

MATEMATIKA FANIDAN BILIMLARNI MUQOBIL JAVOBLI VA YOZMA ISH TOPSHIRIQLARI ASOSIDA ZAMONAVIY TEST NAZARIYALARI BILAN BAHOLASH

ASSESSMENT OF MATHEMATICS KNOWLEDGE USING MODERN TEST THEORIES BASED ON MULTIPLE-CHOICE AND WRITTEN RESPONSE TASKS

Z.M.To'lanov, Q.P.Murodov

Namangan Davlat Universiteti akademik litseyi, 160107, Namangan shahar, Boburshoh ko'chasi, 161

Qisqacha mazmuni. Test natijalarini baholashda politomik elementlar uchun zamonaviy psixrometrik model — *umumlashtirilgan qisman kredit modeli* (ing. Generalized partial credit model - GPCM) qo'llanilishi va u orqali yozma ishlarning bosqichma-bosqich baholash mezonlarini integratsiya qilish imkoniyatlari o'rganiladi. Tadqiqot doirasida matematik yozma ishlarining adolatli baholash tizimini rivojlantirish maqsadida, har bir yozma topshiriq uchun uch bosqichli mezonlar: yechim strategiyasini aniqlash, hisoblash aniqligi va taqdim etish uslubi alohida ball toifalari sifatida qabul qilindi. Muqobil tanlovli testlar uchun *ikki parametrli logistik model* (2PL) tanlab olindi. Tadqiqot va adabiyotlarni [4-8] o'rganish jarayonida *umumlashtirilgan qisman kredit modeli* yordamida yozma ishlarni baholashda har bir ball toifasi orqali ishtirokchilarning bilim va kompetensiyalarini yanada aniqroq ajratish imkonini berishi aniqlandi. Modelning olinayotgan ishga nisbatan mosligini tekshirish usullari yoritildi, yozma ish uchun modelga mos baholash mezonlari o'rganildi. Shuningdek, aralash formatdagi imtihonlarda umumiy natijani aniqlash maqsadida *birgalikda konkurent kalibrlash* usuli haqida qo'yilgan maqsadga mos ma'lumotlar berildi.

Tadqiqotning yakunida o'tkazilgan virtual sinov test natijalari va xulosalari bayon qilindi. Shuningdek, yozma ishlar va muqobil javobli savollar natijalari o'rtasidagi o'zaro korrelyatsiyani o'rganishda, yozma ishning har bir bosqichini alohida test elementi (savol parametri) sifatida qabul qilish orqali, ularning statistik ko'rsatkichlarini tahlil qilish mumkinligi aniqlandi.

Kalit so'zlar: ikki parametrli logistik model, qisman kredit model, chegara, birgalikda konkurent kalibrlash, baholash mezonlari, issiqlik xaritasi, simulyatsiya;

1. Kirish

Ma'lumki matematika fanining ayrim masalalari shunday xususiyatga egaki, talabgor bilimining bir paytda bir nechta qirralarini ochib bera oladi. Bu kabi masalalarni yechish asnosida talabgor o'z bilim va qobiliyatini ishga solib savolning yechish uchun to'g'ri usulni ya'ni yaxshi algoritmni va strategiyani tanlab oladi, chizmalarni to'g'ri chizadi va h.k., Shu vaziyatda talabgorlar ichida mavzularni yaxshi o'zlashtirgan, kognitiv bilim ko'nikmalarini olgan, ammo qisqa

vaqt sharoitida arifmetik xatoliklarga, yoki ayrim masalalarni xususiyatidan kelib chiqib ko'p arifmetik hisob-kitob talab qiladigan masalalarni xato ishlashi yoki oxiriga yetkazmaslik holatlari yuzaga kelishi mumkin. Bunday vaziyatlarda talabgorning ayrim kognitiv qobiliyatlari to'liq baholanmaslik ehtimoli paydo bo'ladi.

Shuningdek, yopiq test topshiriqlari (jumladan, muqobil javobli va qisqa javobli ochiq savollar) asosan Blum taksonomiyasining [26] pastki kognitiv

darajalarini (bilish, tushunish, qo'llash) baholash imkonini beradi. Biroq, yuqori darajadagi kognitiv kompetensiyalarni (tahlil, sintez, baholash) aniqlash va baholash uchun esa yozma ishlar muhim ahamiyat kasb etadi [27]. Tabiiyki, imtihonlarda yuqori darajadagi kognitiv kompetensiyalarni aniqlashga yo'naltirilgan savollarni kiritish, kelgusida tayyorgarlik ko'rayotgan talabgorlar uchun ijobiy washback (ta'sir) effektini yuzaga keltiradi [28].

2. Aralash formatdagi test variantlari bilan baholash

2.1 Muqobil javobli yopiq testlar uchun

Demak, bizning oldimizda yozma ish shaklidagi ishlarni baholash va uni muqobil javobli test savollari bilan birgalikda baholash usulini ishlab chiqish masalasi paydo bo'ladi. Buni amalga oshirishda amaliy xarakterga ega bo'lgan usullar bilan tanishib chiqamiz. Qulaylik uchun quyidagi imtihon formatini misol tariqasida olamiz: Muqobil javobli savollar - 35 ta va yozma ish topshiriqlari - 10 ta. 1. Ko'p tanlovli test natijalarini baholash uchun ikki parametrli model (2PL) tanlanadi. 2PL model har bir savolning ikki parametrini baholaydi: ajratish (discrimination) va qiyinchilik (difficulty). Ajratish parametri savolning respondent qobiliyatiga qanchalik sezgir ekanini ko'rsatadi; qiyinchilik parametri esa savolning qanchalik qiyin ekanini (qobiliyat shkalasida) bildiradi. Bu yerda parametrlarni baholash maqsadida *marginal maksimal ehtimollikka o'xshashlik* (ing. maximum marginal likelihood - MML) usulidan foydalanish misol tariqasida olingan.

Elementning javob funksiyasi (2PL uchun)

$$P_i(\theta_n) = \frac{e^{a_i(\theta_n - b_i)}}{1 + e^{a_i(\theta_n - b_i)}}$$

bu yerda:

Matematika fani va boshqa predmetlar bo'yicha yozma ishlarni zamonaviy psixometrik yondashuvlar asosida baholash tizimini joriy etish, baholash jarayonining adolatli, ishonchli va pedagogik maqsadga mos bo'lishiga erishish uchun muhim ilmiy va amaliy tadqiqotlarni o'rganib chiqish va ulardan foydalanish uchun dasturiy ta'minot scriptlarini ishlab chiqish zarurati paydo bo'ladi.

- a_i - elementning diskriminatsiya (farqlash) parametri ($a_i > 0$);
- b_i - elementning qiyinchilik (difficulty) parametri;
- θ_n - n -talabgorning latent qobiliyati.

Olingan ma'lumotlar asosida yakuniy test natijalarini olish uchun R dasturining 'ltm', 'mirt', 'TAM' kabi paketlarida kerakli funksiyalar to'plami bor, misol uchun *maksimal likelihood* modeli har bir paketda har xil funksiyada uchraydi, turlicha iteratsiyalar bilan hisoblaydi ammo juda bir-biriga yaqin natijalar beradi. Bularni amalga oshirish quyidagi buyruq funksiyalarga asoslanadi va yechim iteratsion optimizatsiya orqali topiladi [3]:

- `library(mirt)`
- `mirt::fscores(model, method=ML)`
- `model <- mirt(data, 1, elementtype=2PL)`

ML usulida Log-ehtimollik funksiyasi maksimal qiymat beradigan θ_n ning qiymati n - ishtirokchining yakuniy qobiliyat ko'rsatkichi bo'ladi. Maksimal qiymat beradigan θ_n ni topish uchun iteratsiya va boshqa hisoblash usullari qo'llaniladi.

2.2 Yozma ish topshiriqlari uchun

Yozma ish topshiriqlari 10 ta - bular ochiq javobli yoki qisman ball beriladigan topshiriqlar. Har bir topshiriq bir nechta ball darajasiga ega (masalan: 0 dan 3 gacha ball). Bunday politomik (ko'p kategoriyali) elementlar uchun *umumlashtirilgan qisman kredit modeli* (ing. Generalized partial credit model - GPCM) qo'llaniladi [4]. Bu model Masters [6] tomonidan taklif qilingan *qisman kredit modeli* (PCM)ning umumlashtirilgan ko'rinishi bo'lib, har bir topshiriq uchun diskriminatsiya (ajratish) parametri va har bir ball kategoriyasi uchun saralangan qiyinchilik *chegara (ing. threshold) parametrlarini baholaydi.

***Chegara** - bu politomik (ko'p balli) IRT modellarida, masalan, *umumlashtirilgan qisman kredit modeli* (GPCM) yoki *darajalangan javob modeli* (ing. *graded response model* - GRM) da, har bir element uchun har bir ball pog'onasidan yuqoriroq pog'onaga o'tish ehtimolini belgilovchi parametr.

Umumlashtirilgan qisman kredit modeli - ko'p kategoriyali (politomik) elementlar uchun latent qobiliyatga bog'liq bo'lgan ehtimollik modelidir. *i*-element bo'yicha *j* nomerli ishtirokchi tomonidan *k* ball olingan ehtimollik quyidagicha ifodalanadi [6]:

$$P(X_{ij}=k|\theta_n) = \frac{\exp\left(\sum_{m=1}^k a_i(\theta_j - b_{jm})\right)}{\sum_{h=0}^{M_i} \exp\left(\sum_{m=1}^h a_i(\theta_j - b_{jm})\right)}$$

bu yerda:

- $X_{ij} - j$ -respondentning i -savolga olgan balli ($0 \leq k \leq M_i$)
- $a_i - i$ -savolning diskriminatsiya parametri
- $\theta_j - j$ -respondentning latent qobiliyati
- $b_{jm} - i$ -savol uchun m -bosqich (chegara) qiyinchilik parametri

- $M_i - i$ -savol uchun maksimal ball.

2PL modeli dixotomik savollar uchun mos, GPCM esa har bir politomik savolga alohida diskriminatsiya parametr berib, turli topshiriqlarning testdagi hissasini adolatli aks ettiradi. Shu tariqa, har ikkala turdagi savollar bir xil shkalada baholanadi.

Yozma ish savollarida har bir band bo'yicha kognitiv bilimlarni baholash uchun GPCM usulini tanlab olingan. Bu modeldagi asosiy talab: kategoriyalar monoton ortib borishi kerak, bizning holatimizda bu 0, 1, 2 va 3 balldan iborat. Yozma ishlarni ushbu model yordamida baholash jarayonida avvalo ishtirokchilarning javoblari ekspertlar tomonidan belgilangan mezonlar asosida baholanadi va tegishli ballar qo'yiladi. Natijada, har bir topshiriq uchun politomik natijalar matritsasi shakllanadi.

Bu ballarni qo'yish uchun quyidagicha baholash mezoni maqsadga muvofiq [23][1][7]:

1-mezon. Yechim strategiyasini aniqlash. Masala shartiga muvofiq zarur matematik obyektlar va munosabatlar (tenglamalar, grafiklar, geometrik belgilar va h.k.) to'g'ri aniqlangan va yechim algoritmi bosqichma-bosqich mantiqan loyihalangan bo'lishi. Misol va masalalarni yechishga to'g'ri yondashilgan, yechim bosqichlari rejalashtirilgan, ammo hisoblash jarayoni oxiriga yetkazilmagan taqdirda ham ushbu me'yor bo'yicha 1 ball beriladi.

2-mezon. Hisoblash aniqligi. Arifmetik va algebraik operatsiyalar nazariy me'yor va qoidalarga qat'iy rioya qilgan holda bosqichma-bosqich izchil bajarilgan, jarayonda xatoliklarga yo'l qo'yilmagan va yakuniy natija to'g'ri topilgan bo'lishi.

3-mezon. Taqdim etish uslubi. Yechim matni rasmiy matematika tili bilan, aniq belgilashlar, izchil

chizmalar, to'g'ri ramzlash va yechimni asoslaydigan mulohazalar va teoremlar bilan ifodalangan bo'lishi.

Bu mezonlar asosida yozma ishlar tekshiriladi va 1-jadvaldagi holat asosida ballar qo'yiladi [23].

(1-jadval)

№	Yechimning umumiy holati	Ball
1	Masala to'liq yechilgan; Yechim mukammal va aniq: yozma ifoda, hisoblar va yechim strategiyasi hammasi to'g'ri, mantiqan izchil va matematik standartlarga to'liq mos.	3
2	Masala to'liq yechilgan, ammo yechimni taqdim etishda ayrim kamchiliklar va noaniqliklar bor.	2
3	Masalaning javobi xato, yechimda muhim kamchiliklar bor, yoki masala chala yechilgan, ammo yechimning asosiy g'oyasi to'g'ri belgilangan, lekin boshqa yo'lga burib ketilgan, yechimga olib boradigan strategiya to'g'ri tanlagani alomatlari bor: chizma, formula yoki teorema kiritilgan.	1
4	Masalani yechishga mutlaqo urinib ko'rilmagan yoki masala yechish uchun yozuvlar yozilgan lekin yechim sari yurilmagan.	0

Baholash mezonining amalda tekshirish uchun (yechim strategiyasini aniqlash misolida) quyidagi ishlar bajarilishi kerak bo'ladi:

1. Avvalo turli darajadagi ishtirokchilar (yaxshi, o'rtacha va past ko'rsatkichli) namunaviy test savollariga yozma javob beradi. Savollarga masalaning yechim strategiyasini ko'rsatishni alohida shart sifatida qo'shish mumkin (faqat aprobatsiya uchun). 2. Har bir ishtirokchining yechimlari yuqoridagi baholash mezoniga muvofiq tekshiriladi. Baholovchilar har bir yechimda strategiyani tanlash va asoslash sifatiga alohida e'tibor qaratadilar. 3. Olingan natijalar statistik tahlil qilinadi. Eng muvaffaqiyatli yechimlar, xatoliklar, qiyinchiliklar aniqlanadi. Yakunda aprobatsiya natijasiga asoslangan takomillashtirish olib boriladi, zarur hollarda mezon shakli, ifodasi yoki ball taqsimoti qayta ko'rib chiqiladi. O'quvchilardan va baholovchilardan fikr-mulohazalar olinadi.

Aprobatsiya natijalari mezonning samaradorligini, ya'ni o'quvchining haqiqiy yechim strategiyasini to'g'ri baholash imkoniyatini, aniqlik va obyektivlik darajasini aniqlashga xizmat qiladi.

Aprobatsiya va muvofiqlashtirish jarayonida test ishtirokchilari uchun vaqt reglamenti ham ko'rib chiqiladi. Matematika fanining yozma imtihonlarida vaqt taqsimotini to'g'ri belgilash — har bir ishtirokchi yechim strategiyasini aniqlash va asoslashga yetarli vaqt ajrata olishi uchun muhim. Har bir topshiriq uchun o'rtacha hisobda 4–6 daqiqa (soddaroq topshiriq) yoki 8–10 daqiqa (murakkab, ko'p bosqichli topshiriq) vaqt ajratiladi. Murakkab topshiriqlar, odatda, yechim strategiyasini asoslash va ko'rsatuvchi bosqichlar ko'proq bo'ladi.

Odatda, xalqaro matematik imtihonlarda (Cambridge, IB) 10 ta yozma topshiriq uchun 60–90 daqiqa ajratiladi. Bu o'rtacha 6–9 daqiqa har bir topshiriq uchun, shundan 2–4 daqiqasi aynan strategiyani asoslashga to'g'ri keladi.

3. Birgalikda konkurent kalibrlash

Endi oldimizda yana bir amaliy masala yuzaga keladi, ya'ni bu ikki ko'p tanlovli va yozma ishlardan olingan qobiliyatlarni umumlashgan yakuniy natijasini hisoblash, buning uchun *bir vaqtda konkurent kalibrlash* modelidan foydalanamiz [8].

Birgalikda (concurrent) kalibrlash deganda, testdagi barcha elementlar – dixotomik ham, politomik ham – bir vaqtning o'zida, yagona modelda kalibrovkalanishi tushuniladi. Bu yondashuv har ikki turdagi savollarni bitta latent o'lchov bo'yicha bog'laydi va umumiy natijani chiqaradi [8].

Birgalikda kalibrlashning afzalligi shundaki, muqobil javobli savollar va yozma topshiriqlarning parametrlarini alohida-alohida emas, balki birgalikda baholab, umumiy qobiliyat bahosini aniqroq va adolatli shakllantirish mumkin. Ikkala turdagi elementlar bitta modelda bo'lgani sababli, model ularning har biriga tegishli og'irlikni (ajratish parametri orqali) "o'zi" beradi – natijada, test natijalarida muvozanatli ulush ta'minlanadi. Masalan, agar yozma topshiriqlar juda yuqori ajratish parametrlariga ega bo'lsa, model ular orqali qobiliyatni ko'proq aniqlaydi, lekin muqobil javobli savollar soni ko'proq bo'lsa, ular ham o'z hissasini qo'shadi. Yagona o'lchovli (ing. unidimensional) IRT modeli faraziga ko'ra, barcha elementlar bitta qobiliyatni o'lchayotgan bo'lsa, birgalikda kalibrlash to'g'ri natija beradi [9].

Avval aytganimizdek, GPCM modelida har bir element uchun diskriminatsiya (ajratish) parametri mavjud bo'lib, bu xususiyati bilan 2PL modeliga o'xshaydi, PCM modelida

esa diskriminatsiya parametri hisobga olinmaydi va birga teng deb olinadi, bu esa uni 1PL (Rasch) modeliga mos deyishimiz mumkin, shu bois, politomik elementlar uchun GPCM va PCM mos ravishda dixotomik elementlar uchun 2PL va 1PL modellariga ekvivalent hisoblanadi.

Bularni amalga oshirish uchun albatta R dasturidan foydalanish maqsadga muvofiq, chunki bundan boshqa dasturlash tillarida iteratsiya muammosiga duch kelamiz. Shuningdek, R dasturida testlarni baholash uchun kerakli funksiyalar shay holatga keltirilgan.

Avvalam bor, R dasturida birgalikda konkurent kalibrlash uchun qaysi raqamli elementlar 2PL, qaysilari GPCM ekanini belgilab olishimiz kerak. Masalan, R dagi MIRT paketi yoki TAM paketi bunga imkon beradi: mirt paketida mirt() funksiyasiga *elementtype* parametri orqali har bir elementning turini ko'rsatish mumkin. Masalan, 35 ta dixotomik savollar uchun 2PL, 10 ta politomik savollar uchun gpcm deb ko'rsatamiz [14].

Birgalikda kalibrlash vaqtida, barcha elementlar bitta latent o'lchovga keltiriladi. Bu testning umumiy natijasi o'lchovi bo'lib, har ikki turdagi elementlar shunga nisbatan parametrlanadi.

Maximum marginal ehtimollik (MML) orqali baholash MIRT paketida standart bo'lib, Expectation-Maximization (EM) algoritmi yordamida bajariladi. Buning nazariy asoslariga to'xtalmaymiz.

Konkurent kalibrlash usulini unga muqobil bo'lgan modellar bilan taqqoslangan adabiyotlar [10][11][12][20] mavjud.

4. Modelni testga mosligini tekshirish

Endi muqobil javobli va yozma savollar ishtirokchining yakuniy ballida qanchalik rol o'ynayotganini tahlil qilish kerak bo'ladi.

Agar barcha muqobil javobli savollarning ajratish ko'rsatkichlari 0.5 atrofida, lekin yozma topshiriqlar ajratishi 1.5 atrofida bo'lsa, demak

yozma topshiriqlar qobiliyatni aniqlashda kuchliroq bo'lib chiqadi. Bunday holatda, test natijasi asosan yozma qismga tayangan bo'ladi. Aksincha, agar muqobil javobli savollar juda ko'p va har biri ham ma'lum darajada ajratuvchan bo'lsa, ularning umumiy ta'siri katta bo'lishi mumkin [22].

Ma'lumki IRT modelida har bir elementning hissasi uning ajratish parametri va testdagi nisbiy ma'lumotiga qarab avtomatik tartibga solinadi, bu bizga qo'shimcha savollarga og'irliklar berib chiqishimiz shart emasligini bildiradi.

Testning ishonchliligini tekshirishning yana bir usuli *testning ma'lumot chizig'i* (ing. test information curve) qaysi qobiliyat darajalarini qanchalik aniq o'lchashini ko'rsatadi. Bu muhim tahlil bo'lib, bizga qo'shimcha tekshirish imkonini beradi. *TAM::IRT.informationCurves(mod)* [15] yoki mirt paketida *testInfo* funksiyalari orqali test ma'lumot grafigini olish mumkin. Maqsad – ko'p tanlovli va yozma elementlar birga qaysi diapazonda eng ko'p ma'lumot berishini ko'rish. Ideal holda testning

ma'lumot chizig'i keng diapazonda yuqori bo'lib, bu test natijasining ishonchliligini bildiradi [22].

Tekshirishning yana biri *Kronbax alfa* klassik usul bo'lib, elementlarning ichki mutanosiblik darajasini o'lchaydi. R da ltm paketida *cronbach.alpha()* funksiyasi mavjud bo'lib, to'g'ridan-to'g'ri ma'lumotlar matritsasini kiritib, alfa ko'rsatkichini hisoblab beradi [15]. Masalan, $\alpha=0.85$ bo'lsa, demak test "yaxshi" ishonchlilikka ega deyish mumkin [15].

Bizning testda elementlar politomik bo'lsa ham, *cronbach.alpha* funksiyasi ularni sonli shkala sifatida qabul qilib, umumiy α ni hisoblaydi. *Kronbax alfa* yuqori bo'lishi uchun elementlar bir xil qobiliyatni o'lchashi va o'zaro korrelyatsiyasi baland bo'lishi lozim [19].

Agar ishonchlilik kutilganidan past bo'lsa (masalan $\alpha<0.7$), test tarkibini qayta ko'rib chiqish kerak bo'ladi, ya'ni quyi diskriminatsiya ko'rsatkichiga ega savollarni chiqarib tashlash yoki savollar sonini ko'paytirish haqida tahlillar qilish kerak.

5. Virtual sinov test natijalari tahlili

Tadqiqotda ilgari surilgan faraz, asosan, testning tarkibiy tuzilishi bilan bog'liq bo'lib, ya'ni 35 ta muqobil javobli va 10 ta uch bosqichli yozma topshiriqlarning birgalikda baholanishi uchun maxsus baholash mezonlari (rubrikalar) va konkurent kalibrlash usulidan foydalanish, shuningdek, mos ravishda umumlashgan qisman kredit modeli (ing. generalized partial credit model, GPCM) hamda ikki parametrli logistik model (2PL) yordamida texnik jihatdan amalga oshirish imkoniyatini va bu usulning (IRT) talablariga muvofiqligini aniqlashdan iborat edi.

Tanlangan modellar o'z tabiati jihatidan test formatiga moslashuvchan bo'lib, yozma

shakldagi topshiriqlar sonining boshqa qiymatlarida ham baholashning adolatlilik xususiyatini saqlash qobiliyatiga ega ekanligi taxmin qilinadi. Ammo amaliy jihatdan aynan 35 ta muqobil javobli savol va 10 ta yozma topshiriqdan iborat test formati samaraliroq hisoblanadi. Chunki bu holatda, 35 ta savol uchun 35 ta qiyinchilik parametri mavjud bo'lsa, 10 ta yozma topshiriq uchun 30 ta qiyinchilik parametri aniqlanadi, bu esa vaqt taqsimoti nuqtayi nazaridan teng ahamiyatga ega bo'lgan topshiriqlar nisbatini ta'minlaydi.

Mazkur test formatining samaradorligi va tanlangan modellarning maqsadga

muvofigligini ilmiy faraz sifatida qabul qilish orqali, *virtual sinov* (ing. pilot test) amalga oshirildi. Buning uchun R dasturida 1000 kishilik natijalar simulyatsiya qilindi. Bunda,

2-jadvalda R dasturida olingan har bir elementning a – δ diskriminatsiya va b – δ qiyinchilik darajasi berilgan:

(2-jadval)

simulyatsiya ko'rsatkichlari normal taqsimotga ega bo'lishi sharti qo'yildi, aks holda yakuniy natijalarning haqiqiy vaziyatga yaqinligi ta'minlanmasligi mumkin.

element	a	b	b1	b2	b3
Q1	0,86658774	0,658718169			
Q2	0,997153481	1,629812473			
Q3	1,007521444	0,687428291			
Q4	1,27327176	-1,109027961			
Q5	1,339077109	0,725797846			
Q6	1,429495896	1,451829682			
Q7	1,483363176	-0,862793724			
Q8	0,929896614	-0,028596928			
Q9	0,784131451	-1,441308327			
Q10	1,129691613	-0,87400089			
Q11	1,742594141	-1,143907296			
Q12	1,078108372	0,817178633			
Q13	1,734192256	0,454082513			
Q14	1,553106153	1,065862375			
Q15	1,714948485	1,084915604			
Q16	1,02480534	-2,415201908			
Q17	1,845951262	-0,630206434			
Q18	0,923977019	1,913676162			
Q19	1,359300263	1,290314558			
Q20	0,745658714	-0,21424683			
Q21	1,960291408	0,73191696			
Q22	1,332017145	-0,205355146			
Q23	1,977564567	-1,761482393			
Q24	1,002560968	0,505986804			
Q25	0,860310797	-1,288347708			
Q26	1,150152209	0,329411531			
Q27	1,557248685	2,366752947			
Q28	1,451192966	-0,557606932			
Q29	1,243633239	-0,423649285			
Q30	1,745055344	-1,917247687			
Q31	1,45649564	0,674819324			
Q32	1,509674889	-0,252603067			
Q33	1,639213698	-1,367618895			
Q34	1,028292741	-0,592970072			
Q35	0,909161822	1,86271282			
Q36	1,651978129		-0,87486	-0,20741	0,409335
Q37	1,616417269		-2,01032	-2,21172	-2,0791
Q38	1,586050982		1,513111	-1,59443	-0,92229
Q39	1,338216939		0,693882	1,524821	0,277478

Q40	1,249694416		-0,63362	-1,63144	-0,13702
Q41	1,39867135		0,049627	0,797303	-0,97271
Q42	1,532446271		1,837742	-0,38913	1,175658
Q43	0,995728516		-0,33497	-1,51113	2,39291
Q44	1,405102922		-1,24956	-1,03023	1,191753
Q45	0,8559046		0,030692	1,776988	0,324969

3-jadvalda 1000 kishilik sinov testida olingan yakuniy T-ballarga mos darajalar quyidagicha ko'rsatkichlarga ega bo'ldi.

(3-jadval)

Model	2PL + GPCM	1PL + PCM	
daraja	soni	soni	farqi
A+	25	26	1
A	44	48	4
B+	92	74	18
B	140	157	17
C+	204	194	10
C	145	158	10
-	350	343	7

3-jadvaldagi ma'lumotlar tahlili shuni ko'rsatadiki, darajalar soni bo'yicha sezilarli tafovutlar mavjud emas. Biroq, aynan bir xil ishtirokchilar turli modellarda har xil T-ball natijalarini olganlar. Ushbu tafovutning asosiy sababi yakuniy natijalarni hisoblash jarayonida test elementlarining diskriminatsiya parametrlari hisobga olinishi bilan izohlanadi.

Kronbax alfa ko'rsatkichlari 4-jadvalda berilgan.

(4-jadval)

turi	2PL + GPCM	1PL + PCM
Umumiy	0,928236	0,916982659
Muqobil javobli	0,887777	0,862552729
Yozma ish	0,878905	0,878408864

6. Test natijalarining ichki tuzilmasini tahlil qilish, korrelyatsion issiqlik xaritasi.

Ma'lumki, Test topshiriqlarining chuqur tahlili uchun elementlar o'rtasidagi o'zaro bog'liqliklarni aniqlash hamda vizual ko'rsatish muhim hisoblanadi [19][21]. Endi test tarkibida yozma ish va muqobil javobli elementlarning bog'liqligini qanday o'lchash mumkin degan o'rinli savol tug'iladi. Ushbu tadqiqotda, har bir elementining har bir chegara (bosqich) parametrini alohida ko'rsatkich sifatida qabul qilib oldik. Shuningdek, barcha dioxotomik elementlar (2PL modeli asosida) ham o'ziga xos mustaqil qiyinchilik

parametrlari sifatida tahlilga kiritildi. Natijada, jami 65 ta ko'rsatkich (35 ta dioxotomik savol va 10 ta ko'p bosqichli savol uchun 3 tadan chegara, jami 30 ta) har bir ishtirokchi bo'yicha 0 yoki 1 ko'rinishida kodlandi.

Har bir chegara (bosqich) alohida qiyinchilik parametri sifatida talqin qilinishi, test natijalarini an'anaviy faqat umumiy ball yoki umumiy element ko'rinishida emas, balki har bir bosqichdagi muvaffaqiyat imkoniyatlarini ham mustaqil tarzda tahlil qilish imkonini beradi.

Korrelyatsiya koeffitsiyentining turini tanlashda ma'lumotlarning tabiatini hisobga olish muhim ahamiyatga ega.

Ma'lumki, dixotomik (0/1) ma'lumotlar uchun Pirson korrelyatsiyasi mos keladi va bu holatda u phi-koeffitsiyent deb ataladi. Phi-koeffitsiyent aynan dixotomik o'zgaruvchilar o'rtasidagi bog'liqlikni o'lchaydi va Pirson koeffitsiyentining xususiy holi hisoblanadi [29].

Shuningdek, ushbu yondashuv test tuzilmasini vizual va statistik tahlil qilish uchun universal va hisoblashda samarali vosita hisoblanadi. Shu sababli, tadqiqotda indikatorlar o'rtasidagi o'zaro bog'liqliklarni aniqlash uchun Pirson korrelyatsiya koeffitsiyenti tanlandi [25].

Bu yondashuvda barcha element va chegaralar o'zaro korrelyatsiya

koeffitsiyenti matritsasiga o'tkazildi va "issiqlik xarita" (ing. heatmap) ko'rinishida vizuallashtirildi (1-rasm).

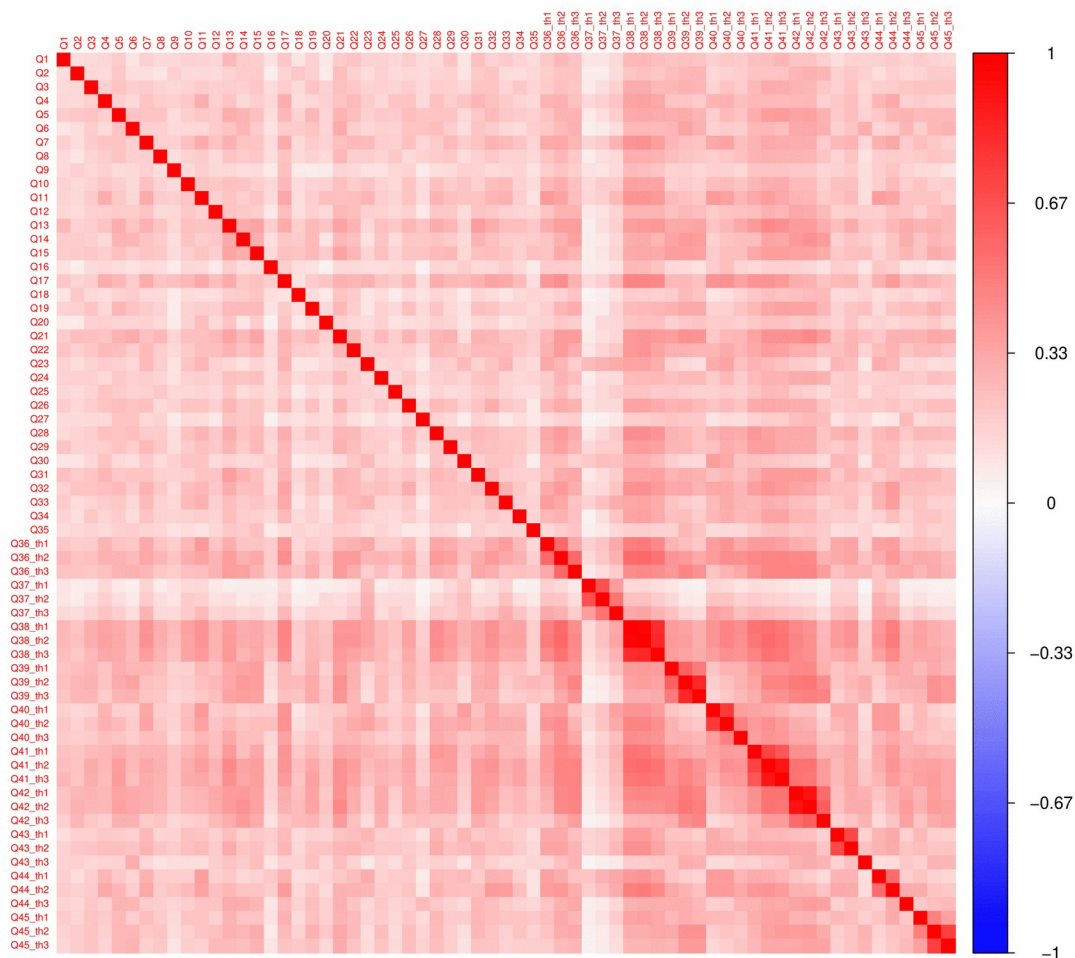
Ushbu vizual tahlil testning o'zaro yaqin yoki o'xshash element-bosqichlarni aniqlash imkonini berdi. Natijalardan shuni kutsak bo'ladiki, ayrim chegaralar va elementlar yuqori o'zaro bog'liqlikka ega bo'lib, bu test tarkibini optimallashtirish va har bir bosqichda baholashni yanada aniqroq amalga oshirish uchun muhim vosita bo'lib xizmat qilishi mumkin.

Buni amalga oshirish uchun R dasturida

`cor & cor (data, method = pearson, use = pairwise.comp`
buyruqdan foydalanish mumkin.

Amaliy nuqtayi nazardan, ushbu metodologiya testlarni loyihalash yoki validlikni aniqlashda keng imkoniyatlar ochadi [19].

(1-rasm)



7. Xulosa

Tadqiqot jarayonida R dasturi yordamida ishlab chiqilgan skript orqali 35 ta muqobil javobli va 10 ta yozma topshiriqdan iborat testning integrativ tarzda IRT modellari asosida baholanishi har bir ishtirokchining qobiliyat darajasini ishonchli va xolis baholash imkonini berishi haqidagi faraz ilgari surilgan edi. Mazkur tadqiqot doirasida farazlarni to'liq tekshirish uchun real ishtirokchilar bilan aprobeatsiya sinovlarini o'tkazish talab etiladi. Ammo olib borilgan faoliyatlar dasturiy ta'minot skriptlarini ishlab chiqish, mos adabiyotlarni o'rganish va simulyatsion sinov testlarini amalga oshirish bilan cheklanildi.

Olingan natijalar shuni ko'rsatadiki, **ushbu turdagi test formatida olingan natijalarni tanlangan modellar orqali texnik jihatdan baholash mumkin**. Bunda R dasturining mirt, tam va ltm kabi paketlari 2PL va GPCM modellarini qo'llashda samarali vosita sifatida xizmat qiladi, hamda zamonaviy psixometrik tahlillarni amalga oshirish uchun zarur funktsionallikka ega.

Shuningdek, tadqiqot mualliflari tomonidan foydalanishga tavsiya etilgan modellar va uslublar ilmiy adabiyotda mavjud bo'lgan ishonchli va empirik tekshiruvlardan o'tgan manbalar asosida shakllantirildi [4][5][6][7][9][10]. Shunga qaramasdan, kelgusidagi real aprobeatsiyalar orqali olib boriladigan tekshiruvlar asosida zarur tuzatish va qo'shimchalar kiritilishi mumkin.

Birgalikda konkurent kalibrlashning asosiy afzalligi barcha turdagi test shakllarini avtomatik ravishda yagona o'lchov shkalasiga keltirish imkoniyatini ta'minlashdan iboratdir.

Xulosa sifatida, quyidagi vazifalar keyingi tadqiqot bosqichida amalga oshirilishi zarur:

- yozma topshiriqlarni baholashda vaqt reglamenti hisobga olingan va IRT modellari bilan mos keluvchi aniq baholash mezonlarini ishlab chiqish;
- olingan testning ma'lumotlar bazasidan toki yakuniy ballargacha hisoblaydigan yagona R dasturi skriptini ishlab chiqish;
- testlarning model bilan mosligini bir nechta bosqichli tekshiruv orqali tasdiqlash.

Shuni ham ta'kidlash joizki, savollar test formati va vaqt reglamentiga moslashtirilgan taqdirda ham, ularning yuqori va past qobiliyatli ishtirokchilarni samarali ajratish xususiyati saqlanishi lozim. Bunday moslashtirish jarayonida yozma ish savollarini ishlab chiqadigan ekspertlarga zarur metodik va statistik ma'lumotlarni taqdim etish, shuningdek, testologlar bilan samarali hamkorlikni ta'minlash ustuvor ahamiyatga ega.

Foydalanilgan adabiyotlar

- [1] A'zamov A., Haydarov B.Q. Matematika sayyorasi. -T.: O'qituvchi, 1993.- 312 b.
- [2] Behavior Research Methods, Instruments, Computers, 1985, 17 (4), 513-514
- [3] Chalmers, R. P. (2012). mirt: A multidimensional element response theory package for the R environment. Journal of Statistical Software, 48(6), 1-29.
- [4] Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. Applied Psychological Measurement, 16(2), 159-176.
- [5] M.Dj. Ermamatov, A. Abbosov, A.A. Baratov. Test topshiriqlarini kalibrovkalash va qobiliyatlarni tenglashtirish, "Axborotnoma" ilmiy-uslubiy jurnali, 3-4/2022, Toshkent

- [6] Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174.
- [7] Burkner, P.-C., Schwabe, R., & Holling, H. (2018). Optimal Designs for the Generalized Partial Credit Model. *British Journal of Mathematical and Statistical Psychology*, 72(4), 271–293. <https://doi.org/10.1111/bmsp.12148>
- [8] Miyazaki, K., Hoshino, T., Mayekawa, S., & Shigemasu, K. (2009). A New Concurrent Calibration Method for Nonequivalent Group Design Under Nonrandom Assignment. *Psychometrika*, 74(1), 1–19. <https://doi.org/10.1007/s11336-008-9076-x>
- [9] Robert W. Lissitz and Xiaodong Hou University of Maryland
testpublibshers.org
- [10] Kim, S.-H., & Cohen, A. S. (1998). A Comparison of Linking and Concurrent Calibration Under Element Response Theory. *Applied Psychological Measurement*, 22(2), 131–143. <https://doi.org/10.1177/01466216980222003>
- [11] Hanson, B. A., & B'eguín, A. A. (1999). Separate Versus Concurrent Estimation of IRT Element Parameters in the Common Element Equating Design. ACT Research Report Series. https://www.act.org/content/dam/act/unsecured/documents/ACT_RR99-08.pdf
- [12] Kang, T., & Petersen, N. S. (2009). Calibrating and Linking Element Parameters to a Base Scale. Educational Testing Service. <https://files.eric.ed.gov/fulltext/ED510480.pdf>
- [13] Thompson, N. A. (2021). Three Approaches for IRT Equating. Assessment Systems Corporation. <https://assess.com/irt-equating/>
- [14] PhilChalmers, JoshuaPritikin, AlexanderRobitzsch, Package 'mirt' January 28, 2025, project.org/web/packages/mirt
- [15] Priyank Mishra, How to Calculate Cronbach's Alpha in R? <https://www.geeksforgeeks.org/how-to-calculate-cronbachs-alpha-in-r/>
- [16] Robitzsch, A. (2024). TAM: Test Analysis Modules (Version 4.2-21) [R package]. Retrieved from <https://cran.r-project.org/web/packages/TAM/index.html>
- [17] Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and element response theory analyses. *Journal of Statistical Software*, 17(5), 1–25.
- [18] Robitzsch, A., et al. (2020). TAM: Test analysis modules for MML-based element analysis. R package version 3.4–39.
- [19] A.B. Normurodov, M.Dj. Ermamatov, KORRELYATSIYA TAHLILI: KIMYO VA BIOLOGIYA FANLARIDAN TEST NATIJALARI, "Axborotnoma" ilmiy-uslubiy jurnali, 1/2025.
- [20] Kyung Yong Kim. A Comparison of the Separate and Concurrent Calibration Methods for the Full-Information Bifactor model. {pubmed.ncbi.nlm.nih.gov}
- [21] Wilkinson, L., & Friendly, M. (2009). The History of the Cluster Heat Map. *The American Statistician*, 63(2), 179–184.

- [22] Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55.
- [23] Cambridge Assessment International Education. (2024). *Mathematics Syllabus: Assessment Objectives and Mark Schemes*. Cambridge University Press.
- [24] International Baccalaureate Organization. (2021). *Mathematics: Analysis and Approaches Guide*. IB Publishing Ltd.
- [25] Joost C F de Winter, Samuel D Gosling, Jeff Potter, Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data
- [26] Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*.
- [27] Haladyna, T. M. (2004). *Developing and Validating Multiple-choice Test Items* (3rd ed.).
- [28] Sabio, Balagtas & David (2015) — “Backwash Effects of Testing on Learning Mathematics”
- [29] Ekström, Joakim (2009) — “The Phi-Coefficient, the Tetrachoric Correlation Coefficient, and the Pearson-Yule Debate”