy ball (T-ball) taqsimoti gistogrammasi keltirilgan. Bu yerda ham umumiy holda Rash modeli bilan aniqlangan haqiqiy ballar

(T-ball) taqsimoti gistogrammasi normal taqsimotga yaqin.

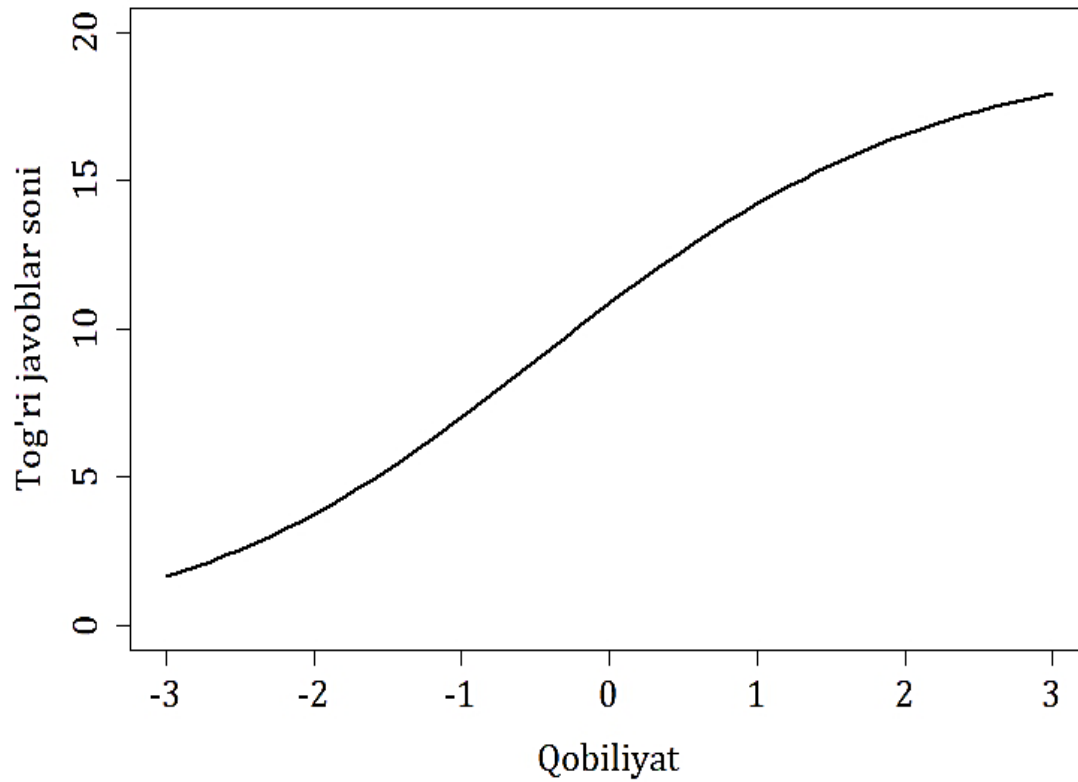
Gistogrammalarning normal taqsimotga yaqinligi testning sifatini, test sinovlarining obyektiv o'tkazilganligini bildiradi.



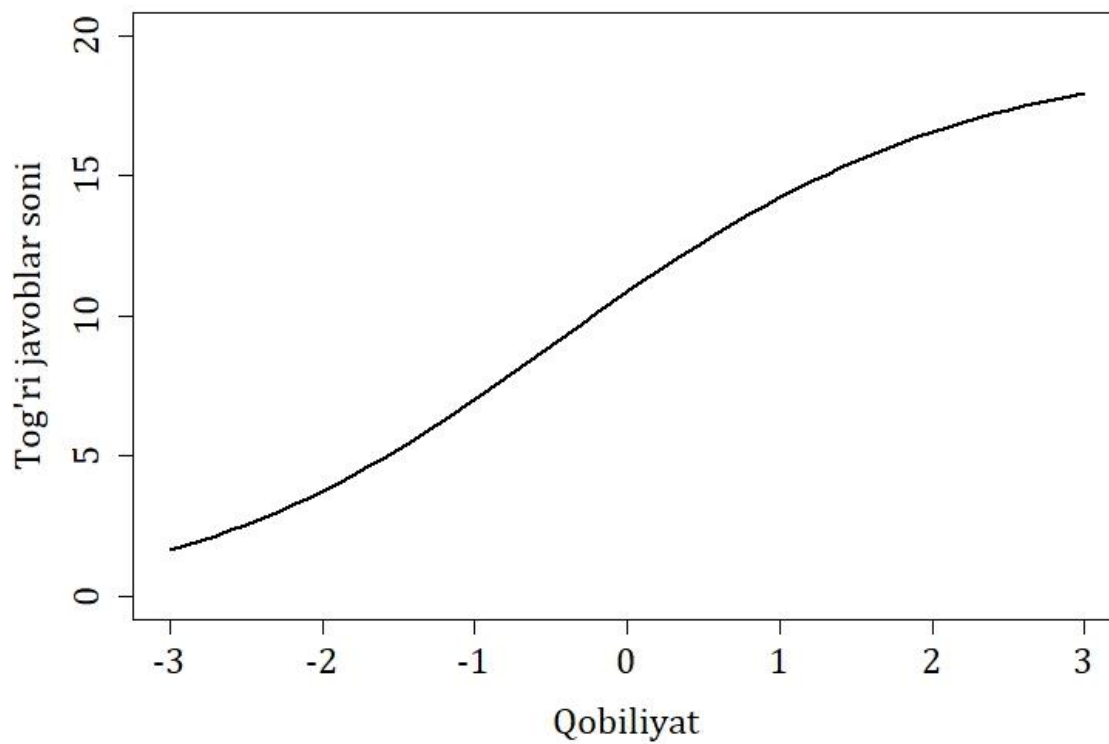
4-rasm. O'qib tushunish ko'nikmasi bo'yicha Rash modeli bilan aniqlangan haqiqiy ball (T-ball) taqsimoti

5-rasmda tinglab tushunish ko'nimasi bo'yicha test xarakteristikasi chizig'i (TXCh) keltirilgan. Rasmdagi test xarakteristikasi chizig'idan to'g'ri javoblar soni qobiliyat darajalariga mos holda ortib borayotganligini va har bir talabgorning qobiliyat darajalariga mos bo'lgan ballar berilganligini ko'rish mumkin.

6-rasmda esa o'qib tushunish ko'nimasi bo'yicha test xarakteristikasi chizig'i keltirilgan. Bunda ham to'g'ri javoblar soni qobiliyat darajalariga mos holda ortib borayotganligini va har bir talabgorning qobiliyat darajalariga mos bo'lgan ballar berilganligini test xarakteristikasi chizig'idan ko'rish mumkin.



5-rasm. Tinglab tushunish ko'nikmasi bo'yicha test xarakteristikasi chizig'i



6-rasm. O'qib tushunish ko'nikmasi bo'yicha test xarakteristikasi chizig'i

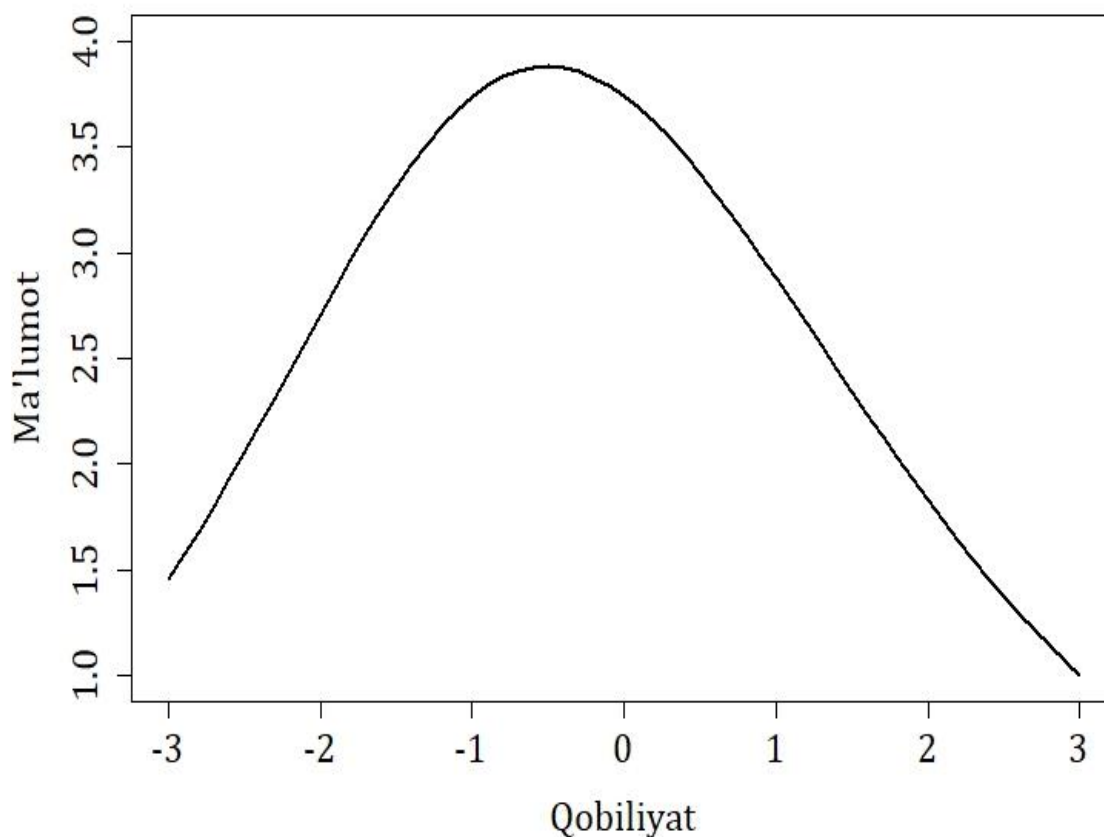
7-rasmda O'zbek tilini xorijiy til sifatida bilish darajasini baholash uchun o'tkazilgan test sinovida tinglab tushunish ko'nikmasi bo'yicha foydalanilgan test variantining test ma'lumoti chiziqlari (TMCh) keltirilgan.

Umuman olganda test ma'lumot chizig'i cho'qqisining nolga nisbatan chap tomonga surilganligi ushbu test varianti qobiliyat darajasi quyi bo'lgan talabgorlar to'g'risida ko'proq ma'lumot berishini, aksincha, ma'lumot chizig'i cho'qqisining nolga nisbatan o'ng tomonga surilganligi esa ushbu test varianti qobiliyat darajasi yuqori

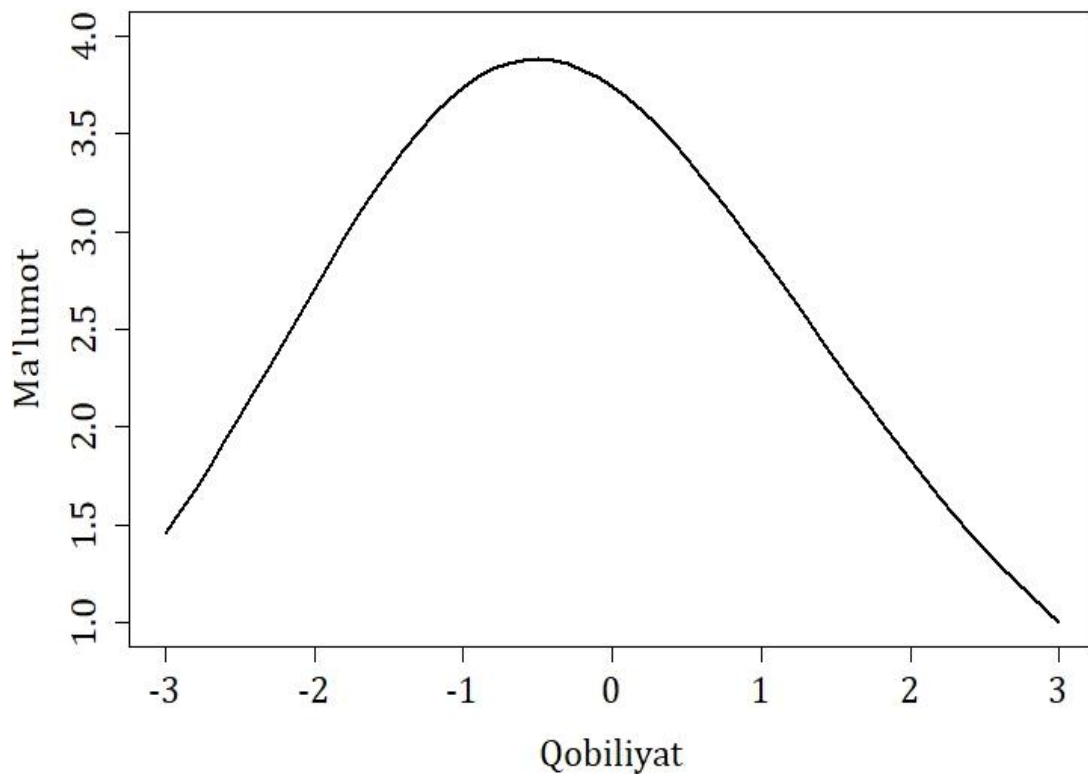
bo'lgan talabgorlar to'g'risida ko'proq ma'lumot berishini anglatadi.

O'zbek tilini xorijiy til sifatida bilish darajasini baholash uchun o'tkazilgan (tinglab tushunish ko'nikmasi bo'yicha) test sinovida foydalanilgan ushbu variant uchun ma'lumot chizig'i (7-rasm) cho'qqisining nolga nisbatan chap tomonga surilganligi ushbu test variantlari qobiliyat darajasi past bo'lgan talabgorlar haqida ko'proq ma'lumot olish mumkinligini anglatadi.

8-rasmda esa o'qib tushunish ko'nikmasi bo'yicha foydalanilgan test variantining test ma'lumoti chiziqlari (TMCh) keltirilgan.



7-rasm. Tinglab tushunish ko'nikmasi bo'yicha test ma'lumoti chizig'i



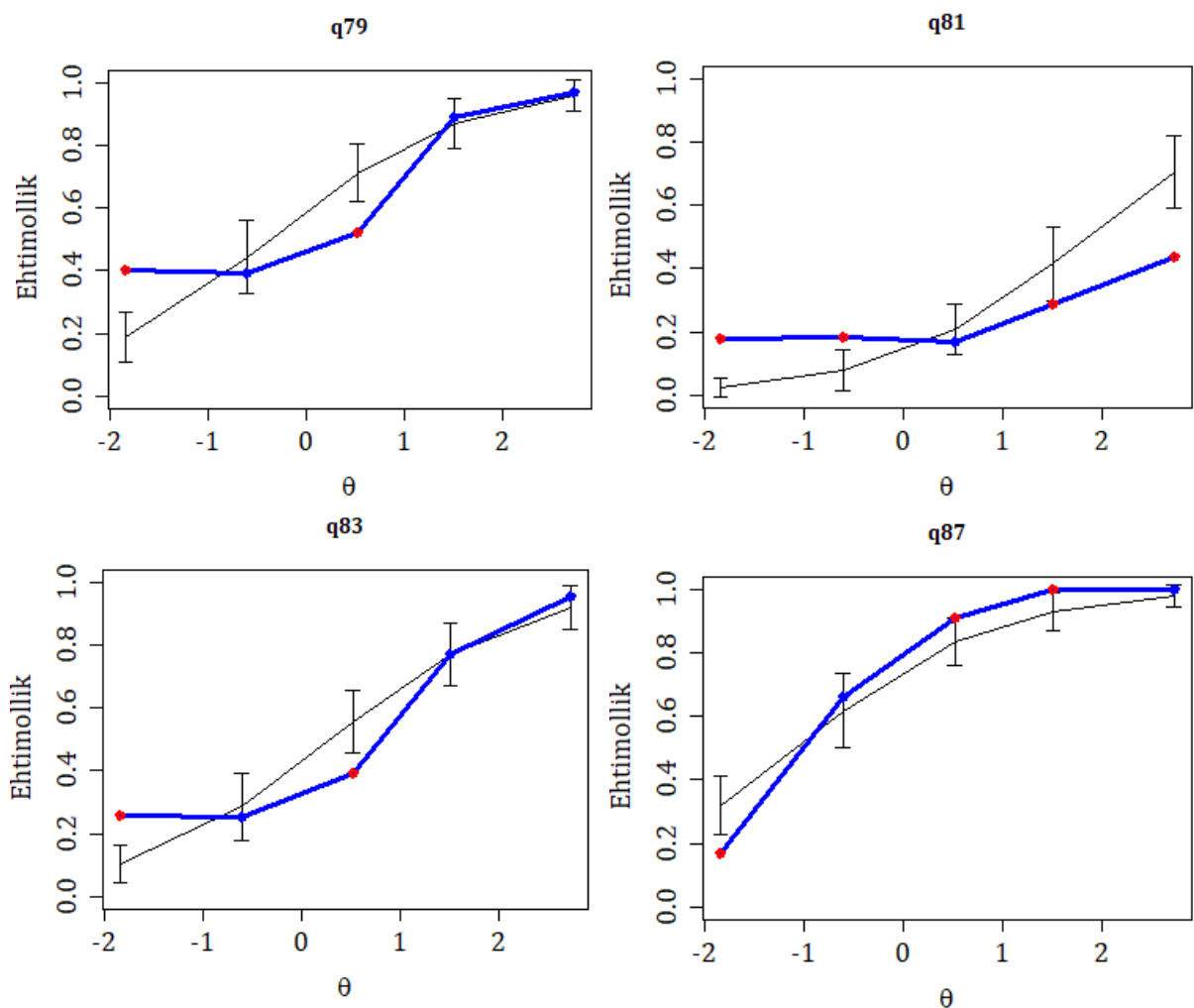
8-rasm. O'qib tushunish ko'nikmasi bo'yicha test ma'lumoti chizig'i

O'zbek tilini xorijiy til sifatida bilish darajasini baholash uchun o'tkazilgan test sinovida o'qib tushunish ko'nikmasi bo'yicha foydalanilgan variant uchun ma'lumot chizig'i (8-rasm) cho'qqisining nolga nisbatan chap tomonga surilganligi ushbu test variantlari qobiliyat darajasi quyi bo'lgan talabgorlar haqida ko'proq ma'lumot olish mumkinligini anglatadi.

Turli xil qobiliyat darajalarini aniqlash uchun kalibrovkalanagan test topshiriqlari bazasini yaratishda test

topshiriqlarining Rash modeli bilan moslik statistikalari o'rganilishi lozimdir.

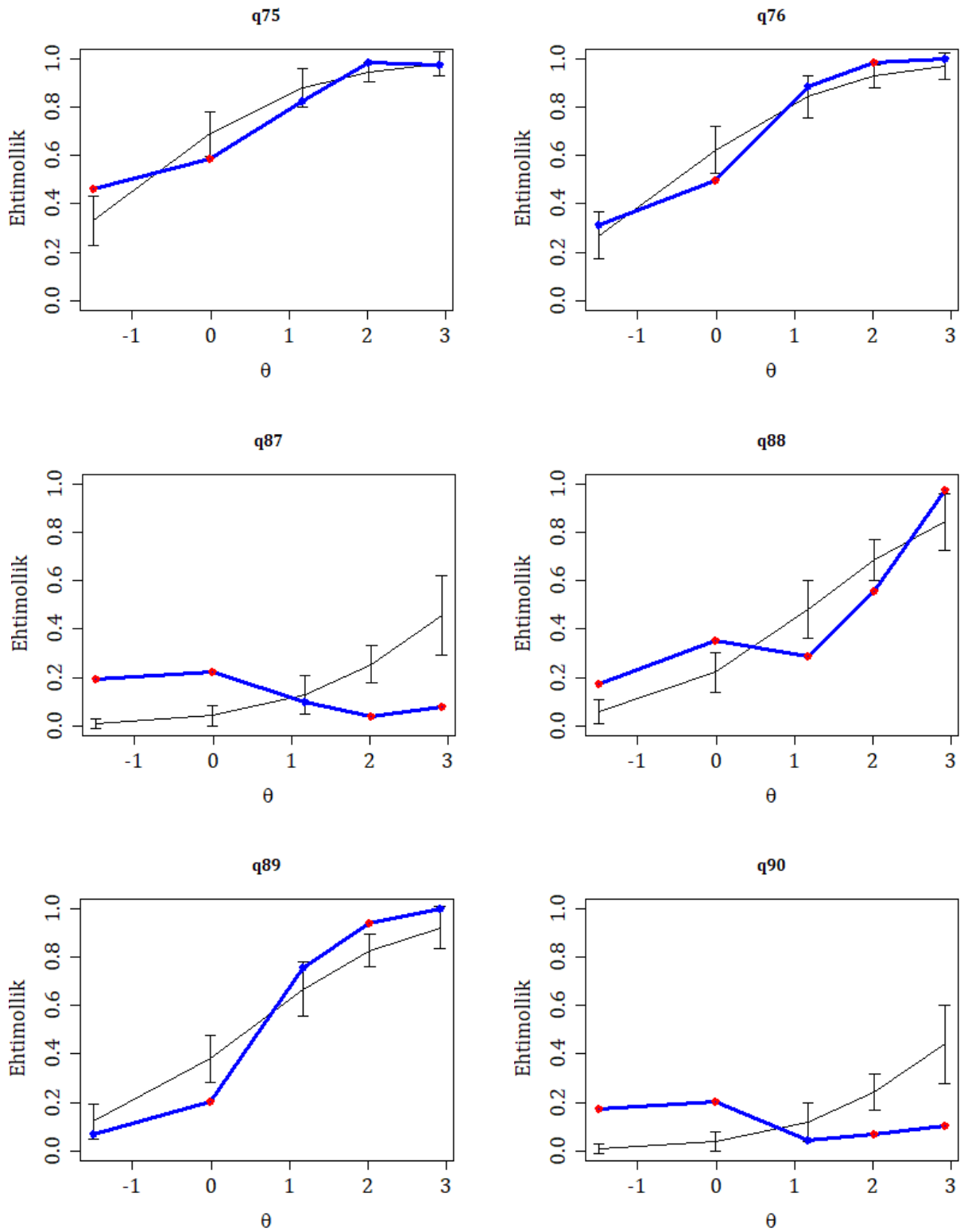
Tahlil natijalariga ko'ra tinglab tushunish ko'nikmasi bo'yicha quyidagi q79, q81, q83, q87 va q88 – ID raqamli (9-rasm) test topshiriqlarining Rash modeli bilan moslik darajasi yaxshi emasligini, ya'ni ajratilgan qobiliyat guruhlarining barchasi bilan mos tushmaganligini bildiradi va bunday test topshiriqlarini test bazasiga qo'shishdan oldin qayta ko'rib chiqish maqsadga muvofiq bo'ladi .



9-rasm. Tinglab tushunish ko'nikmasi bo'yicha test topshiriqlari natijalarining Rash modeli bilan mosligi

Tahlil natijalariga ko'ra o'qib tushunish ko'nikmasi bo'yicha quyidagi q75, q76, q87, q88, q89 va q90 - ID raqamli (10-rasm) test topshiriqlarining Rash modeli bilan moslik darajasi yaxshi emasligini, ya'ni

ajratilgan qobiliyat guruhlarining barchasi bilan mos tushmaganligi aniqlandi. Bunday test topshiriqlarini ham test bazasiga qo'shishdan oldin qayta ko'rib chiqish maqsadga muvofiq bo'ladi.



10-rasm. O'qib tushunish ko'nikmasi bo'yicha test topshiriqlari natijalarining Rash modeli bilan mosligi

### 2.3. Ko'p qirrali Rash modeli bo'yicha natijalar

Ko'p qirrali Rash modeli odatda esse yozish imtihonlarida qo'llaniladi. Bunda tartiblangan  $M + 1$  ta kategoriyali reyting shkalasi ishlatiladi. Bunda eng minimal kategoriya 0 maksimal kategoriya esa  $M$  bo'ladi [13]:

$$\log \left( \frac{P_{nijk}}{P_{nijk-1}} \right) = B_n - D_i - C_j - F_k,$$

bu yerda  $P_{nijk}$  -  $n$ -sinaluvchiga  $i$ -topshiriqqa  $j$ -baholovchi tomonidan  $k$ -ball berish ehtimoli;  $P_{nijk-1}$  -  $n$ -sinaluvchiga  $i$ -topshiriqqa  $j$ -baholovchi tomonidan  $k-1$  ball berish ehtimoli;  $B_n$  - sinaluvchining qobiliyati;  $D_i$  -  $i$ -chi topshiriqning qiyinlik darajasi;  $C_j$  -  $j$ -chi baholovchining qat'iyiligi;  $F_k$  -  $k - 1$  kategoriyadan  $k$  va  $k = 1, \dots, M$  kategoriyaga ko'tarilgandagi qiyinlik darajasi.

1978-yilda Andrich [14] Rashning reyting shkalasi modelini taklif qildi. Bunda sinaluvchilar tartiblangan reyting shkalasiga qanday javob berishi modellashtiriladi. Standart Andrich modelida reytingi aniqlanayotgan sinaluvchining qobiliyati va elementlarning qiyinlik darajalari yoki reyting kategoriyasining chegaraviy qiymati hisoblanadi.

Ko'p qirrali Rash modeli - Andrich taklif qilgan modelga o'xshash, lekin sinaluvchilarning yozma ish bo'yicha

natijalarini baholashda baholovchining qat'iyiligi va kategoriyadan kategoriyaga o'tish qiyinlik darajasini hisobga oladi.

Ko'p qirrali Rash modelida turli xil qirralarni hisobga olish imkoniyati mavjud. Ushbu ilmiy tadqiqotda talabgorning qobiliyat darajasi, element (savol) qiyinlik darajasi, baholovchilarning qat'iyiligi, bir kategoriyadan boshqasiga o'tish qadamlarining qiyinlik darajasi hisobga olinadi.

O'zbek tilini xorijiy til sifatida bilish darajasini baholash uchun o'tkazilgan aprobatsiya test sinovining yozish va gapirish ko'nikmalarini baholash uchun ko'p qirrali Rash modelidan foydalanildi [13]. Hisoblashlar R dasturining TAM paketi bilan amalga oshirildi [15].

Aprobatsiya test sinovida yozish ko'nikmasini baholash uchun 2 ta essedan foydalanilgan. Har bir esse 4 ta *11-grammatik savodxonlik*, *12-mantiqlik va izchillik*, *13-so'z boyligi*, *14-topshiriqning bajarilganligi* kabi baholanadigan elementlardan iborat. 1-esse 6 ta kategoriyali (0, 1, 2, 3, 4, 5), 2-esse esa 7 ta kategoriyali (0, 1, 2, 3, 4, 5, 6) mezon asosida baholangan. Quyidagi 9-jadvalda birinchi va ikkinchi esse elementlarining qiyinlik darajalari keltirilgan.

## 9-jadval

Birinchi va ikkinchi esse elemenlarining ko'p qirrali Rash modeli bilan aniqlangan qiyinlik darajalari

Element	Qiyinlik darajasi (1-esse)	Qiyinlik darajasi (2-esse)
I1	-0,03	0,05
I2	-0,11	-0,08
I3	-0,43	-0,15
I4	-0,40	-0,60

9-jadvaldan har ikkala esselarning grammatik savodxonlik hamda mantiqiylik va izchillik elementlari talabgorlar uchun nisbatan qiyinroq, so'z boyligi va topshiriqlarning bajarilganligi elementlari esa nisbatan oson bo'lganligini ko'rish mumkin. Elementlar orasida qiyinlik darajasi farqlari mavjud, lekin juda katta emas. Bu yozma ishda qobiliyatlarining darajasi 4 ta element bo'yicha turli xil ekanligini ko'rsatadi. Misol uchun, esse qandayligidan qat'iy nazar grammatik

savodxonlik hamda mantiqiylik va izchillik talabgorlar uchun so'z boyligi va topshiriqlarning bajarilganligiga nisbatan qiyinroq.

Kategoriyalar orasidagi qadamlar - bir kategoriyadan ikkinchi kategoriyaga o'tishning qiyinlik darajalarini tahlil qilish imkonini beradi. Quyidagi 10-jadvalda birinchi va ikkinchi esselar uchun bir kategoriyadan keyingi kategoriya o'tish qadamining qiyinlik darajalari keltirilgan.

## 10-jadval

Bir kategoriyadan keyingisiga o'tish qadamlarining ko'p qirrali Rash modeli bilan aniqlangan qiyinlik darajalari

Qadamlar	O'tish qadamlarining qiyinlik darajalari (esse1)	O'tish qadamlarining qiyinlik darajalari (esse2)
1	-1,26	-0,33
2	-1,31	-1,54
3	-0,23	-0,69
4	0,74	-0,04
5	2,07	0,39
6	-	2,22

10-jadvaldan birinchi esseda past kategoriyalar ( $0 \rightarrow 1$ ,  $1 \rightarrow 2$ ) juda osonligi hamda birinchi va ikkinchi qadamlar deyarli bir xil, ya'ni 0 balldan 1 ballga o'tish, 1 balldan 2 ballga o'tish qiyinlik darajasi farqi kichik ekanligi ko'rinadi. 5-qadam qiyinlik darajasi 4-qadam qiyinlik darajasiga nisbatan juda katta, ya'ni 4 balldan 5 ballga o'tish qiyinlik darajasi juda yuqori. Kategoriyalar tartibli ishlashi uchun  $qadam1 < qadam2 < qadam3 < qadam4 < qadam5$  bo'lishi maqsadga muvofiq. Umuman olganda, bunday tartib deyarli amalga oshmoqda, biroq birinchi qadamdan 2-chisiga o'tishida bir oz nomuvofiqlik bor.

Ikkinchi esseda qadam qiyinlik darajasi juda kichik, bu 1 balldan 2 ballga o'tish juda osonligini bildiradi. Shuningdek 10-jadvaldan  $qadam1 > qadam2$  ekanligini ko'rish mumkin. Kategoriya tartibi mantiqan ideal emas, chunki qadam 6 juda katta, bu 6 ballni olish juda qiyin ekanligini bildiradi. 0, 1, 2, 3, 4, 5, 6 ballardan iborat 7 ta kategoriyali shkala murakkab, lekin barcha bosqichlar uchun  $qadam1 < qadam2 < qadam3 < qadam4 < qadam5 < qadam6$  bo'lishi - qadam 1 dan qadam 2 ga o'tishdan boshqa barcha bosqichlar ishlayotganligini ko'rsatadi.

Hisoblashlar baholovchilarning qat'iyligi deyarli bir xil va 0 ga juda

yaqin ekanligini ko'rsatdi. Birinchi esseda  $R1 = -0.03$  va  $R2 = +0.03$  birinchi baholash biroz qat'iy emas, ikkinchi baholash esa biroz qat'iyroq. Ularning farqi juda kichikligi - baholovchilar muvofiqligi yaxshi ekanligini bildiradi. Ikkinchi esseda baholovchilar qat'iyligi farqi birinchi esseni baholashdagiga qaraganda biroz kattaroq ( $R1 = -0.08$ ,  $R2 = +0.08$ ). Bu baholash subyektivligi biroz kuchliroq ekanligini ko'rsatadi. 2-esseda baholovchilar qat'iyligi o'rtasidagi farq 1-essega qaraganda kattaroqdir.

Quyidagi 11-jadvalda baholovchilarning qat'iyligi kategoriyalar bo'yicha keltirilgan.

11-jadvaldan birinchi esseni baholagan birinchi baholovchi past ballarda qat'iy emas, yuqori ballarda esa qat'iyroq, ikkinchisi esa buning aksi. Qat'iylilik qiymatlari haddan tashqari katta emas.

Ikkinchi esseda faqat 5-qadamda 1- va 2-baholovchilarning qat'iylilik qiymatlari ko'p farq qilmaydi, qolgan qadamlarda esa 1- va 2-baholovchilarning farqi katta. Bu 2-esseni baholashda baholovchilar mutanosib baholay olmayotganligini bildiradi. Shuningdek, bu holat baholash mezonlari bilan ham bog'liq bo'lishi mumkin.

## 11-jadval

Baholovchilarning kategoriyalar bo'yicha qat'iyliigi

Qadam:Baholovchi	Kategoriyalar bo'yicha qat'iylik (esse1)	Kategoriyalar bo'yicha qat'iylik (esse2)
1:1	0.30	0.47
2:1	0.04	-0.75
3:1	0.19	0.16
4:1	-0.12	0.34
5:1	-0.41	-0.02
6:1	-	-0.21
1:2	-0.30	-0.47
2:2	-0.04	0.75
3:2	-0.19	-0.16
4:2	0.12	-0.34
5:2	0.41	0.02
6:2	-	0.21

Gapirish ko'nikmasi ham 4 ta (*I1-grammatik savodxonlik, I2-mantiqiylik va izchillik, I3-so'z boyligi, I4-talaffuz*) elementdan tashkil topgan bo'lib, 1 ta topshiriqdan iborat. Baholovchilar 7 ta kategoriyadan (0, 1, 2, 3, 4, 5, 6) iborat bo'lgan baholash mezoni asosida talabgorlarni baholaganlar. Topshiriq

tuzilmasi yozma ish bilan bir xil bo'lganligi uchun gapirish ko'nikmasini ham ko'p qirrali Rash modeli bilan tahlil qilish mumkin. Quyidagi 12-jadvalda gapirish ko'nikmasi bo'yicha berilgan test topshiriqlarining qiyinlik darajalari berilgan.

## 12-jadval

Gapirish ko'nikmasi bo'yicha berilgan topshiriq elementlarining ko'p qirrali Rash modeli bilan aniqlangan qiyinlik darajalari

Elementlar	Qiyinlik darajasi
I1	-2,63
I2	-2,46
I3	-2,36
I4	-2,63

12-jadvaldan barcha elementlarning qiyinlik darajasi manfiy va absolyut qiymatlari katta, ya'ni ular juda oson ekanligi ko'rinadi. Elementlarning qiyinlik darajasi barqaror bo'lib, so'z boyligi elementi nisbatan eng qiyin, grammatik

savodxonlik va tallafuz elementlari esa eng oson va deyarli bir xildir.

Quyidagi 13-jadvalda gapirish ko'nikmasi bo'yicha berilgan topshiriq uchun bir kategoriyadan keyingisiga o'tish qadamlarining qiyinlik darajalari keltirilgan.

**13-jadval**

Bir kategoriyadan keyingisiga o'tish qadamlarining ko'p qirrali Rash modeli bilan aniqlangan qiyinlik darajalari

Qadamlar	O'tish qadamlarining qiyinlik darajalari
1	-10,15
2	0,98
3	1,67
4	2,48
5	1,81
6	3,21

13-jadvaldan 1-qadam qiyinlik darajasi juda kichik (**-10,15**) ekanligini ko'rish mumkin. Bu esa eng past kategoriya juda oson tanlanishini, ya'ni juda kam axborot beruvchi kategoriyadir. Qadamlar, asosan, o'sib boradi, lekin 5-qadam qiyinlik darajasi 4-qadam qiyinlik darajasidan kichik, bu esa reyting shkalasining o'rtayuqori qismida (4-5-qadamlar) tuzilmaviy muammo mavjudligini va baholovchilar bu ikki kategoriyani aniq farqlay olmayotganligini bildiradi. Bu kamchilik baholash mezonini yoki topshiriqning berilishi bilan bog'liq

bo'lishi mumkin. Shuning uchun baholovchilar har ikkala holatni tekshirib ko'rish va muammoni yechimini aniqlashlari lozim bo'ladi.

Hisoblashlar baholovchilarning qat'iyligi bir-biridan taxminan **0,57** logit birligi bilan farq qilishini ko'rsatdi:  $R1=+0.29$  va  $R2=-0.29$  - birinchi baholovchining qat'iyligi kattaroq va ikkinchi baholovchining qat'iyligi kichik.

Quyidagi 14-jadvalda baholovchilarning qat'iyligi kategoriyalar bo'yicha keltirilgan.

## Baholovchilarning kategoriyalar bo'yicha qat'iyligi

Qadam:Baholovchi	Kategoriyalar bo'yicha qat'iylilik
1:1	0,29
2:1	0,16
3:1	0,34
4:1	0,42
5:1	-0,22
6:1	-0,99
1:2	-0,29
2:2	-0,16
3:2	-0,34
4:2	-0,42
5:2	0,22
6:2	0,99

1-baholovchining 1-4-qadamlarda qat'iyligi katta va 5-6-qadamlarda esa kichik, 2-baholovchining qat'iyligi 1-baholovchi qat'iyligiga nisbatan to'liq teskari. Baholovchilar orasidagi eng katta farq 6-qadamda bo'lib, 1-baholovchining kategoriyalar bo'yicha qat'iylilik qiymati  $-0,99$ , 2-baholovchining kategoriyalar bo'yicha qat'iylilik qiymati  $+0,99$  ga teng ekanligini ko'rish mumkin. Eng yuqori ball (6 ball) kategoriyasi baholovchilar tomonidan keskin farqli qo'llanilishi – kuchli differensial baholovchi ta'siri (*ing. differential rater function*) mavjudligini ko'rsatadi. Bu esa baholovchilar reyting shkalasining qadamlarini har xil tushunishini

anglatadi. Ushbu ta'sirni hisobga olmaslik baholashning adolatliligiga salbiy ta'sir qilishi mumkin.

O'zbek tilini xorijiy til sifatida bilish darajasini baholash uchun yozish ko'nikmasi bo'yicha o'tkazilgan aprobatsiya test sinovi natijalari ko'p qirrali Rasch modeli asosidagi tahlil natijalarining umumiy xulosalari 15-jadvalda keltirilgan. Tahlil natijalariga ko'ra, gapirish ko'nikmasi bo'yicha 4- va 5-qadamlarni birlashtirish yoki tartiblash imkoniyatini ko'rib chiqish, baholovchilar uchun, ayniqsa, yuqori baholar (6-qadam) bo'yicha qo'shimcha trening o'tkazish lozimligini aytish mumkin.

1- va 2-esselarni baholash bo'yicha umumiy xulosalar

Jihatlar	1-esse (0-5)	2-esse (0-6)
Element (savol)ning farqlanishi	O'rtacha	Kuchliroq
Kategoriyalar soni	6	7
Kategoriyalar tartibi	Yaxshi	Muammoli
Eng yuqori ball	Qiyin	Juda qiyin
Baholovchining qat'iylik qiymati farqi	Juda kichik	Kattaroq
Baholovchilar va o'tish qadamlarining o'zaro ta'siri	Barqaror	Beqaror

Bundan tashqari baholovchilar baholash mezonlarini tahlil asosida o'rganib chiqishlari, baholovchilar uchun berilgan tavsiyalalarni o'rganib,

kerakli o'rinlarda o'zlarining baholash usullarini yaxshilashi va seminar-treninglarda malakalarini oshirishlari lozim.

## Xulosa

O'zbek tilini xorijiy til sifatida bilish darajasini aniqlash va sertifikatlash uchun o'tkazilgan ap-robatsiya test sinovi natijalari bo'yicha test variantining ishonchlilik koef-fitsiyenti (Kronbax alfa koefitsiyenti) tinglab tushunish ko'nikmasi uchun 0,87 va o'qish ko'nikmasi uchun 0,88 ga tengligi mezonlarga ko'ra variantning ichki muvofiqligi "yaxshi" ekanligini bildiradi.

O'lchashning standart xatoligi tinglab tushunish ko'nikmasi uchun 1,76 ga va o'qib tushunish ko'nikmasi uchun 1,67 teng ekanligi aniqlandi. Natijalarni talqinida bu hisobga olinishi lozim.

Qobiliyat darajasi yuqori va past bo'lgan talabgorlar haqida teng ma'lumot olish uchun 0 logit birligidan pastdagi va yuqoridagi test top-shiriqlarini teng taqsimlanishiga erishish lozim.

Taqsimotdagi takrorlangan bir xil qiyinlikdagi test topshiriqlari hisobiga bo'sh joylarga mos keladigan qiyinlik darajasidagi test topshiriqlaridan qo'yish taqsimotni yanada yaxshi-lashga imkon yaratadi.

Barcha test topshiriqlariga to'g'ri yoki noto'g'ri javob bergan talabgorlar

qobiliyatlarining xatoliklari juda katta bo'lishiga olib keladi. Bu ma'lumotni test tuzuvchi mutaxassilar hisobga olishi zarur.

Rash modeli bilan moslik darajasi yaxshi bo'lmagan o'qib tushunish ko'nikmasi bo'yicha q75, q76, q87, q88, q89 va q90-ID raqamli, tinglab tushunish bo'yicha q79, q81, q83, q87 va q88- ID raqamli test topshiriqlarini kalibirovkalangan test bazasiga qo'shishdan oldin qayta ko'rib chiqish lozim.

Yozish ko'nikmasi bo'yicha tadqiqot natijalariga ko'ra, element (savol)ning farqlanishi 1-esseda o'r-tacha, 2-esseda esa kuchliroq, baho-lovchining qat'iylik qiymati farqi 1-es-seda juda kichik, 2-esseda esa kattaroq ekanligi ko'rsatib berildi.

Gapirish ko'nikmasi bo'yicha tadqiqot natijalariga ko'ra, ele-mentlarning qiyinlik darajasi barqaror va muammosiz ekanligi aniqlandi. Bundan tashqari, reyting shkalasida qadamlar (4-5-qadamlar) tartibsiz-ligining mavjudligi, baholovchilarning umumiy qat'iyligi realistik baho-langanligi va baholovchilar ayrim qadamlarni keskin farqli qo'llayot-ganligi aniqlandi.

## ADABIYOTLAR

1. C. Spearman. Correlation from faulty data, *British J. of Psychology*. 3, 271-295, 1910.
2. М. Б. Чельшкова. Теория и практика конструирования педагогических тестов: учеб. Пособие М.: Логос, 2002, с. 432.
3. C. Spearman. Demonstration of formulae for true measurement of correlation. *Am. J. of Psychology*. 18, 160-169, 1907.
4. C. Spearman, "General intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 201-293, 1904.
5. C. Spearman,. The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101, 1904.
6. Rasch G. Probabilistic models for some intelligence and attainment tests, Copenhagen, Danish Institute for Educational research, 1960.
7. R. K. Hambleton and W. J. Van der Linden. Advances in item response theory and applications: An introduction, *Advanced Psychological Measurement*, 6, 373-378, 1982.
8. Hambleton, R.K., Swaminathan, H.,& Rogers, H.J. Fundamentals of item response theory. Newbury Park, CA: Sage, 1991.
9. Baker, Frank. The Basics of Item Response Theory, ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD, 2001.
10. Ivailo Partchev. A visual guide to item response theory, Friedrich Schiller-Universitat Jena, 2004.
11. Gunter Maris, Timo Bechger, Jesse Koops and Ivailo Parchev, *Data Management and Analysis of Tests*, p. 1-49, 2022.
12. B. D. Wright and M. H. Stone. Best Test Design. MESA Press, Chicago, 1979.
13. J. M. Linacre. Many-facet Rasch measurement. Chicago: MESA Press, 1989.
14. D. Andrich. A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573, 1978.
15. A. Robitzsch, T. Kiefer, & M. Wu. TAM: Test Analysis Modules. R package version 3.5-19. <https://CRAN.R-project.org/package=TAM> , 2020.

## ASSESSMENT OF LANGUAGE SKILLS: CLASSICAL TEST THEORY, RASCH AND MANY-FACET RASCH MODELS

**A.B. Normurodov, M.J. Ermamatov, A.R. Sattiyev, K.A. Amonov**

*Scientific and Educational Practical Center Under the Agency for Assessment of Knowledge and Competences, [a.normurodov@uzbmb.uz](mailto:a.normurodov@uzbmb.uz)*

**Abstract.** This article examines the possibilities of analyzing language skill assessment results based on Classical Test Theory, the Rasch model, and the Many-Faceted Rasch model. For the study, the results of an experimental (pilot) test conducted to determine and certify the proficiency level of Uzbek as a foreign language were used. Initially, the test results were analyzed using traditional Classical Test Theory, including descriptive statistics of test items, difficulty levels, and the correlation between individual test items and the total score. Differences between the distribution of test items across three conditional difficulty levels and the normal distribution, as well as the impact of correlation values with the total score on test quality, were demonstrated through calculations. Based on the Rasch model, item difficulty levels, the Wright map, model fit, and test characteristic curves were examined. It was determined that there are relatively few test items for high- and low-ability levels, and the extent to which the difficulty of the test form and the relationship between responses and ability conform to the Rasch model was identified. Writing and speaking skills were analyzed using the Many-Faceted Rasch model. According to the results for writing skills, item discrimination was moderate in Essay 1 and stronger in Essay 2, while the difference in rater severity was very small in Essay 1 but larger in Essay 2. For speaking skills, the analysis showed that item difficulty levels were stable and unproblematic, there was some disorder in the rating scale steps (particularly steps 4–5), overall rater severity was well calibrated, and certain rating scale categories were applied inconsistently by raters.

**Keywords:** Item difficulty, Wright map, Test Information Function, Test Characteristic Curve, Classical Test Theory, Many-Facet Rasch Model.