O‘ZBEKISTON RESPUBLIKASI OLIY TA’LIM, FAN VA INNOVATSIYALAR VAZIRLIGI HUZURIDAGI BILIM VA MALAKALARNI BAHOLASH AGENTLIGI

AXBOROTNOMA

BULLETIN

3

2023

Toshkent

“AXBOROTNOMA”
O‘zbekiston Respublikasi Oliy ta’lim, fan va innovatsiyalar vazirligi huzuridagi Bilim va malakalarni baholash agentligi ilmiy-uslubiy jurnali.
Bir yilda 4 marta chiqadi.
O‘zbekiston Matbuot va axborot agentligida 2007-yil 19-aprelda qayta ro‘yxatdan o‘tkazilgan.
Guvohnoma № 0247

Ta’sischi:
Bilim va malakalarni baholash agentligi

Bosishga ruxsat etildi: 24.10.2023
Shartli bosma tabog‘i: 5,4
Nashriyot hisob tabog‘i: 5,3
Adadi 100. Buyurtma № 3 - A - 23

Bilim va malakalarni baholash agentligi bosmaxonasida chop etildi.

Nashrga tayyorlovchilar:
A. Baratov
Muharrir:
Z.Shakarov
Kompyuterda sahifalovchi:
Y. To‘raev
Tahririyat manzili:
100084, Toshkent shahri,
Bog‘ishamol ko‘chasi, 12-uy.

MUNDARIJA

CONTENTS

KIRISH

“Axborotnoma” ilmiy-uslubiy jurnalining ushbu sonida pedagogik o‘lchovlar bo‘yicha ilmiy-uslubiy tadqiqotlar olib boruvchi mutaxassislarning ilmiy ishlari natijalari bo‘yicha beshta maqola berilgan.

Birinchi maqolada biologiya fanidan umumiy o‘rta ta’lim maktablarining 9-sinf bitiruvchi o‘quvchilaridan ilmiy tadqiqot uchun olingan test sinovi natijalari R dasturining Dexter to‘plamida hisoblashlarga asoslangan holda Rash modeli bilan tahlil qilindi va model bilan moslik o‘rganildi.

Ikkinchi maqolada baholovchilarni tayyorlashning ahamiyati ko‘rib chiqilgan va ularning aniqligi va ishonchliligini oshirish uchun samarali strategiyalar o‘rganilgan. Baholovchilar duch keladigan qiyinchiliklarni, samarali baholash uchun zarur bo‘lgan ko‘nikmalarni va ularning ish faoliyatini yaxshilash uchun qo‘llaniladigan o‘qitish usullari muhokama qilingan.

Uchinchi maqolada umumiy elementli test variantlari bilan qobiliyatni baholashning obyektivlik uchun muhimligi 9-sinf o‘quvchilaridan biologiya fani bo‘yicha olingan test sinovlarining klassik test nazariyasi va Rash modeli bo‘ycha olingan qiyinlik darajalarini taqqoslash va test xarakteristikasi chiziqlari ko‘rsatilgan.

To‘rtinchi 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 muhokama qilingan. Test sinovida ishlatilgan variantdagi test topshiriqlarining qiyinlik darajasi tahlil qilingan. Rash modeli bilan hisoblangan qobiliyat va qiyinlik darajalaridan foydalanib Rayt xaritasi olingan va u bilan qobiliyat va qiyinlik darajalari mosligi, ichki va tashqi moslik statistikalari, element xarakteristikasi, element va test ma‘lumoti chiziqlari muhokama qilingan.

Beshinchi maqolada umumiy o‘rta ta’lim maktablari, ixtisoslashtirilgan davlat umumta’lim maktablari va akademik litseylarning 11-sinf bitiruvchi o‘quvchilaridan ilmiy tadqiqot maqsadida fizika fanidan olingan test sinovi natijalarining klassik test nazariyasi va Rash modeli asosidagi tahlillari berilgan.

Jurnal ta’lim sohasida pedagogik o‘lchovlar bo‘yicha faoliyat olib borayotgan barcha mutaxassislar, pedagoglar, doktarantlar, shuningdek, talabalarga mo‘ljallangan.

Jurnal pedagogika fanlari bo‘yicha O‘zbekiston Respublikasi Oliy attesttatsiya komissiyasining dissertatsiyalar asosiy ilmiy natijalarini chop etish tavsiya etilgan ilmiy nashrlar ro‘yxatiga 2018-yildan kiritilgan.

RASH MODELI BILAN MOSLIK: BIOLOGIYA FANIDAN O‘TKAZILGAN TEST SINOVI NATIJALARI

A.B. Normurodov, M.Dj. Ermamatov, I.A. Boyxonov

Bilim va malakalarni baholash agentligi huzuridagi Ilmiy-o‘quv amaliy markazi, 100084, Toshkent sh., Bog‘ishamol k., 12

Qisqacha mazmuni. Ushbu maqolada biologiya fanidan umumiy o‘rta ta’lim maktablarining 9-sinf bitiruvchi o‘quvchilaridan ilmiy tadqiqot uchun olingan test sinovi natijalari R dasturining Dexter to‘plamida hisoblashlarga asoslangan holda Rash modeli bilan tahlil qilindi va model bilan moslik o‘rganildi. Rash modeli ehtimollarga asoslangan model bo‘lib, qobiliyat darajasi va test topshiriqlari qiyinlik darajalarining o‘zaro ta‘sirini tavsiflaydi hamda o‘lchov jarayonlarini paralellashtiradi. Bunda qobiliyat va test topshiriqlari qiyinlik darajalarini hisoblash chiziqli o‘lchash orqali amalga oshiriladi. Bir o‘lchovli konstruktlar uchun tuzilgan test topshiriqlarining qiyinlik darajalari tanlanma guruhlarga va qobiliyat darajalari test variantiga bog‘liq bo‘lmasligi ta‘minlanadi. Test topshiriqlari xususiyatlarini Rash modeliga moslashtirishni tahlil qilish muammoli elementlarni aniqlash va o‘ziga xos xususiyatlarga ega bo‘lgan qobiliyatlarni ajratish imkonini beradi.

Kalit so‘zlar: Rash modeli, qiyinlik darajalari, qobiliyat darajalari, Rash modeli bilan moslik.

1. Kirish

Bilimlarni baholashda asosiy muammolardan biri bu turli xil sinaluvchilarning qobiliyat darajalarini imkon qadar kichik xatoliklar bilan aniqlashdir. Buning uchun test bazalarida ham turli xil sinaluvchilarning qobiliyat darajalariga mos bo‘lgan turli xil qiyinlik darajadagi test topshiriqlari mavjud bo‘lishi maqsadga muvofiqdir. Bunday darajadagi test topshiriqlari taqsimoti to‘g‘ri bo‘lishi uchun qiyinlik darajasi bo‘yicha kalibrovkalangan test bazalari yaratilishi va bu test bazalarida test topshiriqlarining sifati ham statistik usullar orqali tahlil qilinishi lozimdir. Bu tahlillar uchun Rash modelidan [1-4] keng foydalaniladi. Rash modelida invariantlik va bir o‘lchovlilik xususiyatlari hisobga olingan [5-6].

Bundan tashqari o‘lchashlar obyektivligiga erishish uchun bir o‘lchovli shkalalar talab etiladi, ya’ni aniqlanayotgan xususiyat ajratib olingan holda talqin qilinishi lozim. Bu tarkibiy qismlarni butunlay ajratib olish imkoni yo‘q, shuning uchun bir

o‘lchovlilikni ta‘minlash deganda barcha test topshiriqlari birgalikda ahamiyati katta bo‘lgan qobiliyat darajalarining tarkibiy qismlarini yaxshi ajratib olishi tushunilishi lozim [2].

Bir o‘lchovlilikni ta‘minlash mushkul bo‘lishiga qaramasdan, uni ta‘minlash uchun oldindan tayyorgarlik ko‘rish va bu qanchalik amalga oshirilganini empirik usullar bilan tekshirish imkoni mavjud. Rash modelining muhim xususiyati u shunchaki ma’lumotlarni tahlil qilish uchun statistik usul emas, balki o‘lchovning nimaligini, ta’lim tizimida o‘lchovlarni qanday sifatli amalga oshirish imkoniyatini beradi [7-9]. Har bir test topshiriqlarining Rash modeli bilan standart xatolik doirasida mosligining statistik tahlillarini ko‘rish orqali ularning umumiy ball bilan qiyinlik darajasi bo‘yicha korrelyatsiyasi yoki sifati haqida gapirish mumkin bo‘ladi [10-11].

Ushbu maqolada biologiya fanidan umumiy o‘rta ta’lim maktablarining 9-sinf bitiruvchi o‘quvchilaridan ilmiy tadqiqot uchun olingan test sinovi natijalarining Rash modeli bilan mosligi ko‘rib chiqildi. Hisoblashlar R dasturining Dexter to‘plamida amalga oshirilgan [12]. O‘rta ta’lim maktablarining 9-sinf bitiruvchilari uchun biologiya fanidan akademik litseylarning 179 ta, o‘rta maktablarining 185 ta va o‘quv markazlarining 60 ta, jami 423 ta sinaluvchidan olingan test sinovlarida ishlatilgan test topshiriqlarining natijalari tahlil qilindi.

2. Rash modeli bilan moslik

Biologiya fanidan umumiy o‘rta ta’lim maktablarining 9-sinf bitiruvchi o‘quvchilaridan ilmiy tadqiqot uchun olingan test sinovi natijalarining Rash modeli asosida tahlili orqali Rayt xaritasi, ichki (infit) va tashqi (outfit) moslik statistikalari o‘rganilgan [13]. Test topshiriqlarining qiyinlik darajalari va qobiliyat darajalarining o‘zaro mosligini Rayt xaritasi bilan tahlil qilish va mo‘ljallangan guruh uchun test topshiriqlarini tanlash mumkinligi ko‘rsatilgan.

Hisoblashlarga ko‘ra test sinovlarida qobiliyat darajalari $-\mathbf{2},\mathbf{61}$ va $\mathbf{2},\mathbf{63}$ logit birligi oralig‘ida, test topshiriqlari qiyinlik darajalari esa $-\mathbf{2},\mathbf{63}$ va $\mathbf{2},\mathbf{41}$ logit birligi oralig‘ida taqsimlangan. Ushbu test varianti orqali sinaluvchilardan olinadigan umumiy ma’lumot miqdori 47,99 ga teng bo‘lib, shundan $(-\mathbf{3}:\mathbf{3})$ oralig‘idagi qobiliyatga ega bo‘lganlar uchun ma’lumot miqdori 40,9 (85,3 foiz) ga teng. $(-\mathbf{3}:\mathbf{0})$ va $(\mathbf{0}:\mathbf{3})$ oraliqlardagi qobiliyatga ega bo‘lganlar miqdori esa mos ravishda 23,8 (49,6 foiz) va 17,1 (35,6 foiz) ga to‘g‘ri keladi. Bu natijalar esa ushbu test varianti qobiliyat darajasi

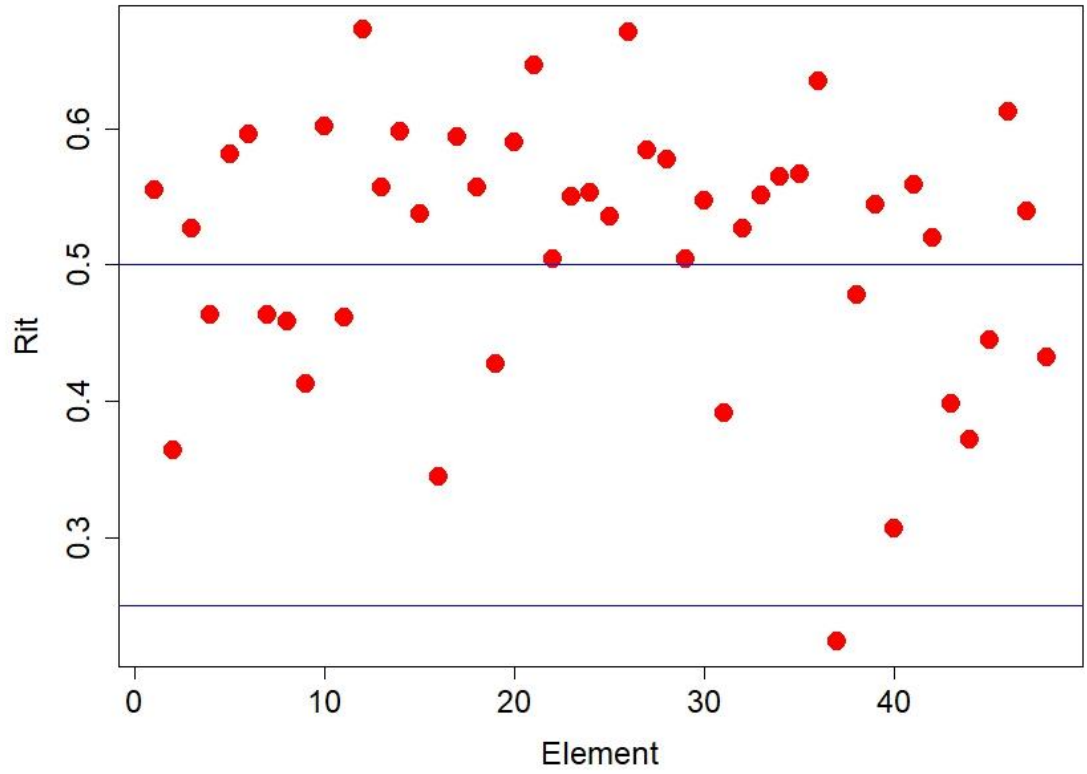
o‘rtachadan past bo‘lgan sinaluvchilar o‘rtachadan yuqori bo‘lgan sinaluvchilarga nisbatan ko‘proq ma‘lumot berishini ko‘rsatadi [13].

Quyidagi 1- va 2-rasmlarda test topshirig‘i balining umumiy ball bilan korrelyatsiyasi (Rit) va test topshirig‘i chiqarib tashlangandagi umumiy ball bilan korrelyatsiyasi (Rir) qiymatlari keltirilgan.

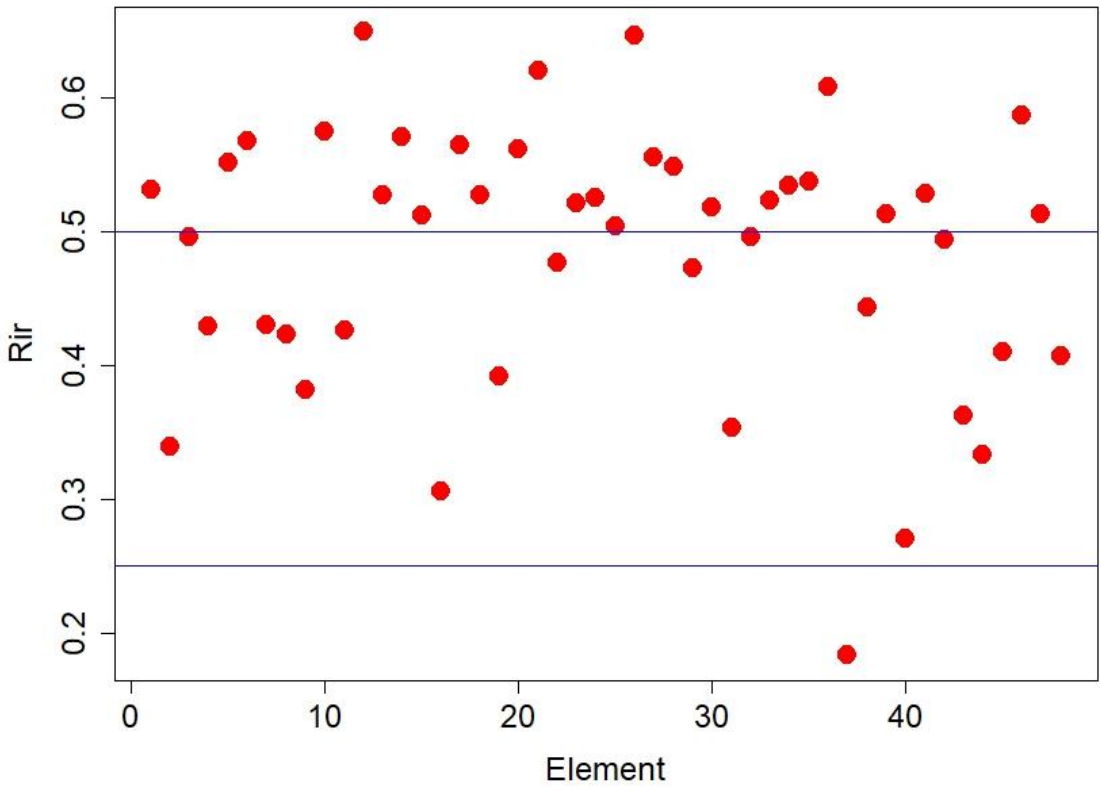
3-rasmda sinaluvchilar 5 ta qobiliyat guruhiga bo‘lingan hamda qalin ko‘k chiziqlar bilan test sinovlaridan olingan natijalar, ingichka qora chiziq bilan kutiladigan qiymatlar esa vertikal standart xatolik chiziqlari bilan birga ko‘rsatilgan. Standart xatolik chegarasidan chiqib ketgan nuqtalar qizil doiralar bilan ko‘rsatilgan. Rasmlarning yuqorisidagi raqamlar test topshiriqlari qiyinlik darajasi bo‘yicha tartiblanganda nechanchi o‘rinda turganligini ko‘rsatadi (1- eng oson, …, 48-eng qiyin). Qiyinlik darajasi bo‘yicha ID raqamlari tartibi 1-jadvalda berilgan.

1- va 2-rasmlardan B0000037 raqamli test topshirig‘i korrelyatsiya koeffitsiyenti eng past 0,25 dan kichik ekanligi ko‘rinadi. Bu test topshirig‘ini qiyinlik darajasi bo‘yicha kalibirovkalangan test bazalariga kiritishdan oldin qayta ko‘rib chiqish tavsiya etiladi. Buni 3-rasmdagi 42 raqamli grafikdan ham ko‘rish mumkin.

Grafikda B0000037 raqamli test topshirig‘i natijalari (qalin ko‘k chiziq) Rash modeli bilan kutiladigan qiymatlar (ingichka qora chiziq) mos tushmayotganligini, vertikal standart xatolik chiziqlaridan ham chiqib ketganligini, ya’ni qizil nuqtalarni kuzatish mumkin. Bu test topshirig‘i faqatgina o‘rtacha qobiliyatdagi guruhlar uchun mos ekanligini aytish mumkin. B0000037 raqamli test topshirig‘ining tashqi (outfit) mosligi belgilangan mezondan katta ekanligini [13] va distraktorlar ham yaxshi ishlamaganligini ko‘rish mumkin [14]. Distraktorlar tahlili haqidagi ma’lumotlar bilan [15-16] havolalarda tanishish mumkin.



1-rasm. Test topshirig‘i balining umumiy ball bilan korrelyatsiyasi (Rit)

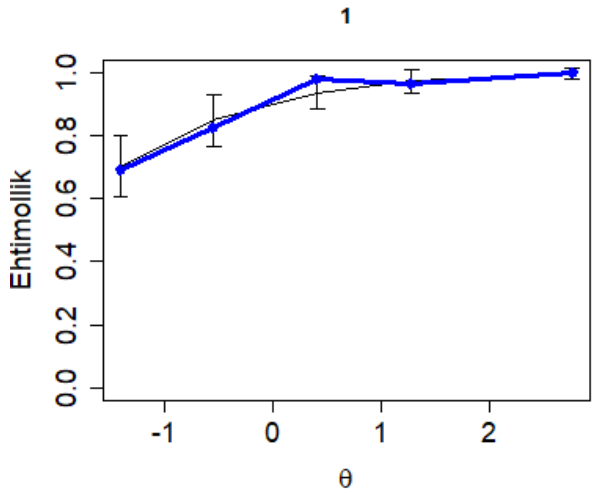


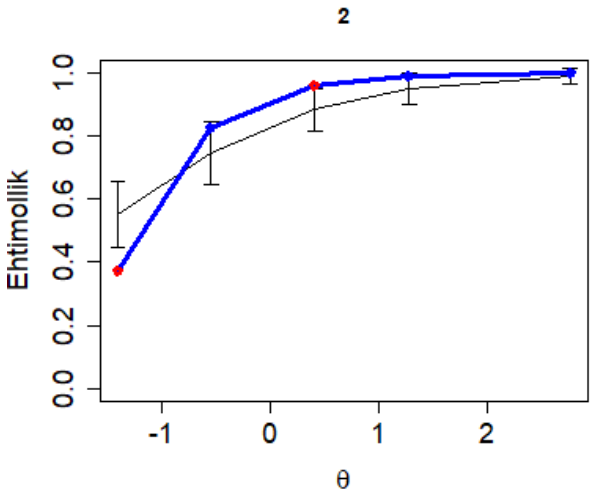
2-rasm. Test topshirig‘i chiqarib tashlangandagi umumiy ball bilan korrelyatsiyasi (Rir)

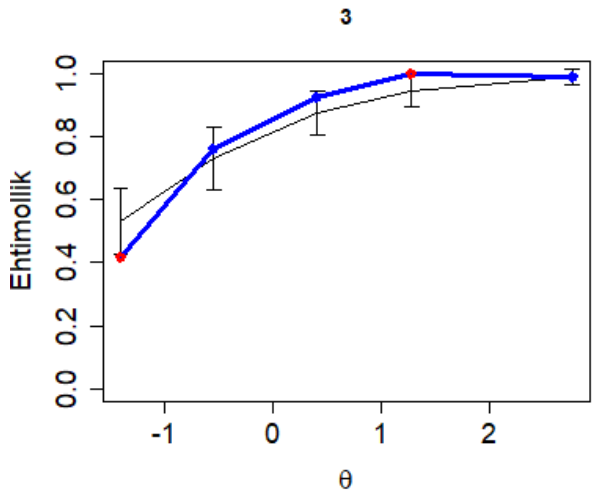
1-jadval

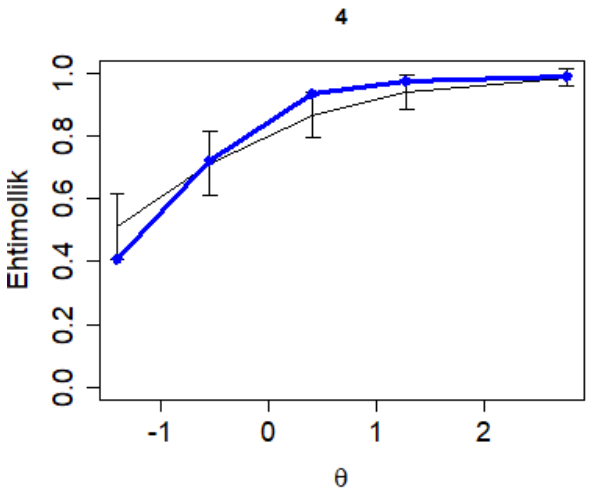
Test topshiriqlarining qiyinlik darajalari va ID raqamlari

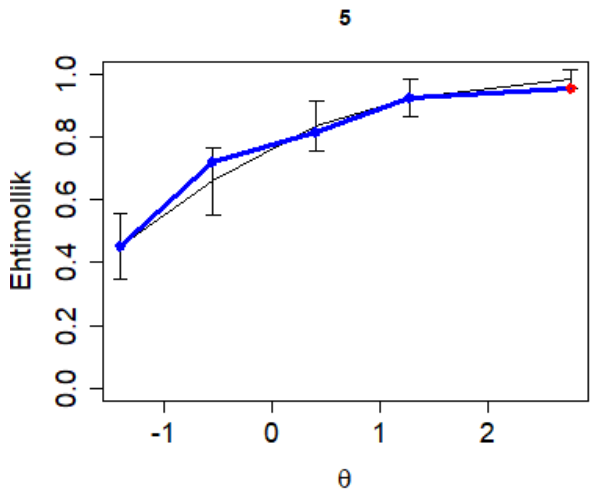
№	ID	**β**	№	ID	**β**
1	B0000002	-2,63	25	B0000020	-0,61
2	B0000001	-1,98	26	B0000034	-0,55
3	B0000015	-1,90	27	B0000008	-0,55
4	B0000042	-1,81	28	B0000041	-0,53
5	B0000009	-1,57	29	B0000044	-0,49
6	B0000024	-1,41	30	B0000045	-0,47
7	B0000033	-1,28	31	B0000004	-0,45
8	B0000023	-1,17	32	B0000026	-0,18
9	B0000010	-1,14	33	B0000017	-0,12
10	B0000030	-1,13	34	B0000021	0,00
11	B0000003	-1,03	35	B0000038	0,02
12	B0000035	-0,95	36	B0000011	0,09
13	B0000012	-0,95	37	B0000005	0,14
14	B0000014	-0,92	38	B0000031	0,20
15	B0000006	-0,92	39	B0000036	0,30
16	B0000029	-0,86	40	B0000016	0,53
17	B0000007	-0,85	41	B0000028	0,72
18	B0000032	-0,85	42	B0000037	0,81
19	B0000018	-0,84	43	B0000043	0,84
20	B0000013	-0,78	44	B0000040	1,18
21	B0000039	-0,73	45	B0000046	1,24
22	B0000027	-0,69	46	B0000047	1,68
23	B0000019	-0,64	47	B0000022	1,76
24	B0000025	-0,62	48	B0000048	2,41

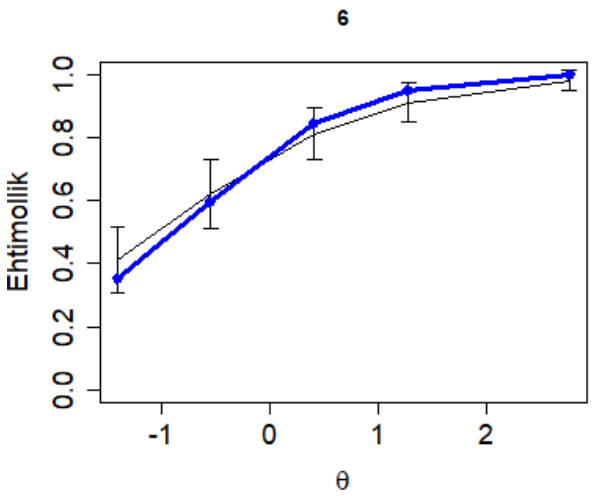

1
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

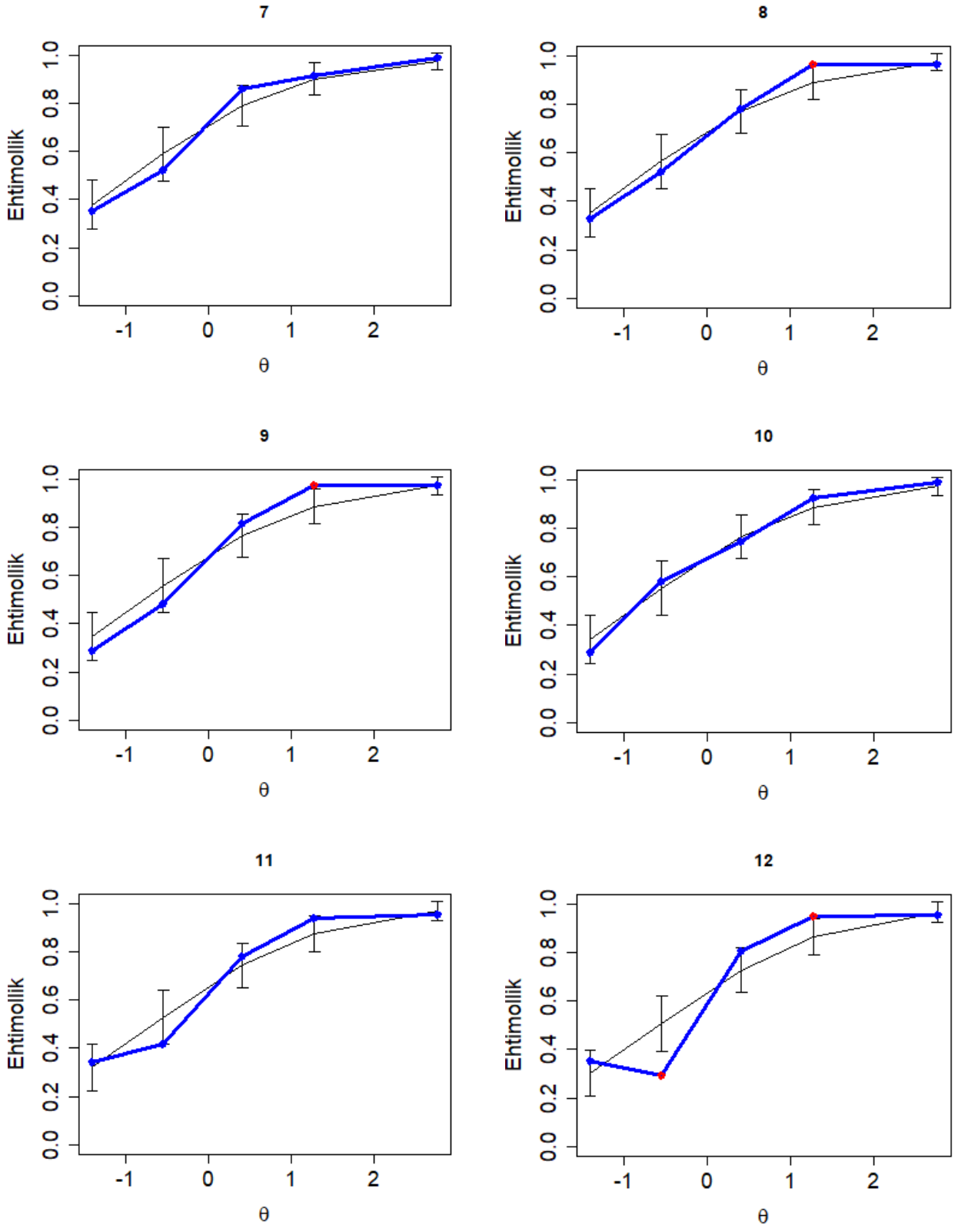

2
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

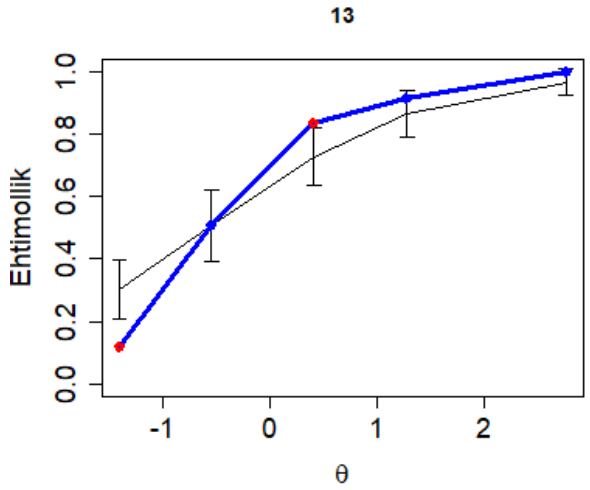

3
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

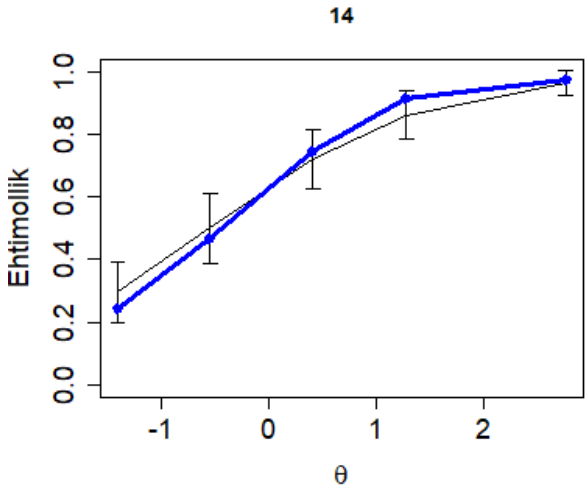

4
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

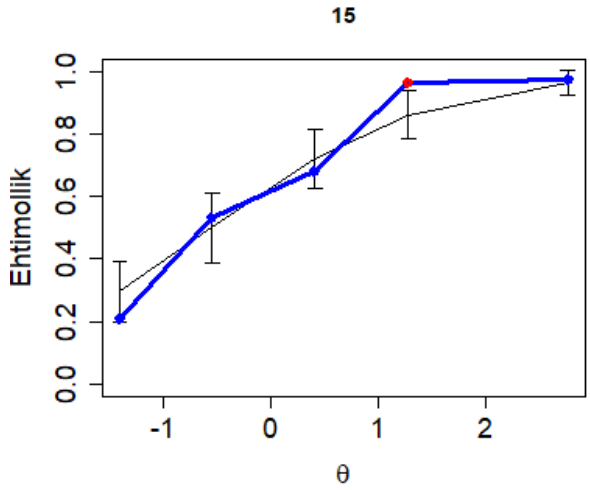

5
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

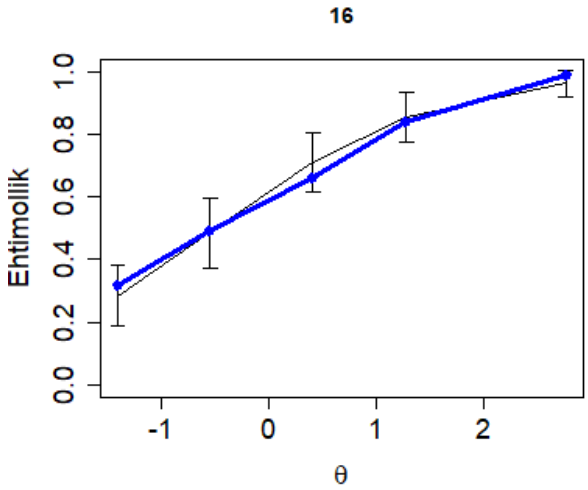

6
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

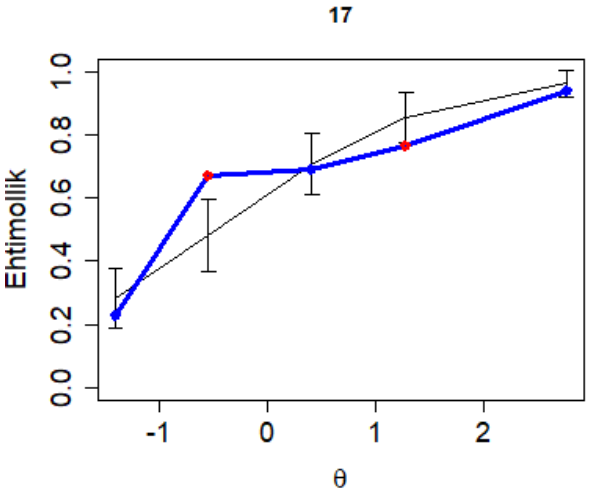

7
8
9
10
11
12
Ehtimollik
θ
0.0
0.2
0.4
0.6
0.8
1.0
-1
0
1
2

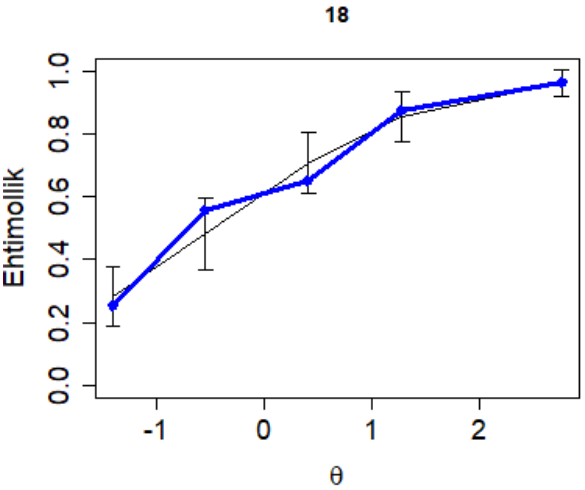

13
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

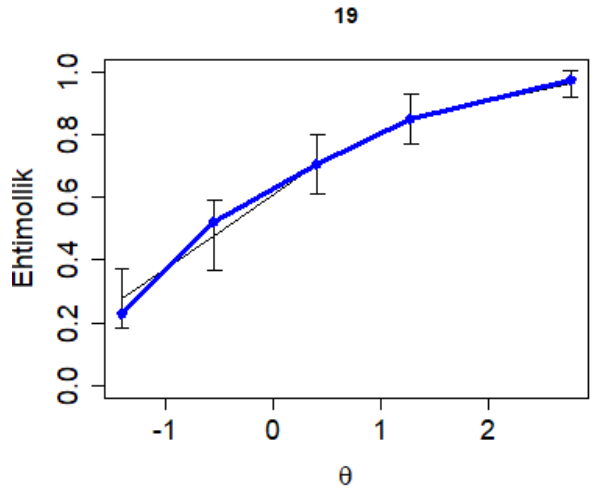

14
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

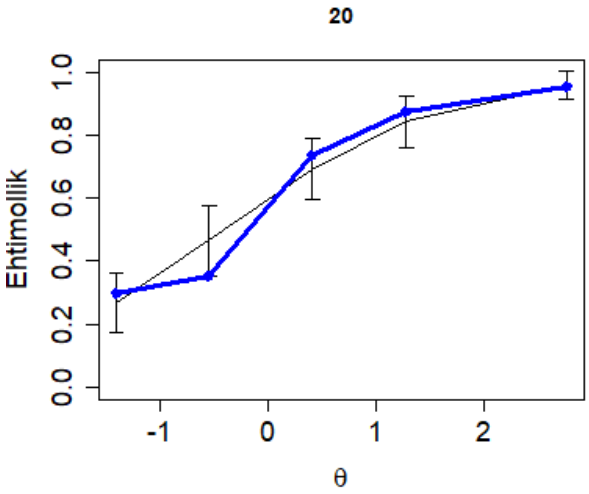

15
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

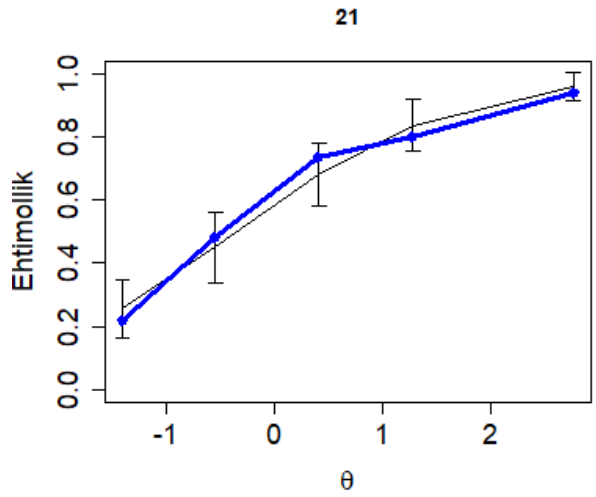

16
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

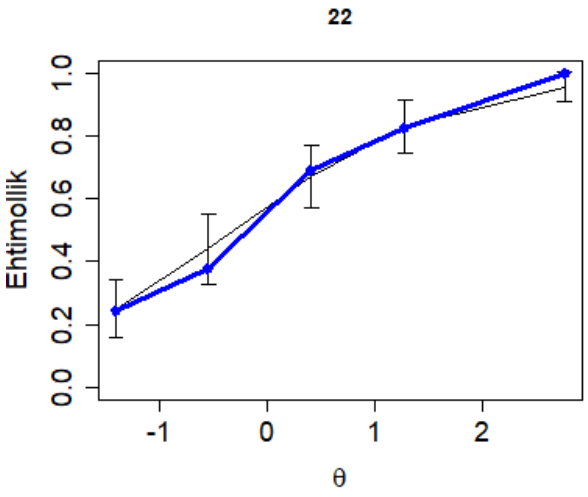

17
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

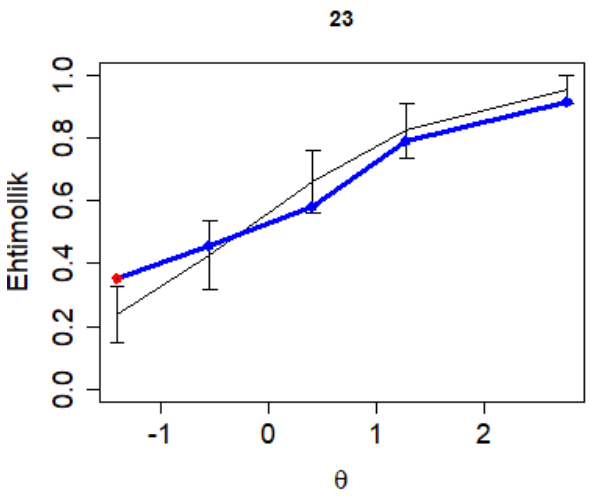

18
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

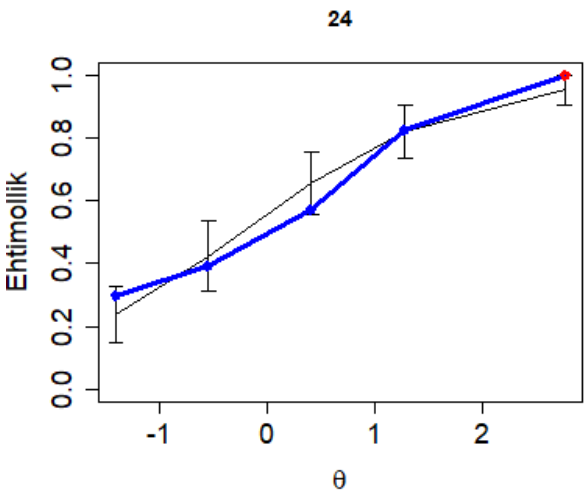

19
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

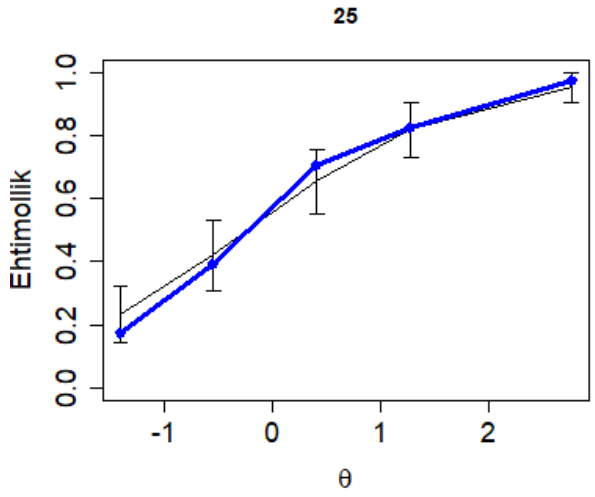

20
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

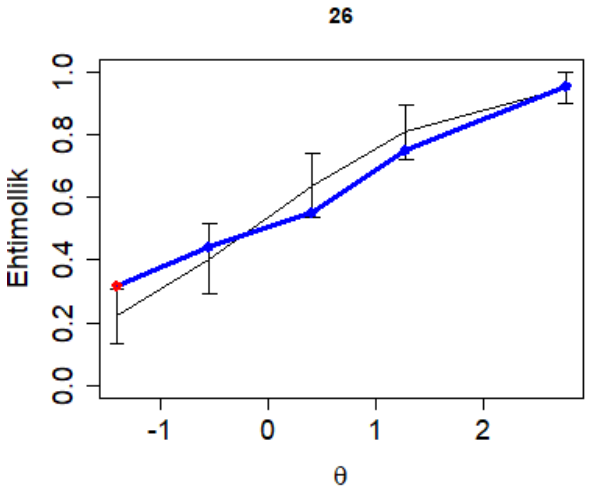

21
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

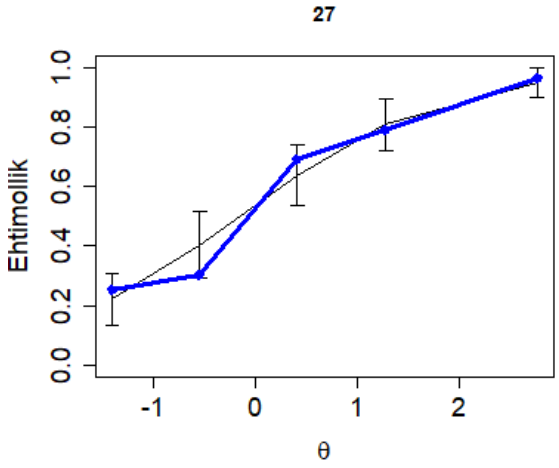

22
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

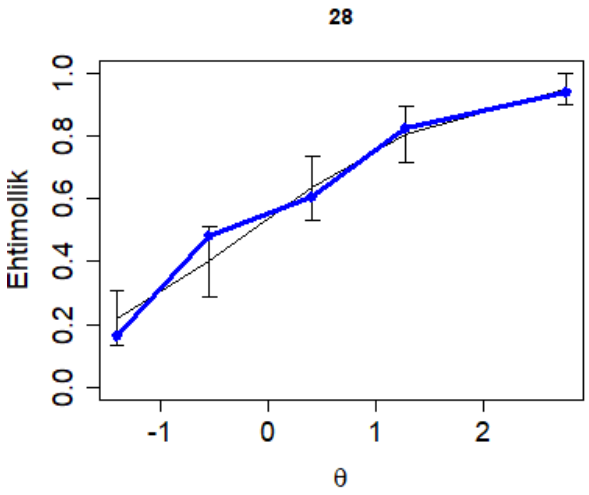

23
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

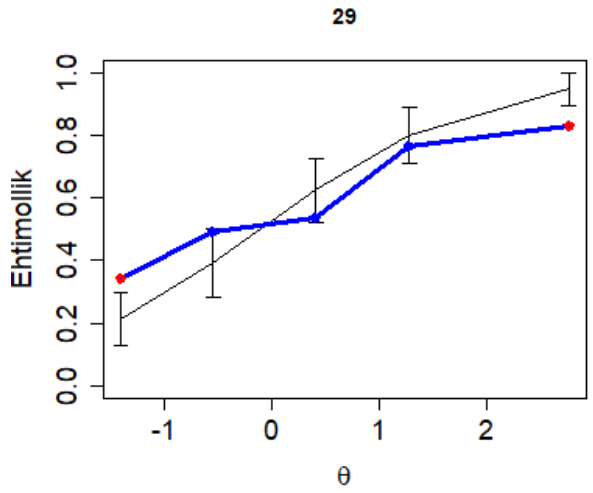

24
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

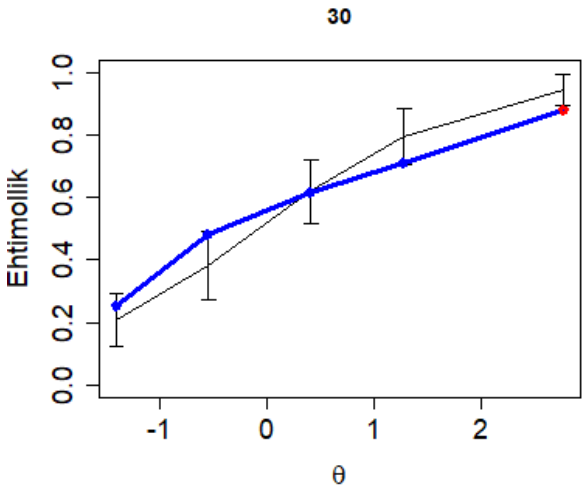

25
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

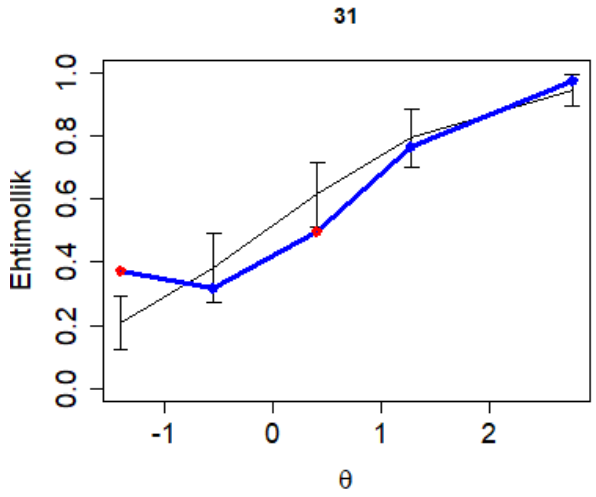

26
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

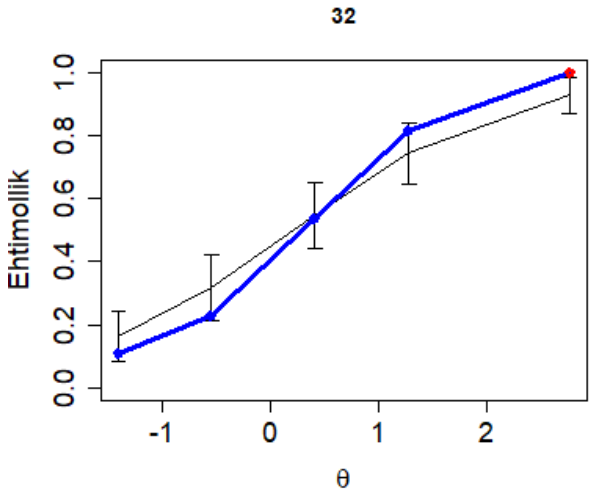

27
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

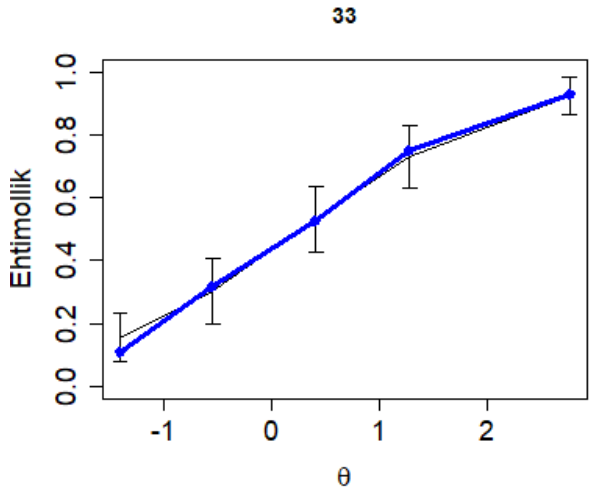

28
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

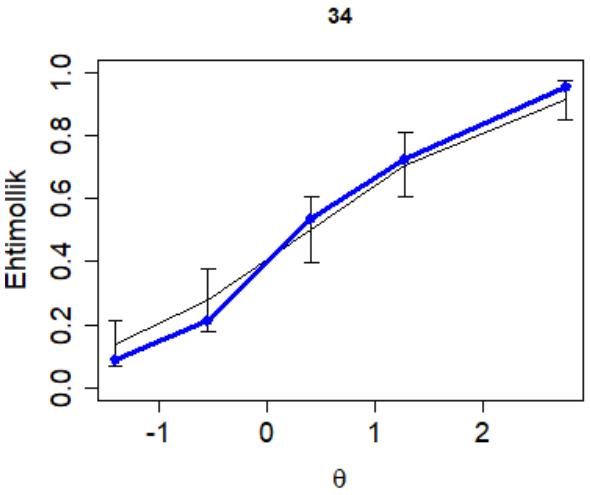

29
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

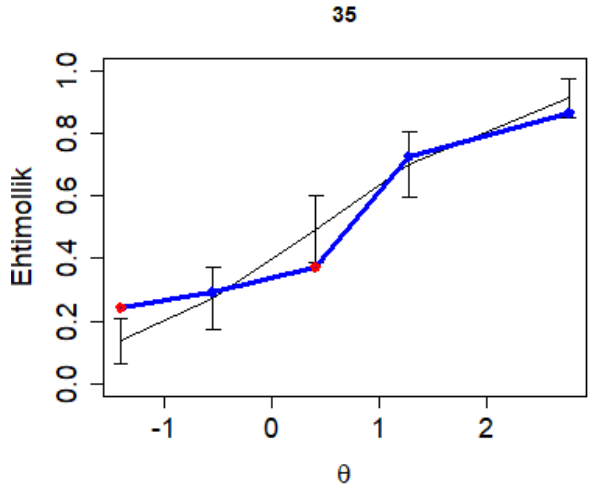

30
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

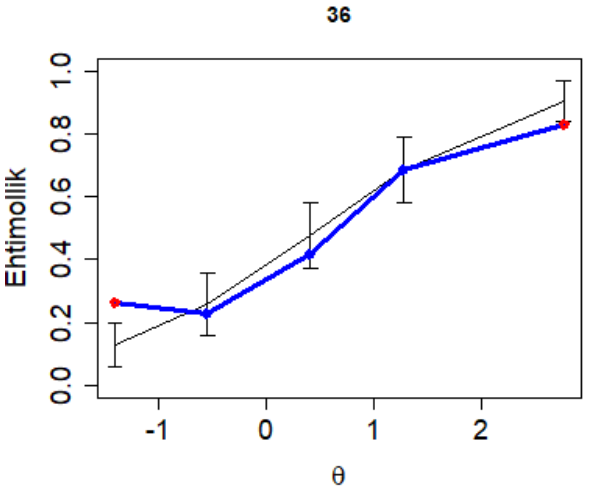

31
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

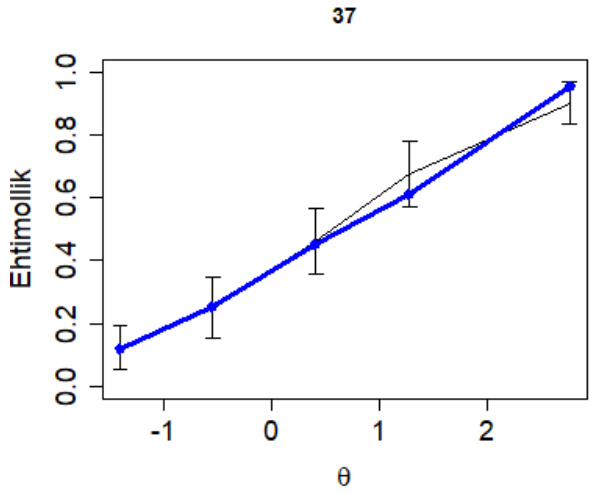

32
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

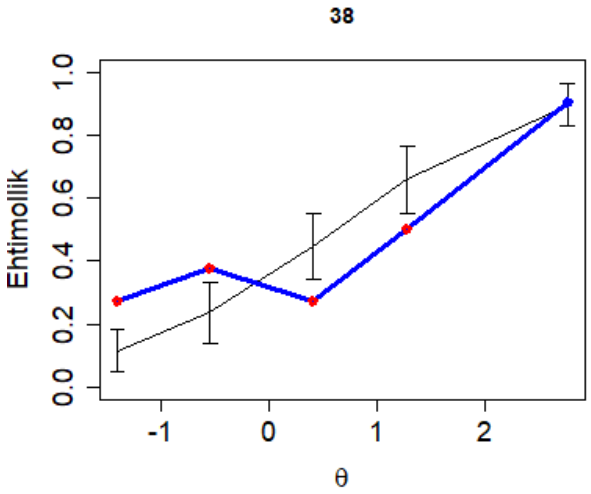

33
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

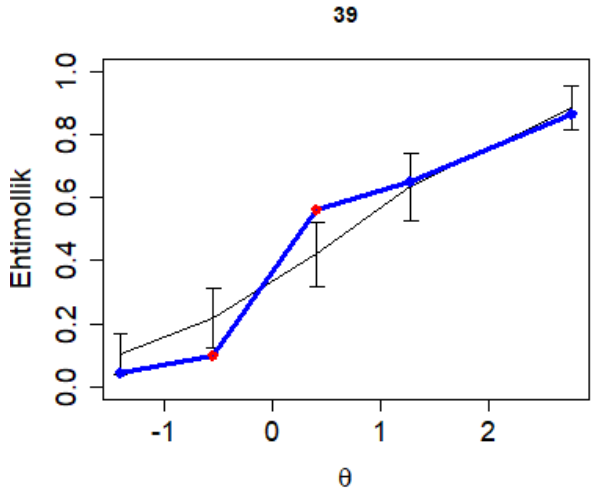

34
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

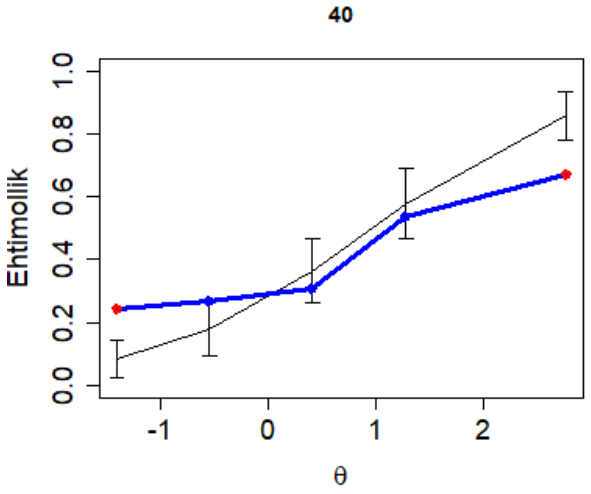

35
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

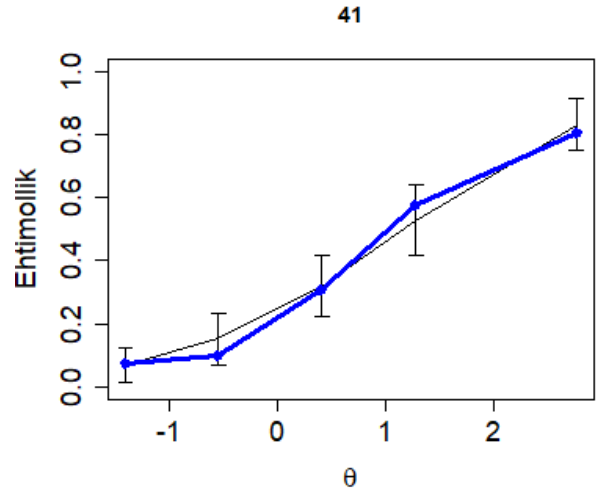

36
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

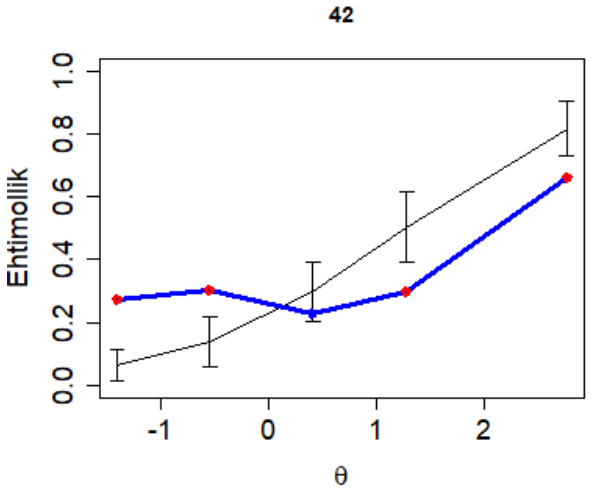

37
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ

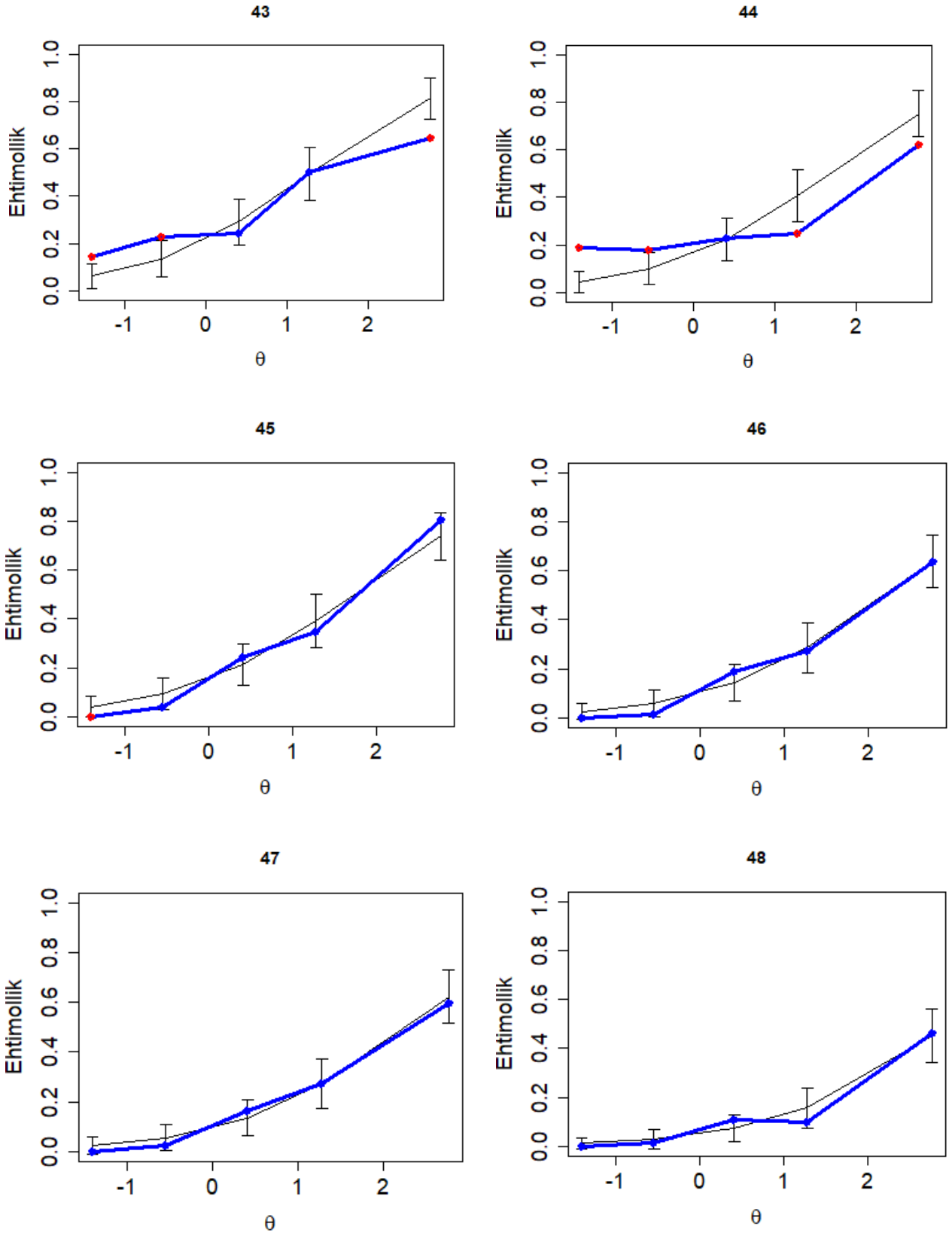

38
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ


39
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ


40
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ


41
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ


42
Ehtimollik
0.0 0.2 0.4 0.6 0.8 1.0
-1 0 1 2
θ



3-rasm. Test topshiriqlari natijalarining Rash modeli bilan mosligi

Umuman olganda qiyinlik darajasi yuqori bo‘lgan test topshiriqlarining korrelyatsiya koeffitsiyenti pastroq, aksincha, qiyinlik darajasi past bo‘lganlar test topshiriqlarining korrelyatsiya koeffitsiyenti yuqori bo‘ladi. 38(B0000031), 40(B0000016), 43(B0000043), 44(B0000040) test topshiriqlarining Rash modeli bilan moslik darajasi yaxshi emasligini, ya’ni ajratilgan qobiliyat guruhlarining hammasi bilan mos tushmaganligini ko‘rish mumkin. Bu test topshiriqlarini ham kalibirovkalangan test bazasiga qo‘shishdan oldin qayta ko‘rib chiqish lozim bo‘ladi. Lekin 38 (B0000031) test topshirig‘ini yuqori qobiliyat darajalarini, 40(B0000016), 43(B0000043), 44(B0000040) test topshiriqlarining esa o‘rtacha qobiliyat darajalarini ajratish imkonini beri-shini, rasmda shu qobiliyat guruhlari bilan mos tushganligini ko‘rish mumkin. Umumiy holda test top-shiriqlari xususiyatlarini Rash mode-liga moslashtirish muammoli elementlarni va o‘ziga xos xususiyatlarga ega bo‘lgan qobiliyatlarni ajratish imkonini beradi.

Xulosa

Biologiya fanidan umumiy o‘rta ta’lim maktablarining 9-sinf bitiruvchi o‘quvchilaridan ilmiy tadqiqot uchun olingan test sinovi natijalarining Rash modeli bilan mosligi o‘rganildi.

Turli xil qobiliyat darajalarini aniqlash uchun kalibrovkalangan test topshiriqlari bazasini yaratishda test sinovlari natijalarini Rash modeli bilan moslik darajalarini o‘rganish lozimligi ko‘rsatildi. Test sinovlari natijalarining Rash modeli bilan mosligi umumiy holda yaxshi ekanligi aniqlandi. B0000037-ID raqamli test topshirig‘ining Rash modeli bilan mosligi yaxshi emasligini, standart xatolik chiziqlaridan ham chiqib ketganligini va test topshirig‘iga javoblar mutanosib emasligi ko‘rsatildi. Bu test topshirig‘ining tashqi(outfit) mosligi faqatgina o‘rtacha qobiliyatdagi guruhlar uchun belgilangan mezonlarga mos ekanligi, boshqa qobiliyatdagi guruhlar uchun katta ekanligi hamda uning distraktorlari ham yaxshi ishlamaganligi aniqlandi.

ID raqamlari 38(B0000031), 40(B0000016), 43(B0000043), 44(B0000040) test topshiriqlarining Rash modeli bilan moslik darajasi yaxshi emasligini, ya’ni ajratilgan qobiliyat guruhlarining barchasi bilan mos tushmaganligi ko’rsatildi. Bu test topshiriqlarini ham kalibirov-

kalangan test bazasiga qo‘shishdan oldin qayta ko‘rib chiqishga to‘g‘ri keladi.

Umumiy holda test topshiriqlari xususiyatlarini Rash modelga mos-lashtirish muammoli elementlarni va o‘ziga xos xususiyatlarga ega bo‘lgan qobiliyatlarni ajratish imkonini beradi.

Muallif A. B. Normurodov ALTE tashkiloti mutaxassisi Wobbe Zijlstraga test tahlili bo‘yicha onlayn seminar–trening va ilmiy maslahatlari uchun minnatdorlik bildiradi.

ADABIYOTLAR

1. Rasch G., Probabilistic models for some intelligence and attainment tests, Copenhagen, Danish Institute for Educational research. 1960.
2. Baker, Frank. The Basics of Item Response Theory, ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD. P. 122-597. 2001.
3. Hambleton, R.K.,Swaminathan, H.,& Rogers, H.J., Fundamentals of item response theory. Newbury Park, CA: Sage. 1991.
4. Ivailo Partchev. A visual guide to item response theory, Friedrich-Schiller-Universitat Jena. 2004.
5. 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.
6. M. Dj. Ermamatov, Zamonaviy test nazariyasi asoslari. Uslubiy qo‘llanma, Toshkent sh., 2020-y., 68 bet.
7. Marianne Mueller. Item fit statistics for Rasch analysis: can we trust them? Journal of Statistical Distributions and Applications (2020) 7:5.
8. Smith, R., Schumacker, R., Bush, M.: Using item mean squares to evaluate fit to the Rasch model. J. Outcome Meas. 2, 66–78 (1998).
9. Wang, W., Chen, C.: Item parameter recovery, standard error estimates, and fit statistics of the winsteps program for the family of Rasch models. Educational and phychologicak measurement. 65, 376–404 (2005).
10. Richard M. Smith. The distributional properties of Rasch item fit statistics. Educational and phychologicak measurement. v. 51, p. 541-565, 2015.
11. 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 b., 2023.
12. Gunter Maris, Timo Bechger, Jesse Koops and Ivailo Parchev, Data Management and Analysis of Tests, p. 1-49, 2022.
13. A.B. Normurodov, I.A. Boyxonov. Rayt xaritasi, ichki va tashqi moslik statistikalari: biologiya fanidan o‘tkazilgan test sinovi natijalari, Axborotnoma, №2, 56-83 b., 2023.
14. M.Dj. Ermamatov, I.A. Boyxonov. Distraktorlar tahlili: biologiya fanidan o‘tkazilgan test sinovi natijalari, Axborotnoma, №2, 4-19 b., 2023.

15. Cizek GJ, O'Day DM: Further investigations of nonfunctioning options in multiple-choice test items. Educational and phychologicak measurement v.54(4), p. 861-872. 1994.
16. Marie Tarrant, James Ware and Ahmed M. Mohammed, An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis, BMC Medical Education 2009, 9:40.

FIT TO RASCH MODEL: RESULTS OF BIOLOGY TEST

A.B. Normurodov, M.DJ. Ermamatov, I.A. Boykhonov

Scientific and Educational Practical Center Under the Agency for Assessment of Knowledge and Competences, anormurodov@gmail.com

Abstract. In this paper based on the calculation of the dexter package of R program the results of the test on biology, which is taken for the investigation from the 9th class, are analysed within the Rasch model and the model fit is studied. Rash model is the model based on probability which describes the interaction between person ability and difficulty of items and parallelizes measurement process. In this case ability and difficuly of items are performed by linear measurements. The independence of item difficulties from the samples and the abilities from the test forms are provided. Analysing fit to the Rash model can give the possibility of determine problelmentic items and identify idiosyncratic abilities.

Keywords: Rasch model, item difficulty, ability, fit to the Rasch model.

TRAINING RATERS FOR ASSESSING PRODUCTIVE SKILLS: ENHANCING ACCURACY AND RELIABILITY

A.A. Abbosov

Agency for Assessment of Knowledge and Competences under the Ministry of Higher Education, Science and Innovations of the Republic of Uzbekistan,
100084, Tashkent., Bogishamol str., 12

Abstract. In various domains, such as language proficiency testing, writing assessments, or public speaking evaluations, the role of raters is crucial in maintaining fairness, consistency, and reliability. The accurate assessment of productive skills, which involve generating language or demonstrating practical abilities, can significantly impact a person's academic, professional, or personal growth. In this article, we will delve into the importance of training raters and explore effective strategies to enhance their accuracy and reliability. We'll discuss the challenges faced by raters, the skills required for effective rating, and the training methods employed to improve their performance.

Keywords: productive skills, scoring, rater, rating scales, rater training, reliability, bias, halo effect.

I. Introduction

In the realm of education and language assessment, the accurate evaluation of productive skills plays a crucial role in measuring a learner's proficiency. Therefore, the process of training raters to assess these skills effectively becomes paramount to ensuring fair and reliable evaluations. Structured training programs are essential for equipping raters with the necessary knowledge and skills to assess productive skills accurately. These programs typically provide training on the criteria and scoring system, as well as opportunities for practice and feedback. Additionally, rater calibration exercises are crucial in ensuring consistency among raters by establishing a common understanding of the assessment criteria. Lastly, the use of rubrics provides a clear and standardized framework for evaluating learners' performance, enabling raters to make objective and reliable judgments. Overall, an effective training process ensures that assessments are fair, valid, and ultimately promote the

development of learners' productive skills.

Furthermore, ongoing training and professional development for raters is essential to maintain the accuracy and reliability of assessments. This can include regular workshops and seminars to update raters on any changes in assessment criteria or techniques. It is also important for raters to receive feedback on their own performance in order to identify areas for improvement and ensure that they are consistently applying the assessment criteria. This feedback can come from supervisors, peers, or even the learners themselves through self-assessment or feedback surveys. By continually refining their skills and knowledge, raters can ensure that their assessments are consistently fair and valid.

II. Challenges Faced by Raters

Assessing productive skills, such as speaking or writing, is a complex task that heavily relies on the expertise and judgment of raters. However, this process is not without its challenges. Raters face various difficulties that can impact the consistency, objectivity, and fairness of assessments. In order to ensure reliable and valid results, it is crucial to address and mitigate these challenges.

One of the major challenges faced by raters is **the subjective nature of assessing** creativity within the context of productive skills. Different raters may have different interpretations of what constitutes creativity or how to evaluate it. This subjectivity can lead to inconsistencies and discrepancies in scores assigned to the same performance. To address this challenge, raters must be provided with clear guidelines and training that help them recognize and evaluate creative elements within the defined rating criteria [1].

Another challenge lies in maintaining **consistency in applying rating criteria** across different performances. Raters may interpret the criteria differently, resulting in variations in scores. To mitigate this challenge, training programs should focus on providing raters with standardized examples and clear rubrics that offer specific guidance on how to evaluate different aspects of the assessed skills. This can help ensure greater consistency and fairness in assessments [2].

Raters must contend with the challenge of **personal bias and the halo effect**, both of which can impact the objectivity of assessments.

Personal biases, whether conscious or unconscious, may result in unfair evaluations based on factors such as accent, appearance, or background of the candidates. Additionally, the halo effect, where initial positive or negative impressions influence subsequent evaluations, can introduce bias into the assessment process. To address this challenge, training programs should emphasize the importance of objectivity, provide awareness of biases, and offer strategies to mitigate their influence [3].

Inter-rater reliability refers to the consistency of ratings between different raters evaluating the same performance. However, achieving high inter-rater reliability can be challenging. Raters may differ in their expertise, experience, or understanding of the rating criteria, leading to inconsistencies in assessments. To enhance inter-rater reliability, training programs should include calibration sessions where raters discuss and align their interpretations of rating criteria, identify areas of discrepancy, and establish a shared understanding [4].

Raters often face challenges related to **fatigue and cognitive load** during prolonged rating sessions. Extended periods of rating can lead to reduced attention span, diminishing the accuracy and consistency of evaluations. To combat this challenge, training programs should provide strategies and guidelines for managing rater fatigue, including scheduled breaks, rotation of raters, and setting appropriate time limits for rating sessions. These measures can help maintain rater engagement and ensure more reliable assessments.

The key to overcoming these challenges lies in thorough training, clear guidelines, and ongoing professional development. By addressing these challenges, we can enhance the accuracy, reliability, and objectivity of assessments, ultimately providing fair and valid evaluations for candidates. It is crucial to continuously explore and implement best practices in rater training and assessment to create a robust and effective evaluation system. By acknowledging and addressing these challenges, we pave the way for improved assessments that contribute to the growth and development of individuals in various domains.

III. Skills Required for Effective Rating:

Raters play a key role in evaluating performance and assigning scores based on established criteria. However, this task requires a specific set of skills and competencies to maintain consistency, objectivity, and accuracy. One of the fundamental skills for effective rating is a deep **understanding of the assessment framework**. Raters need to be knowledgeable about the specific criteria, rubrics, and standards that guide the evaluation process. This includes a clear understanding of the performance indicators, task requirements, and expected outcomes. A solid grasp of the assessment framework empowers raters to evaluate performances in a consistent and objective manner [5].

Raters must also possess the skill to **provide objective and impartial evaluations**. They need to avoid personal biases and ensure that evaluations are free from any influence related to personal preferences, stereotypes, or external factors. Objective evaluations require raters to focus solely on the performance and the established criteria without being swayed by irrelevant attributes or circumstances. This skill is vital to maintain the fairness and integrity of the assessment process [6].

Another essential skill for raters is **effective communication and feedback**. Raters should be able to provide constructive and specific feedback to candidates, highlighting strengths and areas for improvement in a clear and concise manner. This skill supports the development and growth of individuals by offering meaningful insights that facilitate skill enhancement. Moreover, effective communication ensures that candidates understand their performance evaluation and can make informed decisions for future improvement [7].

Raters need to possess **strong critical thinking and analytical skills** to assess performances accurately. They should be able to analyze and evaluate complex tasks, identify strengths and weaknesses, and make informed judgments based on the established criteria. Critical thinking skills enable raters to assess performances from multiple angles, considering various perspectives and nuances. This skill is essential for maintaining a high level of reliability in the assessment process [8].

By emphasizing and nurturing these skills among raters, educational institutions, language testing agencies, and other assessment bodies can ensure that evaluations accurately

reflect individuals' abilities and promote their growth. Ongoing research and best practices in rater training and development contribute to the continuous improvement of assessment systems.

One ought to remember that effective rating is not only about assigning scores. It is about providing meaningful feedback and supporting individuals in their learning journeys. By cultivating and recognizing the skills required for effective rating, we can create assessment processes that benefit both candidates and the broader educational community.

IV. Training and Standardization

Raters should undergo comprehensive training programs to develop and enhance their assessment skills. These programs should focus on clarifying rating criteria, providing examples, and offering practice opportunities to familiarize raters with the assessment process. Additionally, standardization sessions that involve discussion and calibration among raters can help align their interpretations of criteria and establish a shared understanding. Training and standardization instill confidence in raters and contribute to consistent and reliable assessments [9].

Training programs are crucial for equipping raters with the necessary skills and enhancing their accuracy and reliability. According to Thompson [3], an effective training program should incorporate a combination of theoretical knowledge, practical exercises, and calibration sessions, where raters discuss and align their interpretations of rating criteria. Mock assessments and standardized samples can also be valuable tools to familiarize raters with the expected performance levels [4].

The training process for raters requires careful consideration to enhance their understanding of the assessment framework and promote consistent evaluations. **Rater training workshops** are a popular and effective method for training raters. These workshops provide an opportunity for raters to learn about the assessment criteria, rubrics, and standards. They involve interactive sessions, presentations, discussions, and practical exercises that help raters gain a comprehensive understanding of the assessment process. Rater training workshops offer a collaborative learning environment where raters can engage with experts,

share experiences, and clarify any uncertainties [10].

Training raters through **rating practice** is a valuable method to enhance their skills and promote consistency. This method involves providing raters with a set of practice performances or recordings that align with the assessment criteria. Raters evaluate these performances independently and then come together for facilitated discussions to compare and justify their ratings. This process allows raters to calibrate their judgments, align their interpretations, and establish a shared understanding of the rating criteria [11].

Video-based training is an effective method that utilizes recorded performances for training raters. Raters are provided with videos that display a range of performances representing different levels of proficiency. They watch these videos individually, evaluate the performances according to the established criteria, and then engage in group discussions to ensure consistent interpretations and ratings. Video-based training allows raters to visualize actual performances, making it easier to understand and apply the assessment criteria [12].

Peer feedback and calibration sessions involve the active participation of raters in evaluating performances and providing feedback to one another. During these sessions, raters rate the same set of performances independently and then engage in a structured discussion to compare and discuss their ratings. This process enables raters to reflect on their evaluations, discuss differences in interpretations, identify areas of agreement, and establish a consensus for consistent evaluations. Peer feedback and calibration sessions foster a culture of collaboration and continuous improvement among raters [13].

Ongoing professional development is a vital component of training raters. It includes continuous training sessions, workshops, conferences, or webinars that focus on refining assessment skills, addressing specific challenges, and staying updated with the latest research and best practices. Ongoing professional development opportunities provide raters with opportunities to enhance their knowledge, learn from experts in the field, and engage in discussions that promote consistent and reliable assessments [14].

By utilizing a combination of these training methods, educational institutions, language testing agencies, and other assessment bodies can establish a robust training framework that empowers raters to conduct fair

and valid assessments. Continuous research and innovation in rater training contribute to the ongoing improvement of assessment processes, ultimately benefitting both candidates and the broader educational community.

V. Conclusion

Training raters for the assessment of productive skills is a complex and necessary endeavor to ensure valid and reliable results. By acknowledging and addressing the challenges faced by raters, emphasizing the required skills, and implementing effective training methods, we can enhance the accuracy and reliability of assessments. A well-trained rater not only improves the fairness of the evaluation but also fosters a conducive environment for learners to develop their productive skills. Raters need to be trained to objectively evaluate and provide constructive feedback on the learners' performance. They should be aware of potential biases and be able to assess the learners' abilities accurately. Moreover, training should focus on developing effective communication and interpersonal skills, as raters need to interact with learners in a supportive and encouraging manner. By investing in comprehensive training programs for raters, we can ensure that assessments provide a valuable learning experience for the learners and contribute to their overall skill development. Further research and improvement in rater training techniques are needed to continually enhance the accuracy and reliability of assessments.

REFERENCES

1. Smith, R., & Johnson, M. (2018). Challenges in rater training: An empirical study. Journal of Language Assessment, 13(3), 215-235. doi:10.1075/jla.00012.smi.
2. Brown, A., & Lee, J. (2019). Enhancing rater training through professional development workshops. Language Testing, 36(4), 519-542. doi:10.1177/0265532218798199.
3. Thompson, L. (2017). Training raters for performance-based assessments. Applied Linguistics, 38(2), 207-226. doi:10.1093/applin/amw008.
4. Johnson, S. (2020). Best practices in rater training: A comprehensive review. Assessing Skills Quarterly, 24(2), 45-63.
5. Johnson, M. (2018). Understanding the assessment framework: A guide for raters. Journal of Language Testing, 23(1), 45-62. doi:10.1177/0265532218787544.
6. Brown, J., & Smith, R. (2019). The role of impartiality in assessment: Challenges and strategies. Journal of Educational Measurement, 56(3), 356-374. doi:10.1111/jedm.12264.
7. Lee, A. (2020). Effective communication and feedback in assessment: Enhancing the rater's role. Language Teaching Research, 24(2), 167-184. doi:10.1177/1362168819881509.
8. Thompson, E. (2017). Critical thinking in rating: Enhancing accuracy and reliability. Applied Linguistics, 42(3), 211-226. doi:10.1093/applin/amw008.
9. Smith, L. (2019). Rater training and standardization: Best practices for maintaining reliability in assessments. Assessing Skills Quarterly, 34(4), 123-137.
10. Johnson, M. (2018). Rater training workshops: Strategies for enhancing assessment expertise. Language Testing, 32(4), 439-461. doi:10.1177/0265532219852132.
11. Brown, J., & Smith, R. (2020). Enhancing rater training through rating practice. Journal of Language Testing, 25(3), 245-264. doi:10.1177/0265532218787544.
12. Lee, A. (2019). Video-based training for raters: A practical approach. Assessing Skills Quarterly, 38(2), 125-142.

13. Thompson, E. (2017). Peer feedback and calibration sessions: Fostering collaboration among raters. Applied Linguistics, 42(3), 211-226. doi:10.1093/applin/amw008.

14. Smith, L. (2021). Ongoing professional development for raters: Building expertise and consistency. Journal of Educational Measurement, 58(2), 189-205. doi:10.1111/jedm.12264.

PRODUKTIV KO'NIKMALARNI BAHOLASH UCHUN MUTAXASSISLARNI TAYYORLASH: ANIQLIK VA ISHONCHLILIKNI OSHIRISH

A.A. Abbosov

Bilim va malakalarni baholash agentligi, 100084, Toshkent sh., Bog‘ishamol k., 12

Qisqacha mazmuni. Turli sohalarda, masalan, tilni bilish darajasini tekshirish, yozish yoki og‘zaki nutqni baholashda baholovchilarning roli adolat, izchillik va ishonchlilikni saqlashda hal qiluvchi ahamiyatga ega. Til ko‘nikmalari yoki amaliy qobiliyatlarni namoyish qilishni o‘z ichiga olgan produktiv ko‘nikmalarini to‘g‘ri baholash insonning akademik, kasbiy yoki shaxsiy o‘sishiga sezilarli ta'sir ko‘rsatishi mumkin. Ushbu maqolada biz baholovchilarni tayyorlashning ahamiyatini ko‘rib chiqamiz va ularning aniqligi va ishonchliligini oshirish uchun samarali strategiyalarni o‘rganamiz. Biz baholovchilar duch keladigan qiyinchiliklarni, samarali baholash uchun zarur bo‘lgan ko‘nikmalarni va ularning ish faoliyatini yaxshilash uchun qo‘llanadigan o‘qitish usullarini muhokama qilamiz.

Kalit so‘zlar: produktiv ko‘nikmalar, baholash, baholovchi, baholash mezonlari, baholovchini tayyorlash, ishonchlilik, tarafkashlik, galo effekti.

UMUMIY ELEMENTLI TEST VARIANTLARI: BIOLOGIYA FANIDAN OLINGAN TEST NATIJALARI

M.Dj. Ermamatov, I.A. Boyxonov

Bilim va malakalarni baholash agentligi huzuridagi Ilmiy-o‘quv amaliy markazi, 100084, Toshkent sh., Bog‘ishamol k., 12

Qisqacha mazmuni. Ushbu maqolada umumiy elementli test variantlari bilan qobiliyatni baholashning obyektivlik uchun muhimligi 9-sinflardan biologiya fani bo‘yicha olingan test sinovlaridan klassik test nazariyasi va Rash modeli bilan olingan qiyinlik darajalarini taqqoslash hamda test xarakteristikasi chiziqlari bilan ko‘rsatilgan.

Kalit so‘zlar: Klassik test nazariyasi, Rash modeli, qiyinlik darajalari, umumiy elementli test variantlari.

1. Kirish

Qiyinlik darajalarini kalibrovkalash va sinaluvchilar qobiliyatini tenglashtirishda umumiy elementli test variantlaridan foydalaniladi [1-2]. Bunday usul kalibrovkalangan test topshiriqlari bazasini to‘ldirish, sinaluvchilarning bo‘sh to‘plami uchun statistik ma‘lumotlar olish va eng muhimi obyektiv baholash uchun katta ahamiyatga ega.

Zamonaviy test nazariyasidagi kalibrovkalash va tenglashtirish ingliz tilidan o‘tkazilgan ko‘p darajali test sinovlarining tinglab tushunish qismi bo‘yicha natijalarini Rash modeli bilan [3] havolada tadqiq qilgandik. Bunda sinaluvchilar qobiliyat darajalarini aniqlashda qiyinlik darajalari hisobga olinishi tufayli xatolik kam bo‘lishi hamda “kompyuter adaptive” testlari uchun bu jarayonning ahamiyati ko‘rsatilgandi. [4] havolada esa kalibrovkalangan test topshiriqlari sharq tillaridan o‘tkazilgan test sinovlari tahlili misolida bayon qilingandi.

[5] havolada ta’lim jarayonida standart testlardan foydalanish orqali turli xil test topshiruvchilar guruhlarining natijalarini solishtirish, o‘quvchilarning o‘quv dasturini qanday darajada o‘zlashtirishini va pedagoglar, shu bilan birga, ta’lim muassasalari samaradorligini baholash o‘rganilgan edi. Turli xil guruhlarning qobiliyat darajalari va turli xil test variantlaridagi test topshiriqlarining qiyin-lik darajalarining bir xil shkalada bo‘lishini ta‘minlash muhimligi test sinovi natijalarini Rash modeli bilan tahlil qilish orqali o‘rganilgandi. Bunda biologiya fanidan 9-sinf bitiruvchilaridan 1 ta variantdan foydalanib olingan natijalar tahlil qilingan edi. [6] havolada esa shu variantning distraktorlari tahlil qilingan edi.

Nazariy nuqtayi nazardan barcha olinadigan test sinovlarida bir xil

variant ishlatish mukammal bo‘lar edi. Ammo qayta test olinganda test topshiriq-larining mazmuni chiqib ketishini hisobga olinganda amaliyot-da ko‘p variant ishlatiladi.

Ko‘p variant ishlatilganda vari-antdagi test topshiriqlari taqsimoti muammosi yuzaga chiqadi. Buning oldini olish uchun test variantlari uchun spetsifikatsiyalar ishlab chiqila-di, lekin qobiliyatni baholashda xatolikni kamaytirishning har doim ham imkoni bo‘lmaydi.

Ushbu maqolada biologiya fani-dan 9-sinf bitiruvchilaridan 4 ta umumiy elementli variantlardan foydalanib olingan test sinovlari natijalarini klassik test nazariyasi va Rash modeli doirasida tahlil qilinadi.

2-bo‘limda klassik test nazariyasi bilan hisoblangan qiyinlik darajalari, element-umumiy ball korrelyatsiyasi va element-qoldiq ball korrelyatsiyasi muhokama qilinadi.

3-bo‘limda obyektiv baholashda klassik test nazariyasi va Rash modeli taqqoslanadi.

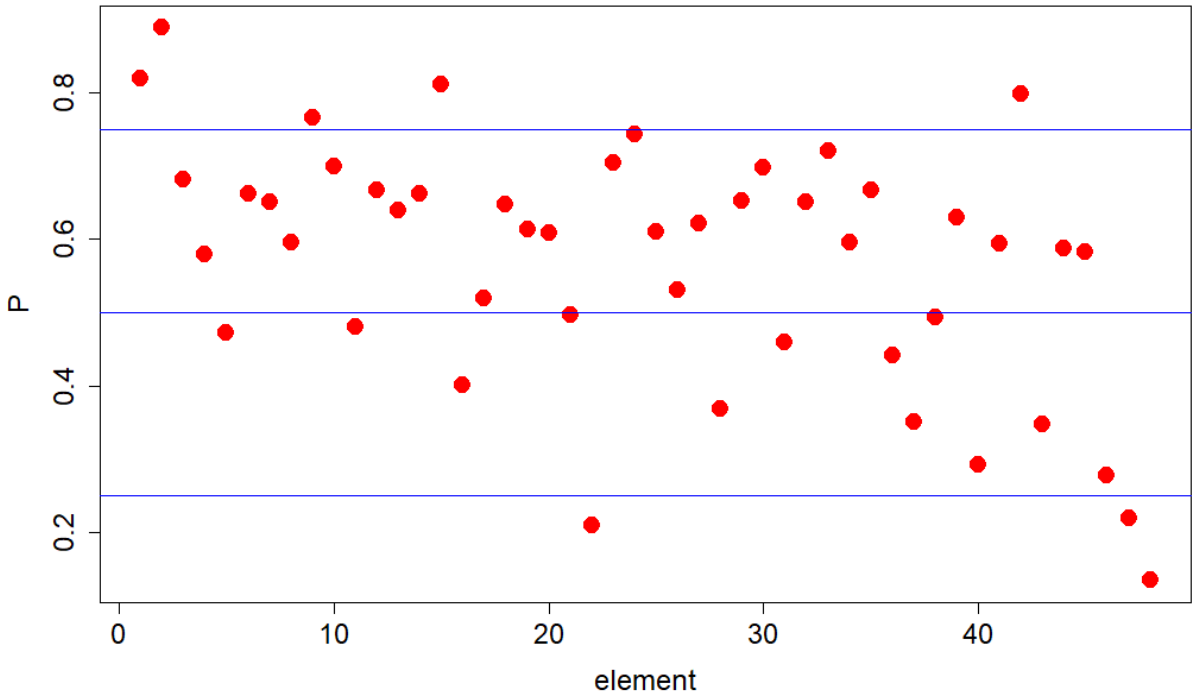
Hisoblashlarda R dasturida ishla-tiladigan dexter dastur to‘plamidan foydalaniladi [7].

2. Klassik test nazariyasi bo‘yicha qiyinlik darajalari

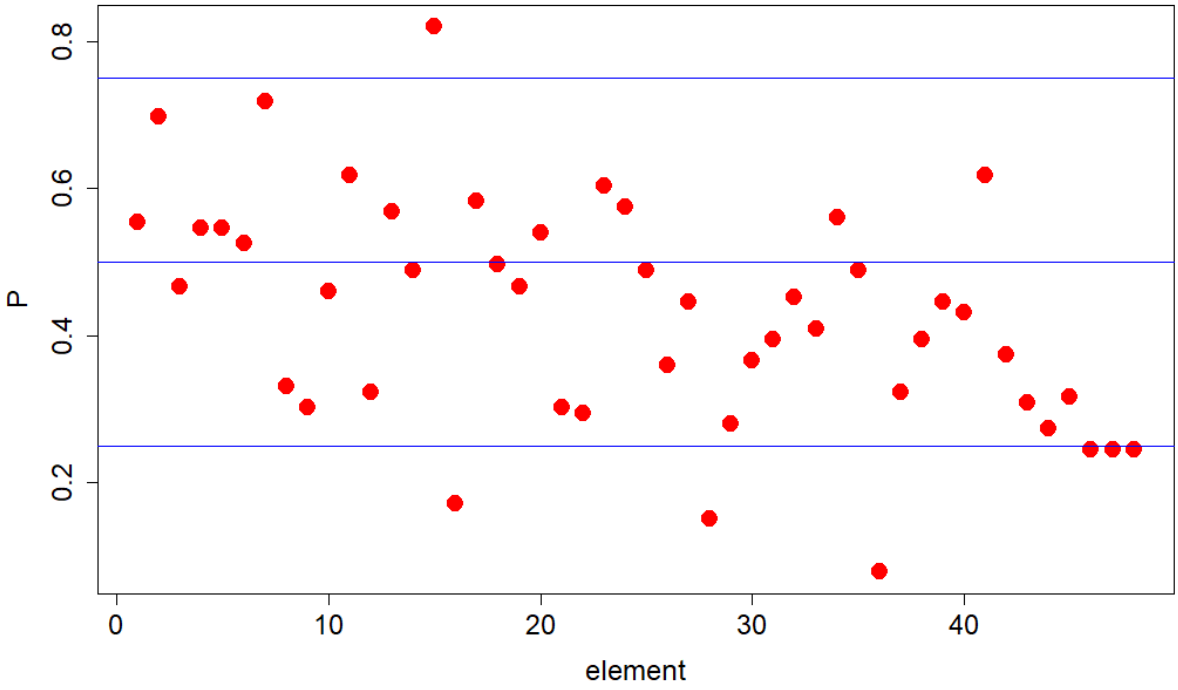
O‘rta ta’lim muassasalarida 9-sinf bitiruvchilari uchun biologiya fani bo‘yicha bilimlarni baholaydigan 4 ta test variantlarining har biri 38 ta yopiq, 7 ta qisqa javobli va 1 ta kengaytirilgan (3 ta tarkibiy qismdan iborat) javobli ochiq test topshiriqlaridan iborat bo‘lib, jami 48 ta test topshirig‘idan iborat [5].

1-(a)-(d) rasmlarda mos ravishda 1-4-variantlardagi test topshiriqlariga to‘g‘ri javob berganlar sonining $\boldsymbol{n_t}$ ushbu test topshirig‘iga javob bergan sinaluvchilar umumiy soni $\boldsymbol{n}$ ga nisbati $\boldsymbol{P = \frac{n_t}{n}}$ berilgan.

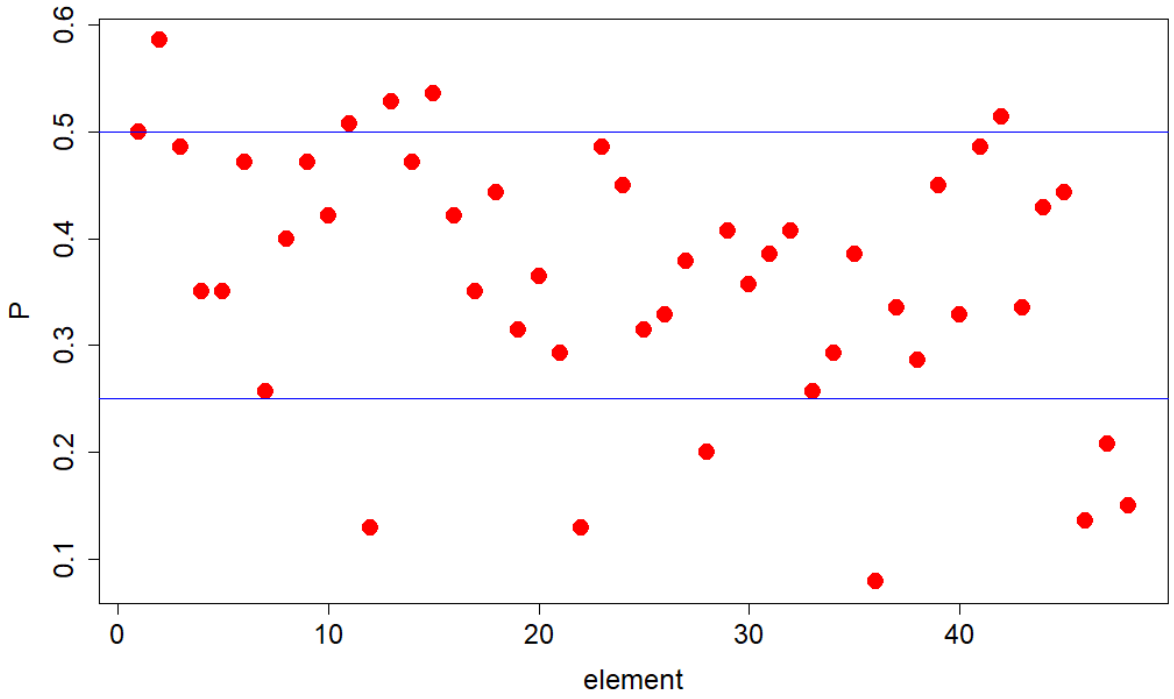
Variantning qiyinlik darajasini tasavvur qilish uchun rasmda P=0,25; 0,50 va 0,75 qiymatlarga to‘g‘ri keladigan gorizontal chiziqlar chi-zilgan. Bunday shartli bo‘lishda P=0,25 ga to‘g‘ri keladigan chiziqdan pastda turgan test topshiriqlarini eng yuqori qiyinlik darajasida, 0,25-0,5 orali-g‘idagi test topshiriqlarini yuqori o‘rta qiyinlik darajasida, 0,5-0,75 orali-g‘idagi test topshiriqlarini past o‘rta qiyinlik va tog‘ri javoblar ulushi 0,75 dan katta bo‘lgan test topshiriqlarini eng past qiyinlik darajasida deb hisoblash mumkin.


P
0.2
0.4
0.6
0.8
0
10
20
30
40
element

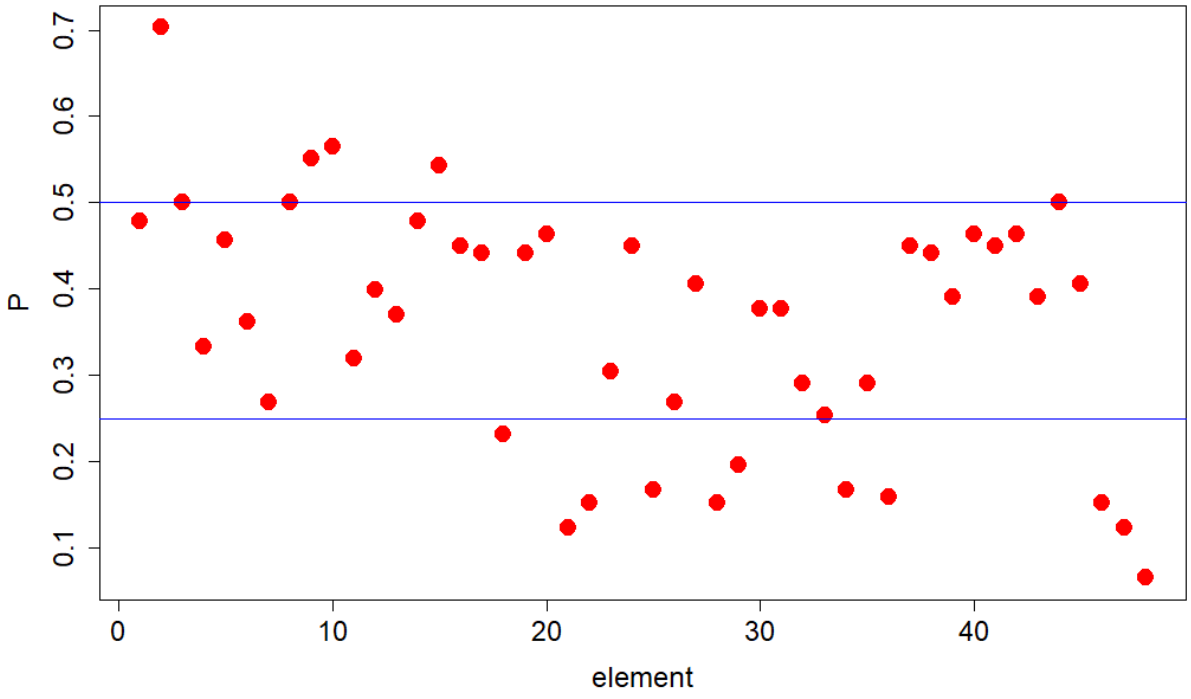
(a)


P
0.2
0.4
0.6
0.8
0
10
20
30
40
element

(b)



(c)



(d)

1-4-variantlardagi test topshiriqlariga tog‘ri javob berganlar ulushi (a)-(d).

1 (a) rasmdan P=0,75 dan yuqorida 4 ta test topshirig‘i, 0,5-0,75 oralig‘idagi test topshiriqlari soni 0,25-0,50 oralig‘idagi test topshiriqlari sonidan ko‘p ekanligi va P=0,25 dan kichik bo‘lgan test topshiriqlari 3 ta ekanligi ko‘rinadi.

1 (b) rasmdan esa P=0,75 dan yuqorida 1 ta test topshirig‘i, 0,5-0,75 oralig‘idagi test topshiriqlari soni 0,25-0,50 oralig‘idagi test topshiriqlari sonidan kam ekanligi va P=0,25 dan kichik bo‘lgan test topshiriqlari 3 ta ekanligi ko‘rinadi.

1 (c) va (d) rasmlardan eng past qiyinlik darajasidagi test topshiriqlari mavjud bo‘lmasdan, eng yuqori darajali test topshiriqlarining soni oshgan-ligini, past o‘rta qiyinlik darajadagi test topshiriqlari soni kamayib, yuqori o‘rta qiyinlik darajasi-dagi test topshiriqlari oshganini ko‘rish mumkin. Agar barcha test varigantlarini bir xil guruh yechganda, albatta, variantdagi test topshiriqlari-ning qiyinlik darajasi oshib boryapti degan xulosaga kelish to‘g‘ri bo‘lardi, ammo bizning tadqiqotimizda 4 ta variantni turli xil guruhlar yechishgan. Bu holatda variantdagi test topshiriqlarining qiyinlik darajasi test topshiriqlari qiyinligining o‘zgarishi tufayli yuzaga keldimi yoki turli qobiliyat darajalari tufaylimi degan savolga klassik test nazariyasi doirasida javob berish mushkul. Bu pedagogik o‘lchovlarda yuzaga keladigan qiyinlik va qobiliyat darajalari invariantligi degan muam-moga olib keladi [8-10]. Invariantlik muammolarini hal qilishga bo‘lgan urinishlar zamonaviy test nazariya-larining rivojlanishiga olib keldi. Bizning tadqiqotimizda ushbu maqola-ning 3-bo‘limida zamonaviy test nazariyalardan biri bo‘lgan Rash modeli [11] bilan muammoni yechish imkoniyati ko‘rsatiladi.

Klassik va zamonaviy test naza-riyalari bir-birini rad etmasligi, balki to‘ldirishini ta‘kidlab o‘tish lozim [8].

3. Rash modeli bilan hisoblangan qiyinlik darajalari

Rash modeli bo‘yicha qiyinlik darajalarini hisoblash, bunda varian-tlarning qiyinlik darajalarini hisobga olish uchun umumiy guruh va umumiy elementli variantlar usullaridan foyda-lanish mumkin [9-11]. Biz ushbu tadqiqotimizda umumiy elementli 4 ta variantdan foydalanamiz. Bunda har bitta variantning 6 ta test topshirig‘i qolganlari uchun umumiy bo‘ladi.

Qobiliyatlar va qiyinlik dara-jalarini baholash uchun turli xil usullar ishlatiladi: maksimal haqiqatga o‘x-shashlik, birgalikda maksimal haqi-qatga o‘xshashlik, shartli maksimal haqiqatga o‘xshahshlik, marginal maksimal haqiqatga o‘xshashlik v.k. usullar ishlatiladi. Masalan, birga-likdagi maksimal haqiqatga o‘xshash-lik usullaridagi muammolar test topshiriqlari va qobiliyat para-metrlarini birgalikda baholash tufayli kelib chiqadi. Agar parametrlar qobiliyatning ishtirokisiz baholana-digan bo‘lsa, bunday muammolar o‘z-o‘zidan yo‘qoladi.

Biz ushbu tadqiqotda Rash modeli bilan hisoblashlarda R dasturida ishlatiladigan dexter dasturlar to‘p-lamidan [7] foydalanamiz. Bu dasturda test topshiriqlarining qiyinlik darajasi shartli maksimal o‘xshashlik usuli [8-

11] bilan aniqlanadi. Yuqorida ta‘kidlab o‘tganimizdek, koordinata boshini tanlashda ikki xil an‘ana bor: birinchisi koordinata boshi sifatida qobiliyat darajalarining o‘rtacha qiymatini olish va ikkinchisi test topshiriqlari qiyinlik darajalarining o‘rtacha qiymatini olish. Dexter dastur to‘plamida koordinata boshi sifatida test topshiriqlarining qiyinlik darajalari olingan [7].

1-4-jadvallarda mos ravishda 1-4-test variantida ishlatilgan test topshiriqlarining qiyinlik darajalari keltirilgan.

1-jadval

1-variantdagi test topshiriqlarining ID raqami (ID), Rash modeli bo‘yicha hisoblangan qiyinlik darajalari ($\boldsymbol{\beta}$), qiyinlik darajalarini hisoblashdagi standart xatolik ($\boldsymbol{SX_{\beta}}$), variant nomeri va elementlarning variantdagi o‘rni

ID	β	SX_{β}	variant	element
B0000001	-1,35	0,11	1	1
B0000002	-2,30	0,17	1	2
B0000003	-0,70	0,13	1	3
B0000004	-0,26	0,10	1	4
B0000005	0,51	0,12	1	5
B0000006	-0,58	0,13	1	6
B0000007	-0,26	0,10	1	7
B0000008	-0,20	0,12	1	8
B0000009	-1,16	0,11	1	9
B0000010	-0,81	0,13	1	10
B0000011	0,40	0,11	1	11
B0000012	-0,15	0,11	1	12
B0000013	-0,44	0,12	1	13
B0000014	-0,58	0,13	1	14
B0000015	-1,57	0,14	1	15
B0000016	0,60	0,11	1	16
B0000017	0,24	0,12	1	17
B0000018	-0,50	0,13	1	18
B0000019	-0,34	0,10	1	19

B0000020	-0,23	0,11	1	20
B0000021	0,36	0,12	1	21
B0000022	2,21	0,15	1	22
B0000023	-0,89	0,11	1	23
B0000024	-0,95	0,11	1	24
B0000025	-0,28	0,12	1	25
B0000026	0,17	0,12	1	26
B0000027	-0,35	0,12	1	27
B0000028	1,18	0,11	1	28
B0000029	-0,53	0,13	1	29
B0000030	-0,80	0,13	1	30
B0000031	0,57	0,12	1	31
B0000032	-0,51	0,13	1	32
B0000033	-0,95	0,13	1	33
B0000034	-0,35	0,10	1	34
B0000035	-0,61	0,13	1	35
B0000036	0,68	0,12	1	36
B0000037	0,97	0,11	1	37
B0000038	0,39	0,12	1	38
B0000039	-0,33	0,10	1	39
B0000040	1,60	0,13	1	40
B0000041	-0,19	0,12	1	41
B0000042	-1,48	0,14	1	42
B0000043	1,11	0,11	1	43
B0000044	-0,15	0,12	1	44
B0000045	-0,23	0,11	1	45
B0000136	1,70	0,14	1	46
B0000137	2,08	0,13	1	47
B0000138	2,92	0,17	1	48

2-jadval

2-variantdagi test topshiriqlarining ID raqami (ID), Rash modeli bo‘yicha hisoblangan qiyinlik darajalari (β), qiyinlik darajalarini hisoblashdagi standart xatolik (SX_{β}), variant nomeri va elementlarning variantdagi o‘rni

ID	β	SX_{β}	variant	element o‘rni
B0000001	-1,35	0,11	2	1
B0000047	-1,40	0,14	2	2
B0000048	-0,55	0,14	2	3
B0000004	-0,26	0,10	2	4
B0000050	-0,61	0,14	2	5
B0000051	-0,66	0,14	2	6
B0000052	-1,60	0,21	2	7
B0000053	0,50	0,22	2	8
B0000054	0,69	0,22	2	9
B0000055	-0,25	0,20	2	10
B0000056	-1,06	0,20	2	11
B0000057	0,54	0,22	2	12
B0000058	-0,92	0,14	2	13
B0000059	-0,40	0,20	2	14
B0000060	-2,24	0,23	2	15
B0000061	1,70	0,27	2	16
B0000062	-0,88	0,20	2	17
B0000063	-0,51	0,14	2	18
B0000064	-0,29	0,20	2	19
B0000065	-0,66	0,20	2	20
B0000066	0,69	0,22	2	21
B0000067	0,73	0,23	2	22
B0000023	-0,89	0,11	2	23
B0000069	-0,84	0,20	2	24
B0000070	-0,40	0,20	2	25
B0000071	0,43	0,15	2	26

B0000072	-0,17	0,20	2	27
B0000073	1,92	0,28	2	28
B0000074	0,83	0,23	2	29
B0000075	0,28	0,21	2	30
B0000076	0,00	0,14	2	31
B0000077	-0,21	0,20	2	32
B0000078	0,03	0,21	2	33
B0000034	-0,35	0,10	2	34
B0000080	-0,40	0,20	2	35
B0000081	2,73	0,25	2	36
B0000082	0,54	0,22	2	37
B0000083	-0,18	0,14	2	38
B0000039	-0,33	0,10	2	39
B0000085	-0,34	0,14	2	40
B0000086	-1,06	0,20	2	41
B0000087	0,23	0,21	2	42
B0000043	1,11	0,11	2	43
B0000089	0,89	0,23	2	44
B0000090	0,59	0,22	2	45
B0000140	1,10	0,24	2	46
B0000141	1,10	0,24	2	47
B0000142	1,68	0,19	2	48

3-jadval

3-variantdagi test topshiriqlarining ID raqami (ID), Rash modeli bo‘yicha hisoblangan qiyinlik darajalari (β), qiyinlik darajalarini hisoblashdagi standart xatolik (SX_{β}), variant nomeri va elementlarning variantdagi o‘rni

ID	β	SX_{β}	**variant**	**element o‘rni**
B0000046	-0,89	0,20	3	1
B0000047	-1,40	0,14	3	2
B0000048	-0,55	0,14	3	3

B0000049	-0,04	0,22	3	4
B0000095	-0,04	0,22	3	5
B0000051	-0,66	0,14	3	6
B0000097	0,61	0,24	3	7
B0000098	-0,55	0,14	3	8
B0000099	-0,74	0,20	3	9
B0000100	-0,77	0,14	3	10
B0000101	-0,92	0,20	3	11
B0000012	-0,15	0,11	3	12
B0000058	-0,92	0,14	3	13
B0000104	-0,74	0,20	3	14
B0000105	-1,01	0,14	3	15
B0000016	0,60	0,11	3	16
B0000107	-0,04	0,22	3	17
B0000063	-0,51	0,14	3	18
B0000109	0,19	0,22	3	19
B0000020	-0,23	0,11	3	20
B0000111	0,34	0,23	3	21
B0000112	1,67	0,20	3	22
B0000068	-0,81	0,20	3	23
B0000024	-0,95	0,11	3	24
B0000115	0,19	0,22	3	25
B0000116	0,10	0,22	3	26
B0000117	-0,22	0,21	3	27
B0000118	1,09	0,26	3	28
B0000119	-0,38	0,21	3	29
B0000120	-0,09	0,22	3	30
B0000121	-0,26	0,21	3	31
B0000122	-0,38	0,21	3	32
B0000123	0,62	0,17	3	33
B0000079	0,34	0,23	3	34

B0000125	-0,26	0,21	3	35
B0000081	2,73	0,25	3	36
B0000037	0,97	0,11	3	37
B0000128	0,39	0,23	3	38
B0000084	-0,62	0,20	3	39
B0000130	0,10	0,22	3	40
B0000131	-0,81	0,20	3	41
B0000132	-0,96	0,20	3	42
B0000088	0,05	0,22	3	43
B0000134	-0,63	0,14	3	44
B0000045	-0,23	0,11	3	45
B0000144	1,77	0,30	3	46
B0000145	1,02	0,26	3	47
B0000146	1,60	0,29	3	48

4-jadval

4-variantdagi test topshiriqlarining ID raqami (ID), Rash modeli bo‘yicha hisoblangan qiyinlik darajalari (β), qiyinlik darajalarini hisoblashdagi standart xatolik (SX_β)variant nomeri va elementlarning variantdagi o‘rni

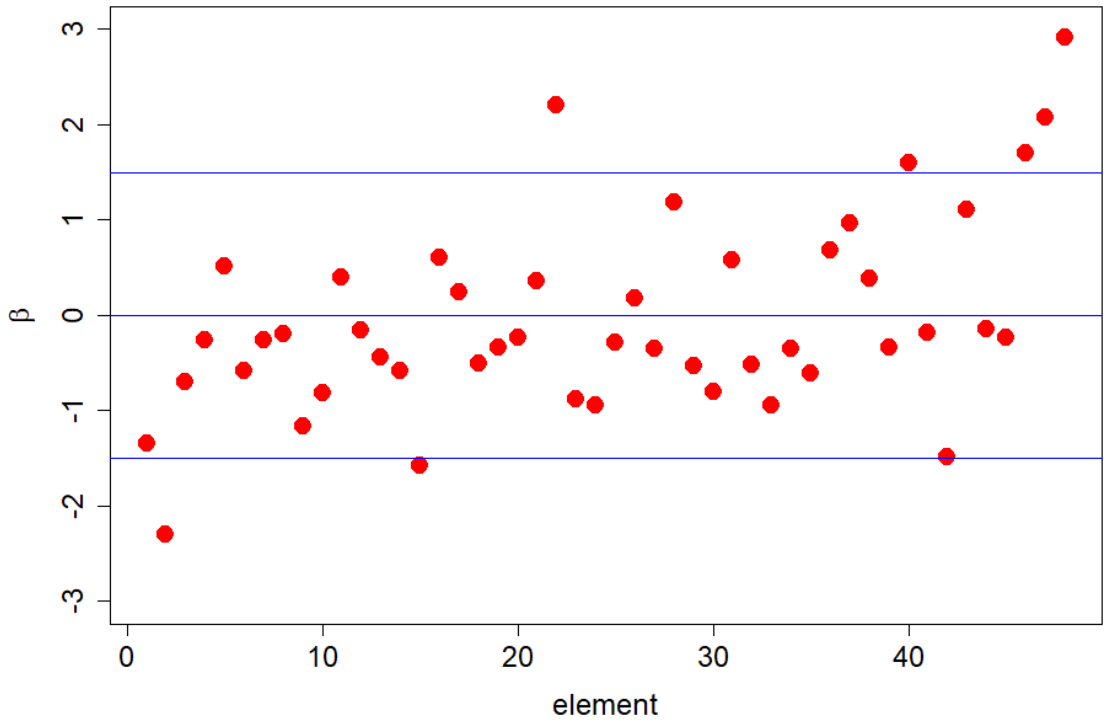
ID	**β**	SE_β	**variant**	**element o‘rni**
B0000091	-0,63	0,19	4	1
B0000092	-1,74	0,20	4	2
B0000093	-0,74	0,19	4	3
B0000094	0,13	0,21	4	4
B0000050	-0,61	0,14	4	5
B0000096	-0,03	0,21	4	6
B0000007	-0,26	0,10	4	7
B0000098	-0,55	0,14	4	8
B0000009	-1,16	0,11	4	9
B0000100	-0,77	0,14	4	10
B0000011	0,40	0,11	4	11

B0000102	-0,23	0,20	4	12
B0000103	-0,07	0,20	4	13
B0000161	-0,63	0,19	4	14
B0000105	-1,01	0,14	4	15
B0000106	-0,49	0,20	4	16
B0000164	-0,45	0,20	4	17
B0000108	0,80	0,24	4	18
B0000019	-0,34	0,10	4	19
B0000110	-0,56	0,20	4	20
B0000168	1,82	0,30	4	21
B0000112	1,67	0,20	4	22
B0000113	0,31	0,22	4	23
B0000114	-0,49	0,20	4	24
B0000172	1,35	0,27	4	25
B0000071	0,43	0,15	4	26
B0000174	-0,27	0,20	4	27
B0000028	1,18	0,11	4	28
B0000176	1,09	0,25	4	29
B0000177	-0,11	0,20	4	30
B0000076	0,00	0,14	4	31
B0000179	0,40	0,22	4	32
B0000123	0,62	0,17	4	33
B0000124	1,35	0,27	4	34
B0000182	0,40	0,22	4	35
B0000126	1,42	0,27	4	36
B0000127	-0,49	0,20	4	37
B0000083	-0,18	0,14	4	38
B0000129	-0,19	0,20	4	39
B0000085	-0,34	0,14	4	40
B0000188	-0,49	0,20	4	41
B0000189	-0,56	0,20	4	42

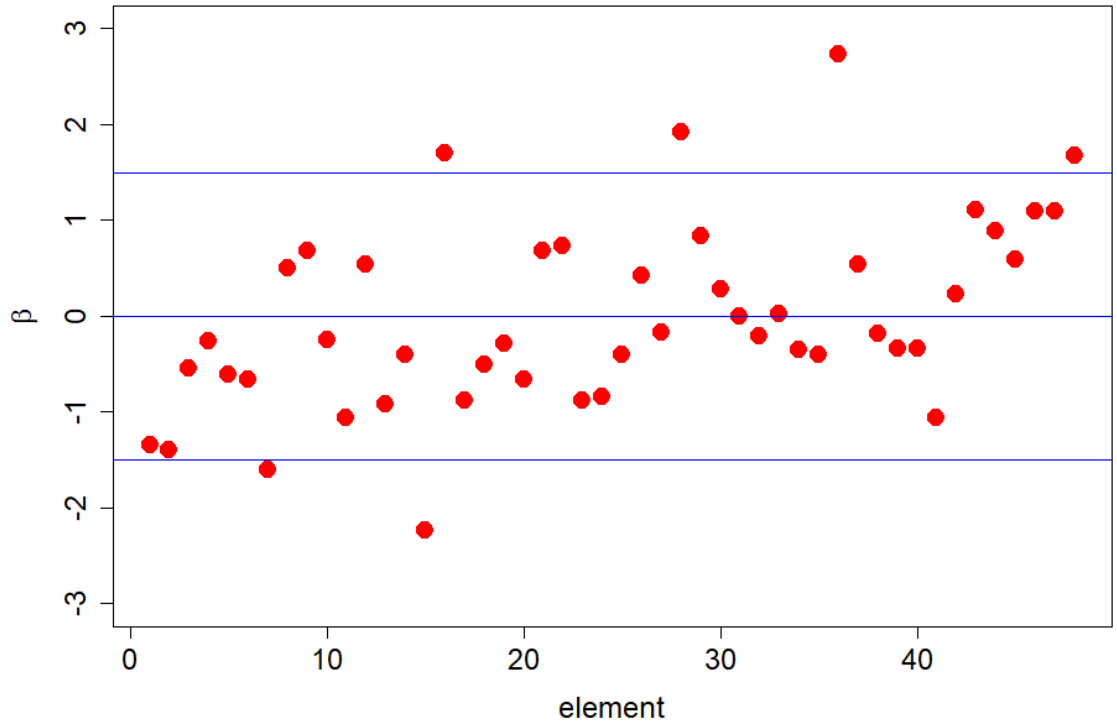
B0000133	-0,19	0,20	4	43
B0000134	-0,63	0,14	4	44
B0000135	-0,27	0,20	4	45
B0000193	1,50	0,28	4	46
B0000137	2,08	0,13	4	47
B0000142	1,68	0,19	4	48

1-4-jadvallarda bir xil ID raqamli test topshiriqlarining qiyinlik darajalari bir xil ekanligiga e'tibor berish lozim. Bunday test topshiriqlari har bir juft test variantlari uchun 6 tani tashkil qiladi. Bu umumiy test topshiriqlari test topshiriqlarining qiyinlik darajalarini bir xil shkalada bo'lishini ta'minlab beradi.

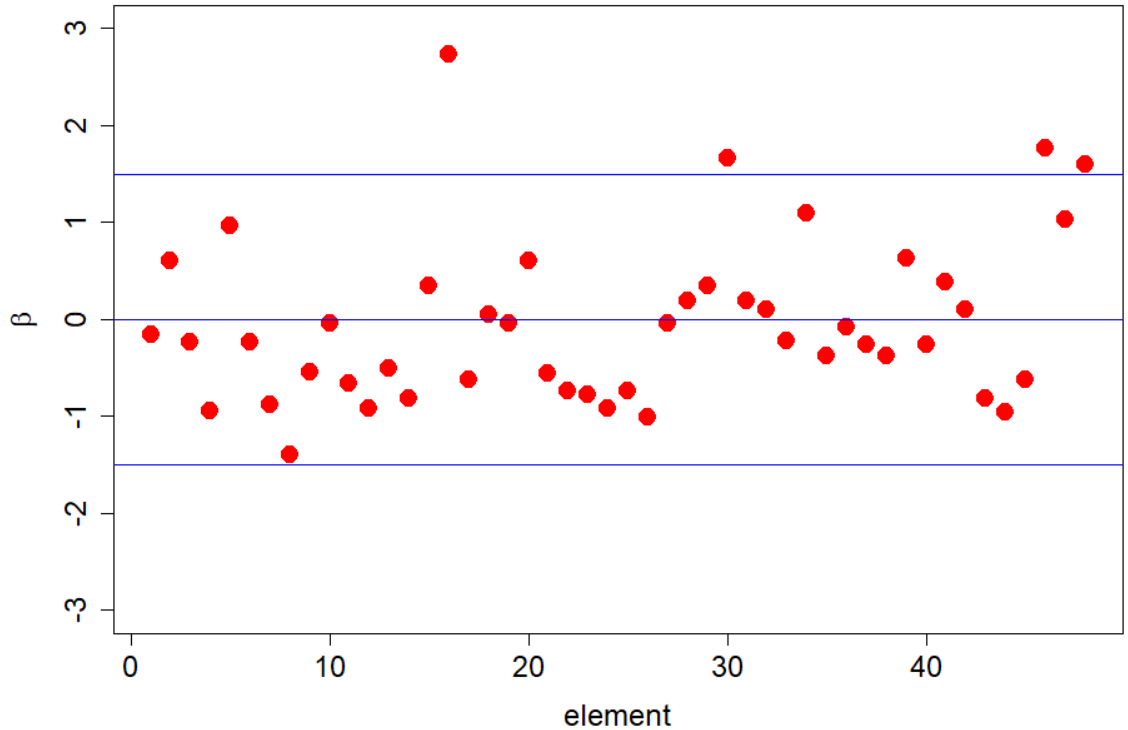
Klassik test nazariyasida chizilgan 1-(a)-(d)-rasmlarning Rash modelidagi analogini 2-(a)-(d) rasmlarda keltiramiz. Bu yerda ham qiyinlik darajalarini shartli ravishda 4 taga ajratib mulohaza yuritish mumkin: qiyinlik darajalari 1,5 logit birligidan katta bo'lgan topshiriqlarini yuqori qiyinlik darajasi, qiyinlik darajalari 1,5 va 0 logit birligi oralig'idagi test topshiriqlarini yuqori o'rta, qiyinlik darajalari 0 va -1,5 logit birligi oralig'idagi test topshiriqlarini past o'rta hamda qiyinlik darajalari -1,5 logit birligidan kichik bo'lgan test topshiriqlarini quyi qiyinlik darajalari deb hisoblash mumkin.



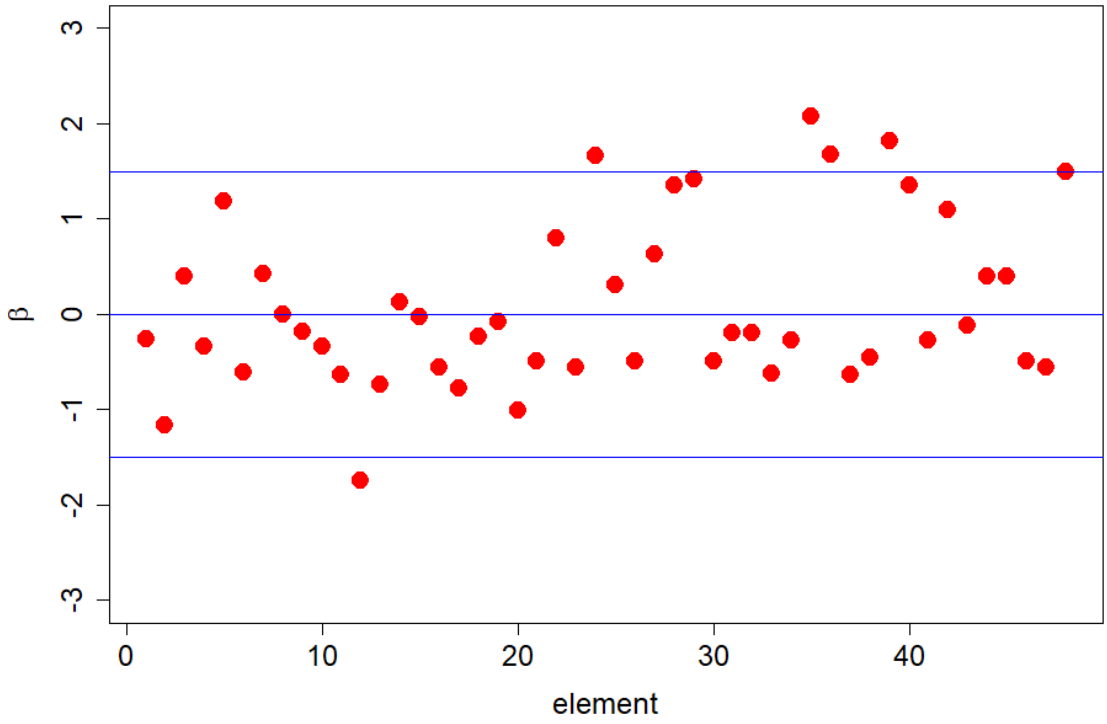
(a)



(b)



(c)



(d)

2-rasm. 1-4-variantlardagi test topshiriqlari qiyinlik darajalari taqsimoti (a)-(d)

1-rasmdagi klassik nazariya bilan aniqlangan qiyinlik darajalarini 2-rasmdagi Rash modeli natijalari bilan solishtirish qiyinlik darajalarining bir xil shkaladalgini ko‘rsatadi. Bunda umumiy elementlarning qiyinlik darajalari 4 ta variant uchun bir xil, bu esa barcha 4 ta variantdagi test topshiriqlari bir xil shkaladaligini bildiradi.

2-(a)-(d) rasmlardan yuqori qiyinlik darajasidagi test topshiriqlari soni barcha variantlarda bir-biridan kam farq qilishi ko‘rinadi. Barcha variantlarda -1,5 va 0 logit birligi oralig‘idagi test topshiriqlari 0 va 1,5 logit birligidagi test topshiriqlaridan ko‘p. Shuningdek, quyi qiyinlik darajasidagi test topshiriqlari soni ham variantlarda bir-biridan keskin farq qilmasligini ko‘rish mumkin.

2-bo‘limdagi klassik nazariya bo‘yicha qiyinlik darajalari tahlilidan variantlardagi qiyinlik darajalari taqsimotining keskin farq qilishi variantdagi test topshiriqlarining qiyinlik darajalari tufaylimi yoki guruhlar qobiliyat darajalarining farqi tufaylimi degan savolga javob berishning imkoni yo‘q edi.

Rash modeli bo‘yicha hisoblar esa variantlarni tuzishda mutaxassislar spetsifikatsiyaga amal qilganligini bildiradi. Bu esa klassik nazariyada aniqlangan farq guruhlardagi qobiliyatlar farqi tufayli paydo bo‘lganligini bildiradi. Ammo bu variantlardagi qiyinlik darajalari aynan bir xilligini bildirmaydi. Shuningdek, variantlardagi test topshiriqlarining qiyinlik darajalari taqsimotini aynan bir xil

qilish imkoni ham yo‘q, lekin zamonaviy test nazariyalarida qobiliyatni yetarlicha aniqlikda aniqlash va qiyinlik darajalarini qobiliyat darajalarini hisoblashda hisobga olish imkoni mavjud.

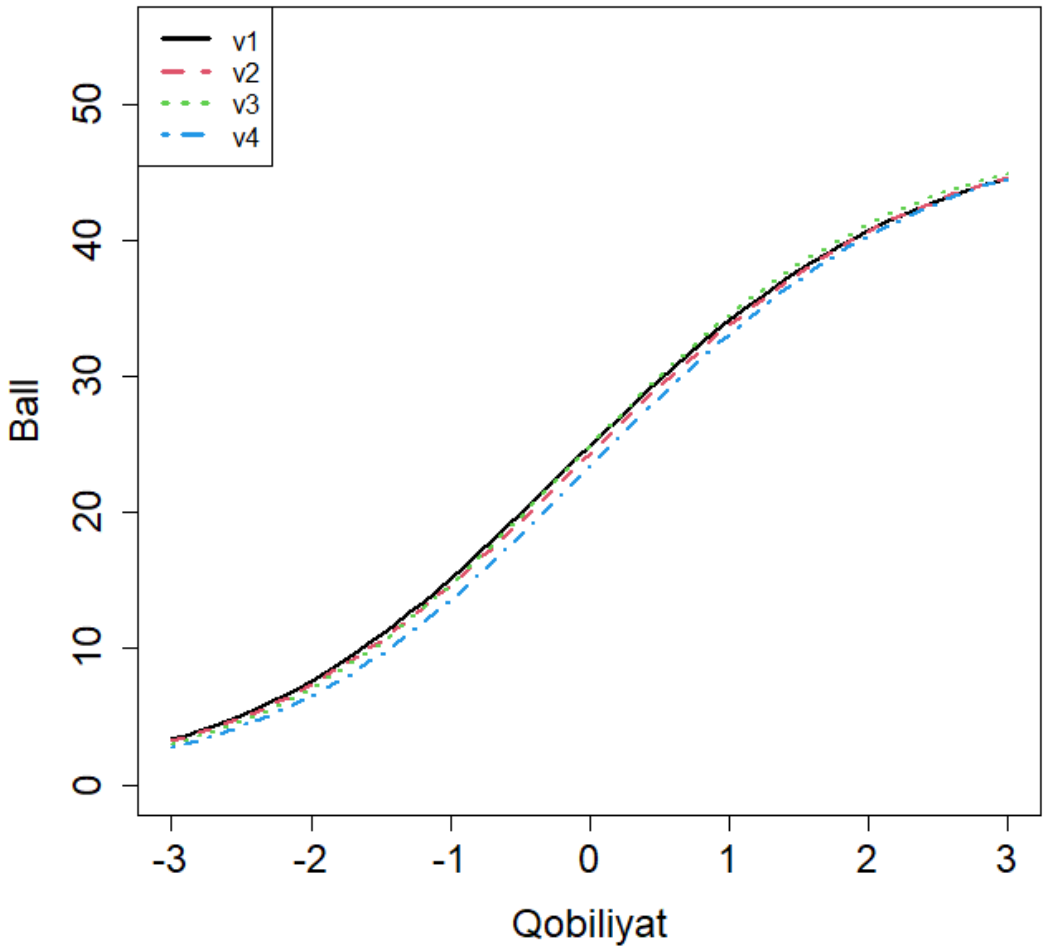
Umuman olganda 4 ta variantdan olingan xom ballar qobiliyat darajalarini ko‘rsatmaydi. Masalan, Rash modelida 1-4-variantlarda bir xil sondagi test topshiriqlarini tog‘ri yechgan sinaluvchilarning ballari variantlar qiyinligidagi farq tufayli bir xil bo‘lmasligi mumkin.

Bu holatni ko‘rsatish uchun 3-rasmda ushbu tadqiqot uchun ishlatilgan, umumiy elementli 4 ta test variantini ishlatib olingan test ma‘lumoti chiziqlarini bitta grafikda ko‘rsatamiz.

3-rasmda test xarakteristikasi chiziqlarining bir-birga juda yaqinligi variantlarning qiyinlik darajalari bir-biriga juda yaqinligini bildiradi. Shunday bo‘lsa-da rasmdan variantlarning qiyinligida kichik farq borligini ko‘rish mumkin. Quyi darajali qobiliyatlar uchun uzluksiz (v1) chiziq eng tepada joylashgan, undan keyin uzuq chiziq (v2), keyin nuqtali uzuq chiziq (v3) va nihoyat nuqtali-chiziqli uzuq chiziq (v4) joylashgan. v3 chiziq qobiliyat darajasi oshib borishi bilan avval v2 chiziq bilan mos tushadi, keyin unga nisbatan teparoqda joylashadi, keyin v1 chiziq bilan mos keladi va nihoyat v1 chiziqdan ham tepada joylashadi. v4 chiziq eng pastda joylashgan. Bu Rash modeli bilan bilimlar baholanganda quyi qobiliyat uchun eng oson variant 1-variant ekanligini, qobiliyat darajasi oshib borishi bilan 1- va 2-variantlar bir xil qiyinlik darajasida qobiliyat darajalariga erishilishini, yuqori qobiliyatlar uchun esa 2-variant 1-variantga nisbatan oson bo‘lib qolishini bildiradi. Barcha qobiliyat darajalari uchun 4-variant eng qiyin variant bo‘ladi. Bunda Rash modeli hisoblariga ko‘ra, misol uchun 4-variantda 10-test topshirig‘ini to‘g‘ri yechgan sinaluvchilarning ballari variant qiyinroq bo‘lgani uchun qolgan variantlarda 10-test topshirig‘ini yechgan sinaluvchilarning ballaridan yuqori bo‘ladi.

4. Xulosa

Ushbu tadqiqotda biologiya fanidan 9-sinf bitiruvchilaridan olingan 4 ta umumiy elementli variantlarni ishlatib olingan test natijalaridan klassik test nazariyasi va Rash modeli bilan aniqlangan qiyinlik darajalari taqqoslandi. Klassik nazariyada test topshiriqlari va qobiliyat darajalarining invariantligini ta‘minlash imkoniyati, variantlardagi test topshiriqlarini bir shkalaga o‘tkazish va sinaluvchilarning bilimlarini obyektiv baholashda Rash modeli qudratli model ekanligi ko‘rsatildi.



3-rasm. 1-4-variantlarning test xarakteristikasi chiziqlari

ADABIYOTLAR

1. Kolen, M. J., & Brennan, R. L. (1995). Test equating. New York: Springer.
2. Yu, Chong Ho and Osborn-Popp, Sharon E. (2005) "Test Equating by Common Items and Common Subjects: Concepts and Applications," Practical Assessment, Research, and Evaluation: Vol. 10 , Article 4.
3. M. Dj. Ermamatov, A. Abbosov, A. A. Baratov, Kalibrovkalangan test topshiriqlari va tenglashtirish, BMBA “Axborotnoma” ilmiy-uslubiy jurnali, 2022-yil, 3-4-son, 3-14-betlar.
4. M. Dj. Ermamatov, D. M. Alimov, A. A. Sulaymonov, A. R. Sattiyev, Kalibrovkalangan test topshiriqlari: sharq tillaridan o‘tkazilgan test sinovi natijalari tahlili, BMBA “Axborotnoma” ilmiy-uslubiy jurnali, 2022-yil, 3-4-son, 6-83-betlar.
5. A.B. Normurodov, M.Dj. Ermamatov, A.A. Baratov, I.A. Boyxonov, Umumiy o‘rta ta’lim maktablarining 9–sinf bitiruvchilari uchun biologiya fanidan bilimlarni baholashda standart testlardan foydalanish, BMBA “Axborotnoma” ilmiy-uslubiy jurnali, 2023-yil, 1-son, 63-77-betlar.
6. M.Dj. Ermamatov, I.A. Boyxonov, Distraktorlar tahlili: biologiya fanidan o‘tkazilgan test sinovi natijalari, BMBA “Axborotnoma” ilmiy-uslubiy jurnali, 2023-yil, 2-son, 4-19-betlar.
7. Gunter Maris, Timo Bechger, Jesse Koops and Ivailo Parchev, Data Management and Analysis of Tests, 2022, p. 1-49.
8. Baker, Frank (2001). The Basics of Item Response Theory, ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.
9. Hambleton, R.K.,Swaminathan, H.,& Rogers, H.J. (1991), Fundamentals of item response theory. Newbury Park, CA: Sage.
10. Ivailo Partchev (2004), A visual guide to item response theory, Friedrich-Schiller-Universitat Jena.
11. Rasch G. (1960), *Probabilistic models for some intelligence and attainment tests*, Copenhagen, Danish Institute for Educational research.

COMMON ITEM TEST FORMS: TEST RESULTS ON BIOLOGY

M.Dj. Ermamatov, I.A. Boykhonov

Scientific-Study Practical Center under the Agency for Assessment Knowledge and Competences, Tashkent 100084, Bogishamol 12

Abstract. In this paper importance of the ability estimation with common item test forms for the objectriviy is shown by the comparison of the difficulties determined by classical test theory and Rasch model from the test results of the test administered among the 9th grade students and test characteristic curves.

Keywords: Classical test theory, Rasch model, difficulty, common item test forms.

TARIX FANIDAN MILLIY SERTIFIKAT UCHUN O‘TKAZILGAN TEST SINOVI NATIJALARINING KLASSIK TEST NAZARIYASI VA RASH MODELI 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 statistikalri, 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 bilish 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 beri-ladi.

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 xarakteristikalari 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.

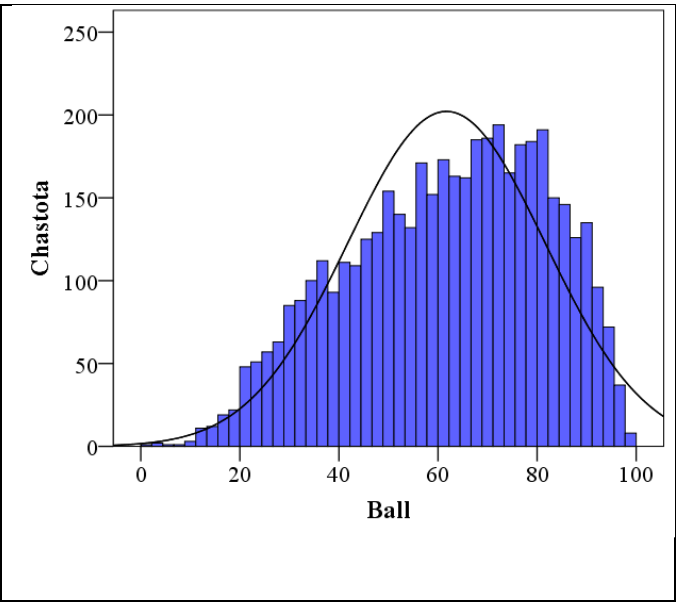
Shuningdek, test variantlari va test topshiriqlarining asosiy statistik xarakteristikalari qatoriga o‘rta qiymat, gistogrammani qurish, moda va mediana kabi ko‘rsatkichlarni hisoblash hamda test ballarining umumiy dispersiyasi (standart og‘ish) ham kiradi [2-5]. Test ballari (yoki to‘g‘ri javoblar)ning 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 koeffitsiyentining 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 javoblarining 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 topshirig‘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 vari-antdan 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	O37B	0,340
55	O38A	0,328

Element-umumiy ball korrelya-tsiyasi 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 javob-larnigina 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 instru-mentdagi 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 kotinuumdagi 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 topshi-riqlarining xarakteristikalar xususiyati bo‘yicha test topshiriqlari bazasiga joylashtirishda foydalaniladi.

IRT parametrlari tajriba (eksperi-ment) 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 ehti-moli shaxsning qobiliyat va element qiyinlik darajalari bilan aniqlanadi.

Buni quyidagi matematik formula orqali ifodalanadi:

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

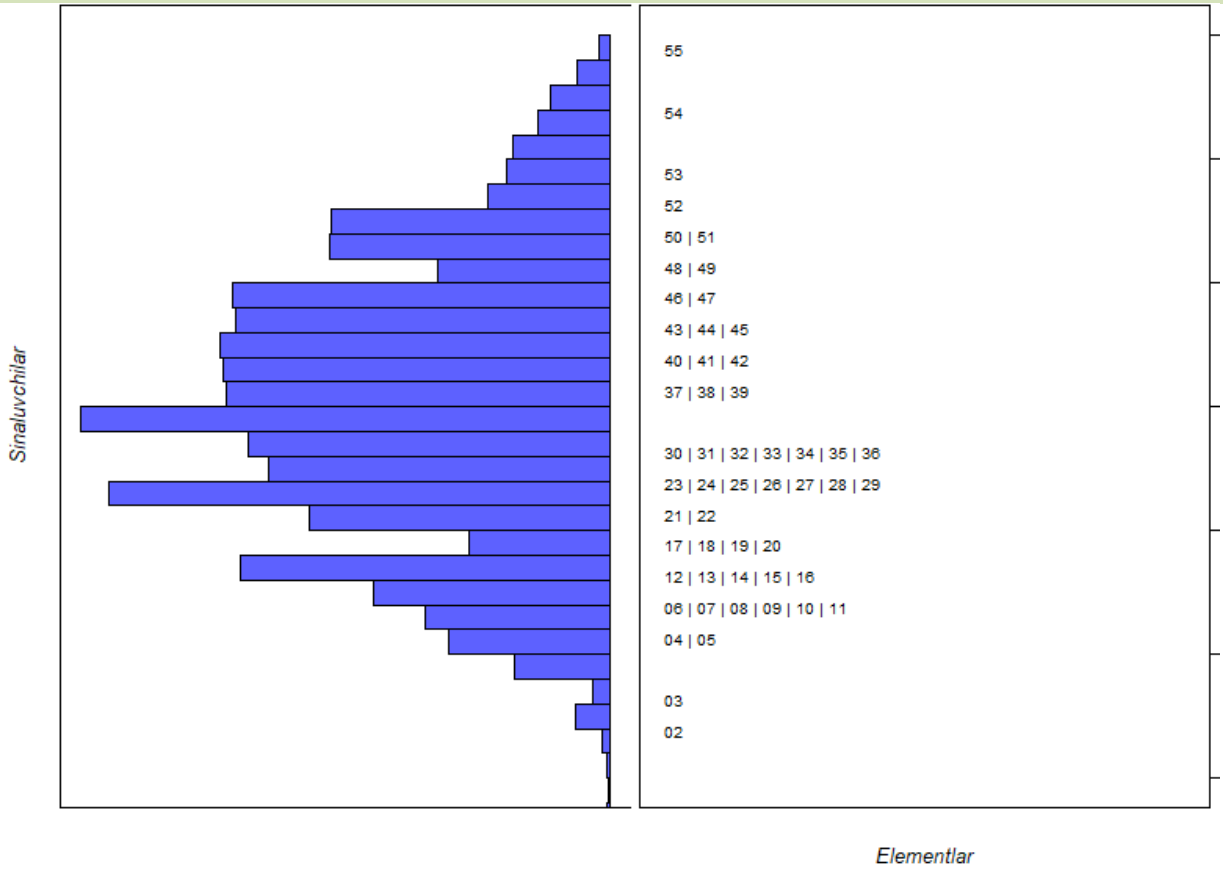
bu yerda, $X_{is} = 1$ *s*-o‘quvchining *i* elementga to‘g‘ri javob berish ehtimolligi, θ_s-qobiliyat o‘zgaruvchisi, b_i–topshiriq qiyinlik darajasi, *e*-natural logarifm asosi (*e*=2,7182818 ...).

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 $(-\mathbf{3,48}:\mathbf{2,88})$ logit birligi orasida, test topshiriqlari qiyinlik darajalari esa $(-\mathbf{3,72}:\mathbf{2,83})$ 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 $(-\mathbf{3}:\mathbf{3})$ 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‘xshahshlik, 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 qobilyat 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 $\boldsymbol{\chi^2}$ usuliga asoslangan. U test topshiriqlari qiyinlik darajalari qobiliyat darajasidan va, aksincha, qobiliyat darajasi qiyinlik darajasidan tashqaridaligini 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]:

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

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

bu yerda $\boldsymbol{E(X_{si})}$ - $\boldsymbol{X_{si}}$ning matematik kutilishi, $\boldsymbol{Var(X_{si})}$ – dispersiya, $\boldsymbol{n_i}$ – test topshiriqlari soni, $\boldsymbol{n_s}$ – 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 $\mathbf{1} \pm \frac{6}{\sqrt{n}}$ va $\mathbf{1} \pm \frac{2}{\sqrt{n}}$ dan foydalanish tavsiya qilinadi. Shunindek, [20] havolada ichki va tashqi mosliklar test tosphsirqlari soni va qiyinlik darajasiga bo‘g‘liqligi ko‘rsatilgan.

4-jadvalda milliy test tizimida fizika fanidan o‘tkazilgan test sinovlarida 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

№	ID	β	Infit	Outfit
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	O38B	1,084	1,057	1,028
50	O44B	1,275	0,870	0,785
51	O39A	1,343	0,898	0,808
52	O36A	1,543	0,989	1,066
53	O39B	1,836	0,930	0,827
54	O37B	2,282	0,978	0,890
55	O38A	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 topshriqlarining talab qilingan me’yorga juda yaqinligi, ushbu test topshiriqlarining to‘g‘ri tuzilganligi hamda tanlangan variantning tanlangan to‘plamga mos ekanligidan dalolat beradi.

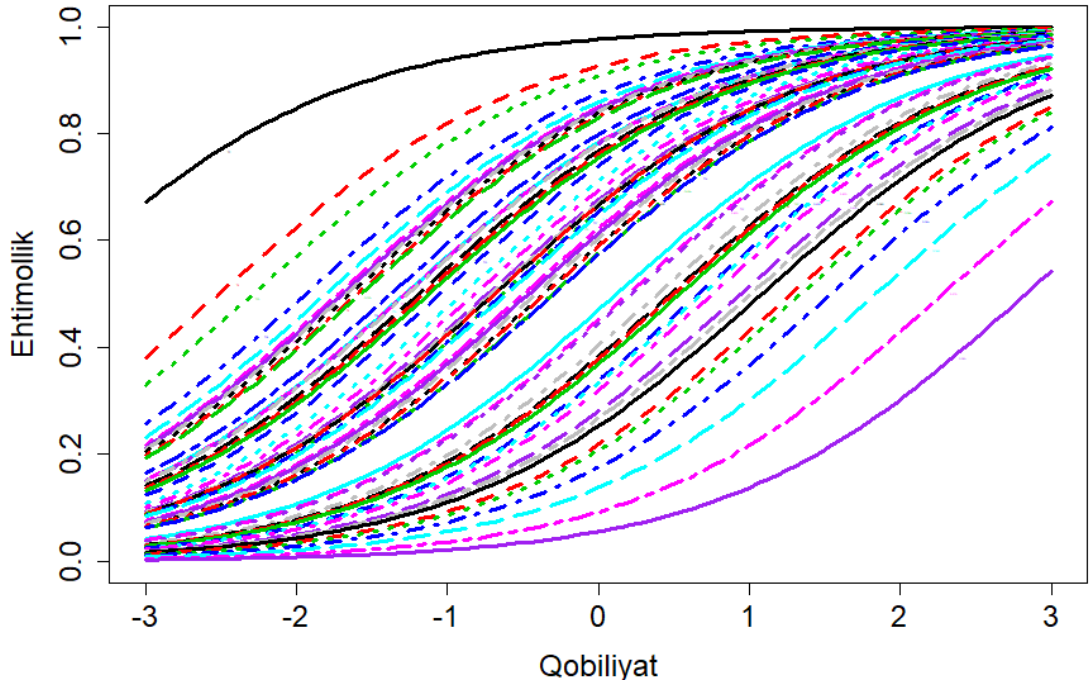
Test topshiriqlarining ichki va tashqi mosligiga o‘xshash jarayonni amalga oshirib, qobiliyatning ichki va tashqi moslik statistikasini tahlil qilish mumkin.

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 O38A 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 (O38A 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.

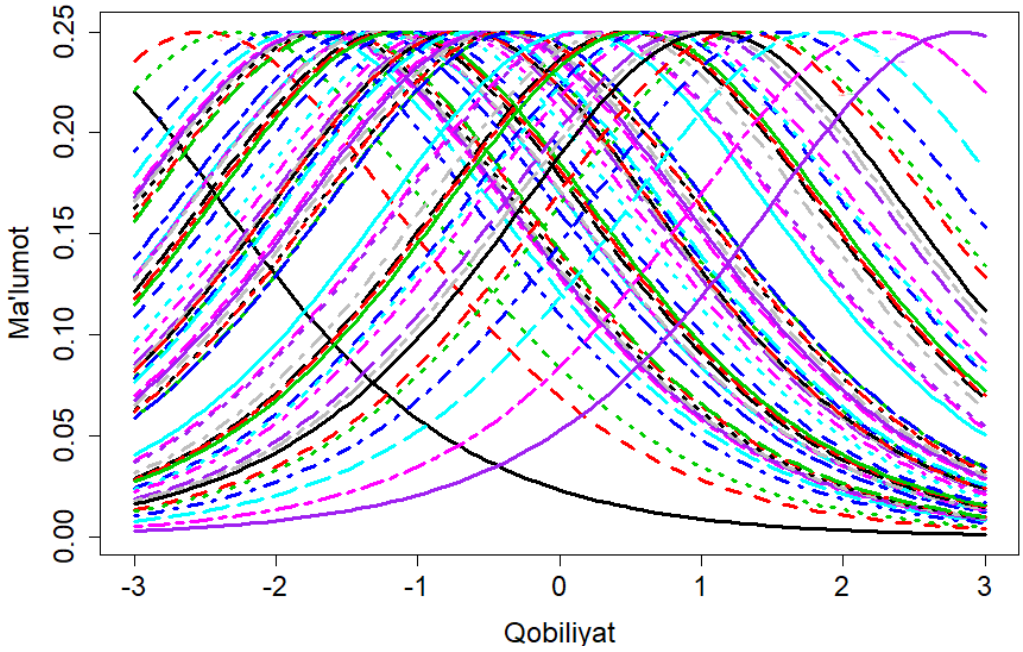
Bu holat zamonaviy test nazariyasidagi tenglashtirish va kalibrovkalash orqali yechiladi. Amaliyotda bu xalqaro tajribada ham [21] respublikamizda chet tillari bo‘yicha “multilevel” testlari natijalarini baholashda ham ishlatilmoqda [22]. Obyektiv baholashda xatoliklarni kamaytirish tartibli, kalibrovkalangan 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 ichonchliligi 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 qobi-liyatni 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 oralig'idagi ma'lumotlardir. Normal taqsimotga ko'ra sinaluvchilar qobiliyat darajalarining 99,7 foizi $(-\mathbf{3}:\mathbf{3})$ oralig'ida bo'ladi. Shuning uchun qiyinlik darajalari $(-\mathbf{3}:\mathbf{3})$ 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 $(\mathbf{0}:\mathbf{3})$ qobiliyat oralig'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, $(-\mathbf{3}:\mathbf{0})$ 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.

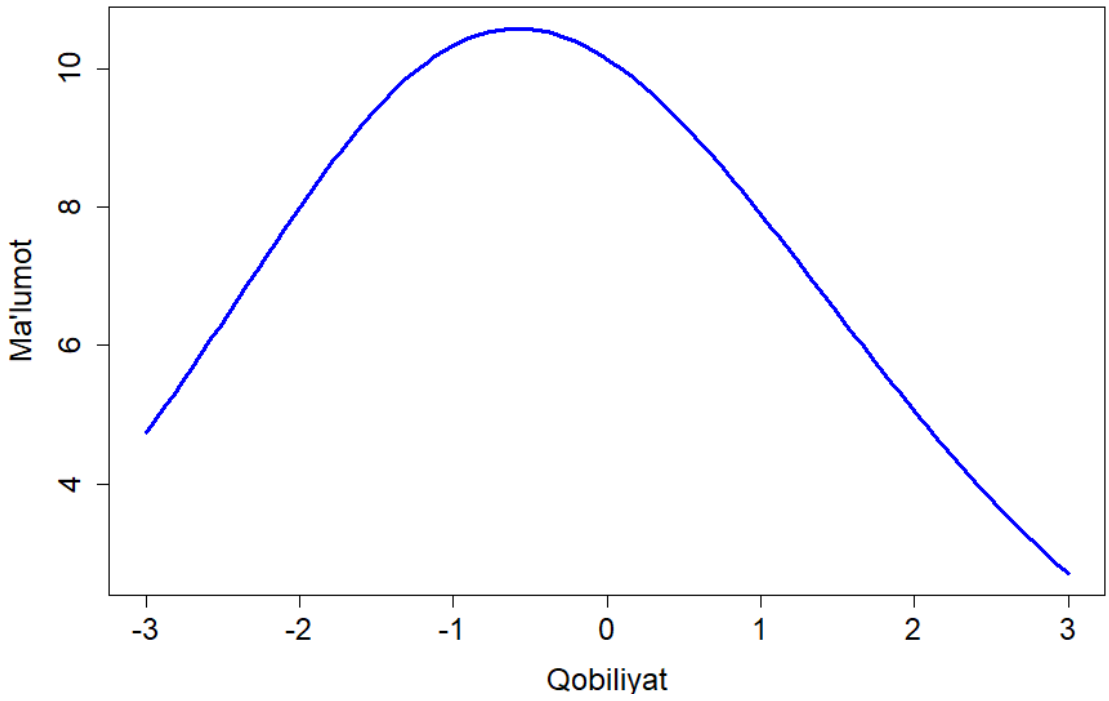


3-rasm. Element ma'lumoti chiziqlari

4-rasmda test ma'lumoti chizig'i (TMCh) keltirilgan. Testning umumiy ma'lumot miqdori 55 ga teng bo'lib, shundan $(-\mathbf{3}:\mathbf{3})$ oralig'idagi qobiliyatga ega bo'lganlar uchun ma'lumot miqdori 45,57 (82,88 foiz) ga teng. $(-\mathbf{3}:\mathbf{0})$ va $(\mathbf{0}:\mathbf{3})$ oraliqlardagi qobiliyatga 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 javoblarining 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 korrelya-tsiyasi 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 kalibrovkalan-gan 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.
7. Gunter Maris, Timo Bechger, Jesse Koops and Ivailo Parchev, 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 Rash 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 Rash 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.

FIZIKA FANIDAN TEST SINOVI NATIJALARINING STATISTIK TAHLILI

Q.A. Amonov, A.A. Baratov

O‘zbekiston Respublikasi Oliy ta’lim, fan va innovatsiyalar vazirligi huzuridagi Bilim va malakalarni baholash agentligi huzuridagi Ilmiy–o‘quv amaliy markazi, 100084, Toshkent sh., Bog‘ishamol k., 12

Qisqacha mazmuni. Ushbu maqolada umumiy o‘rta ta’lim maktablari, ixtisoslashtirilgan davlat umum ta’lim maktablari va akademik litseylarning 11-sinf bitiruvchi o‘quvchilaridan ilmiy tadqiqot maqsadida fizika fanidan olingan test sinovi natijalarining klassik test nazariyasi va Rash modeli asosidagi tahlillari o‘rganildi. Test sinovlarining statistik tahlillari asosida ularning o‘rta qiymati, moda va medianasi test ballarining umumiy dispersiyasi aniqlandi. Rash modeli asosida test topshiriqlarining qiyinlik darajalarining test topshiruvchilarning qobiliyat darajasiga mosligi, Rayt xaritasi o‘rganildi.

Kalit so‘zlar: Test topshiriqlari, Moda, Mediana, Standart tafovut, Dispersiya, Kronbax alfa koeffitsiyenti, validlik, qiyinlik darajasi, Rash modeli, Rayt xaritasi, qobiliyat darajalari.

I. Kirish

O‘quvchilarning bilimlarini uzluksiz nazorat qilish va baholash fan sohalarining o‘zlashtirish qiyin bo‘lgan bo‘limlarini va mavzularini aniqlashda hamda ta’lim sifatini oshirishda muhim ahamiyatga ega. O‘quvchilarning bilimlarini baholash uchun o‘tkaziladigan barcha pedagogik o‘lchovlar shaffof, xolis, imkon qadar obyektiv, tanlangan obyektdan va o‘lchov vositalaridan xoli bo‘lishi kerak.

O‘quvchilarning bilimlarini baholashda bu kabi mezonlarga qat’iy rioya qilish pedagog o‘qituvchilarimizdan qat’iyatlilikni talab qiladi. Bitta test varianti bilan bir nechta guruh o‘quvchilarning bilimlari tekshirilganda bu test tarkibidagi topshiriqlar har xil qiyinlik darajasiga ega ekanligini ko‘rsatadi, ya’ni test topshiriqlarining qiyinlik darajalari sinaluvchilar bilim va qobiliyatiga bog‘liq bo‘lib qoladi. Agarda bitta guruhda ishtirok etayotgan sinaluvchilarning bilim darajasi tekshirilayotganda bir qancha test variantlaridan foydalanilsa, sinaluvchilarning bilim va qobiliyatlari test topshiriqlarining qiyinlik darajasiga bog‘liq bo‘lib qoladi. Bundan tashqari sinaluvchilarning bilim va ko‘nikmalarini aniqlash uchun o‘tkaziladigan pedagogik o‘lchashlar obyektivligiga erishish uchun bir o‘lchovli shkalalardan foydalanish talab etiladi [1-3].

II. Test sinovlari natijalarining klassik test nazariyasi bo‘yicha tahlili

Test sinovlari natijalarini tahlil qilishning bir qancha usullari mavjud bo‘lib, ulardan biri klassik test nazariyasi hisoblanadi. Klassik test nazariyasi asosida aniqlangan statistik kattaliklar orqali test topshiriqlariga qo‘yilgan asosiy talab va ko‘rsatkichlar aniqlanadi. Ushbu ko‘rsatkichlar – test topshirig‘ining qiyinlik darajasi, test ballarining dispersiyasi (test topshirig‘ining boshqa test topshiriqlari bilan farqlanishi, o‘zgaruvchanligi), shuningdek, umumiy ballar yig‘indisi bilan korellyatsiyasidan iborat bo‘ladi. Test topshiriqlarining qiyinlik darajasini aniqlash usullaridan biri test topshirig‘ini empirik sinovdan o‘tkazish va to‘g‘ri javoblar ulushini aniqlashdan iboratdir. Test ballari (yoki to‘g‘ri javoblar)ning dispersiyasi test topshiruvchilarning tayyorgarlik darajasini aniqlashga, biladiganlarni bilmaydiganlardan ajratishga imkon beradi.

Pedagogik o‘lchovlarda klassik test nazariyasining asosiy statistik kattaliklari qatoriga o‘rta qiymat, gistogrammani qurish, moda va mediana kabi ko‘rsatkichlarni hisoblash hamda test ballarining umumiy dispersiyasi (standart tafovut) ko‘rsatkichlari ham kiradi [4-8].

Test ballari (yoki to‘g‘ri javoblar)ning o‘rta arifmetik qiymati va bir qancha statistik tahlil orqali aniqlangan kattaliklari fanlar, ta’lim muassasalari va boshqa muhim belgilar kesimida ham o‘rganiladi. Bu ko‘rsatkichlar test ballari o‘rtasidagi tafovutni umumlashtiradi, ularning o‘ziga xos bo‘lgan qonuniyatni ochib beradi. Test sinovi natijalari asosida aniqlangan test ballari taqsimotining gistogrammasi quriladi va u normal taqsimotga yaqin yoki uzoqligi baholanadi. Gistogrammaning normal taqsimotga yaqinligi test sifatining yaxshiligini va test sinovlarining obyektiv o‘tkazilganligini bildiradi. Test ballarining eng ko‘p takrorlanadigan qiymati statistik tahlilda 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 bir-biridan qanchalik ko‘p farq qilsa, ballar taqsimoti normal taqsimotdan shuncha uzoqda bo‘ladi [9].

Ushbu maqolada fizika fanidan umumiy o‘rta ta’lim maktablarining 11-sinf o‘quvchilaridan ilmiy tadqiqot uchun olingan test sinovi natijalarining klassik test nazariyasi va Rash modeli asosidagi tahlillari keltirilgan.

Fizika fanidan o‘tkazilgan test sinovlarida foydalanilgan test varianti 48 ta (38 ta yopiq va 10 ta ochiq test topshiriqlari) test topshiriqlaridan iborat bo‘lib, ajratilgan vaqt javoblar varaqasini bo‘yash bilan birgalikda 120 daqiqani tashkil etdi. Test sinovida Respublikamiz miqyosida jami 415 nafar 11-sinf o‘quvchilari ishtirok etdi. Sinaluvchilarning mazkur test topshiriqlarini yechishda to‘plangan ballarning o‘rta qiymati:

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad (1)$$

ifoda orqali hisoblandi (1-jadval), bu yerda *N*- test topshiruvchilar soni, X_i- sinaluvchilarning to‘plagan ballari.

Test topshiruvchilarning tayyorgarlik darajasini aniqlash uchun ularning individual to‘plagan ballari o‘zaro bir-biridan farq qilish kerak, agar barcha individual ballar mos kelsa, u holda farq nolga teng bo‘ladi. Agar individual ballar o‘zaro mos kelmasa, u holda sinov natijalarining o‘zgarishi o‘rtacha qiymatidan tafovutni yuzaga keltiradi. Bu tafovutlar ijobiy yoki salbiy bo‘lishi mumkin. Barcha tafovutlarning yig‘indisi esa nolga teng bo‘ladi. Shuning uchun test ballarining o‘zgarishini tavsiflash uchun kvadrat tafovutlardan foydalaniladi. Kvadrat tafovutlar yig‘indisi sinaluvchilar soniga (*N*) bog‘liq bo‘lib bu bog‘liqlikdan xalos bo‘lish uchun *N* ga teskari bo‘lgan bog‘liqlikdan foydalanamiz. Test topshirig‘ining boshqa test topshiriqlaridan farqlanishi, ya’ni test ballarining dispersiyasi uchun quyidagi ifodadan foydalanamiz [10]:

$$s_x^2 = \frac{1}{N(N-1)}\left(N\sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2\right) \qquad (2).$$

Test ballarining dispersiyasi $s_x^2 = 57,205$ ga teng.

Dispersiya bilan bog‘liq bo‘lgan yana bir muhim parametrlardan biri bu – standart tafovut bo‘lib u quyidagiga teng $s_x = \sqrt{s_x^2}$ [10]. Aynan bizning holatimizda standart tafovut 7,56 ga teng.

Test ballari dispersiyasining qiymati kichik bo‘lsa, sinaluvchilarning tayyorgarlik darajasini farqlash qiyin bo‘ladi va ularni maqbul aniqlik bilan reytingga kiritishga imkon bermaydi. Juda katta dispersiya qiymati esa test jarayonining mumkin bo‘lgan buzilishlarini, topshiriqlarning yetarli darajada aniq emasligini va boshqalarni ko‘rsatadi.

Statistik tahlil natijalari asosida fizika fanidan test sinovi ballarining o‘rta qiymati, medianasi, modasi, standart xatoligi, dispersiyasi, diapazoni, maksimum, minimum qiymatlari va test sinovi ballarining taqsimoti (gistogrammada normal taqsimot bilan) 1-jadvalda keltirilgan. Shuningdek, eng muhim ko‘rsatkichlardan biri ishonchlilik koeffitsiyenti ya’ni, Kronbax alfa koeffitsiyenti ham taqdim etilgan.

Statistik tadqiqot natijalariga ko‘ra, fizika fanidan test variantining ishonchlilik koeffitsiyenti, ya’ni Kronbax alfa koeffitsiyenti quyidagi ifoda orqali hisoblanadi [11]:

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum s_i^2}{s_x^2}\right) \quad (3)$$

bu yerda *K*-test topshiriqlar soni, s_i^2 –alohida olingan test topshirig‘i dispersiyasi s_x^2 –butun test dispersiyasi.

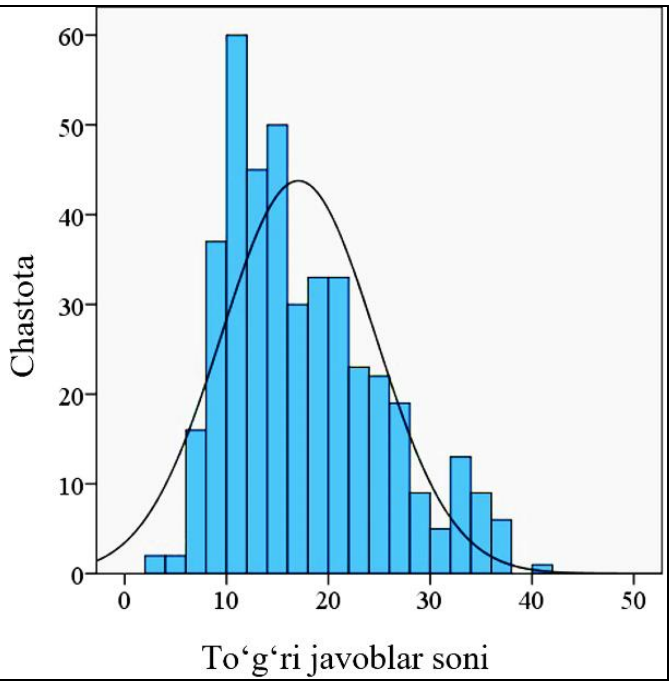
Statistik tahlil natijalariga asosan Kronbax alfa koeffitsiyentining qiymati 0,85 ga teng ekanligi aniqlandi. Kronbax alfa koeffitsiyentining 0,7 dan kattaligi ekanligi ushbu test variantining ishonchliligi yaxshiligini ko‘rsatmoqda [10].

Gistogrammadan ko‘rish mumkunki, test topshiriqlarining individual ballari taqsimoti normal taqsimotdan farq qiladi (moda – 10 va o‘rta qiymat - 17,05). Sinaluvchilarning 53,7 foizi 15 ball va undan past ball to‘plagan (medianadan pastda), 46,3 foiz test topshiruvchilar 15 balldan yuqori ball to‘plagan (medianadan yuqorida). Kichik hajmdagi tanlovlarda normal taqsimot markaziga o‘rta qiymatning yaqinligi, qo‘llanilgan test variantining maqsadiga ko‘ra to‘g‘ri tuzilganligini ko‘rsatadi.

1-jadval

Fizika fanidan test sinovi natijalarining tavsif statistikasi ma’lumotlari va natijalari bo‘yicha to‘g‘ri javoblar taqsimoti

Test topshiruvchilar soni	415
O‘rta qiymat	17,05
Mediana	15
Moda	10
Standart tafovut	7,56
Dispersiya	57,205
Asimmetriya	0,756
Ekstsess	-0,091
Diapazon	38
Minimum	3
Maksimum	41
Kronbax alfasi	0,85



Test topshiriqlarining qiyinlik darajalari tahlil qilinganda 48 ta test topshirig‘idan bittasi 1-qiyinlik darajasidagi, 33 tasi 2-qiyinlik darajasidagi va 14 tasi 3-qiyinlik darajasidagi test topshirig‘idan iborat ekanligi aniqlandi (2-jadval).

2-jadval

Fizika fanidan o‘tkazilgan test sinovi natijalari bo‘yicha test topshiriqlarining aniqlangan qiyinlik darajalari

№	ID	X_{max}	X_i	Ans (foizda)	V
1	F000001	415	312	75,18	1
2	F000003	415	276	66,51	2
3	F000018	415	269	64,82	2
4	F000047	415	259	62,41	2
5	F000021	415	244	58,80	2
6	F000009	415	228	54,94	2
7	F000002	415	226	54,46	2
8	F000017	415	224	53,98	2
9	F000046	415	217	52,29	2
10	F000007	415	213	51,33	2
11	F000005	415	204	49,16	2
12	F000012	415	199	47,95	2
13	F000011	415	191	46,02	2
14	F000004	415	186	44,82	2
15	F000014	415	183	44,10	2
16	F000033	415	183	44,10	2
17	F000045	415	181	43,61	2
18	F000035	415	180	43,37	2
19	F000037	415	169	40,72	2
20	F000031	415	168	40,48	2
21	F000048	415	163	39,28	2
22	F000032	415	158	38,07	2
23	F000008	415	154	37,11	2
24	F000028	415	150	36,14	2
25	F000026	415	138	33,25	2
26	F000006	415	133	32,05	2
27	F000015	415	133	32,05	2

28	F000041	415	131	31,57	2
29	F000043	415	121	29,16	2
30	F000016	415	118	28,43	2
31	F000030	415	117	28,19	2
32	F000019	415	112	26,99	2
33	F000024	415	111	26,75	2
34	F000040	415	110	26,51	2
35	F000013	415	103	24,82	3
36	F000010	415	100	24,10	3
37	F000025	415	96	23,13	3
38	F000036	415	80	19,28	3
39	F000022	415	77	18,55	3
40	F000034	415	72	17,35	3
41	F000042	415	72	17,35	3
42	F000039	415	71	17,11	3
43	F000023	415	58	13,98	3
44	F000027	415	58	13,98	3
45	F000044	415	45	10,84	3
46	F000029	415	34	8,19	3
47	F000038	415	34	8,19	3
48	F000020	415	16	3,86	3

Test topshiriqlari natijalarining klassik test nazariyasida element – umumiy ball korrelyatsiya koeffitsiyentini hisoblash orqali validlik baholanadi. Ushbu koeffitsiyentni aniqlash uchun quyida keltirilgan ifodadan foydalanamiz [10]:

$$r_{kk} = \frac{\bar{X}_1 - \bar{X}_0}{s_x}\sqrt{\frac{n_1 \cdot n_0}{n(n-1)}} \qquad (4)$$

bu yerda $\bar{X}_1$-berilgan test topshirig‘ini bajargan sinaluvchilarning o‘rtacha individual bali, $\bar{X}_0$- berilgan test topshirig‘ini bajara olmagan sinaluvchilarning o‘rtacha individual bali, n_1- test topshirig‘iga to‘g‘ri javob bergan sinaluvchilarning soni, n_0-test topshirig‘iga noto‘g‘ri javob bergan

sinaluvchilar soni, $n=n_1+n_0$- sinaluvchilarning umumiy soni, s_x – standart tafovut.

Bunda har bir test topshirig‘iga berilgan javoblarning umumiy test bali bilan korrelyatsiyasi element–umumiy ball korrelyatsiya koeffitsiyentiga teng bo‘ladi. Umuman olganda, element–umumiy ball korrelyatsiya koeffitsiyenti 0,5 va undan katta bo‘lgan topshiriq valid hisoblanadi. Statistik tadqiqot natijalariga ko‘ra element–umumiy ball korrelyatsiya koeffitsiyenti manfiy bo‘lgan topshiriqlar mazkur variantdan chiqarilishi belgilangan.

Element–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. 4-jadvalda Element–umumiy ball korrelyatsiya koeffitsiyenti qiymati manfiy bo‘lganda tayyorgarlik darajalari past bo‘lgan sinaluvchilar g‘olib bo‘lib, bilim darajalari yuqori bo‘lgan sinaluvchilar test topshiriqlarini yechishda noto‘g‘ri javobni tanlaydilar yoki ularni o‘tkazib yuboradilar.

3-jadval

Individual test topshiriqlariga berilgan javoblarning umumiy ball bilan korrelyatsiyalari

№	*ID*	EUBKK	*№*	*ID*	EUBKK
1	F000036	-0,132	25	F000048	0,391
2	F000022	-0,055	26	F000019	0,392
3	F000039	-0,053	27	F000046	0,396
4	F000023	-0,023	28	F000001	0,404
5	F000024	-0,021	29	F000014	0,425
6	F000040	0,007	30	F000011	0,434
7	F000025	0,019	31	F000045	0,439
8	F000026	0,045	32	F000035	0,449
9	F000030	0,086	33	F000047	0,455
10	F000043	0,153	34	F000002	0,465
11	F000015	0,161	35	F000032	0,469
12	F000028	0,214	36	F000007	0,472
13	F000034	0,276	37	F000033	0,478
14	F000020	0,279	38	F000017	0,511
15	F000012	0,286	39	F000038	0,534

16	F000031	0,315	40	F000009	0,536
17	F000037	0,324	41	F000005	0,575
18	F000041	0,325	42	F000029	0,578
19	F000018	0,356	43	F000044	0,624
20	F000003	0,369	44	F000042	0,649
21	F000021	0,38	45	F000013	0,661
22	F000027	0,386	46	F000016	0,672
23	F000006	0,387	47	F000004	0,692
24	F000008	0,388	48	F000010	0,708

Olib borilgan statistik tadqiqot natijalariga ko‘ra test sinovida foydalanilgan test topshiriqlarining 5 tasi (F000022, F000023, F000024, F000036 va F000039–ID raqamli test topshiriqlari) umumiy ball bilan korrelyatsiya koeffitsiyenti manfiy qiymatini qabul qilgan. 3- jadvaldan ko‘rinib turibdiki, ID raqami F000040, F000025, F000026, F000030, F000043 va F000015 hamda F000028 bo‘lgan test topshiriqlarining element–umumiy ball korrelyatsiya koeffitsiyenti kichik ekanligi aniqlandi.

III. Test sinovlari natijalarining Rash modeli asosida tahlili

Jahonning ko‘plab rivojlangan davlatlarida pedagogik o‘lchov vositalarining sifatini aniqlashda Rash modeli asosida matematik-statistik tadqiqotlar olib borilmoqda [12-14]. Rash modeli asosidagi matematik-statistik tahlillar o‘lchanayotgan xususiyatlarga obyektiv va xolis yondashuvni ta’minlab beradi. Ushbu model daniyalik olim Jorg Rash tomonidan yaratilgan bo‘lib, bir o‘lchovlilikni ta’minlash uchun oldindan tayyorgarlik ko‘rish bu ishlar qanchalik amalga oshirilganini empirik usullar bilan tekshirish imkoni beradi. Rash modelining muhim xususiyati u shunchaki ma’lumotlarni tahlil qilish uchun statistik usul emas, balki o‘lchovning nimaligini, ta’lim tizimida o‘lchovlarni sifatli amalga oshirish imkoniyatini ham beradi [12,13].

Test topshiruvchilarning yashirin qobiliyati va test topshiriqlarining qiyinlik darajasi kabi parametrlarini ham Rash modeli orqali ochib berish mumkin. Bu ikki kattalikdan birinchisi o‘zgaruvchi sifatida, ikkinchisi esa parametr sifatida kiritiladi. Chunki test topshiruvchilarning qobiliyati (bilimi) bu modelda elementlarga, ya’ni topshiriqlarga berilgan javoblarga qarab belgilanadi, shuning uchun topshiriqlarning qiyinlik darajasini parameter sifatida qarash qulay. Rash modeliga ko‘ra, dixotomik elementlarga individual javoblar shaxsning qobiliyat darajasi va element qiyinligi bilan aniqlanadi. Ma’lum bir

qobiliyatga ega bo‘lgan shaxsning ma’lum bir qiyinlikdagi elementga to‘g‘ri javob berish ehtimolligini aniqlaydi. Bu quyidagi matematik formula orqali ifodalanadi [13]:

$$P(X_{is} = 1|\theta_s, b_i) = \frac{e^{\theta_s - b_i}}{1 + e^{\theta_s - b_i}} \quad (5)$$

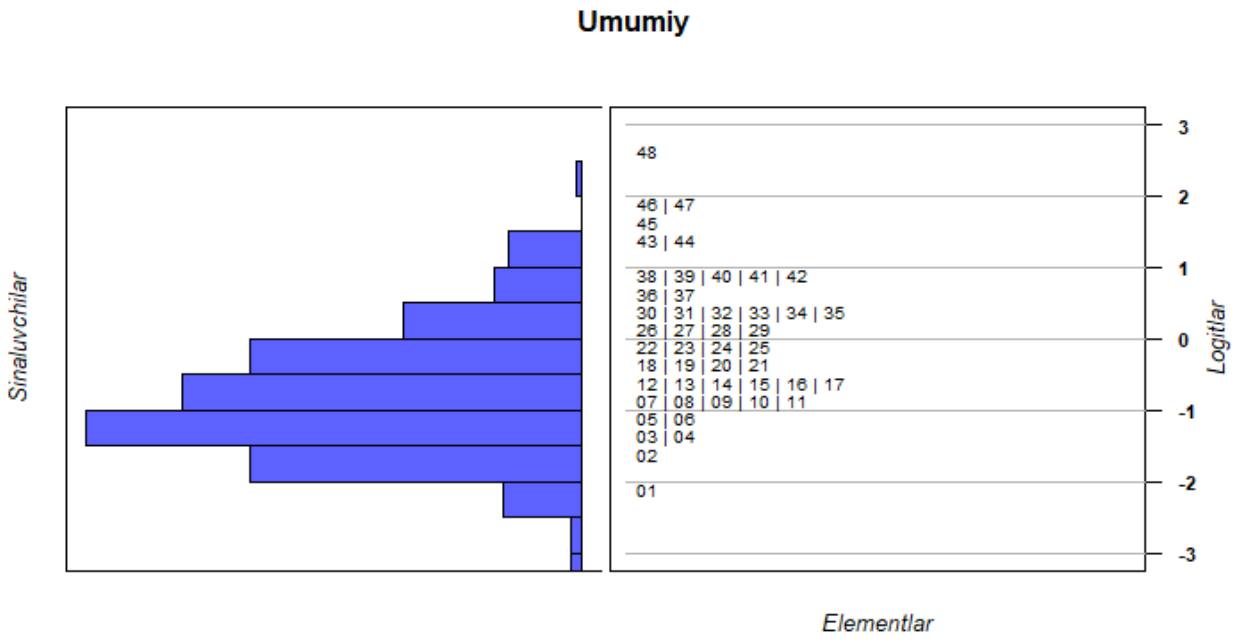
Bu yerda X_{is} = 1, s-o‘quvchining i-elementga to‘g‘ri javob berish ehtimolligi, θ_s-qobiliyat o‘zgaruvchisi, b_i–topshiriq qiyinlik darajasi.

Ushbu (5) ifodaga asosan to‘g‘ri javoblar ehtimolligi qobiliyat va qiyinlik kabi o‘zgaruvchilarining farqiga bog‘liq, bu esa qobiliyat va qiyinlik o‘zgaruvchilariga ixtiyoriy o‘zgarmas son qo‘shilganda ehtimollik o‘zgarmasligini bildiradi. Rash modeli asosida aniqlangan qiyinlik darajalarini sinaluvchilar qobiliyatlariga qanchalik mosligini Rayt xaritasi [15] yordamida tahlil qilish mumkin [16-20].

Rayt xaritasi – test topshiriqlarining qiyinlik darajalari va sinaluvchilarning qobiliyat darajalarining o‘zaro mos kelishini aniqlovchi diagramma hisoblanadi [15].

Fizika fanidan test sinovlari natijalarining Rash modeli bo‘yicha tahlilini maxsus dastur asosida amalga oshirish uchun ishlab chiqilgan turli xil dasturiy ta’minotlardan foydalanildi.

1-rasmda fizika fanidan o‘tkazilgan test sinovlarining Rash modeli asosida aniqlangan qiyinlik darajalarini sinaluvchilar qobiliyat darajalariga mosligi (Rayt xaritasi) ko‘rsatilgan.



1-Rasm. Fizika fanidan topshirilgan test natijalari asosida aniqlangan qobiliyat va qiyinlik darajalarining mosligi (Rayt xaritasi).

1-rasmdan ko‘rinib turibdiki, Rayt xaritasida test topshiriqlarining ko‘p sondagi qismi talabgorlarning ko‘p sondagi qismiga deyarli mos keladi, bu esa sinaluvchilarning ko‘p qismi uchun qobiliyatga mos test topshiriqlari mavjud ekanligini bildiradi. Sinaluvchilar haqida olinadigan ma‘lumot miqdori test topshiriqlari soniga ham bog‘liq bo‘ladi. Shuning uchun test topshiriqlari shakllantirilganda imkon qadar har bir sinaluvchining qobiliyat darajasiga mos test topshiriqlarini ishlab chiqish maqsadga muvofiqdir.

Rayt xaritasidan foydalanib test topshiriqlarini -3 va 3 oralig‘ida bir xil taqsimlash va bu orqali test variantida test topshiriqlarining qiyinlik darajalari hamda qobiliyat darajalari taqsimotlarini bir-biriga yaqin bo‘lishiga erishish mumkin. 1–rasmdagi keltirilgan Rayt xaritasidan fizika fanidan test topshiriqlarining (1:3) va (-1:-3) logit birligi oralig‘idagi qobiliyat darajalari uchun test topshiriqlari soni juda kam ekanligini ko‘rish mumkin.

Hisoblashlarga ko‘ra qobiliyat dar ajalari -2,02 va 2,74 logit birligi oralig‘ida taqsimlanganligini ko‘rish mumkin.

Quyida fizika fanidan o‘tkazilgan test sinovi natijalarining Rash modeli bilan aniqlangan qiyinlik darajalari keltirilgan (4-jadval):

4-jadval

Rash modeli bilan aniqlangan qiyinlik darajalari

№	ID	b	№	ID	b
1	F000001	-2,02	25	F000026	-0,01
2	F000003	-1,56	26	F000006	0,06
3	F000018	-1,48	27	F000015	0,06
4	F000047	-1,37	28	F000041	0,08
5	F000021	-1,20	29	F000043	0,21
6	F000009	-1,02	30	F000016	0,25
7	F000002	-1,00	31	F000030	0,26
8	F000017	-0,98	32	F000019	0,33
9	F000046	-0,90	33	F000024	0,35
10	F000007	-0,86	34	F000040	0,36
11	F000005	-0,76	35	F000013	0,46
12	F000012	-0,71	36	F000010	0,51
13	F000011	-0,62	37	F000025	0,57
14	F000004	-0,56	38	F000036	0,83
15	F000014	-0,53	39	F000022	0,88

16	F000033	-0,53	40	F000034	0,97
17	F000045	-0,51	41	F000042	0,97
18	F000035	-0,50	42	F000039	0,99
19	F000037	-0,37	43	F000023	1,25
20	F000031	-0,36	44	F000027	1,25
21	F000048	-0,30	45	F000044	1,57
22	F000032	-0,25	46	F000038	1,90
23	F000008	-0,21	47	F000029	1,90
24	F000028	-0,15	48	F000020	2,74

Rash modeli asosida nafaqat sinaluvchilarning qobiliyat va test topshiriqlarining qiyinlik darajalarini, balki sinaluvchilar to‘plagan xom ballarini yagona shkaladagi qiymatlarini ham quyidagicha aniqlashi mumkin:

$$T = 50 + 10\left(\frac{X - M}{SD}\right) \qquad (6)$$

bu yerda X – xom ball, M – testda qayd etilgan o‘rtacha ball, SD – esa standart tafovut.

Quyidagi 5–jadvalda Rash modelida aniqlangan sinaluvchilarning xom ballari, qobiliyat darajasi – β, va T-ballari keltirilgan.

5-*jadval*

Sinaluvchilarning Rash modelida aniqlangan xom ballari-X, qobiliyat darajasi – β, va T-ballari

№	ID	X	β	T	№	ID	X	β	T
1	0000367	3	-3,13	23,27	209	000112	15	-0,93	48,56
2	0000368	3	-3,13	23,27	210	000182	15	-0,93	48,56
3	0000366	4	-2,78	27,27	211	000192	15	-0,93	48,56
4	0000036	5	-2,54	30,00	212	000199	15	-0,93	48,56
5	0000055	6	-2,34	32,29	213	0000041	16	-0,71	50,98
6	0000220	6	-2,34	32,29	214	0000049	16	-0,71	50,98

7	0000233	6	-2,34	32,29	215	000016	16	-0,71	50,98
8	0000266	6	-2,34	32,29	216	000022	16	-0,71	50,98
9	0000031	7	-2,11	34,98	217	0000270	16	-0,71	50,98
10	0000059	7	-2,11	34,98	218	0000303	16	-0,71	50,98
11	000006	7	-2,11	34,98	219	0000308	16	-0,71	50,98
12	0000219	7	-2,11	34,98	220	0000313	16	-0,71	50,98
13	0000232	7	-2,11	34,98	221	0000319	16	-0,71	50,98
14	0000236	7	-2,11	34,98	222	0000320	16	-0,71	50,98
15	0000238	7	-2,11	34,98	223	0000323	16	-0,71	50,98
16	0000268	7	-2,11	34,98	224	0000325	16	-0,71	50,98
17	0000369	7	-2,11	34,98	225	0000328	16	-0,71	50,98
18	0000401	7	-2,11	34,98	226	0000393	16	-0,71	50,98
19	000099	7	-2,11	34,98	227	0000413	16	-0,71	50,98
20	000101	7	-2,11	34,98	228	0000414	16	-0,71	50,98
21	000003	8	-1,85	37,91	229	000080	16	-0,71	50,98
22	0000033	8	-1,85	37,91	230	000116	16	-0,71	50,98
23	0000035	8	-1,85	37,91	231	000124	16	-0,71	50,98
24	0000043	8	-1,85	37,91	232	000174	16	-0,71	50,98
25	0000057	8	-1,85	37,91	233	0000327	17	-0,59	52,42
26	000010	8	-1,85	37,91	234	0000344	17	-0,59	52,42
27	0000235	8	-1,85	37,91	235	0000396	17	-0,59	52,42
28	0000262	8	-1,85	37,91	236	0000417	17	-0,59	52,42
29	0000301	8	-1,85	37,91	237	000078	17	-0,59	52,42
30	0000361	8	-1,85	37,91	238	000115	17	-0,59	52,42
31	0000363	8	-1,85	37,91	239	000138	17	-0,59	52,42

32	0000375	8	-1,85	37,91	240	000141	17	-0,59	52,42
33	000197	8	-1,85	37,91	241	000184	17	-0,59	52,42
34	0000032	9	-1,66	40,13	242	000193	17	-0,59	52,42
35	0000034	9	-1,66	40,13	243	0000050	18	-0,53	53,09
36	0000046	9	-1,66	40,13	244	000013	18	-0,53	53,09
37	000005	9	-1,66	40,13	245	000026	18	-0,53	53,09
38	0000056	9	-1,66	40,13	246	0000310	18	-0,53	53,09
39	0000058	9	-1,66	40,13	247	0000311	18	-0,53	53,09
40	0000060	9	-1,66	40,13	248	0000338	18	-0,53	53,09
41	000015	9	-1,66	40,13	249	0000342	18	-0,53	53,09
42	0000211	9	-1,66	40,13	250	0000360	18	-0,53	53,09
43	0000212	9	-1,66	40,13	251	0000403	18	-0,53	53,09
44	0000229	9	-1,66	40,13	252	0000405	18	-0,53	53,09
45	000023	9	-1,66	40,13	253	0000418	18	-0,53	53,09
46	0000237	9	-1,66	40,13	254	000068	18	-0,53	53,09
47	0000240	9	-1,66	40,13	255	000069	18	-0,53	53,09
48	0000244	9	-1,66	40,13	256	000083	18	-0,53	53,09
49	0000263	9	-1,66	40,13	257	000086	18	-0,53	53,09
50	0000267	9	-1,66	40,13	258	000097	18	-0,53	53,09
51	0000316	9	-1,66	40,13	259	000109	18	-0,53	53,09
52	0000371	9	-1,66	40,13	260	000117	18	-0,53	53,09
53	000091	9	-1,66	40,13	261	000118	18	-0,53	53,09
54	000102	9	-1,66	40,13	262	000126	18	-0,53	53,09
55	000104	9	-1,66	40,13	263	000143	18	-0,53	53,09
56	000110	9	-1,66	40,13	264	000156	18	-0,53	53,09

57	000111	9	-1,66	40,13	265	000161	18	-0,53	53,09
58	0000044	10	-1,55	41,38	266	0000337	19	-0,50	53,39
59	0000045	10	-1,55	41,38	267	0000357	19	-0,50	53,39
60	0000052	10	-1,55	41,38	268	0000395	19	-0,50	53,39
61	0000053	10	-1,55	41,38	269	0000397	19	-0,50	53,39
62	0000214	10	-1,55	41,38	270	0000406	19	-0,50	53,39
63	0000224	10	-1,55	41,38	271	000123	19	-0,50	53,39
64	0000230	10	-1,55	41,38	272	000137	19	-0,50	53,39
65	0000239	10	-1,55	41,38	273	000144	19	-0,50	53,39
66	0000241	10	-1,55	41,38	274	000170	19	-0,50	53,39
67	0000243	10	-1,55	41,38	275	000198	19	-0,50	53,39
68	0000246	10	-1,55	41,38	276	0000302	20	-0,49	53,60
69	0000253	10	-1,55	41,38	277	0000312	20	-0,49	53,60
70	0000255	10	-1,55	41,38	278	0000340	20	-0,49	53,60
71	0000256	10	-1,55	41,38	279	0000341	20	-0,49	53,60
72	0000259	10	-1,55	41,38	280	0000392	20	-0,49	53,60
73	0000260	10	-1,55	41,38	281	0000415	20	-0,49	53,60
74	0000261	10	-1,55	41,38	282	000088	20	-0,49	53,60
75	0000269	10	-1,55	41,38	283	000136	20	-0,49	53,60
76	0000305	10	-1,55	41,38	284	000151	20	-0,49	53,60
77	0000362	10	-1,55	41,38	285	000153	20	-0,49	53,60
78	0000374	10	-1,55	41,38	286	000162	20	-0,49	53,60
79	0000378	10	-1,55	41,38	287	000166	20	-0,49	53,60
80	0000411	10	-1,55	41,38	288	000177	20	-0,49	53,60
81	000103	10	-1,55	41,38	289	000178	20	-0,49	53,60

82	000106	10	-1,55	41,38	290	000179	20	-0,49	53,60
83	000114	10	-1,55	41,38	291	000186	20	-0,49	53,60
84	000183	10	-1,55	41,38	292	000203	20	-0,49	53,60
85	000191	10	-1,55	41,38	293	000207	20	-0,49	53,60
86	000194	10	-1,55	41,38	294	000011	21	-0,46	53,91
87	000195	10	-1,55	41,38	295	0000321	21	-0,46	53,91
88	000206	10	-1,55	41,38	296	0000324	21	-0,46	53,91
89	000209	10	-1,55	41,38	297	0000343	21	-0,46	53,91
90	0000038	11	-1,49	42,06	298	000066	21	-0,46	53,91
91	000008	11	-1,49	42,06	299	000125	21	-0,46	53,91
92	000009	11	-1,49	42,06	300	000130	21	-0,46	53,91
93	0000215	11	-1,49	42,06	301	000132	21	-0,46	53,91
94	0000218	11	-1,49	42,06	302	000150	21	-0,46	53,91
95	0000221	11	-1,49	42,06	303	000155	21	-0,46	53,91
96	0000222	11	-1,49	42,06	304	000159	21	-0,46	53,91
97	0000223	11	-1,49	42,06	305	000167	21	-0,46	53,91
98	0000225	11	-1,49	42,06	306	000171	21	-0,46	53,91
99	0000226	11	-1,49	42,06	307	000176	21	-0,46	53,91
100	0000227	11	-1,49	42,06	308	000180	21	-0,46	53,91
101	0000228	11	-1,49	42,06	309	0000329	22	-0,40	54,61
102	0000231	11	-1,49	42,06	310	0000330	22	-0,40	54,61
103	0000242	11	-1,49	42,06	311	0000359	22	-0,40	54,61
104	0000247	11	-1,49	42,06	312	0000416	22	-0,40	54,61
105	0000257	11	-1,49	42,06	313	000142	22	-0,40	54,61
106	0000258	11	-1,49	42,06	314	000152	22	-0,40	54,61

107	0000365	11	-1,49	42,06	315	000154	22	-0,40	54,61
108	0000373	11	-1,49	42,06	316	000160	22	-0,40	54,61
109	0000380	11	-1,49	42,06	317	000163	22	-0,40	54,61
110	0000409	11	-1,49	42,06	318	000165	22	-0,40	54,61
111	000090	11	-1,49	42,06	319	000181	22	-0,40	54,61
112	000092	11	-1,49	42,06	320	000187	22	-0,40	54,61
113	000094	11	-1,49	42,06	321	000279	22	-0,40	54,61
114	000100	11	-1,49	42,06	322	0000358	23	-0,27	56,08
115	000105	11	-1,49	42,06	323	0000412	23	-0,27	56,08
116	000205	11	-1,49	42,06	324	000075	23	-0,27	56,08
117	000210	11	-1,49	42,06	325	000082	23	-0,27	56,08
118	0000039	12	-1,44	42,70	326	000128	23	-0,27	56,08
119	000004	12	-1,44	42,70	327	000164	23	-0,27	56,08
120	0000040	12	-1,44	42,70	328	000168	23	-0,27	56,08
121	0000042	12	-1,44	42,70	329	000173	23	-0,27	56,08
122	0000048	12	-1,44	42,70	330	000175	23	-0,27	56,08
123	0000051	12	-1,44	42,70	331	000189	23	-0,27	56,08
124	000018	12	-1,44	42,70	332	000019	24	-0,06	58,54
125	0000216	12	-1,44	42,70	333	0000394	24	-0,06	58,54
126	000025	12	-1,44	42,70	334	0000399	24	-0,06	58,54
127	0000264	12	-1,44	42,70	335	0000408	24	-0,06	58,54
128	0000306	12	-1,44	42,70	336	000079	24	-0,06	58,54
129	0000351	12	-1,44	42,70	337	000085	24	-0,06	58,54
130	0000377	12	-1,44	42,70	338	000121	24	-0,06	58,54
131	0000379	12	-1,44	42,70	339	000127	24	-0,06	58,54

132	0000382	12	-1,44	42,70	340	000147	24	-0,06	58,54
133	0000388	12	-1,44	42,70	341	000158	24	-0,06	58,54
134	0000389	12	-1,44	42,70	342	000169	24	-0,06	58,54
135	000096	12	-1,44	42,70	343	000172	24	-0,06	58,54
136	000098	12	-1,44	42,70	344	000200	24	-0,06	58,54
137	000119	12	-1,44	42,70	345	0000332	25	0,18	61,26
138	000185	12	-1,44	42,70	346	0000333	25	0,18	61,26
139	000190	12	-1,44	42,70	347	0000334	25	0,18	61,26
140	000204	12	-1,44	42,70	348	0000347	25	0,18	61,26
141	0000037	13	-1,34	43,80	349	000076	25	0,18	61,26
142	0000054	13	-1,34	43,80	350	000131	25	0,18	61,26
143	000007	13	-1,34	43,80	351	000133	25	0,18	61,26
144	000012	13	-1,34	43,80	352	000145	25	0,18	61,26
145	000017	13	-1,34	43,80	353	000157	25	0,18	61,26
146	000020	13	-1,34	43,80	354	0000353	26	0,35	63,20
147	000021	13	-1,34	43,80	355	000070	26	0,35	63,20
148	0000213	13	-1,34	43,80	356	000081	26	0,35	63,20
149	0000254	13	-1,34	43,80	357	000129	26	0,35	63,20
150	0000265	13	-1,34	43,80	358	000135	26	0,35	63,20
151	0000304	13	-1,34	43,80	359	000149	26	0,35	63,20
152	0000314	13	-1,34	43,80	360	000278	26	0,35	63,20
153	0000355	13	-1,34	43,80	361	0000251	27	0,44	64,21
154	0000370	13	-1,34	43,80	362	0000309	27	0,44	64,21
155	0000381	13	-1,34	43,80	363	0000336	27	0,44	64,21
156	0000420	13	-1,34	43,80	364	000061	27	0,44	64,21

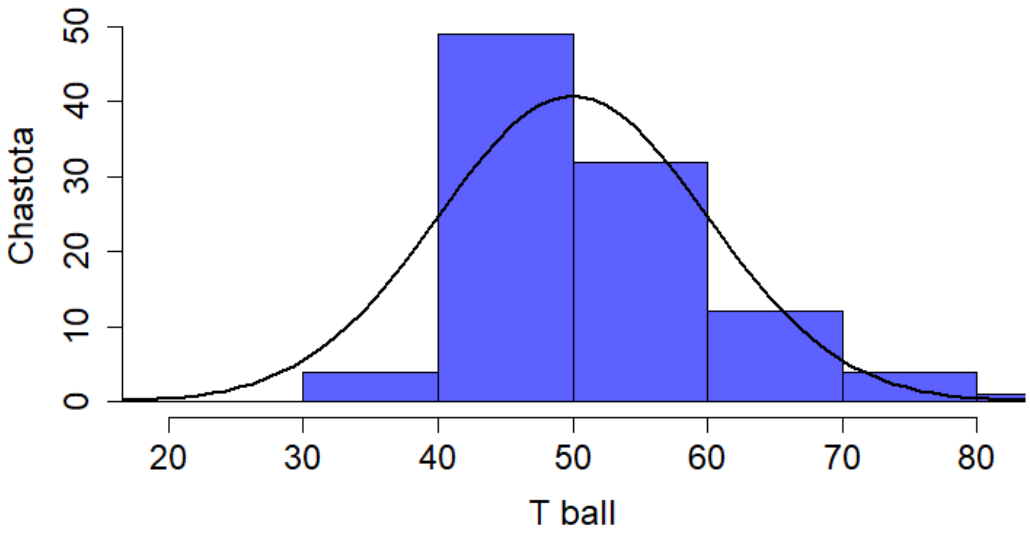
157	000062	13	-1,34	43,80	365	000067	27	0,44	64,21
158	000087	13	-1,34	43,80	366	000074	27	0,44	64,21
159	000093	13	-1,34	43,80	367	000188	27	0,44	64,21
160	000113	13	-1,34	43,80	368	000201	27	0,44	64,21
161	000122	13	-1,34	43,80	369	000202	27	0,44	64,21
162	000196	13	-1,34	43,80	370	000286	27	0,44	64,21
163	000001	14	-1,16	45,82	371	000292	27	0,44	64,21
164	000002	14	-1,16	45,82	372	000298	27	0,44	64,21
165	000014	14	-1,16	45,82	373	0000250	28	0,48	64,66
166	0000217	14	-1,16	45,82	374	000064	28	0,48	64,66
167	000024	14	-1,16	45,82	375	000071	28	0,48	64,66
168	0000245	14	-1,16	45,82	376	000084	28	0,48	64,66
169	0000249	14	-1,16	45,82	377	000291	28	0,48	64,66
170	0000252	14	-1,16	45,82	378	0000345	29	0,50	64,90
171	0000317	14	-1,16	45,82	379	000077	29	0,50	64,90
172	0000322	14	-1,16	45,82	380	000146	29	0,50	64,90
173	0000326	14	-1,16	45,82	381	000297	29	0,50	64,90
174	0000349	14	-1,16	45,82	382	0000271	30	0,52	65,17
175	0000354	14	-1,16	45,82	383	0000352	30	0,52	65,17
176	0000376	14	-1,16	45,82	384	0000346	31	0,57	65,72
177	0000383	14	-1,16	45,82	385	000072	31	0,57	65,72
178	0000385	14	-1,16	45,82	386	000277	31	0,57	65,72
179	0000386	14	-1,16	45,82	387	0000272	32	0,67	66,92
180	0000387	14	-1,16	45,82	388	0000273	32	0,67	66,92
181	0000398	14	-1,16	45,82	389	0000274	32	0,67	66,92

182	0000402	14	-1,16	45,82	390	0000276	32	0,67	66,92
183	0000404	14	-1,16	45,82	391	000148	32	0,67	66,92
184	0000410	14	-1,16	45,82	392	000280	32	0,67	66,92
185	0000419	14	-1,16	45,82	393	000282	32	0,67	66,92
186	000073	14	-1,16	45,82	394	000290	32	0,67	66,92
187	000095	14	-1,16	45,82	395	000294	32	0,67	66,92
188	000108	14	-1,16	45,82	396	000295	32	0,67	66,92
189	000120	14	-1,16	45,82	397	000300	32	0,67	66,92
190	000139	14	-1,16	45,82	398	000283	33	0,87	69,11
191	000140	14	-1,16	45,82	399	000293	33	0,87	69,11
192	000208	14	-1,16	45,82	400	0000331	34	1,11	71,89
193	0000047	15	-0,93	48,56	401	0000356	34	1,11	71,89
194	0000248	15	-0,93	48,56	402	000065	34	1,11	71,89
195	0000307	15	-0,93	48,56	403	000134	34	1,11	71,89
196	0000315	15	-0,93	48,56	404	000296	34	1,11	71,89
197	0000318	15	-0,93	48,56	405	0000275	35	1,31	74,16
198	0000339	15	-0,93	48,56	406	0000348	35	1,31	74,16
199	0000364	15	-0,93	48,56	407	000287	35	1,31	74,16
200	0000372	15	-0,93	48,56	408	000289	35	1,31	74,16
201	0000384	15	-0,93	48,56	409	000281	36	1,42	75,49
202	0000390	15	-0,93	48,56	410	000284	36	1,42	75,49
203	0000391	15	-0,93	48,56	411	000285	36	1,42	75,49
204	0000400	15	-0,93	48,56	412	0000335	37	1,49	76,25
205	0000407	15	-0,93	48,56	413	000288	37	1,49	76,25
206	000063	15	-0,93	48,56	414	000299	37	1,49	76,25

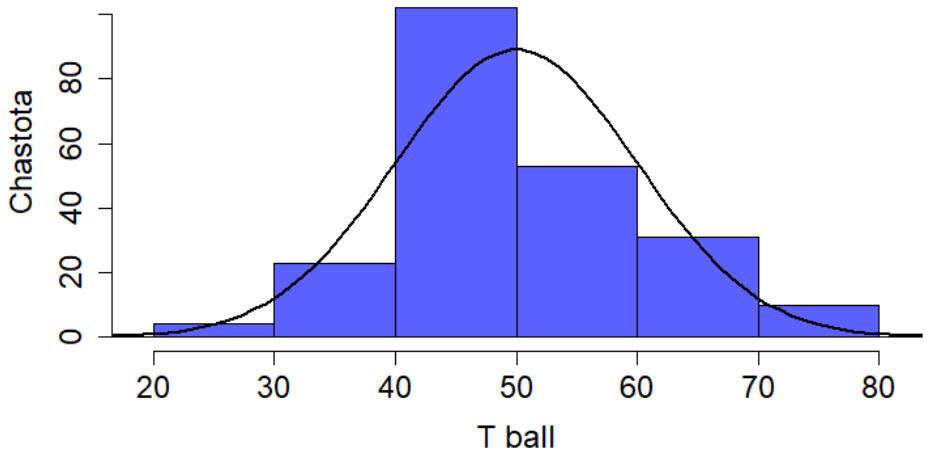
207	000089	15	-0,93	48,56	415	0000350	41	2,13	83,62
208	000107	15	-0,93	48,56					

Ushbu jadvalda keltirilgan ma'lumotlarga tayanib sinaluvchilarning *T* – ball shkalasida o'rta ta'lim maktablari, ixtisoslashtirilgan davlat umumta'lim maktablari va akademik litseylar kesimidagi taqsimoti aniqlandi.

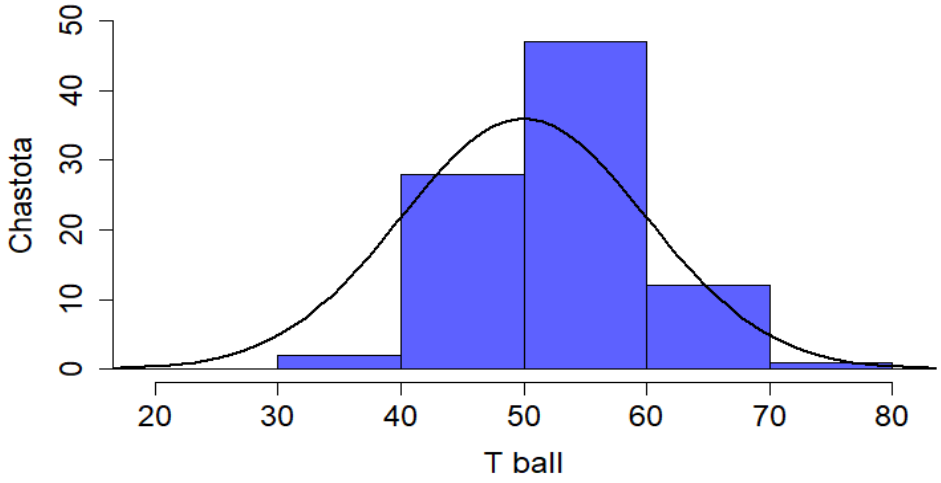
2–rasmda o'rta ta'lim maktablari va ixtisoslashtirilgan davlat umum ta'lim maktablarining *T* – ball shkalasidagi taqsimoti normal taqsimotga nisbatan chapga, aksincha, akademik litseylarda esa o'ngga siljiganligi ko'rinadi, bu esa o'z navbatida akademik litsey o'quvchilarining qobiliyat darajasi o'rta ta'lim maktablari va ixtisoslashtirilgan davlat umumta'lim maktablari o'quvchilarining qobiliyat darajasiga nisbatan yuqori ekanligini ko'rsatadi.



a)



b)



c)

2-rasm. Sinaluvchilarning T – bali shkaladagi ballarining taqsimoti, o‘rta ta’lim maktablari – *a,* ixtisoslashtirilgan davlat umum ta’lim maktablari –*b*, akademik litseylar - *c*

Xulosa

Ilmiy tadqiqot natijalari baholashlarda xom ballardan voz kechib, zamonaviy test nazariyasi bilan baholash usullariga o‘tish juda muhimligini ko‘rsatadi. Shuningdek, xom ball bilan baholashda o‘zgaruvchilar orasidagi chiziqli bog‘liqlik buzilishi va buning natijasida shkalalar nomutanosib bo‘lishi va eng muhimi sinaluvchilarning qobiliyatiga to‘g‘ri baho berilmasligi ham mumkin. Fizika fanidan umumiy o‘rta ta’lim maktablarining 11-sinf bitiruvchi o‘quvchilaridan ilmiy tadqiqot uchun olingan test sinovi natijalari klassik test nazariyasi va Rash modeli asosida tahlil qilindi va Rayt xaritasi o‘rganildi. Rayt xaritasidan foydalanib test variantida Rash modeli bilan hisoblangan test topshiriqlarining qiyinlik darajalari hamda qobiliyat darajalari taqsimotlarini aniqroq o‘lchash va bu orqali sinaluvchilar uchun test topshiriqlarini tanlash mumkinligi ko‘rsatildi. Bilimlarni baholashda xom ballardan voz kechib, zamonaviy test nazariyasi bilan baholash usullariga o‘tish juda muhimdir.

ADABIYOTLAR

1. Baker, Frank. The Basics of Item Response Theory, ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD. 2001.
2. Hambleton, R.K.,Swaminathan, H.,& Rogers, H.J., Fundamentals of item response theory. Newbury Park, CA: Sage. 1991.
3. Ivailo Partchev. A visual guide to item response theory, Friedrich-SchillerUniversitat Jena. 2004.
4. Spearman, C. (1904a). “General intelligence,” objectively determined and measured. American Journal of Psychology, 15, 201-293.
5. Spearman, C. (1904b). The proof and measurement of association between two things. American Journal of Psychology, 15, 72-101.
6. Spearman, C. 1907. Demonstration of formulae for true measurement of correlation. Am. J. of Psychology. 18, 160-169.
7. Spearman, C. 1910. Correlation from faulty data, British J. of Psychology. 3, 271-295.
8. Avanesov V.S. Teoriya i metodika pedagogicheskix izmereniy. sT i MKO UGTU-UPU, 2005.
9. A.R. Sattiyev, M.Dj. Ermamatov. Matematika fanidan milliy sertifikat uchun o‘tkazilgan test sinovlari natijalari tahlili, Axborotnoma, №1, 2023, 4-62
10. Glas Dj., Stenli Dj. Statisticheskie metodы v pedagogike i psixologii. -M.: Progress, 1976, 495.
11. Cronbach L. J. Coefficient alpha and the internal structure of tests, Psychometrika, 1951, T.16, №3, 297-334.
12. Rasch G., Probabilistic models for some intelligence and attainment tests, Copenhagen, Danish Institute for Educational research. 1960.
13. Rasch G. "An item analysis which takes individual differences into account." British journal of mathematical and statistical psychology 19.1 1966, 49-57.
14. Hattie J. Methodology review: assessing unidimensionality of tests and ltenls, Applied psychological measurement, 1985, T, 9, №2, 139-164.
15. B.D. Wright and M.H. Stone, Best Test Design, MESA Press, Chicago, 1979.
16. M.D. Ermamatov, M.D. Alimov, A.A. Sulaymonov, A.R. Sattiyev. Kalibrovkalangan test topshiriqlari: Sharq tillaridan o‘tkazilgan test sinovi natijalarining statistik tahlili, Axborotnoma №. 3-4, 16-83 b., 2022.
17. M.D. Ermamatov, A. Abbosov, A.A. Baratov, Test topshiriqlarini kalibrovkalash va qobiliyatlarini tenglashtirish, Axborotnoma №. 3-4, 4-16 b., 2022.
18. 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, 2023, 4-62.
19. Dimitris Rizopoulos, ltm: An R package for Latent Variable Modelling and Item, Response Theory Analyses, Journal of Statistical Software, V.17, 1-15, 2006.
20. David Torres Irribarra and Rebecca Freund, Wright Map: IRT item-person map with ConQuest integration, 2014, 1-36.

STATISTICAL ANALYSIS OF PHYSICS TEST RESULTS

K.A. Amonov, A.A. Baratov

Scientific-study Practical Center under the Agency for Assessment of Knowledge and Competences under the ministry of higher education, science and innovation of the republic of Uzbekistan, Tashkent 100084, Bogishamol st. 12

Abstract. In this article, a physics test was conducted for scientific research with 11th grade students of general secondary schools, specialized state general education schools and academic lyceums. The obtained test results were analyzed based on classical test theory and Rash model. Based on the statistical analysis of the tests, their mean value, mode and median, the total dispersion of the test scores were determined. Based on the Rash model, the compatibility of the difficulty levels of the test tasks with the ability level of the test takers, Wright's map was studied.

Keywords: Test items, Mode, Median, Standard deviation, Variance, Cronbach's alpha coefficient, validity, level of difficulty, Rasch model, Wright-map, ability levels