

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

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 matematika fanidan 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 javobli testlar uchun *ikki parametrli logistik model* (2PL) tanlab olindi. Tadqiqot va adabiyotlarni 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 tadqiqot natijalariga nisbatan mosligini tekshirish usullari yoritildi, yozma ish uchun modelga mos baholash mezonlari o'rganildi. Shuningdek, aralash formatdagi imtihonlarda umumiy natijani aniqlash maqsadida *birgalikda (konkurrent) kalibrovkalash* 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 (konkurrent) kalibrovkalash, baholash mezonlari, issiqlik xaritasi, simulyatsiya.

1. Kirish

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 savolni 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 masalalarning 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 [1] quyi kognitiv darajalarini (bilish, tushunish, qo'llash) baholash imkonini beradi. Biroq yuqori darajadagi kognitiv

kompetensiyalarni (tahlil, sintez, baholash) aniqlash va baholash uchun yozma ishlar muhim ahamiyat kasb etadi [2]. Tabiiyki, imtihonlarda yuqori darajadagi kognitiv kompetensiyalarni aniqlashga yo'naltirilgan savollarni kiritish, kelgusida tayyorgarlik ko'rayotgan talabgorlar uchun ijobiy ta'sir effektini (ing. positive washback effect) yuzaga keltiradi [3].

Matematika fani va boshqa predmetlar bo'yicha yozma ishlarni zamonaviy psixrometrik 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 ketma-ketligini ishlab chiqish zarurati paydo bo'ladi.

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. Muqobil javobli test natijalarini baholash uchun ikki parametrlil model (2PL) tanlanadi. 2PL model har bir

savolning ikki parametrini baholaydi: ajratish (ing. discrimination) va qiyinlik darajasi (ing. difficulty). Ajratish parametri savolning talabgor qobiliyatiga qanchalik sezgir ekanini ko'rsatadi; qiyinlik parametri esa savolning qanchalik qiyin ekanini (qobiliyat shkalasida) bildiradi. Bu yerda parametrlarni baholash maqsadida *maksimal chegaraviy 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:

- a_i – elementning diskriminatsiya (farqlash) parametri ($a_i > 0$);
- b_i – elementning qiyinlik (difficulty) parametri;
- $\theta_n - n$ – talabgorning latent qobiliyati.

Olingan ma'lumotlar asosida yakuniy test natijalarini olish uchun R dasturining 'ltm', 'mirt', 'TAM' kabi

• *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.

paketlarida kerakli funksiyalar to'plami bor, misol uchun maksimal o'xshashlik modeli har bir paketda har xil funksiyada uchraydi, turlicha iteratsiyalar bilan hisoblaydi, ammo bir-biriga yaqin natijalar beradi. Bularni amalga oshirish quyidagi buyruq funksiyalarga asoslanadi va yechim iteratsion optimizatsiya orqali topiladi [4]:

Maksimal qiymat beradigan θ_n ni topish uchun iteratsiya va boshqa hisoblash usullari qo'llaniladi.

2.2. Yozma ish topshiriqlari uchun

Ma'lumki, yozma ish topshirig'i – bu ishtirokchidan matematik masala yoki muammoni yechim bosqichlarini batafsil bayon qilgan holda yozma ravishda ifodalashni talab qiluvchi imtihon shaklidir. Ushbu topshiriqda har bir yechim bosqichi – strategiyani tanlash, hisoblash jarayonini asoslash va natijani to'g'ri yozish alohida baholanadi. Yozma ish topshirig'i o'quvchining matematik fikrlash, muammoni tahlil qilish, izchil va aniq

yozma ifoda hamda hisoblash tartibli baholash imkonini beradi.

Yozma ish topshiriqlarini 10 ta deb qabul qilgan edik. Har bir topshiriq bir necha ball darajasiga ega (masalan: 0 dan 3 gacha). Bunday politomik (ko'p kategoriyali) elementlar uchun *umumlashtirilgan qisman kredit modeli* (ing. Generalized partial credit model - GPCM) qo'llaniladi [5]. Bu model Masters [6] tomonidan taklif qilingan *qisman kredit modeli* (PCM)ning

umumlashtirilgan ko'inishi bo'lib, har bir topshiriq uchun diskriminatsiya (ajratish) parametri va har bir ball kategoriyasi uchun saralangan qiyinlik *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.

Chegara - bu qobiliyat shkalasida bir kategoriyadan ikkinchi kategoriyaga o'tishda talabgorning 50 foiz imkoniyati borligini ifodalaydi. Masalan, 3 kategoriyali elementni

olaylik (0,1,2): talabgorning 1-chegara 0 ball olishga nisbatan 1 ball olishda 50 foiz imkoniyati, 2-chegara esa talabgorning 1 ga nisbatan 2 ball olishda 50 foiz imkoniyati mavjudligini ifodalaydi. Qisman kredit modellarida qiyinlik darajalari chegaralar bo'yicha taqsimlanib ketadi. Ammo, umumlashtirilgan qisman kredit modellarida ehtimollik har doim ham 50 foizga teng bo'lmasligi mumkin.

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 [5]:

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

bu yerda:

- $X_{ij} - j -$ talabgorning i -savolga olgan balli ($0 \leq k \leq M_i$)
- $a_i - i$ -savolning diskriminatsiya parametri
- $\theta_j - j -$ talabgorning yashirin qobiliyati
- $b_{im} - i$ -savol uchun m -bosqich (chegara) qiyinlik 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 [7].

Bu ballarni qo'yish uchun quyidagicha baholash mezonini maqsadga muvofiq [8-9]:

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 lozim. 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 lozim.

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 lozim.

Bu mezonlar asosida yozma ishlar tekshiriladi va 1-jadvaldagi holat asosida ballar qo'yiladi [9].

1-jadval

Yozma ishni baholash mezonlari

№	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, qiyinliklar 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 (konkurrent) kalibrovkalash

Endi oldimizda yana bir amaliy masala yuzaga keladi, ya'ni bu ikki muqobil javobli va yozma ishlardan olingan qobiliyatlarni umumlashgan yakuniy natijasini hisoblash, buning uchun *bir vaqtda* (concurrent)

kalibrovkalash modelidan foydalanamiz [10].

Birgalikda (konkurrent) kalibrovkalash deganda, testdagi barcha elementlar – dixotomik ham, politomik ham – bir vaqtning o'zida, yagona modelda kalibrovkalanishi tushu-

niladi. Bu yondashuv har ikki turdagi savollarni bitta yashirin (ing. latent) o'lchov bo'yicha bog'laydi va umumiy natijani chiqaradi [11].

Ta'kidlash lozimki, birgalikda (konkurrent) kalibrovkalashni IRT doirasida individual savollarni alohida kalibrovka qilish yoki test natijalarini tenglashtirish (equating) bilan almashtirmaslik zarur. [13] Konkurrent kalibrovkalash – parametrlarni umumiy model asosida bir vaqtning o'zida baholash jarayonidir, tenglashtirish esa – turli test shakllarida olingan ballarni bir xil metrikada ifodalash uchun qo'llaniladigan alohida psixometrik yondashuvdir [12].

Birgalikda kalibrovkalashning 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. Bir o'lchovli (ing. unidimensional) IRT modeli faraziga

ko'ra, barcha elementlar bitta qobiliyatni o'lchayotgan bo'lsa, birgalikda kalibrovkalash to'g'ri natija beradi [14].

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 modellari 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 (konkurrent) kalibrovkalash 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 kalibrovkalash vaqtida, barcha elementlar bitta yashirin (latent) o'lchovga keltiriladi. Bu testning umumiy natijasi o'lchovi bo'lib, har ikki turdagi elementlar shunga nisbatan parametrlanadi.

Maksimal chegaraviy o'xshashlik orqali baholash MIRT paketida standart bo'lib, *maksimal kutilish*

algoritmi (ing. Expectation-Maximization - EM) yordamida bajariladi. Buning nazariy asoslariga to'xtalmaymiz.

Birgalikda (konkurrent) kalibrovkalash usulini unga muqobil bo'lgan modellar bilan taqqoslangan adabiyotlar [15-17] 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 [18].

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.

Bu test ma'lumot grafigini R dasturida

TAM: IRT. information Curves (mod) [19] yoki *mirt* paketida **testInfo** funksiyalari orqali olish mumkin. Maqsad – muqobil javobli 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 [20].

Tekshirishning yana biri *Kronbach alfani* aniqlash orqali aniqlanadigan 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 [21]. Masalan, $\alpha = 0,85$ bo'lsa, demak test "yaxshi" ishonchlilikka ega deyish mumkin [22].

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 [23].

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.

Ixtiyoriy moslashtirish ta'sirini (ing. overfitting) kamaytirish va parametrlarni ishonchli baholash uchun polytomik IRT modellarida kamida 300 nafar ishtirokchini, itemlar soni esa kamida 10–15 ta elementni tashkil qilishi tavsiya etiladi. Optimal natijaga erishish uchun esa, namunalar soni 500 va undan ortiq, itemlar soni esa 20 yoki undan ko'proq bo'lishi maqsadga muvofiq.

Odatda yuqori malakali test tuzuvchilar tomonidan shakllantirilgan va malakali ekspertizadan o'tkazilganda itemlar soni 26 tadan kamroq bo'lishiga yo'l qo'yiladi.

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 kalibrovkash usulidan foydalanish, shuningdek, mos ravishda umumlashgan qisman kredit modeli (ing. generalized partial credit model, GPCM) hamda ikki parametrlil 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 qiymat-

larida 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 qiyinlik parametri mavjud bo'lsa, 10 ta yozma topshiriq uchun 30 ta qiyinlik parametri aniqlanadi, bu esa vaqt taqsimoti nuqtayi nazaridan teng ahamiyatga ega bo'lgan topshiriqlar nisbatini ta'minlaydi.

Mazkur test formatining samardorligi va tanlangan modellarning maqsadga muvofiqligini ilmiy faraz sifatida qabul qilish orqali, *virtual sinov* (ing. pilot test) amalga oshirildi. Buning uchun R dasturida 1000

kishilik natijalar simulyatsiya qilindi. Bunda, simulyatsiya ko'rsatkichlari normal taqsimotga ega bo'lishi sharti qo'yildi, aks holda yakuniy natijalar-

ning haqiqiy vaziyatga yaqinligi ta'minlanmasligi mumkin.

2-jadvalda R dasturida olingan har bir elementning a –diskriminatsiya va b –qiyinlik darajasi berilgan.

2-jadval

R dasturida simulyatsiya qilingan elementlarning qiyinlik darajalari

element	a	b	b1	b2	b3
Q1	0,867	0,659			
Q2	0,997	1,630			
Q3	1,008	0,687			
Q4	1,273	-1,109			
Q5	1,339	0,726			
Q6	1,429	1,452			
Q7	1,483	-0,863			
Q8	0,930	-0,029			
Q9	0,784	-1,441			
Q10	1,130	-0,874			
Q11	1,743	-1,144			
Q12	1,078	0,817			
Q13	1,734	0,454			
Q14	1,553	1,066			
Q15	1,715	1,085			
Q16	1,025	-2,415			
Q17	1,846	-0,630			
Q18	0,924	1,914			
Q19	1,359	1,290			
Q20	0,746	-0,214			
Q21	1,960	0,732			
Q22	1,332	-0,205			
Q23	1,978	-1,761			
Q24	1,003	0,506			
Q25	0,860	-1,288			
Q26	1,150	0,329			

Q27	1,557	2,367			
Q28	1,451	-0,558			
Q29	1,244	-0,424			
Q30	1,745	-1,917			
Q31	1,456	0,675			
Q32	1,510	-0,253			
Q33	1,639	-1,368			
Q34	1,028	-0,593			
Q35	0,909	1,863			
Q36	1,652		-0,875	-0,207	0,409
Q37	1,616		-2,010	-2,212	-2,079
Q38	1,586		1,513	-1,594	-0,922
Q39	1,338		0,694	1,525	0,277
Q40	1,250		-0,634	-1,631	-0,137
Q41	1,399		0,050	0,797	-0,973
Q42	1,532		1,838	-0,389	1,176
Q43	0,996		-0,335	-1,511	2,393
Q44	1,405		-1,250	-1,030	1,192
Q45	0,856		0,031	1,777	0,325

3-jadval

Turli modellar bilan hisoblangan test natijalarinig Kronbax alfa ko'rsatkichlari:

turi	2PL + GPCM	1PL + PCM
Umumiy	0,928	0,916
Muqobil javobli	0,887	0,862
Yozma ish	0,878	0,878

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

Ma'lumki, test topshiriqlarining chuqur tahlili uchun elementlar o'rtasidagi o'zaro bog'liqliklarni aniqlash hamda vizual ko'rsatish muhim hisoblanadi [24]. Endi test

tarkibidagi 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 dixotomik elementlar (2PL modeli asosida) ham o'ziga xos mustaqil qiyinlik parametrlari sifatida tahlilga kiritildi. Natijada, jami 65 ta ko'rsatkich (35 ta dixotomik 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 qiyinlik 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 phi-koeffitsiyent deb ataladi. Phi-koeffitsiyent aynan dixotomik o'zgaruvchilar o'rtasidagi bog'liqlikni o'lchaydi va Pirson koeffitsiyentining xususiy holi hisoblanadi [25].

Shuningdek, ushbu yondashuv test tuzilmasini vizual va statistik tahlil

qilish uchun universal, hisoblashda samarali vosita hisoblanadi. Shu sababli, tadqiqotda indikatorlar o'rtasidagi o'zaro bog'liqliklarni aniqlash uchun Pirson korrelyatsiya koeffitsiyenti tanlandi [27][28].

Bu yondashuvda barcha element va chegaralar o'zaro korrelyatsiya koeffitsiyenti matritsasiga o'tkazilishi va "issiqlik xarita" (ing. heatmap) ko'rinishida vizuallashtirilishi mumkin.

Ushbu vizual tahlil testning o'zaro yaqin yoki o'xshash element-bosqichlarni aniqlash imkonini beradi. 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 matrix <
– cor(data, method =
"pearson", use =
"pairwise.complete.obs")
```

buyruqdan foydalanish mumkin.

Amaliy nuqtayi nazardan, ushbu metodologiya testlarni loyihalash yoki validlikni aniqlashda keng imkoniyatlar ochadi [24].

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 talabgorning

qobiliyat darajasini ishonchli va xolis baholash imkonini berishi haqidagi faraz ilgari surilgan edi. Mazkur tadqiqot doirasida farazlarni to'liq tekshirish uchun real ishtirokchilar bilan aprobatsiya 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 modellari 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 [4-7, 9, 10]. Shunga qaramasdan, kelgusidagi real aprobatsiyalar orqali olib boriladigan tekshiruvlar asosida

zarur tuzatish va qo'shimchalar kiritilishi mumkin.

Birgalikda (konkurrent) kalibrov-kalashning 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 ustuvor ahamiyatga ega.

ADABIYOTLAR

1. Bloom, B. S. (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain.
2. Haladyna, T. M. (2004). Developing and Validating Multiple-choice Test Items (3rd ed.).
3. Sabio, Balagtas & David (2015) – “Backwash Effects of Testing on Learning Mathematics”.
4. Chalmers, R. P. (2012). mirt: A multidimensional element response theory package for the R environment. Journal of Statistical Software, 48(6), 1–29.
5. Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. Applied Psychological Measurement, 16(2), 159–176.
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. A'zamov A., Haydarov B.Q. Matematika sayyorasi. -T.: O'qituvchi, 1993.-312 b.
9. Cambridge Assessment International Education. (2024). \emph{Mathematics Syllabus: Assessment Objectives and Mark Schemes}. Cambridge University Press.
10. 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>.
11. Thompson, N. A. (2021). Three Approaches for IRT Equating. Assessment Systems Corporation. <https://assess.com/irt-equating/>.
12. M.Dj. Ermamatov, A. Abbosov, A.A. Baratov. Test topshiriqlarini kalibrovkalash va qobiliyatlarni tenglashtirish, “Axborotnoma” ilmiy-uslubiy jurnali, 3–4/2022, Toshkent.
13. 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.

14. PhilChalmers, JoshuaPritikin, AlexanderRobitzsch, Package 'mirt' January 28, 2025, project.org/web/packages/mirt.
15. 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>.
16. Kyung Yong Kim. A Comparison of the Separate and Concurrent Calibration Methods for the Full-Information Bifactor model. {pubmed.ncbi.nlm.nih.gov}.
17. 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>.
18. Robert W. Lissitz and Xiaodong Hou University of Maryland testpublibshers.org.
19. Priyank Mishra, How to Calculate Cronbach's Alpha in R? <https://www.geeksforgeeks.org/how-to-calculate-cronbachs-alpha-in-r/>.
20. 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>.
21. Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and element response theory analyses. *Journal of Statistical Software*, 17(5), 1–25.
22. Robitzsch, A., et al. (2020). TAM: Test analysis modules for MML-based element analysis. R package version 3.4–39.
23. A.B. Normurodov, M.Dj. Ermamatov, Korrelyatsiya tahlili: kimyo va biologiya fanlaridan test natijalari, "Axborotnoma" ilmiy-uslubiy jurnali, 1/2025.
24. Wilkinson, L., & Friendly, M. (2009). The History of the Cluster Heat Map. *The American Statistician*, 63(2), 179–184.
25. Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55.
26. International Baccalaureate Organization. (2021). *Mathematics: Analysis and Approaches Guide*. IB Publishing Ltd.
27. 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.
28. Ekström, Joakim (2009) — "The Phi-Coefficient, the Tetrachoric Correlation Coefficient, and the Pearson-Yule Debate".

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

Z.M. Tulanov, Q.P. Murodov

Academic lyceum of the Namangan State University, 160107, Namangan, Boburshoh st., 161

Abstract. In this article, the application of a modern psychometric model for evaluating test results involving polytomous items—the Generalized Partial Credit Model (GPCM)—is explored, alongside the potential for integrating step-by-step assessment criteria for written assignments using this model. To develop a fair assessment system for mathematical written tasks within this research, three-stage criteria—identification of solution strategy, calculation accuracy, and presentation style—were adopted as distinct scoring categories for each written assignment. For multiple-choice tests, the two-parameter logistic model (2PL) was selected. During the review of literature and study process, it was determined that using the Generalized Partial Credit Model enables a more precise differentiation of participants' knowledge and competencies through each scoring category in evaluating written assignments. Methods to assess model fit to the current tasks were highlighted, and suitable assessment criteria aligning with the model were investigated for written tasks. Additionally, relevant information was provided regarding the concurrent calibration method for determining overall scores in mixed-format exams.

At the conclusion of the research, results and findings from virtual trial tests were presented. Furthermore, in examining the correlation between written assignments and multiple-choice question results, it was demonstrated that statistical analysis could be enhanced by treating each stage of written work as a separate test item (item parameter).

Keywords: 2PL logistic model, partial credit model, threshold, concurrent calibration, assessment criteria, heat map, simulation.