

TARIX FANIDAN MILLIY SERTIFIKAT UCHUN O'TKAZILGAN TEST SINOV NATIJALARINING KLASSIK TEST NAZARIYASI VA RASH MODEL ASOSIDA TAHLILI

A.R.Sattiyev, M.Dj.Ermamatov

Bilim va malakalarni baholash agentligi huzuridagi

Ilmiy-o'quv amaliy markazi, 100084, Toshkent sh., Bog'ishamol k., 12

Qisqacha mazmuni. Ushbu maqolada O'zbekiston Respublikasi bo'yicha ilk bor tarix fanidan Milliy sertifikat uchun o'tkazilgan test sinovlari natijalari klassik test nazariyasi va Rash modeli asosida tahlil qilingan. Tavsif statistikasi va alohida test topshiriqlariga berilgan javoblarning umumiy ball bilan korrelyatsiyasi hamda test sinovida ishlatilgan variantdagi test topshiriqlarining qiyinlik darajasi tahlil qilingan. Rash modeli bilan hisoblangan qobiliyat va qiyinlik darajalaridan foydalanib Rayt xaritasi olingan va shu asosda qobiliyat va qiyinlik darajalari mosligi, ichki va tashqi moslik statistikalari, element xarakteristikasi, element va test ma'lumoti chiziqlari muhokama qilingan.

Kalit so'zlar: Test topshiriqlari, Kronbax alfa koeffitsiyenti, validlik, qiyinlik darajasi, Rash modeli, Rayt xaritasi, qobiliyat darajalari.

Kirish

Avvalgi maqolamizda [1] keltirib o'tilganidek, "Umumta'lim fanlarini o'qitish darajasini baholashning milliy test tizimini joriy etish to'g'risida"gi qaror ijrosini amaliyotga tatbiqi umumta'lim fanlari bo'yicha maktablar va akademik litseylarda ta'lim olayotgan o'quvchilar hamda pedagog o'qituvchilarning tegishli fan bo'yicha bilim darajasini aniqlash imkonini beradi.

Umumta'lim fanlari bo'yicha Milliy sertifikat uchun test jarayoniga tayyorgarlik ko'rish, test topshiriqlarini shakllantirish, ularni ekspertiza qilish

va, shu bilan birga, test sinovlari o'tkazish, natijalarini e'lon qilish Bilim va malakalarni baholash agentligi tomonidan amalga oshirilib kelinmoqda.

Ushbu maqolada tarix fanidan Milliy sertifikat uchun o'tkazilgan test sinovlari natijalarining tahlili beriladi.

Respublika bo'yicha 2023-yilning 21-may kuni tarix fanidan Milliy sertifikat uchun talabgorlar o'rtasida o'tkazilgan test sinovida jami 4554 nafar talabgor qatnashdi. ID raqamlari- 2421814, 2421619, 2409819, 2415907, 2415902 va 2415816 bo'lgan

sinaluvchilar (6 nafar) barcha test topshiriqlariga to'g'ri javob berganligi hamda ID raqami- 2411509 bo'lgan sinaluvchi (1 nafar) barcha test topshiriqlariga noto'g'ri javob berganligi uchun ushbu sinaluvchilarning (jami 7 nafar) natijalari tahlildan chiqarildi va 4547 nafar sinaluvchining

natijalari tahlil qilindi. Tarix fanidan Milliy sertifikat uchun o'tkazilgan test sinovlarida har bir variant 45 ta (36-45-ochiq test topshiriqlarining A va B qismlarga ajratilishi hisobiga 55 ta) test topshirig'idan iborat bo'lib, ajratilgan vaqt javoblar varaqasini bo'yash bilan birgalikda 90 daqiqani tashkil etgan.

1. Test sinovlari natijalarini klassik test nazariyasi bo'yicha tahlili

Pedagogik o'lchovlarning nazariy asoslariga ko'ra har bir test bo'yicha tuzilgan test topshiriqlari mazmuni ekspert tekshiruvidan o'tkazilgandan keyingi bosqichda test sinovlari o'tkaziladi va natijalar asosida testlarning xususiyatlari sifatini tashxislash uchun ularning statistik xarakteristikalarini aniqlanadi. Testlar statistikada tanlanma to'plam deb hisoblanadi. U real testlar to'g'risida xulosalar chiqarish imkonini beradi.

Statistik xarakteristikalar yordamida test topshiriqlariga qo'yilgan asosiy talablarning ko'rsatkichlari aniqlanadi. 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 darajasini aniqlash usullaridan biri topshiriqni empirik sinovdan o'tkazib, to'g'ri javoblar salmog'ini aniqlashdan

iborat. 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'rta qiymat, gistogrammani qurish, moda va mediana kabi ko'rsatkichlarni hisoblash hamda test ballarining umumiy dispersiyasi (standart og'ish) ham kiradi [2-5]. Test ballari (yoki to'g'ri javoblar)ning o'rta 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 obektiv

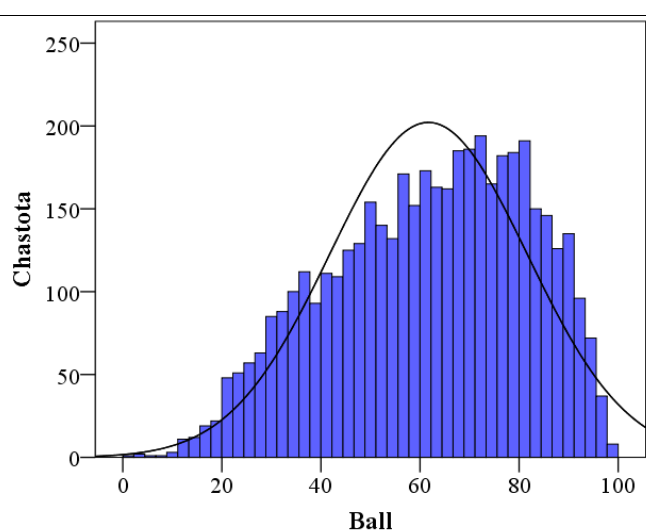
o'tkazilganligini bildiradi. Test ballarining eng ko'p takrorlanadigan qiymati statistikada moda, o'sish tartibida joylashtirilgan test ballari qatorining o'rtasida joylashgan qiymati esa mediana deyiladi. O'rta arifmetik qiymat, moda va mediana qiymatlari

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.

1-jadval

Tarix fanidan milliy sertifikat uchun o'tkazilgan test sinovi natijalarining tavsif statistikasi ma'lumotlari

Test topshiruvchilar soni	4547
O'rta qiymat	61,70
Mediana	63,70
Moda	77,20
Standart tafovut	19,94
Dispersiya	397,65
Asimmetriya	-0,33
Ekstsess	-0,74
Diapazon	97,60
Minimum	1,30
Maksimum	98,90



1-jadvalda tarix fanidan test sinovi natijalari bo'yicha olingan tavsif ma'lumotlari keltirilgan. Gistogrammadan ko'rinib turibdiki, test sinovi natijalari bo'yicha test topshiriqlarining individual ballari taqsimoti normal taqsimotga yaqin ekanligi ko'rinib turibdi. Statistik tadqiqot natijalariga ko'ra, tarix fanidan test variantining ishonchlilik koeffitsiyenti, ya'ni Kronbax alfa koeffitsiyenti test sinovi natijalari bo'yicha 0,92 ga teng ekanligi aniqlandi. Kronbax alfa koeffitsi-

yentining 0,9 va undan kattaligi ushbu test sinovlari uchun tanlab olingan test variantlarining ishonchliligi a'lo darajada ekanligini ko'rsatmoqda [6]. 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.

2. Klassik test nazariyasi asosida aniqlangan qiyinlik darajalari

Quyida tarix fanidan milliy sertifikat uchun o'tkazilgan test sinovi natijalari asosida, test topshiriqlarining - ID raqami, ishtirokchilar soni - N hamda bitta test topshirig'iga barcha talabgorlar to'g'ri javob berganida hosil bo'ladigan ballarning yig'indisi - X_{max} , talabgorlarning har bir test top-

shirig'idan olgan ballari yig'indisi - X_i , $\frac{X_i}{X_{max}} 100\%$ 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 keltirilgan.

2-jadval

Tarix fanidan Milliy sertifikat uchun o'tkazilgan test sinovi natijalari bo'yicha test topshiriqlarining aniqlangan qiyinlik darajalari

№	ID	N	X_{max}	X_i	Ans (foizda)	V
1	T5	4547	14550,4	13952	95,89	1
2	T9	4547	10003,4	8863,8	88,61	1
3	T12	4547	5911,1	5109	86,43	1
4	T19	4547	14550,4	12009,6	82,54	1
5	T6	4547	14550,4	11763,2	80,84	1
6	T18	4547	10003,4	7977,2	79,74	1
7	T10	4547	10003,4	7942	79,39	1
8	T33	4547	10003,4	7911,2	79,09	1
9	T20	4547	14550,4	11443,2	78,65	1
10	T4	4547	10003,4	7803,4	78,01	1
11	O37A	4547	6820,5	5298	77,68	1
12	T29	4547	5911,1	4417,4	74,73	2

13	T30	4547	10003,4	7328,2	73,26	2
14	T11	4547	14550,4	10646,4	73,17	2
15	T17	4547	10003,4	7312,8	73,10	2
16	T16	4547	14550,4	10617,6	72,97	2
17	T2	4547	10003,4	7183	71,81	2
18	T8	4547	10003,4	7121,4	71,19	2
19	T1	4547	5911,1	4175,6	70,64	2
20	T13	4547	14550,4	10089,6	69,34	2
21	T21	4547	5911,1	3949,4	66,81	2
22	T31	4547	10003,4	6538,4	65,36	2
23	T23	4547	10003,4	6384,4	63,82	2
24	T7	4547	10003,4	6338,2	63,36	2
25	T35	4547	10003,4	6300,8	62,99	2
26	O44A	4547	5001,7	3150,4	62,99	2
27	T32	4547	5911,1	3667,3	62,04	2
28	T27	4547	10003,4	6204	62,02	2
29	O40A	4547	5001,7	3075,6	61,49	2
30	T3	4547	5911,1	3511,3	59,40	2
31	O40B	4547	5001,7	2937	58,72	2
32	T22	4547	10003,4	5792,6	57,91	2
33	O45B	4547	5001,7	2845,7	56,89	2
34	T26	4547	10003,4	5671,6	56,70	2
35	O42A	4547	5001,7	2781,9	55,62	2
36	T34	4547	10003,4	5552,8	55,51	2
37	T24	4547	5911,1	2805,4	47,46	2
38	O45A	4547	5001,7	2301,2	46,01	2
39	T25	4547	5911,1	2691	45,52	2
40	O43B	4547	5001,7	2112	42,23	2
41	T28	4547	5911,1	2399,8	40,60	2
42	T14	4547	5911,1	2376,4	40,20	2
43	T15	4547	10003,4	3942,4	39,41	2
44	O42B	4547	5001,7	1855,7	37,10	2

45	O41A	4547	5001,7	1832,6	36,64	2
46	O41B	4547	5001,7	1774,3	35,47	2
47	O43A	4547	5001,7	1610,4	32,20	2
48	O36B	4547	5001,7	1552,1	31,03	2
49	O38B	4547	7729,9	2301,8	29,78	2
50	O44B	4547	5001,7	1334,3	26,68	2
51	O39A	4547	6820,5	1747,5	25,62	2
52	O36A	4547	5001,7	1133	22,65	3
53	O39B	4547	7729,9	1446,7	18,72	3
54	O37B	4547	7729,9	1055,7	13,66	3
55	O38A	4547	6820,5	609	8,93	3

Test topshiriqlarining qiyinlik darajalari bo'yicha (2-jadval) tahlil qiladigan bo'lsak, 55 ta test topshiriqlaridan 11 tasi (20,00 foiz) 1-qiyinlik darajasidagi test topshiriqlaridan, 40 tasi (72,73 foiz) 2-qiyinlik darajasidagi test topshiriqlaridan va 4 tasi (7,27 foiz) 3-qiyinlik darajasidagi test topshiriqlaridan iborat ekanligi aniqlandi.

3-jadvalda keltirilgan ma'lumotlarga ko'ra, ID raqamlari – T5, T9 va T12 bo'lgan yopiq turdagi test topshiriqlarining qiyinlik darajasi bo'yicha juda oson va O38A bo'lgan ochiq turdagi test

topshirig'ining qiyinlik darajasi bo'yicha juda qiyin ekanligini e'tiborga olgan holda ushbu test topshiriqlarini qiyinlik darajasi bo'yicha qayta ko'rib chiqish va sabablarini aniqlash maqsadga muvofiqdir.

Shuningdek, variantda 3-qiyinlik darajasidagi test topshiriqlari me'yordan kamligini hisobga olgan holda ularning sonini 16-25 foizgacha oshirish variantdagi test topshiriqlarining qiyinlik darajasi bo'yicha teng taqsimlanishiga erishiladi.

3. Element-umumiy ball korrelyatsiyalari

Har bir topshirig'iga berilgan javoblarning umumiy test bali bilan korrelyatsiyasi (1, 2, 3, ... 55-test topshiriqlari va umumiy ball orasidagi korrelyatsiya) element-umumiy ball korrelyatsiya deb atalishi mumkin [7].

Umuman olganda, element-umumiy ball korrelyatsiyasi 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. Element-umumiy ball korrelyatsiyasi qiymati 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.

3-jadvalda test natijalari tahlili natijasida olingan test topshiriqlarining biserial korrelyatsiya koeffitsiyenti qiymatlari keltirildi.

3-jadval

Test topshiriqlarining biserial korrelyatsiya koeffitsiyenti

№	ID	BKK
1	T5	0,293
2	T9	0,431
3	T12	0,463
4	T19	0,517
5	T6	0,418
6	T18	0,575
7	T10	0,493
8	T33	0,477
9	T20	0,562
10	T4	0,515
11	O37A	0,522
12	T29	0,442
13	T30	0,518
14	T11	0,441
15	T17	0,463
16	T16	0,523
17	T2	0,454
18	T8	0,519
19	T1	0,391
20	T13	0,488
21	T21	0,487
22	T31	0,492

23	T23	0,386
24	T7	0,504
25	T35	0,565
26	O44A	0,520
27	T32	0,379
28	T27	0,365
29	O40A	0,575
30	T3	0,378
31	O40B	0,593
32	T22	0,526
33	O45B	0,534
34	T26	0,341
35	O42A	0,618
36	T34	0,384
37	T24	0,422
38	O45A	0,446
39	T25	0,354
40	O43B	0,424
41	T28	0,202
42	T14	0,400
43	T15	0,245
44	O42B	0,532
45	O41A	0,528
46	O41B	0,476
47	O43A	0,540
48	O36B	0,466
49	O38B	0,379
50	O44B	0,485
51	O39A	0,478
52	O36A	0,354
53	O39B	0,414

54	037B	0,340
55	038A	0,328

Element-umumiy ball korrelyatsiyasi qiymatlari 0,25 dan kichik bo'lgan (3-jadval, ID raqamlari – T15 va T28 bo'lgan test topshiriqlari) test topshiriqlarini o'rganib chiqib, kerakli o'zgarishlar qilish, lozim bo'lsa, variantdan chiqarib tashlash kerak.

Element-umumiy ball korrelyatsiyasi qiymatlari 0,5 dan kichik bo'lgan test topshiriqlari esa o'rganib chiqilib, qiyinlik darajalarini hisobga olgan holda kerakli o'zgarishlar qilish maqsadga muvofiq bo'ladi.

4. Test sinovlari natijalarini Rash modeli asosida tahlili

Rash modeli – ingliz tilida “item response theory” (IRT) deb nomlanadi. Bu “elementlarga javob nazariyasi” yoki “elementlarga reaksiya” degan ma'noni anglatadi. IRT faqat savollarga javoblarnigina baholamasdan, balki turli xildagi so'rovnomalarni baholashda ham ishlatiladi. Shuning uchun ham bu nazariyada “test topshirig'i” o'rnida “element - (item)” so'zini ishlatish maqsadga muvofiq hisoblanadi. Tahlilda “test topshirig'i” tushunchasi va “element tushunchasi” bir xil ma'noda ishlatiladi.

IRT bu matematik model bo'lib, yashirin xarakteristikalar (bilim, stress, munosabat) uni namoyon qiluvchilari (kuzatilgan natija, javoblar) bilan bog'liqligini ifodalaydi. U instrumentdagi savollarning, bu savollarga javob berayotgan shaxslarning va uning asosida yashirin xarakteristikalar xususiyatlari orasidagi bog'liklikni o'rnatadi. IRT yashirin xarakteristika va o'lchov savollarini kuzatib bo'lmay-

digan kontinuumda tartiblashgan deb hisoblaydi. Shuning uchun uning asosiy vazifasi shaxsning o'sha kontinuumdagi joyini aniqlashdir.

Hozirgi vaqtda IRT modeli ko'p mamlakatlarda, jumladan AQShda GRE, GMAT kabi yuqori darajali testlar natijasini tahlil qilishda hamda bu model bilan aniqlangan test topshiriqlarining xarakteristikalar xususiyati bo'yicha test topshiriqlari bazasiga joylashtirishda foydalaniladi.

IRT parametrlari tajriba (eksperiment) va modellarni moslashtirish orqali topiladi. Buni amalga oshirish uchun bir necha xil usullar ishlab chiqilgandir. IRT modellarida model parametrlarini aniqlash muhim rol o'ynaydi. Rash modeli bo'yicha bu parametrlar – yashirin qobiliyat va elementlar qiyinlik darajasidir [8-9].

Rash modeliga ko'ra, dixotomik elementlarga individual javoblar ehtimoli shaxsning qobiliyat va element qiyinlik darajalari 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, θ_s -qobiliyat o'zgaruvchisi, b_i -topshiriq qiyinlik darajasi, e -natural logarifm asosi ($e=2,7182818 \dots$).

Tarix fanidan test sinovlari natijalarining Rash modeli bo'yicha tahlilini maxsus dastur asosida amalga oshirish uchun ishlab chiqilgan turli xil dasturiy paketlardan foydalanamiz. Qiyinlik darajasi b ni aniqlashda biz ltm dasturiy

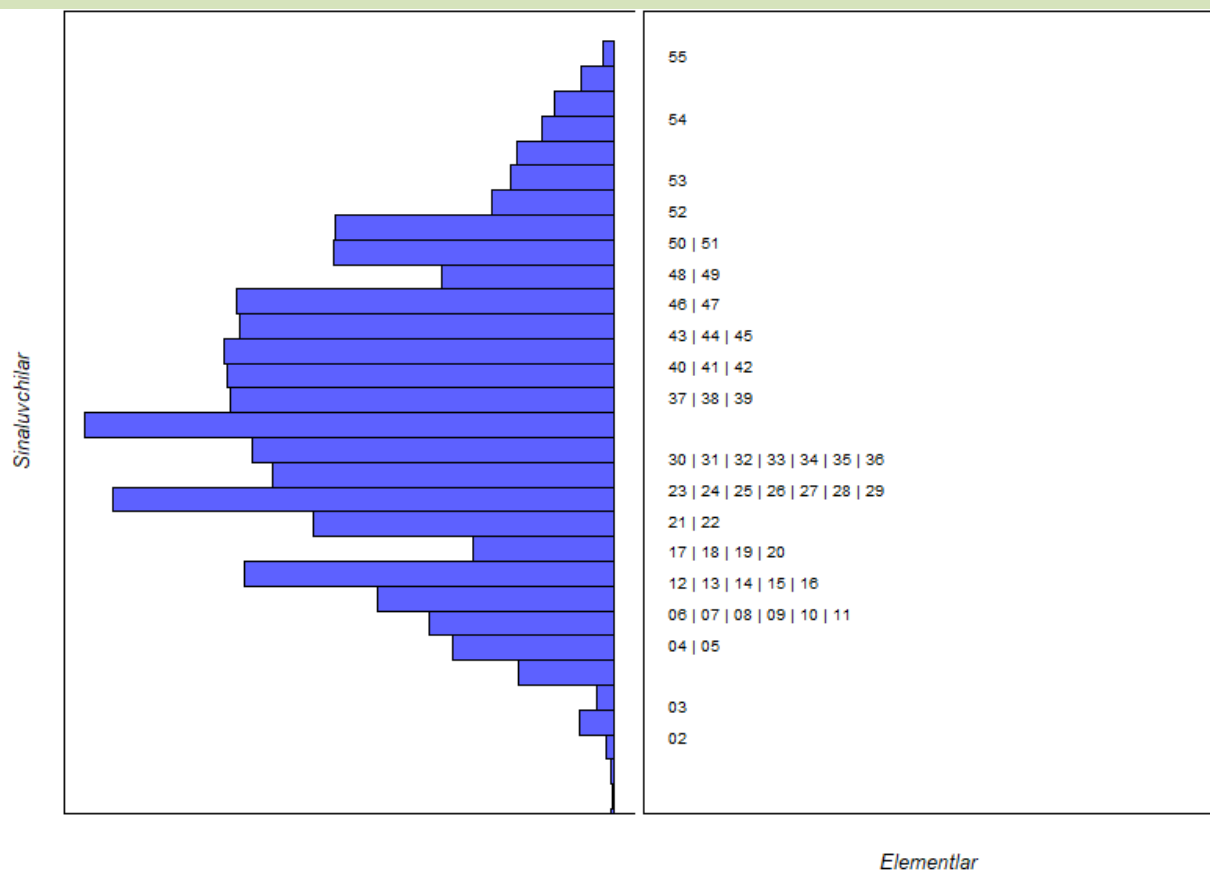
paketidan foydalanamiz [10], chunki Rash modeli uchun bu dasturiy paket yordamida tajribaning (test natijalari) modelga qanchalik mosligini hisoblash mumkin bo'ladi.

Rash modeli asosida aniqlangan qiyinlik darajalarini sinaluvchilar qobiliyatlariga qanchalik mosligini Rayt xaritasi yordamida tahlil qilish mumkin.

5. Rayt xaritasi

Rash modeli asosida aniqlangan qiyinlik darajalarini sinaluvchilar qobiliyatiga qanchalik mosligini Rayt xaritasi yordamida tahlil qilish mumkin. Rayt xaritasi – test topshiriqlarining qiyinlik darajalari va

sinaluvchilarning qobiliyat darajalarining o'zaro mos kelishini aniqlovchi diagrammadir [11]. Tarix fanidan o'tkazilgan test sinovlari natijasi asosida chizilgan Rayt xaritasi 1-rasmda keltirilgan.



1-rasm. Test natijalari asosida aniqlangan qobiliyat va qiyinlik darajalarining mosligi (Rayt xaritasi)

Rasmdan qobiliyat darajalari $(-3,48 : 2,88)$ logit birligi orasida, test topshiriqlari qiyinlik darajalari esa $(-3,72 : 2,83)$ oraliqda taqsimlanganligi, 1-o'rinda turgan test topshirig'i (ID raqami – T5 bo'lgan yopiq turdagi test topshirig'i) esa test topshiriqlarining qiyinlik darajalari bo'yicha

$(-3:3)$ logit birligi oralig'idan tashqarida ekanligi ko'rinadi.

Taqsimotdagi bo'sh joylarga e'tibor berib, bir xil qiyinlikdagi test topshiriqlari o'rniga bo'sh joylarga mos keladigan qiyinlik darajasidagi test topshiriqlaridan qo'yish taqsimotni yanada yaxshilashi mumkin.

6. Rash modeli asosida aniqlangan qiyinlik darajalari, ichki (infit) va tashqi (outfit) moslik statistikasi

Qobiliyatlar va qiyinlik darajalarini baholash uchun turli xil usullar ishlatiladi: maksimal haqiqatga o'xshashlik, birgalikda haqiqatga o'xshashlik, shartli maksimal haqiqatga o'xshashlik, marginal maksimal haqiqatga o'xshashlik va hokazo usullar

ishlatiladi. Masalan, birgalikdagi maksimal haqiqatga o'xshashlik usullaridagi muammolar test topshiriqlari va qobiliyat parametrlarini birgalikda baholash tufayli kelib chiqadi. Agar parametrlar qobiliyatning ishtirokisiz baholanadigan bo'lsa, bunday muam-

molar o'z-o'zidan yo'qoladi. Buning uchun biz ko'rib chiqayotgan tanlanma to'plamni bosh to'plamdan ixtiyoriy tanlab olingan deb hisoblab, qobiliyat darajasi parametrlari bo'yicha integrallab, ularni haqiqatga o'xshashlik funksiyasiga bog'liqlikdan ozod qilishimiz mumkin va hosil bo'lgan haqiqatga o'xshashlik funksiyasining maksimumini topib, qobiliyatni baholashimiz mumkin [7,12-13].

Yuqoridagi baholash jarayoni Bok va Liberman [14] hamda Bok va Aitkinlar [15] tomonidan kiritilgan. Marginal maksimal haqiqatga o'xshashlik usuli raqamli hisob-kitoblar uchun qiyinroq, chunki u raqamli integrallashni o'z ichiga oladi, lekin hozirgi zamonaviy kompyuterlar uchun bu muammo tug'dirmaydi, shuning uchun zamonaviy test nazariyasi bo'yicha kompyuter dasturlarida qobiliyat va element parametrlarini baholash uchun ushbu usulni qo'llash tobora keng tarqalmoqda.

Ichki moslik (infit) statistikasi qobiliyatni mo'ljalga olingan qobiliyat darajalarining javoblar namunasiga va, aksincha, javoblar namunasining qobiliyat darajasiga sezgir bo'ladi. Ichki moslik aniqlangan mezonlardan katta bo'lsa, Gutman namunasi [16] to'g'risida, mezonlardan kichik bo'lsa, muqobil o'quv dasturi bilan bog'liq ma'lumotlar beradi.

Tashqi moslik (outfit) statistikasi an'anaviy χ^2 usuliga asoslangan. U test topshiriqlari qiyinlik darajalari qobiliyat darajasidan va, aksincha, qobiliyat darajasi qiyinlik darajasidan tashqari-daligini ko'rsatadi. Masalan, tashqi moslik uning uchun aniqlangan mezonlardan katta bo'lsa, maqsadga muvofiq bo'lmagan javoblarni, kichik bo'lsa, tasodifiy yoki ehtiyotsizlik bilan berilgan javoblarni bildiradi.

Tashqi va ichki moslik statistikasi quyidagi formulalar bilan hisoblanadi [17]:

$$Outfit_i = \frac{\sum_s (X_{si} - E(X_{si}))^2}{n_i}, Infit_i = \frac{\sum_s (X_{si} - E(X_{si}))^2}{\sum_s Var(X_{si})}, \quad (4)$$

$$Outfit_s = \frac{\sum_i (X_{si} - E(X_{si}))^2}{n_s}, Infit_i = \frac{\sum_i (X_{si} - E(X_{si}))^2}{\sum_{si} Var(X_{si})}, \quad (5)$$

bu yerda $E(X_{si})$ - X_{si} ning matematik kutilishi, $Var(X_{si})$ - dispersiya, n_i - test topshiriqlari soni, n_s - sinaluvchilar soni.

Ko'p tadqiqotlarda tashqi va ichki moslik darajasining quyi va yuqori chegarasi mos ravishda 0,7 va 1,3 oralig'ida qilib olinadi. Linacre [18]

tahlillaridan so'ng bu chegaralarni 0,5 va 1,5 qilib olish ham mumkinligini ko'rsatdi. [19] havolada tashqi va ichki moslik uchun mos ravishda $1 \pm \frac{6}{\sqrt{n}}$ va $1 \pm \frac{2}{\sqrt{n}}$ dan foydalanish tavsiya qilinadi. Shunindek, [20] havolada ichki va tashqi mosliklar test topshiriqlari soni va qiyinlik darajasiga bo'g'liqligi ko'rsatilgan.

4-jadvalda milliy test tizimida fizika fanidan o'tkazilgan test sinov-

larida ishlatilga test topshiriqlarining Rash modeli bilan olingan qiyinlik darajalari hamda (4) formula bilan hisoblangan ichki va tashqi moslik statistikalari keltirilgan. Jadvaldagi ma'lumotlar tartib raqami test topshiriqlarining qiyinlik darajasi oshib borishi tartibida qo'yilgan (1-test topshirig'i eng oson, ..., 55-test topshirig'i eng qiyin).

4-jadval

Test topshiriqlarining qiyinlik darajalari, ichki (*infit*) va tashqi (*outfit*) moslik statistikalari

<i>Nº</i>	<i>ID</i>	β	<i>Infit</i>	<i>Outfit</i>
1	T5	-3,715	0,967	0,727
2	T9	-2,515	0,910	0,676
3	T12	-2,287	0,886	0,665
4	T19	-1,939	0,879	0,742
5	T6	-1,805	0,996	0,939
6	T18	-1,722	0,814	0,641
7	T10	-1,696	0,897	0,781
8	T33	-1,674	0,917	0,872
9	T20	-1,642	0,852	0,681
10	T4	-1,597	0,893	0,751
11	O37A	-1,574	0,851	0,794
12	T29	-1,376	0,970	0,843
13	T30	-1,283	0,902	0,800
14	T11	-1,277	1,013	0,941
15	T17	-1,273	0,964	0,908
16	T16	-1,265	0,911	0,799
17	T2	-1,193	0,976	0,966

18	T8	-1,155	0,911	0,819
19	T1	-1,122	1,027	1,007
20	T13	-1,046	0,968	0,898
21	T21	-0,900	0,938	0,880
22	T31	-0,819	0,956	0,902
23	T23	-0,734	1,079	1,078
24	T7	-0,709	0,948	0,883
25	O44A	-0,688	0,876	0,804
26	T35	-0,688	0,891	0,818
27	T32	-0,637	1,069	1,050
28	T27	-0,636	1,105	1,153
29	O40A	-0,608	0,824	0,766
30	T3	-0,497	1,067	1,060
31	O40B	-0,461	0,807	0,757
32	T22	-0,418	0,923	0,867
33	O45B	-0,365	0,879	0,825
34	T26	-0,355	1,139	1,173
35	O42A	-0,299	0,778	0,714
36	T34	-0,294	1,093	1,096
37	T24	0,120	1,005	1,014
38	O45A	0,195	0,980	0,958
39	T25	0,220	1,090	1,129
40	O43B	0,392	0,999	0,970
41	T28	0,478	1,262	1,367
42	T14	0,499	1,022	1,085
43	T15	0,541	1,242	1,391
44	O42B	0,666	0,850	0,793
45	O41A	0,691	0,857	0,800
46	O41B	0,756	0,935	0,877
47	O43A	0,941	0,821	0,747
48	O36B	1,009	0,919	0,864

49	038B	1,084	1,057	1,028
50	044B	1,275	0,870	0,785
51	039A	1,343	0,898	0,808
52	036A	1,543	0,989	1,066
53	039B	1,836	0,930	0,827
54	037B	2,282	0,978	0,890
55	038A	2,832	0,938	0,743

4-jadvaldan ichki moslik 0,7-1,3 oralig'idan tashqariga chiqmaganligini, bu esa variantda Gutman namunasi va dastur bilan bog'liq muammolar mavjud emasligini ko'rsatadi.

4-jadvalga ko'ra qiyinlik darajasi bo'yicha 41- va 43-o'rinda turgan test topshiriqlarining tashqi mosligi 1,3 dan katta, 2-, 3-, 6- va 9-test topshiriqlarining tashqi mosligi 0,7 dan kichik ekanligini, shuningdek, ular taqsimot chekkalarida turganini ko'rish mumkin. Moslik statistikasi mezonlari doirasida bo'lmagan test topshiriqlari asosan qiyinlik darajasi past va yuqori bo'lgan test topshiriqlariga to'g'ri kelishini ko'rish mumkin. Qiyinlik darajasi

bo'yicha test topshiriqlariga berilgan javoblarning deyarli barchasi 0,7-1,3 oralig'ida joylashganligi javoblar mutanosib ekanligini va javoblar tasodifiy va ehtiyotsizlik tufayli bo'lmaganligini bildiradi. Tahlil natijalariga ko'ra, test topshiriqlarining talab qilingan me'yorga juda yaqinligi, ushbu test topshiriqlarining to'g'ri tuzilganligi hamda tanlangan variantning tanlangan to'plamga mos ekanligidan dalolat beradi.

Test topshiriqlarining ichki va tashqi mosligiga o'xshash jarayonni amalga oshirib, qobiliyatning ichki va tashqi moslik statistikasini tahlil qilish mumkin.

7. Element xarakteristikasi, element va test ma'lumoti chiziqlari

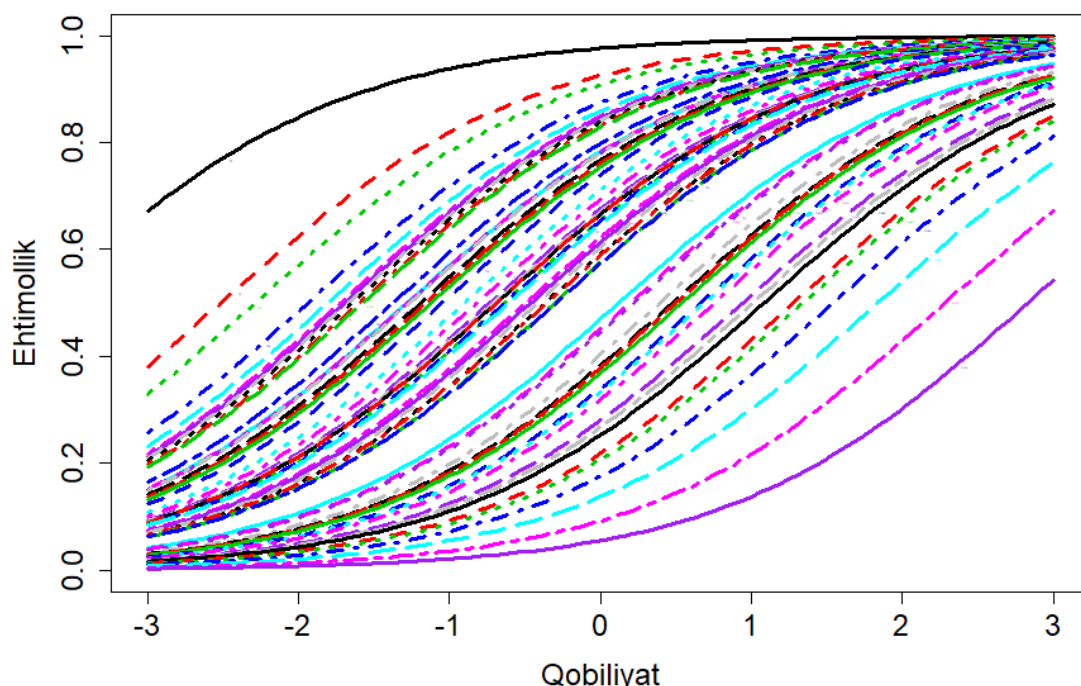
O'tkazilgan test sinovi natijalari asosida olingan element xarakteristikasi, element va test ma'lumoti chiziqlarini ko'rib chiqamiz. 4-jadvaldan, Rash modeli bo'yicha 038A ID raqamli test topshirig'i eng qiyin, T5 ID raqamli test topshirig'i esa eng oson ekanligini ko'rish mumkin. Bu 2-rasmdagi element xarakteristikasi

chiziqlarining (EXCh) o'zaro joylashuvidan ham yaqqol ko'rinadi (038A ID test topshirig'i pastdan birinchi uzluksiz chiziq, T5 ID raqamli test topshirig'i yuqoridan birinchi uzluksiz chiziq). 2- rasmdan shu narsa ko'rinib turibdiki, har xil qobiliyatli test topshiruvchilarning qobiliyatini baho-

lash uchun, albatta, har xil qiyinlikdagi test topshiriqlari bo'lishi kerak.

Bu holat zamonaviy test nazariyasidagi tenglashtirish va kalibrovkalash orqali yechiladi. Amaliyotda bu xalqaro tajribada ham [21] respublikamizda chet tillari bo'yicha "multilevel" testlari natijalarini baholashda ham ishlatilmoqda [22]. Obyektiv baholashda xatoliklarni kamaytirish tartibli, kalibrovkalangan

test topshiriqlaridan iborat bazalarga o'tish uchun shu usulda baholash tavsiya qilinadi. Bu test topshiriqlarini qiyinlik darajalri va qobiliyat darajalari bir xil shkalada bo'ladi va ular test natijalarini hisoblash jarayonidayoq aniqlanadi. Shuningdek, bunday baholashda statistik tahlillar izchilligi ta'minlanadi. Bu esa test natijalarining ichonchililigi va validligi haqida to'g'ri xulosalar chiqarish imkonini beradi.



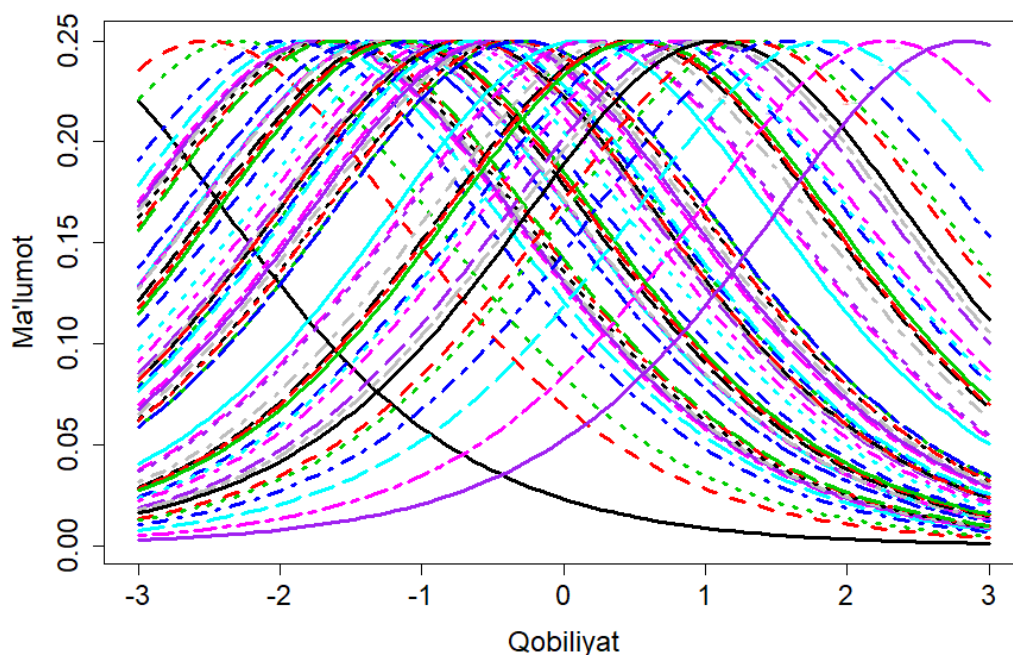
2-rasm. Element xarakteristikasi chiziqlari

3-rasmda tasvirlangan har bir elementning (test topshirig'ining) ma'lumot chiziqlari (EMCh)ni tahlil qilib, bu to'g'risida ko'proq ma'lumot olish mumkin. EMCh kengligi har bir element har xil qobiliyatli test topshiruvchilar haqida qanchalik ko'p

ma'lumot berishi mumkinligini ko'rsatadi. EMCh balandligi esa ma'lumot miqdorini bildiradi. Rash modelida elementning turli qobiliyatni ajratish xususiyati (diskriminativligi) bir xil deb qaraladi. Demak, barcha test topshiriqlarining turli qobiliyatlarni

ajratish xususiyati bir xil deb qaralganda, test topshiriqlarining test topshiruvchilar haqida beradigan ma'lumot miqdori bir xil, lekin ular turli xil qobiliyat oraliq'idagi ma'lumotlardir. Normal taqsimotga ko'ra sinaluvchilar qobiliyat darajalarining 99,7 foizi ($-3:3$) oraliq'ida bo'ladi. Shuning uchun qiyinlik darajalari ($-3:3$) oraliqdan tashqarida bo'lgan test topshiriqlari bu oraliqda ko'p ma'lumot bermaydi. Masalan, O38A ID raqamli test topshirig'i qobiliyatli test topshiruvchilar haqida ko'p ma'lumot beradi, T5 ID raqamli test topshirig'i esa

qobiliyati pastroq test topshiruvchilar haqida ko'proq ma'lumot beradi: umumiy ma'lumot miqdorini 1 deb olsak (100 foiz) O38A ID raqamli test topshirig'idan ($0:3$) qobiliyat oraliq'idagilar uchun beradigan ma'lumot miqdori 0,49 (49 foiz), T5 ID raqamli test topshirig'ining ma'lumot miqdori esa bu oraliqda 0,02 (2 foiz) ga teng. Aksincha, ($-3:0$) oraliqda esa T5 ID raqamli test topshirig'i beradigan ma'lumot miqdori 0,30 (30 foiz), O38A ID raqamli test topshirig'iniki 0,05 (5 foiz)ga teng.



3-rasm. Element ma'lumoti chiziqlari

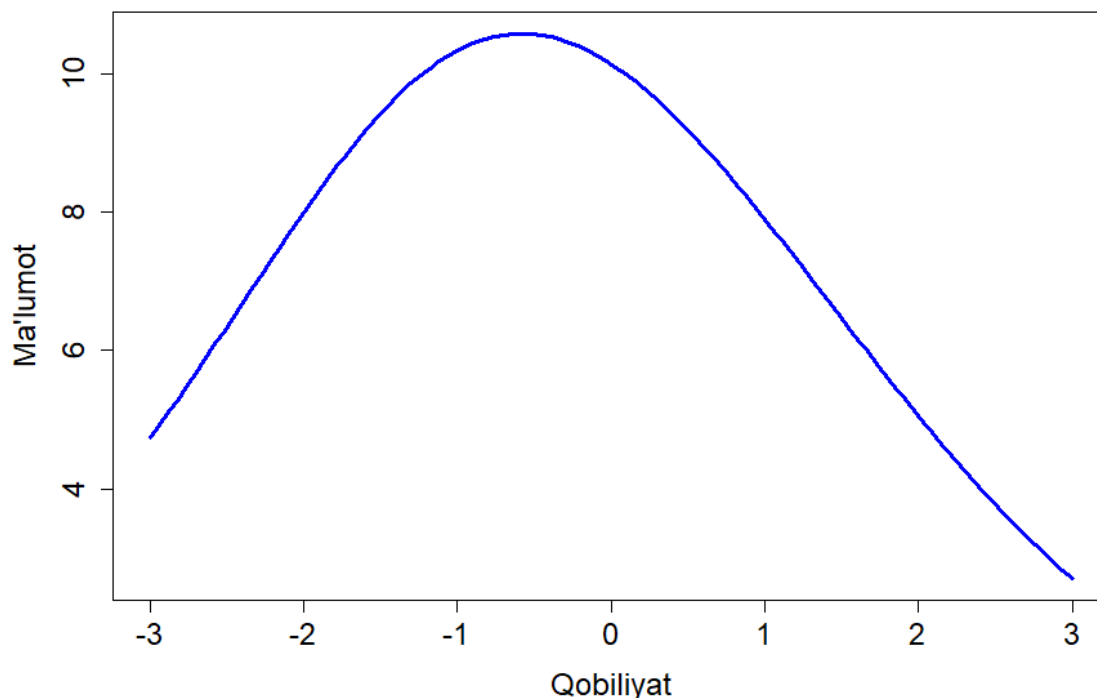
4-rasmda test ma'lumoti chizig'i (TMCh) keltirilgan. Testning umumiy ma'lumot miqdori 55 ga teng bo'lib, shundan ($-3:3$) oraliq'idagi qobiliyatga ega bo'lganlar uchun ma'lumot miqdori 45,57 (82,88 foiz) ga teng. ($-3:0$) va ($0:3$) oraliqlardagi qobi-

liyatga ega bo'lganlar miqdori mos ravishda 26,16 (47,58 foiz) va 19,41 (35,30 foiz) ga mos keladi. Ma'lumot chizig'i cho'qqisining nolga nisbatan chap tomonga surilganligi ushbu test varianti qobiliyat darajasi past bo'lgan

sinaluvchilar to'g'risida ko'proq ma'lumot berishini anglatadi.

Qiyinlik darajasi juda past bo'lgan test topshiriqlarining o'rniga –3 logit birligi atrofidagi test topshiriqlaridan

kiritib, yuqori va past qobiliyat darajalaridan olinadigan ma'lumot miqdori orasidagi tafovutni yanada kamaytirish mumkin.



4-rasm. Test ma'lumoti chizig'i

Xulosa

Tarix fanidan Milliy sertifikat uchun o'tkazilgan test sinovida foydalanilgan test topshiriqlarining ishonchlilik koeffitsiyenti (Kronbax alfa koeffitsiyenti) 0,92 ga teng ekanligi, ishonchlilik koeffitsiyentiga qo'yilgan mezonlarga asosan (2-jadval), test topshiriqlarining o'zaro ichki muvofiqligi "a'lo" darajada ekanligi aniqlandi.

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.

Element-umumiy ball korrelyatsiyasi qiymatlari 0,25 dan kichik bo'lgan test topshiriqlarini o'rganib chiqib, kerakli o'zgarishlar qilish, lozim bo'lsa, variantdan chiqarib tashlash tavsiya etildi. Biserial korrelyatsiya koeffitsiyenti qiymatlari 0,5 dan kichik bo'lgan test topshiriqlari esa o'rganib chiqilib, qiyinlik darajalarini hisobga olgan holda, kerakli o'zgarishlar qilish tavsiya qilingan.

Natijalarini Rash modeli bilan baholash xom ball bilan baholashga nisbatan standartlik, validlik va

ichonchlilikni aniqroq talqin qilish imkonini beradi.

Rash modeli bilan hisoblangan test topshiriqlari qiyinlik darajalari va qobiliyat darajalarining o'zaro mosligi Rayt xaritasi bilan tahlil qilinishi mumkin. Rayt xaritasidan aniqlangan qiyinlik darajalari bilan solishtirish orqali bir xil qiyinlikdagi test topshiriqlari o'rniga bo'sh joylarga mos keladigan qiyinlik darajasidagi test topshiriqlaridan qo'yish taqsimotni yanada yaxshilash uchun imkon yaratadi.

Test topshiriqlarini kalibrovkalgan test bazasiga kiritish uchun maqolada keltirilgan statistik usullar bilan tadqiq qilinishi lozim.

ADABIYOTLAR

1. A.R. Sattiyev, M. Dj. Ermamatov. Matematika fanidan milliy sertifikat uchun o'tkazilgan test sinovlari natijalari tahlili// Axborotnoma, (2023), №2, 35-55-betlar.
2. Spearman, C. (1904a). "General intelligence," objectively determined and measured. American Journal of Psychology, 15, 201-293.
3. Spearman, C. (1904b). The proof and measurement of association between two things. American Journal of Psychology, 15, 72-101.
4. Spearman, C. 1907. Demonstration of formulae for true measurement of correlation. Am. J. of Psychology. 18, 160-169.
5. Spearman, C. 1910. Correlation from faulty data, British J. of Psychology. 3, 271-295.
6. [https://ru.wikipedia.org/wiki/Alfa_Kronbaxa#:~:text=Koeffitsent alfa Kronbaxa,i dlya proverki ix nadyojnosti](https://ru.wikipedia.org/wiki/Alfa_Kronbaxa#:~:text=Koeffitsent%20alfa%20Kronbaxa,i%20dlya%20proverki%20ix%20nadyojnosti).
7. Gunter Maris, Timo Bechger, Jesse Koops and Ivailo Partchev, Data Management and Analysis of Tests, 2022, p. 1-49.
8. Hambleton, R.K., Swaminathan, H., and Rogers, H.J. (1991), Fundamentals of item response theory. Newbury Park, CA: Sage.
9. Ivailo Partchev (2004), A visual guide to item response theory, Friedrich-Schiller-Universitat Jena.
10. Dimitris Rizopoulos (2006). ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses, Journal of Statistical Software, 17 (5), 1-25.
11. B.D. Wright and M.H. Stone, Best Test Design (MESA Press, Chicago, 1979).
12. M. Dj. Ermamatov, A. R. Sattiyev, A.B. Normurodov, Z. O. Olimbekov, A. A. Baratov. Fizika fanidan o'tkazilgan test sinovi natijalari: rayt xaritasi, ichki va tashqi moslik statistikalari, rash modeli bilan moslik// Axborotnoma, №1, 4-62, (2023).
13. Dimitris Rizopoulos, ltm: An R package for Latent Variable Modelling and Item, Response Theory Analyses, Journal of Statistical Software, v.17, p. 1-15, 2006.
14. Bock R. D. and Liebermann, M. (1970), Fitting a response curve model for dichotomiously scored items, Psychometrika, 35, 179-198.

15. Bock R. D. and Aitkin, M. (1981), *Marginal maximum likelihood estimation of item parameters*, Psychometrika 46, 443–459.
16. Guttman, L. (1950). The basis for scalogram analysis. In Stouffer et al. Measurement and Prediction. The American Soldier Vol. IV. New York: Wiley.
17. Marianne Mueller, Item fit statistics for Rasch analysis: can we trust them?, Journal of Statistical Distributions and Applications (2020) 7:5.
18. Linacre [Linacre, M.: Teaching Rasch measurement. Trans. Rasch Meas. 31, 1630–1631 (2017).
19. Smith, R., Schumacker, R., Bush, M.: Using item mean squares to evaluate fit to the Rasch model. J. Outcome Meas. 2, 66–78 (1998).
20. Wang, W., Chen, C.: Item parameter recovery, standard error estimates, and fit statistics of the winsteps program for the family of Rasch models. Educ. Psychol. Meas. 65, 376–404 (2005).
21. Wobbe Zijlstra, ALTE-accreditation Uzbekistan Psychometrics-Online Seminar, Tashkent, 2022-June.
22. A. Abbosov, Analysis of multilevel English language proficiency tests (2023), Axborotnoma, p. 79-91.

ANALYSIS OF RESULTS OF TEST FOR NATIONAL CERTIFICATE ON HISTORY WITHIN CLASSICAL TEST THEORY AND RASH MODEL

A.R. Sattiev, M.Dj. Ermamatov

Scientific-study Practical Center under the Agency for Assessment of Knowledge and Competences, Tashkent 100084, Bogishamol st. 12.

Abstract. In this paper the results of the test on history administered for the first time within the Republic of Uzbekistan for the National certificate are analyzed within classical test theory and Rasch model. Descriptive statistics and element-item correlations are discussed. Difficulties of the items used in the test form are analyzed. Using difficulties and abilities estimated by Rasch model Wrightmap of the results are obtained and discussed compatibility of the item difficulties and abilities, infit and outfit statistics, element characteristics, element and test information curves are discussed.

Keywords: Test items, Cronbach's alpha, validity, difficulty, Rasch model, Wright-map, ability.