

MATEMATIKA FANIDAN KO'P TANLOVLI VA YOZMA ISH SHAKLIDAGI IMTIHON NATIJALARI ASOSIDA ISHTIROKCHILARNING QOBILIYATINI BIRGALIKDA KALIBRLASH

EVALUATING PARTICIPANTS' ABILITIES IN MATHEMATICS EXAMS USING MULTIPLE-CHOICE AND WRITTEN WORK VIA CONCURRENT CALIBRATION

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

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

Qisqacha mazmuni. Test natijalarini adolatli baholashda polixotomik elementlar uchun zamonaviy psixrometrik model — *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 me'zonlar: yechim strategiyasini aniqlash, hisoblash aniqligi va taqdim etish uslubi alohida ball toifalari sifatida qabul qilindi. Ko'p tanlovli testlar uchun *ikki parametrli logistik model* (2PL) tanlab olindi. Tadqiqot va adabiyotlarni [4-8] o'rganish jarayonida GPCM 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 tanlanayotgan savollar ma'lum xususiyatlarga ega bo'lishi lozimligi aniqlandi. Shuningdek, aralash formatdagi imtihonlarda umumiy natijani aniqlash maqsadida birgalikda konkurent kalibrlash usuli haqida qo'yilgan maqsadga mos ma'lumotlar berildi. Shuningdek, yozma ish va ko'p tanlovli savollarning korrelyatsiyasini o'rganishda *yoza ish bosqichlarini alohida savol parametri* deb qabul qilingan holda ko'rsatkichlarini olish ilmiy yangilik sifatida yoritildi.

Kalit so'zlar: Partial Credit Model, Birgalikda konkurent kalibrlash, baholash mezonlari, treshold, dixotomik, polixotomik, heatmap

1. Kirish

Ma'lumki milliy sertifikatlash tizimining asosiy mezonlaridan biri adolatli baholash hisoblanadi, shuningdek matematika fanining ayrim masalalari shunday xususiyatga egaki, talabgor bilimining bir nechta qirralarini ochib beradi, ushbu 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, shunda talabgorlar ichida mavzularni yaxshi o'zlashtirgan kognitiv bilim ko'nikmalarini olgan, ammo qisqa vaqt sharoitida arifmetik

xatoliklarga, yoki ayrim masalalarni xususiyatidan kelib chiqib (shu talabgorga nisbatan) ko'p arifmetik hisob-kitob talab qiladigan masalalarni xato ishlash yoki oxiriga yetkazmaslik holatlari yuzaga kelishi mumkin [1]. Bunday vaziyatlarda talabgorning ayrim kognitiv qobiliyatlari to'liq baholanmasligi ehtimoli paydo bo'ladi. Bu muammo uchun yozma ishlarni tekshirishda ayrim bosqichlarga ham ball berish amaliyoti yechim bo'ladi. 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 scriptlari ishlab chiqilishi zarurati paydo bo'ladi.

2. Aralash formatdagi testlarni baholash tizimi

2.1 Ko'p tanlovli yopiq testlar uchun

Demak, bizning oldimizda yozma ish shaklidagi ishlarni baholash va uni ko'p tanlovli test savollari bilan birgalikda baholash tizimini 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: Ko'p tanlovli 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 murakkab ekanini (qobiliyat shkalasida) bildiradi. Bu modelni Maximum Marginal Likelihood (MML) usulida hisoblashni misol tariqasida tanlab olamiz. Aslida EAP usuli ham mavjud [2].

Elementning javob funksiyasi (1PL va 2PL uchun)

a) 1PL (Rasch) modeli:

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

b) 2PL modeli:

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

bu yerda:

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

R ning 'ltm', 'mirt', 'TAM' kabi paketlari bizga 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, itemtype = "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 xisoblash 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 polixotomik (ko'p kategoriyali) itemlar uchun *umumlashtirilgan qisman kredit modeli* (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 ajratish parametri va har bir ball kategoriyasi uchun saralangan qiyinchilik (*threshold) parametrlarini baholaydi.

Threshold - bu polixotomik (ko'p balli) IRT modellarida, masalan, Generalized Partial Credit Model (GPCM) yoki Graded Response Model (GRM) da, har bir item uchun har bir ball pog'onasidan yuqoriroq pog'onaga o'tish ehtimolini belgilovchi parametr.

Bu model ham MML yordamida baholanadi va 2PL kabi har bir itemning respondent qobiliyatiga ta'sirini hisobga oladi.

GPCM modeli - ko'p kategoriyali (polixotomik) itemlar uchun latent qobiliyatga bog'liq bo'lgan ehtimollik modelidir. *i*-item bo'yicha respondent *j* nomerli ishtirokchi tomonidan *k* ball

olingan ehtimollik quyidagicha ifodalanadi [6]:

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

bu yerda:

- X_{ij} - *j*-respondentning *i*-savolga olgan balli ($0 \leq k \leq M_i$)
- a_i - *i*-savolning diskriminatsiya parametri
- θ_j - *j*-respondentning latent qobiliyati
- b_{im} - *i*-savol uchun *m*-bosqich (threshold) qiyinchilik parametri
- M_i - *i*-savol uchun maksimal ball.

2PL modeli dixotomik savollar uchun mos, GPCM esa har bir polixotomik savolga alohida diskriminatsiya parametr berib, turli topshiriqlarning testdagi hissasini adolatli aks ettiradi. Shu tariqa, har ikkala turdagi savollar yagona latent qobiliyat shkalasida baholanadi.

3. Yozma ishlarni tekshirish va ball qo'yish tizimi

Yozma ish savollarida har bir band bo'yicha kognitiv chuqurlikni 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. Bu ballarni qo'yish uchun quyidagicha baholash mezoni maqsadga muvofiq [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 [1]. *Eslatma: bu yerdagi baholash mezoni va ballash tizimi to'liq mukammal emas.*

1-jadval

No	Yechimning umumiy holati	Ball
1	Masala to'liq yechilgan	3
2	Masala to'liq yechilgan, ammo yechimda ayrim muhim bo'lmagan kamchiliklar 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.	1
4	Masalani yechishga mutlaqo urinib ko'rilmagan, masala yechish uchun yozuvlar yozilgan lekin yechim sari yurilmagan	0

4. 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 itemlar – dixotomik ham, polixotomik ham – bir vaqtning o'zida, yagona modelda kalibrlanishi tushuniladi. Bu yondashuv har ikki turdagi savollarni bitta latent o'lchov bo'yicha bog'laydi va umumiy natijani chiqaradi [8].

Birgalikda kalibrlashning afzalligi shundaki, ko'p tanlovli savollar va yozma topshiriqlarning parametrlarini alohida-alohida emas, balki birgalikda baholab, umumiy qobiliyat bahosini aniqroq va adolatli shakllantirish mumkin. Ikkala turdagi itemlar 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 ko'p tanlovli savollar soni ko'proq bo'lsa, ular ham o'z hissasini qo'shadi. Unidimensional IRT modeli faraziga ko'ra, barcha itemlar bitta qobiliyatni o'lchayotgan bo'lsa, birgalikda kalibrlash to'g'ri natija beradi [9].

Bularni amalga oshirish uchun albatta R dasturidan foydalanish maqsadga muvofiq, chunki bundan boshqa dasturlash tillarida iteratsiya muammosiga duch kelamiz. Birinchidan, R dasturida testlarni baholash uchun kerakli funksiyalar shay holatga keltirilgan va logarifmik tenglamalarni irratsional ildizlarini topish jarayonida iteratsiyaviy siklga tushib qolmaydi.

Avvalo biz modelda qaysi itemlar 2PL, qaysilari GPCM ekanini belgilab olishimiz kerak. Masalan, R dagi *mirt* paketi yoki TAM paketi bunga imkon beradi: *mirt* paketida *mirt()* funksiyasiga *itemtype* parametri orqali har bir itemning turini ko'rsatish mumkin. Masalan, 35 ta dioxotomik savollar uchun "2PL", 10 ta polioxotomik savollar uchun "gpcm" deb ko'rsatamiz [14]. *mirt* paketining afzalligi - u bir necha turdagi itemlarni bir modelda oson baholay oladi.

Birgalikda kalibrlash vaqtida, barcha itemlar bitta latento'lchovni "ability" qobiliyat deb atadik. Bu testning umumiy natijasi o'lchovi bo'lib, har ikki turdagi itemlar shunga nisbatan parametrlanadi.

5. Model va testning mosligini ta'minlash uchun tahlillar

5.1 Muvozanatli ulush

Agar barcha ko'p tanlovli savollar diskriminatsiya 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 ko'p tanlovli 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 itemning hissasi uning ajratish parametri va testdagi nisbiy informatsiyasiga qarab avtomatik tartibga solinadi, bu bizga qo'shimcha savollarga og'irliklar berib chiqishimiz shart emasligini bildiradi.

Alohida alohida olingan "ko'p tanlovli qobiliyat" va "yozma qobiliyat" natijalarini bir baholash shkalasiga o'tkazishning boshqa usullarini ko'rib chiqish va ularni bu yerdagi gipoteza bilan taqqoslash maqsadga muvofiq. Biz taklif

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

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

Endi ko'p tanlovli va yozma savollar ishtirokchining yakuniy ballida qanchalik rol o'ynayotganini tahlil qilish kerak bo'ladi.

etayotgan model test formatiga nisbatan "egiluvchan" bo'lib, yozma qobiliyatlar uchun olinadigan savollar soni 4 - 10 oralig'ida o'zgargan taqdirda ham adolatlilik xususiyati saqlanib qoladi. Ammo amalda 35+10 formati maqsadga muvofiq bo'lishi mumkin, chunki 35 ta savolda 35 qiyinlik parametri bor deb olsak, 10 yozma savol uchun 30 ta qiyinlik parametri bor, bu vaqt taqsimotiga nisbatan teng kuchli miqdorlardir.

Testning ishonchliligini tekshirishning yana bir usuli *testning informatsion egri chizig'i* 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 informatsiyasi grafigini olish mumkin. Maqsad - ko'p tanlovli va yozma itemlar birga qaysi diapazonda eng ko'p ma'lumot berishini ko'rish. Ideal holda testning informatsiya egri chizig'i keng

diapazonda yuqori bo'lib, bu test natijasining ishonchliligini bildiradi [22]. Tabiiyki, bizning gipotezimizni tekshirish uchun muhim ko'rsatkich hisoblanadi.

Tekshirishning yana biri *Kronbax alfa* klassik usul bo'lib, itemlarning 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]. Kronbax alfa 0.8 dan yuqori bo'lsa, test ishonchli hisoblanadi[15].

Bizning testda itemlar polixotomik bo'lsa ham, cronbach.alpha ularni sonli shkala sifatida qabul qilib,

umumiy α ni hisoblaydi. Natijada α qiymati va uning ishonch oralig'i chiqadi. Masalan, $\alpha = 0.85$ bo'lsa, demak test "yaxshi" ishonchlilikka ega deyish mumkin [15]. Kronbax alfa yuqori bo'lishi uchun itemlar 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.

5.2 Test natijalarining ichki tuzilmasini tahlil qilish, korrelyatsion heatmap

Ma'lumki, Test topshiriqlarining chuqur tahlili uchun itemlar o'rtasidagi o'zaro bog'liqliklarni aniqlash hamda vizual ko'rsatish muhim hisoblanadi [19][21]. Endi test tarkibida yozma ish va ko'p tanlovli savollarni qanday bog'liqligini o'lchash mumkin degan o'rinli savol tug'iladi. Ushbu tadqiqotda, test tuzilmasining yangi aspekti sifatida, har bir polikategoriya (GPCM) itemining har bir threshold (bosqich) parametrini alohida ko'rsatkich sifatida qabul qilib oldik. Shuningdek, barcha dixotomik itemlar (2PL modeli asosida) ham o'ziga xos mustaqil qiyinchilik parametrlari sifatida tahlilga kiritildi. Natijada, jami 65 ta ko'rsatkich (35 ta dixotomik savol va 10 ta polikategoriya savol uchun 3 tadan threshold, jami 30 ta) har bir ishtirokchi bo'yicha 0/1 ko'rinishida kodlandi.

Har bir threshold (bosqich) alohida qiyinchilik parametri sifatida talqin qilinishi, test natijalarini an'anaviy faqat umumiy ball yoki umumiy item ko'rinishida emas, balki har bir bosqichdagi muvaffaqiyat

imkoniyatlarini ham mustaqil tarzda tahlil qilish imkonini berdi.

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

Ma'lumki, Pearsonning momentlar bo'yicha korrelyatsiya koeffitsiyenti uzluksiz o'zgaruvchilar o'rtasidagi chiziqli bog'liqlikni o'lchash uchun keng qo'llaniladi. Shu bilan birga, dixotomik (0/1) ma'lumotlar uchun ham Pearson korrelyatsiyasi mos keladi va bu holatda u phi-koeffitsiyent deb ataladi. Phi-koeffitsiyent aynan dixotomik o'zgaruvchilar o'rtasidagi bog'liqlikni o'lchaydi va Pearson koeffitsiyentining xususiy holi hisoblanadi.

Amaliy va nazariy izlanishlar shuni ko'rsatadiki, 0/1 shaklidagi test natijalari uchun Pearson korrelyatsiyasi (ya'ni, phi-koeffitsiyent) indikatorlar o'rtasidagi bog'liqlikni aniq va ishonchli baholash imkonini beradi (Everitt, 1992; DeVellis, 2017). 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 Pearson korrelyatsiya koeffitsiyenti tanlandi. Bu yondashuv savollar bir-biri bilan bog'liqliklarni tahlil qilish imkonini beradi.

Buyondashuvda barcha item va thresholdlar o'zaro korelyatsiya koeffitsiyenti matritsasiga o'tkazildi va "issiqlik xarita" (heatmap) ko'rinishida vizuallashtirildi. Ushbu vizual tahlil testning o'zaro yaqin yoki o'xshash item/bosqichlarni aniqlash imkonini

berdi. Natijalardan shuni kutsak bo'ladiki, ayrim thresholdlar va itemlar 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 [19].

6. Xulosa

Albatta qo'yilgan gipotezani tekshirish ma'lum amaliy ishlarni taqozo qiladi, buning uchun ma'lum R dasturida yozilgan script yordamida 35 ta ko'p tanlovli va 10 ta yozma topshiriqli testni birgalikda IRT asosida baholanishi har bir ishtirokchining qobiliyatini ishonchli va adolatli tarzda aniqlay olishi faraz qilindi. R dagi mirt, TAM, ltm kabi paketlar bu vazifalarni bajarishda katta yordam beradi - ular 2PL, GPCM kabi modellarni qo'llab-quvvatlaydi [17] va zamonaviy psixrometrik tahlillarni amalga oshirishga imkon beradi.

Mualliflar tomonidan qilingan farazlarni ishonchli va tekshirilgan [4][5][6][7][9][10] manbalarga asosanib ishlab chiqilgan bo'lsa-da, uning mutlaq ishonchliligi ustida hukm chiqarish noto'g'ri bo'lishi mumkin, ya'ni ilmiy statistik tekshiruvlardan keyin kerakli to'ldirish yoki o'zgartirish talab etilishi mumkin.

Birgalikda konkurent kalibrlashning afzalliklari: Barcha test shakllari avtomatik ravishda bir xil o'lchov shkalasiga keltiriladi, buning uchun:

- maxsus ishlab chiqilgan baholash mezonini bo'lishi kerak;
- R dasturi uchun to'g'ri ishlaydigan script tayyorlash kerak;
- testlarning modelga mosligini bir nechta bosqichda tekshirish kerak;

Yana shuni xulosa qilib aytishimiz mumkinki, parametrlar bahosining aniqligi uchun umumiy topshiriqlarning soni ko'p bo'lishi talab etiladi.

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 item 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 testpulibshers.org
- [10] Kim, S.-H., & Cohen, A. S. (1998). A Comparison of Linking and Concurrent Calibration Under Item Response Theory. *Applied Psychological Measurement*, 22(2), 131–143. <https://doi.org/10.1177/01466216980222003>
- [11] Hanson, B. A., & B´eguin, A. A. (1999). Separate Versus Concurrent Estimation of IRT Item Parameters in the Common Item 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 Item 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] project.org/web/packages/mirt
- [15] [cronbachs-alpha-in-r](https://cran.r-project.org/web/packages/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 item response theory analyses. *Journal of Statistical Software*, 17(5), 1–25.
- [18] Robitzsch, A., et al. (2020). TAM: Test analysis modules for MML-based item 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] 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.