

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

AXBOROTNOMA

BULLETIN

2

2025

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

Tahririyat hay'ati:

Karimov Madjit
(rais)
Baratov Akmal
(bosh muharrir)
Urayeva Elvira
Mirzayev Farxod
Ermamatov Mirshod
(bosh muharrir o'rinbosari)
Normurodov Asror
(mas'ul kotib)
Sattiyev Abdulaziz
Abbosov Avazbek
Mirvaliyev Zoid

Bosishga ruxsat etildi: 31.07.2025
Shartli bosma tabog'i: 5,4
Nashriyot hisob tabog'i: 5,3
Adadi 100. Buyurtma №2-A - 25

Bilim va malakalarni baholash agentligi
bosmaxonasida chop etildi.

Nashrga tayyorlovchilar:

A. Normurodov

Muharrir:

N.Abduraxmanova

Kompyuterda sahifalovchi:

Y. To'rayev

Tahririyat manzili:

100084, Toshkent shahri,
Bog'ishamol ko'chasi, 12-uy.

© “Axborotnoma”

MUNDARIJA

Kirish	3
M.Dj. Ermamatov, A.B. Normurodov, A.R. Sattiyev Klassik va zamonaviy test nazariyalarida element xususiyatlari	4
Z.M. To'lanov, Q.P. Murodov Matematika fanidan bilimlarni muqobil javobli va yozma ish topshiriqlari asosida zamonaviy test nazariyalari asosida baholash	27
A.A. Abbasov Ko'p darajali ingliz tili testlarida matn-bankli bo'shliqlarni to'ldirish topshiriqlari: dizayn, ishonchlilik va CEFR bilan muvofiqlik	53
K.Q. Jalilov Milliy test tizimi uchun test tuzuvchi mutaxassislarni tayyorlash masalalari (o'zbek tilini bilish darajasi sertifikatini misolida)	54
Q.A. Amonov, A.A. Baratov Kimyo fanidan test natijalarining tafsilotlar asosidagi statistik tahlili	65
M.P. Abdullaeva Yozish va gapirish ko'nikmalarini baholovchi mutaxassislarni tayyorlash tizimi	84

CONTENTS

M.J. Ermamatov, A.B. Normurodov, A.R. Sattiyev Item characteristics in classical and item test theories	26
Z.M. Tulanov, Q.P. Murodov Assessment of mathematics knowledge using modern test theories based on multiple-choice and written response tasks	42
A.A. Abbasov Intra-Text Gap-Filling in Multilevel English Language Tests: Design, Validity, and CEFR Alignment	43
K.K. Djalilov Issues of test item writers training for the national testing system (on the example of the uzbek language proficiency tests)	64
K.A. Amonov, A.A. Baratov Statistical analysis of chemistry test results on the basis of details	83
M.P. Abdullaeva Rater training system for assessing writing and speaking skills	96

KIRISH

“Axborotnoma” ilmiy-uslubiy jurnalining ushbu sonida pedagogik oʻlchovlar boʻyicha ilmiy-uslubiy izlanishlar olib boruvchi mutaxassislar tomonidan olib borilgan tadqiqotlar natijalari haqida oltita maqola berilgan.

Birinchi maqolada 9-sinf oʻquvchilari uchun biologiya fanidan oʻtkazilgan test sinovi natijalarining tadqiqoti amalga oshirilgan. Tadqiqot doirasida elementlar (test topshiriqlari) qiyinlik darajalari zamonaviy psixometrik yondashuvlar – Rash va ikki parametrli logistik (2PL) modellari yordamida baholangan. Maqolada ushbu ikki model bilan hisoblangan qiyinlik darajalari, ular oʻrtasidagi farqlar, elementlarning ajratish darajasi qiymatlari, ularning diagnostik imkoniyatlari va amaliy qoʻllanilishi tahlil qilingan.

Ikkinchi maqolada test natijalarini baholashda politomik elementlar uchun zamonaviy psixometrik model – *umumlashtirilgan qisman kredit modeli* (ing. Generalized partial credit model - GPCM) qoʻllanilishi va u orqali yozma ishlarning bosqichma-bosqich baholash mezonlarini integratsiya qilish imkoniyatlari oʻrganilgan.

Uchinchi maqolada matn-bankli boʻshliqni toʻldirish topshiriqlarini ishlab chiqish prinsiplari, ularning ishonchliligi va Umumiy Yevropa Til Mezonlari (CEFR) bilan qanday muvofiqlashtirilishi muhokama qilinagan.

Toʻrtinchi maqolada testning validligini taʼminlashda test topshiriqlarini tuzishga jalb qilinadigan fan ekspertlari baholash va pedagogik oʻlchovlarning fundamental tamoyil va talablarini qay darajada tushunishlari hamda test tuzish jarayoniga tatbiq qila olishlari muhim ahamiyat kasb etadi. Shu bilan birga, taʼlim tizimidan baholash bilan shugʻullanuvchi tashkilotlarga test topshiriqlarini tuzish uchun jalb qilinadigan aksar mutaxassislar baholash boʻyicha tizimli bilimga ega emasligi va bu mazkur tashkilotlar tomonidan oʻtkaziladigan testlarning validligiga salbiy taʼsir koʻrsatishi haqida soʻz borgan. Shuningdek pedagogik dizaynning “orqaga qarab loyihalashtirish” modeli asosida tuzilgan va gibrid tarzda tashkil etilgan qisqa muddatli kurslar muammoni yechishga yordam berishi mumkinligi tahlil qilingan.

Beshinchi maqolada kimyo fanidan milliy sertifikat uchun oʻtkazilgan test sinovi natijalari asosida talabgorlarning hududlar, yosh va tillar kesimida statistik tahlillari amalga oshirildi. 25–35 yosh oraligʻidagi ishtirokchilarning natijalari boshqa yosh guruhlariga nisbatan yuqorili, aksincha 20 yoshgacha boʻlgan talabgorlarning natijalari boshqa yosh guruhlariga nisbatan past ekanligi aniqlandi. Qiyinlik darajasi yuqori boʻlgan test topshiriqlari test tafsilotiga asosan qaysi tematik mavzularga mos kelishi oʻrganilgan.

Oltinchi maqolada chet tillar boʻyicha yozish va gapirish koʻnikmalari uchun ekspertlarni masofaviy tayyorlash va ularni Bilimlarni baholash agentligining Chet tillarni bilish va egallash darajasini baholash boshqarmasi tomonidan tashkil etiladigan test sinovlarining yozish va gapirish koʻnikmalarini baholash jarayoniga jalb etish boʻyicha metodologiya tahlil qilingan. Maqolada chet tillar boʻyicha fan mutaxassislari boʻlgan nomzodlarni tanlab olish bosqichlari, ularning malakasini oshirish mexanizmi va shuningdek umuman bilimni baholash jarayonining sifatini oshirish va xolisligini taʼminlash choralari eʼtibor qaratilgan.

KLASSIK VA ZAMONAVIY TEST NAZARIYALARIDA ELEMENT XUSUSIYATLARI

M.Dj. Ermamatov, A.B. Normurodov, A.R. Sattiyev

Bilim va malakalarni baholash agentligi huzuridagi Ilmiy-o'quv amaliy markazi, 100084, Toshkent sh., Bog'ishamol k., 12, a.normurodov@uzbmb.uz

Qisqacha mazmuni. Ushbu maqolada 9-sinf o'quvchilari uchun biologiya fanidan o'tkazilgan test sinovi natijalarining tadqiqoti amalga oshirilgan. Tadqiqot doirasida elementlar (test topshiriqlari) qiyinlik darajalari zamonaviy psixometrik yondashuvlar – Rash va ikki parametrlil logistik (2PL) modellari yordamida baholangan. Maqolada ushbu ikki model bilan hisoblangan qiyinlik darajalari, ular o'rtasidagi farqlar, elementlarning ajratish darajasi qiymatlari, ularning diagnostik imkoniyatlari va amaliy qo'llanilishi tahlil qilingan. Shuningdek, klassik test nazariyasi bilan elementlarning nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlari aniqlangan. Ushbu klassik ko'rsatkichlar ikki parametrlil modelda aniqlangan ajratish darajasi parametrlari bilan taqqoslanib, ularning o'zaro bog'liqligi tadqiq qilindi. Ajratish darajasi va nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlarining bir-biriga mos holda o'zgarayotganligi elementlarning ajratish darajasi haqida bu ikkita kattalik ham bir xil ma'lumot berishini ko'rsatdi. Tadqiqot natijalari Rash va ikki parametrlil modellarining elementlar sifatini aniqlashdagi samaradorligini, shuningdek, har bir modelning afzallik va cheklovlarini ochib beradi. Maqola natijalari test sinovlarini ishlab chiqishda hamda baholash tizimini takomillashtirishda foydali bo'ladi.

Kalit so'zlar: nuqtaviy biserial korrelyatsiya koeffitsiyenti, qiyinlik darajasi, Rash modeli, ikki parametrlil model.

1. Kirish

Klassik test nazariyasi va Rash modeli [1-5] asosida amalga oshirilgan bizning oldingi tadqiqotlarimizda [6-9] 9-sinf o'quvchilari uchun biologiya fanidan tadqiqot maqsadida o'tkazilgan test sinovlari natijalari o'rganilgan. Rash modeli faqat qiyinlik darajasi parametri (**b**) ni o'z ichiga oladi. Agar model qiyinlik darajasi (**b**)

va ajratish darajasi (**a**) ni o'z ichiga olsa, u ikki parametrlil model deb nomlanadi. Demak, ikki parametrlil modelda elementning ajratish darajasi (**a**) parametri ham qo'shiladi.

Bu modellar o'rtasidagi farq element parametrlarining soni va ularga beriladigan mazmun bilan aniqlanadi. Modellar test savollarining

sifatini baholash, test ishonchliligini oshirish va har bir talabgorga moslashtirilgan baholash tizimlarini yaratish imkoniyatini yaratadi.

Rash modeli bir parametrli model bo'lib, bir o'lchovlilik (bitta yashirin xususiyat) va lokal mustaqillik farazlarini qabul qilishi uning o'ziga xos xususiyatlaridir. Shuningdek, bu model hamma elementlar uchun ajratish darajasini bir xil deb qabul qiladi. Rash modelining tarafdorlari uning quyidagi:

- modelning murakkab emasligi;
- undagi parametrlarning mazmuni va talqini maqsadga muvofiqligi;
- bu model talabgorlarning xarakteristikalaridan kelib chiqqan holda elementlarning farqli ishlashiga (*ing. differential item functioning*) yo'l qo'ymasligi sababli testda "adolat"ni ta'minlashi;
- uning qobiliyat va qiyinlik darajalarining "interval-daraja o'lchovlari" imkoniyatini yaratishini, bu esa natijalarni guruhlar va vaqt bo'yicha taqqoslashga zamin yaratishi

kabi afzalliklarini etirof etishadi.

Rash modelining tanqidchilari esa barcha elementlar uchun ajratish darajasining bir xilligi ba'zi turdagi ma'lumotlar uchun chegaralash bo'lib qolishini ta'kidlashadi.

Ikki parametrli model birinchi marta F.M. Lord tomonidan [10] 1950-yillarda ishlab chiqilgan va

normal taqsimot funksiyasini qo'llab, aniqroq ehtimolliklarni hisoblashda ishlatilgan. Shundan keyin, Birnbaum [11] ushbu modelni yanada rivojlantirib, test elementlarini tasvirlashda logistik funksiyadan foydalangan. Bu, o'z navbatida, modelning matematik talqinini yanada soddalashtirishga imkon yaratgan.

Normal taqsimot modelida ehtimolliklarni hisoblashda integrallash jarayonlari qo'llaniladi, bu esa hisoblashning murakkabligini oshiradi. Biroq logistik modelda ehtimollik to'g'ridan to'g'ri element va talabgor qobiliyati orasidagi funksiya sifatida ifodalanadi. Bu yondashuvning matematik jihatlari ancha soddalashtirilgan bo'lib, hisoblashlarda integrallashdan foydalanishning hojati yo'q.

Logistik modelning yana bir muhim afzalligi shundaki, uning statistik xususiyatlari yuqori aniqlikka ega va test natijalarining tahlilini osonlashtiradi. Shu sababli ikki parametrli logistik model (2PL) ko'plab zamonaviy test modellari ichida afzalroq hisoblanadi, chunki u test elementlarining qiyinlik darajasini va talabgorlarning qobiliyat darajalarini bir vaqtda hisobga oladi.

Shuning bilan birgalikda matematik nuqtayi nazardan hisoblashlarni amaliyotda olingan natijalar bilan moslashtirib parametrlarni aniqlashda ularning soni kam bo'lishi afzaldir. Tadqiqotlarda bir parametrli va ko'p parametrli modellardan qaysi biri afzalroqligi borasida munozaralar mavjud. Ikki parametrli model

tarafdorlari bu model yaxshiroq moslikni ta'minlab berishini (ayniqsa, elementlar ajratish darajasi bir-biridan sezilarli farq qilganda) aytishadi. Ikki parametrli model tanqidchilari esa modelda parametrlarning ko'pligi ularning talqinini murakkablash-tirishini va "o'ta moslik (ing. overfit)" muammosini keltirib chiqarishini va turli umumiy to'plamlarda o'lchovlar invariantlikni Rash modeli darajasida ta'minlab bera olmasligini ta'kid-lashadi.

Ushbu maqolada 9-sinf o'quv-chilari uchun biologiya fanidan tadqiqot maqsadida o'tkazilgan test sinovlarida foydalanilgan elementlar-

ning qiyinlik darajalarini Rash va ikki parametrli modellar bilan hisoblash orqali ular o'rtasidagi farq va modellarning afzalliklari o'rganildi. Klassik test nazariyasi asosida aniqlangan elementlarining nuqtaviy biserial korrelyasiya koeffitsiyenti va ikki parametrli model orqali aniqlangan elementlarning ajratish darajasi qiymatlari o'rtasidagi bog'liqlik o'rganildi.

Rash modeli doirasida hisob-lashlarda R dasturining dexter to'plamidan [12], ikki parametrli model doirasida hisoblashlarda esa kengaytirilgan dexterMML to'plamidan foydalanildi.

2. Tadqiqot usullari

Zamonaviy test nazariyasining (Item Response Theory - IRT) barcha modellarida talabgorning elementlarga (test topshiriqlariga) to'g'ri javob berish ehtimoli orqali ifodalanadi [1-5]. Har bir talabgor o'ziga xos yashirin (latent) qobiliyat darajasiga (θ_j) ega bo'lib, elementlar esa o'z parametrlari bilan ajralib turadi.

Zamonaviy test nazariyasi faqat savollarga javoblarnigina baholamas-dan, 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.

Umumiy holda ehtimollik quyidagicha belgilanadi:

$$P_{ij} = Pr(y_{ij} = 1 | \theta_j, a_i, b_i, c_i).$$

Bu yerda:

P_{ij} - j -talabgorninng i -elementga to'g'ri javob berish ehtimoli,

θ_j - j -talabgorninng qobiliyat da-rajasi,

a_i - i -elementning ajratish da-rajasi,

b_i - i -elementning qiyinlik da-rajasi,

c_i - i -elementga taxminiy to'g'ri javob berish ehtimoli.

Zamonaviy test nazariyalarining yuqoridagi parametrlarga asoslangan quyidagi modellari mavjud:

bir parametrli model (Rasch):

$$P_{ij}(\theta_j, b_i) = \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)}, \quad (1)$$

ikki parametrlı model (2PL):

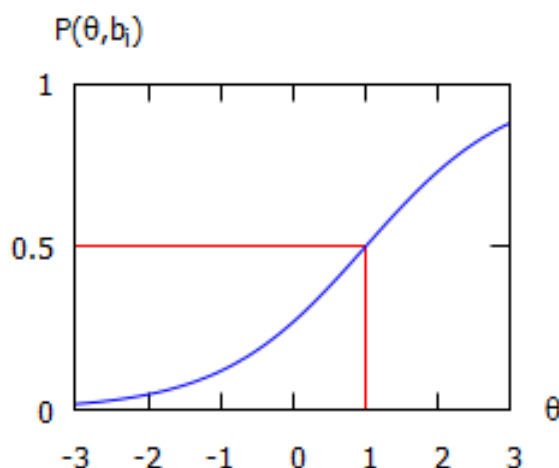
$$P_{ij}(\theta_j, b_i) = \frac{\exp(a_i(\theta_j - b_i))}{1 + \exp(a_i(\theta_j - b_i))}, \quad (2)$$

uch parametrlı model (3PL):

$$P_{ij}(\theta_j, b_i) = c_i + (1 - c_i) \cdot \frac{\exp(a_i(\theta_j - b_i))}{1 + \exp(a_i(\theta_j - b_i))} \quad (3)$$

Eng sodda zamonaviy test nazariyasi modelida dixotomik elementlar bitta parametrga ega. Bunday elementga javob funksiyasi, ya'ni b_i qiyinlik darajasi berilganda qobiliyatning to'g'ri javoblar ehtimoligiga bog'liqligi 1-rasmda berilgan.

1-rasmda ko'rsatilgan grafik bir parametrlı logistik funksiya (logit) deb ataladi. Funksiya argumenti $-\infty$ dan $+\infty$ gacha o'zgarganda ehtimollik 0 bilan 1 orasida o'zgarishi bu funksiyaning o'ziga xos xususiyatidir. Undan tashqari bu funksiya murakkab emas va talqin qilish oson. Modellarda matematik nuqtayi nazardan parametrlar kam bo'lishi ham afzalroqdir [1-5], shuning uchun bu model juda ko'p tadqiq qilinadi va amaliyotda qo'llaniladi.



1-rasm. Bir parametrlı modelda qiyinlik darajasi $b=1$ bo'lgan elementga javob funksiyasi

Yuqorida keltirilgan formulalardagi $\exp(\theta_j - b_i)$ funksiyaga e'tibor berilsa, bir parametrlı logistik modelda to'g'ri javoblarni topish ehtimoli shaxs qobiliyati θ_j va element qiyinligi b_i o'rtasidagi o'zaro ta'sir orqali ifodalanadi. b_i – joylashuv parametri yoki qiyinlik parametri deb ataladi. Qiyinlik parametrining va qobiliyat darajasining o'zaro teng qiymatlariga 0,5 ehtimollik to'g'ri keladi (1-rasm). Bir parametrlı modelni uni rivojlantirgan olim sharafiga Rash modeli deb ham atashadi [1]. Modelning rivojlanishi bilan esa [2-5] manbalar orqali tanishish mumkin.

Ikki parametrlı logistik model (2PL) bir parametrlı modeldan farqli ravishda testga to'g'ri javob berish ehtimoligini faqat talabgorning qobiliyat darajasi (θ) va elementning qiyinlik darajasining faqat bitta parametri (b) bilan emas, balki test

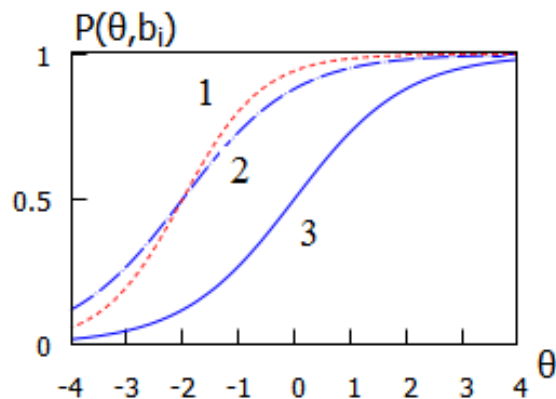
elementi (test topshirig'i)ning ikkita parametriga asoslanib hisoblaydi. Bu parametrlar:

1. b_i – i -elementning qiyinlik darajasi (item difficulty),
2. a_i – i -elementning ajratish darajasi (item discrimination) bo'lib, har bir element qobiliyatlarni qay darajada yaxshi ajratishini ko'rsatadi.

Ikki parametrlı modelning elementga javob funksiyasi (2) formula ko'rinishida bo'ladi [2-5]. Bu formuladan ikki parametrlı modelning bir parametrlı modeldan farqi yaqqol ko'rinib turibdi, $\exp(a_i(\theta_j - b_i))$ - funksiya $\exp(\theta_j - b_i)$ - bilan almashtirilsa bir parametrlı model hosil bo'ladi. Xuddi bir parametrlı modeldagi singari b_i – i -elementning qiyinlik darajasi, yangi parametr a_i – i -elementning ajratish darajasi (ing. discrimination) deb ataladi.

Biz Rash modelida elementga javob funksiyalari doim o'zaro parallel ekanligini va hech qachon kesishmasligini ko'rgan edik. Bu modelda turli xil qiyinlik parametrlari chiziqni o'ngga yoki chapga suradi va uning shakli o'zgarmay qoladi.

Ikki parametrlı modelda bu manzara o'zgaradi.



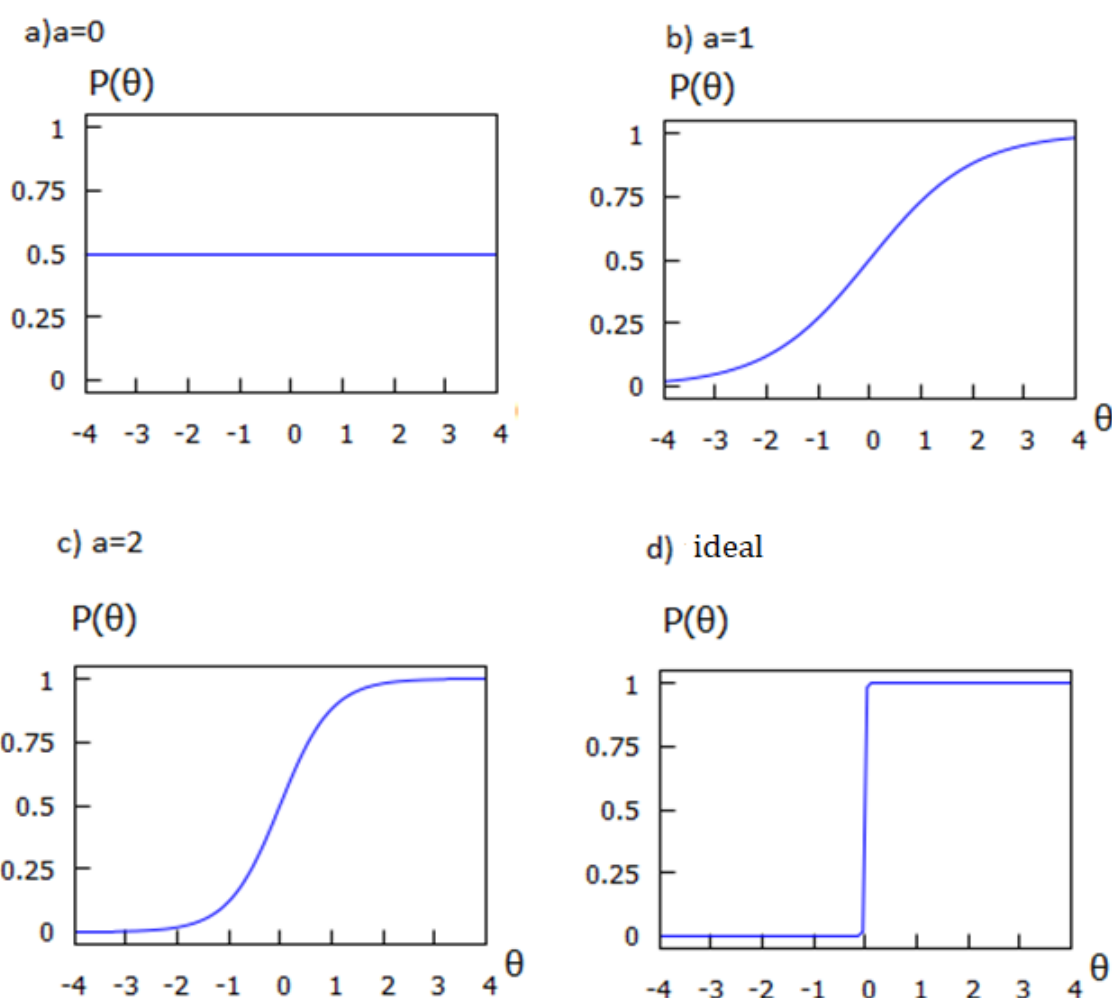
2-rasm. Ikki parametrlı modelning 2 xil ajratish va qiyinlik darajalariga (2- va 3- elementlarning ajratish darajalari bir xil, lekin qiyinlik darajalari har xil, 1- va 2- elementlarning qiyinlik darajalari bir xil, lekin ajratish darajalari farq qiladi) ega elementlarga javob funksiyalari.

2-rasmdan 1- va 2-elementlarning qiyinligi -2 ga teng va ular ikkalasi uchun ham bu qobiliyat darajasida to'g'ri javob berish ehtimoli $0,5$ ga teng, faqat 1-chiziq tikroq ekanligi ko'rinadi. Bunga sabab 1-element ajratish darajasi qiymati ($a_1 = 1,4$) 2-element ajratish darajasi qiymatidan ($a_2 = 1$) katta ekanligidir. 3-element ajratish darajasi qiymati 2-element ajratish darajasi qiymati bilan bir xil ($a_2 = a_3 = 1$), lekin qiyinlik darajasi qiymati bilan farq qiladi ($b_2 = -2; b_3 = 0$). 1-element esa 3-elementdan ajratish darajasi qiymati bilan ham, qiyinlik darajasi qiymati bilan ham farq qiladi. Qiyinlik darajasi parametri b ni joylashuv, ajratish darajasi parametri a ni esa burchak koeffitsiyenti parametri

deb ham atashadi. \mathbf{b} nuqtada burchak ko'effitsiyetining maksimal qiymati $p'(\mathbf{b}) = \frac{a}{4}$ ga tengdir. 1- va 2-chiziqlarning kesishishiga e'tibor qaratamiz. Bu holat bir parametrli modelda umuman sodir bo'lmas edi. 1-chiziq qobiliyatli test topshiruvchilar uchun 2-chiziqqa nisbatan yuqorida joylashgan, qobiliyati past talabgorlar uchun esa unga nisbatan pastda joylashgan. Bu 1-chiziq bilan ifoda-

langan element 2-chiziq bilan ifodlangan elementga nisbatan yuqori qobiliyatli talabgorlar uchun osonroq, quyi qobiliyatli talabgorlar uchun esa qiyinroq ekanligini bildiradi. Demak, 1-chiziq qobiliyatlarni yaxshiroq ajratib beradi.

3-rasmda elementning ajratish darajasi 0 ga, 1 ga va 2 ga teng bo'lgan hamda ideal ajratish darajasiga ega bo'lgan elementga javob funksiyalari ko'rsatilgan.



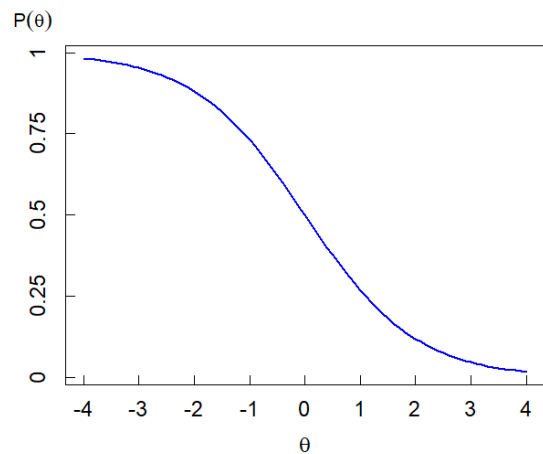
3-rasm. Bir xil qiyinlikdagi ($\mathbf{b}=0$) va turli xil ajratish darajasiga ega bo'lgan ikki parametrli modelning elementga javob funksiyalari

3-a rasmda elementga javob berish ehtimoli har qanday qobiliyat

darajasidagi talabgorlar uchun bir xil, ya'ni bunday elementning ajratish

darajasi nolga teng, 3-d rasmda esa ideallashtirilgan holat tasvirlangan, bunday element o'rtachadan quyi va yuqori qobiliyatlarni aniq ajratib bergan bo'lar edi, ya'ni bu yerda elementning ajratish darajasi qiymati juda katta. 3-b va 3-c rasmlar 3-a va 3-d rasmlar oralig'ida elementning ajratish darajasi qanday o'zgarishini ko'rsatish uchun berilgan. Ba'zan, amaliyotda ajratish darajasi qiymati manfiy chiqib qolishi ham mumkin. Bunday elementlarda jiddiy xato mavjud bo'lib, ular albatta o'rganilishi va xato tuzatilishi lozim, yoki element bazadan chiqarilishi kerak.

4-rasmda qiyinlik darajasi $b=0$ va ajratish darajasi $a=-1$ bo'lgan elementga javob funksiyasi ko'rsatilgan.



4-rasm. Qiyinlik darajasi $b=0$ va ajratish darajasi $a=-1$ bo'lgan elementga javob funksiyasi

Rasmdan quyi qobiliyatli talabgorlarning elementga to'g'ri javob berish ehtimoli yuqori qobiliyatli talabgorlarning bu elementga to'g'ri javob berish ehtimolidan katta ekanligi ko'rinadi. Bu esa element ko'zlangan maqsadda ishlamayotganligini anglatadi.

2. Natija va muhokamalar

9-sinf o'quvchilari o'rtasida biologiya fanidan tadqiqot uchun o'tkazilgan test sinovlarida foydalanilgan umumiy elementli 4 ta variant uchun elementlarning qiyinlik darajalari Rash va ikki parametrlil model bilan hisoblandi. Har bir variantlar uchun elementlarining ajratish darajasi (a) va klassik test nazariyasi asosida nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlari hisoblandi. Har bir elementga berilgan javoblarning umumiy test bali bilan korrelyatsiyasi (1, 2, 3, ... 48- elementlar va umumiy ball orasidagi

korrelyatsiya) nuqtaviy biserial korrelyatsiya koeffitsiyentiga teng bo'ladi.

Umuman olganda, nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymati 2-qiyinlik darajasidagi elementlar uchun 0,5 va undan katta, 1- va 3-qiyinlik darajasidagi elementlar uchun esa 0,25 va undan katta bo'lsa, valid hisoblanadi. Nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymati manfiy bo'lgan elementlar esa variantdan chiqariladi. Aks holda bilim darajasi past bo'lgan talabgorlar g'olib bo'lib, bilim darajasi yuqori bo'lgan

talabgorlar elementlarga javob berishda noto'g'ri javobni tanlaydilar yoki ularni o'tkazib yuboradilar.

1-jadvalda test sinovlarida foydalanilgan umumiy elementli 1-variant uchun elementlarning Rash va ikki parametrli modellar bilan hisoblangan qiyinlik darajalari, ularning farqi, ajratish darajasi va nuqtaviy biserial korrelyatsiya koeffitsiyentlari berilgan. Jadvaldan Rash va ikki parametrli modellar bilan hisoblangan qiyinlik darajalarining farqi unchalik katta emasligi ko'rinadi.

1-jadvaldan B0000037 ID raqamli elementning qiyinlik darajasi ikki parametrli modelda Rash modeli bilan hisoblangan qiyinlik darajasiga nisbatan kattaroq (qiyinlik darajalari farqi **0,769**), B0000138 – ID raqamli element uchun esa aksincha Rash modeli bilan hisoblangan qiyinlik darajasi ikki parametrli model bilan hisoblangan qiyinlik darajasiga nisbatan kattaroq (qiyinlik darajalari farqi **–0,578**) ekanligi ko'rinadi.

1-jadval

1-variantdagi elementlarning Rash va ikki parametrli modellar bilan hisoblangan qiyinlik darajalari, ularning farqi, ajratish darajasi va nuqtaviy biserial korrelyatsiya koeffitsiyentlari

Elementlar	Ajratish darajasi (a)	Nuqtaviy biserial korrelyatsiya koeffitsiyenti	Qiyinlik darajasi (2PL)	Qiyinlik darajasi (Rash)	Qiyinli darajalari farqi (2PL-Rash)
B0000037	0,410	0,228	1,740	0,971	0,769
B0000015	1,432	0,537	-1,233	-1,575	0,342
B0000001	1,548	0,554	-1,014	-1,351	0,337
B0000042	1,334	0,52	-1,210	-1,484	0,274
B0000002	1,154	0,364	-2,044	-2,297	0,253
B0000024	1,443	0,553	-0,730	-0,951	0,221
B0000040	0,665	0,31	1,812	1,599	0,214
B0000010	1,403	0,601	-0,648	-0,813	0,165
B0000016	0,658	0,346	0,764	0,602	0,162
B0000023	1,244	0,55	-0,727	-0,885	0,158
B0000033	1,289	0,551	-0,791	-0,947	0,155

B0000009	1,098	0,413	-1,008	-1,162	0,154
B0000043	0,766	0,401	1,234	1,112	0,122
B0000030	1,227	0,547	-0,690	-0,798	0,108
B0000014	1,322	0,598	-0,484	-0,584	0,100
B0000006	1,320	0,596	-0,485	-0,584	0,099
B0000035	1,256	0,567	-0,524	-0,612	0,088
B0000034	1,242	0,566	-0,276	-0,353	0,077
B0000003	1,162	0,527	-0,627	-0,697	0,070
B0000039	1,188	0,545	-0,265	-0,333	0,068
B0000020	1,360	0,591	-0,169	-0,234	0,065
B0000018	1,202	0,557	-0,444	-0,500	0,056
B0000012	1,366	0,672	-0,104	-0,152	0,049
B0000027	1,279	0,585	-0,300	-0,349	0,048
B0000013	1,196	0,557	-0,399	-0,445	0,046
B0000045	0,959	0,446	-0,188	-0,234	0,046
B0000019	0,895	0,429	-0,294	-0,337	0,043
B0000007	0,885	0,464	-0,215	-0,257	0,042
B0000011	0,908	0,463	0,439	0,400	0,040
B0000032	1,125	0,528	-0,479	-0,514	0,034
B0000004	0,928	0,465	-0,239	-0,263	0,023
B0000029	1,081	0,505	-0,508	-0,528	0,020
B0000025	1,124	0,536	-0,274	-0,281	0,007
B0000041	1,166	0,56	-0,183	-0,186	0,003
B0000026	1,475	0,672	0,134	0,174	-0,040
B0000031	0,783	0,393	0,534	0,575	-0,041
B0000008	0,933	0,459	-0,251	-0,200	-0,051
B0000017	1,223	0,595	0,185	0,241	-0,056
B0000038	0,950	0,48	0,320	0,387	-0,067
B0000021	1,387	0,647	0,280	0,360	-0,080
B0000005	1,187	0,582	0,411	0,507	-0,096

B0000044	0,725	0,373	-0,281	-0,146	-0,135
B0000036	1,367	0,637	0,538	0,683	-0,145
B0000028	1,398	0,58	0,967	1,181	-0,215
B0000022	1,228	0,508	1,811	2,208	-0,397
B0000136	1,494	0,611	1,296	1,696	-0,400
B0000137	1,492	0,542	1,633	2,079	-0,446
B0000138	1,288	0,432	2,338	2,916	-0,578

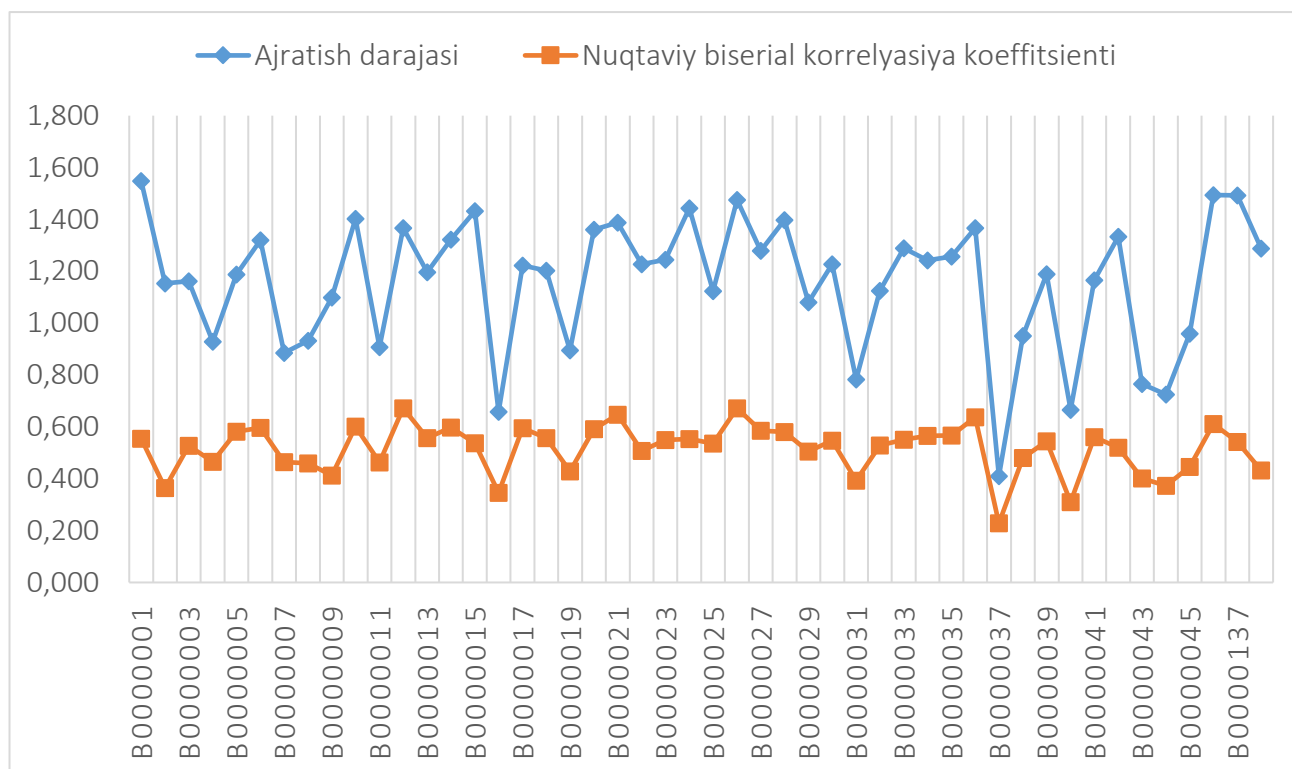
B0000037 ID raqamli elementning ajratish darajasi (0,410) va nuqtaviy biserial korrelyatsiya koeffitsiyenti (0,228) eng kichik, bu esa ushbu elementning ajratish darajasi yaxshi emasligini bildiradi. Elementlarning ajratish darajasi va nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlarining mos holda o'zgarayotganligini 1-jadvaldan ko'rish mumkin.

Bu holatni 1-variantdagi elementlarning ajratish darajasi va nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlari o'rtasidagi bog'liqlikni 5-rasmdan ham ko'rish mumkin.

Demak, nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlari kichik bo'lgan elementlarning ajratish

darajasi qiymati ham kichik bo'ladi. Bundan bu elementlar qobiliyat darajalarini yaxshi ajratib bermaydi degan xulosa qilish mumkin.

2-jadvalda test sinovlarida foydalanilgan umumiy elementli 2-variant uchun elementlarning Rash va ikki parametrlil model bilan hisoblangan qiyinlik darajalari, ularning farqi, ajratish darajalari va nuqtaviy biserial korrelyatsiya koeffitsiyentlari berilgan. Jadvaldan Rash va ikki parametrlil model bilan hisoblangan qiyinlik darajalarining farqi unchalik katta emasligini, faqat B0000061, B0000090, B0000089 elementlarning qiyinlik darajalari farqi katta ekanligini ko'rish mumkin.



5-rasm. 1- variantdagi elementlarning ajratish darajasi (a) va nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlari o'rtasidagi bog'liqlik

2-jadval

2-variantdagi elementlarning Rash va ikki parametrlil model bilan hisoblangan qiyinlik darajalari, ularning farqi, ajratish darajasi va nuqtaviy biserial korrelyatsiya koeffitsiyentlari

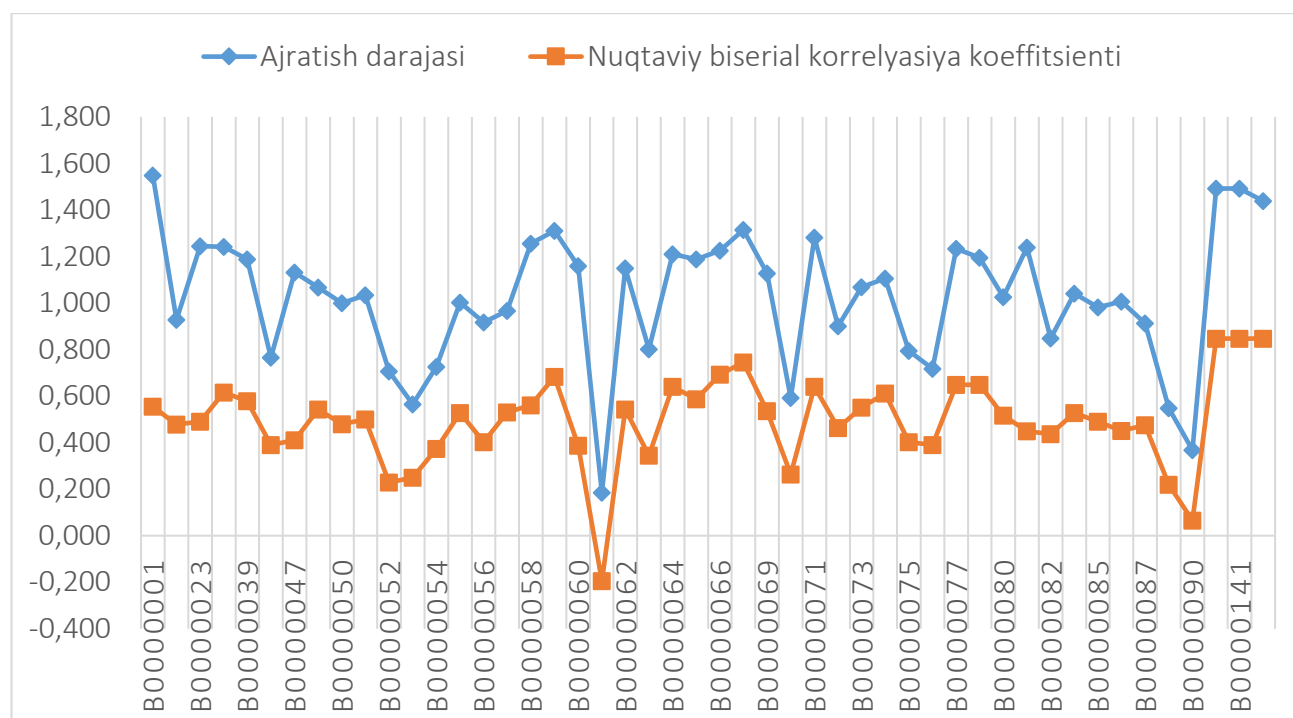
Elementlar	Ajratisht darajasi (a)	Nuqtaviy biserial korrelyatsiya koeffitsiyenti	Qiyinlik darajasi (2PL)	Qiyinlik darajasi (Rash)	Qiyinli darajalari farqi (2PL-Rash)
B0000061	0,184	-0,195	8,472	1,704	6,768
B0000090	0,367	0,065	2,067	0,591	1,476
B0000089	0,548	0,22	1,835	0,885	0,950
B0000053	0,565	0,25	1,240	0,498	0,741
B0000054	0,726	0,373	1,203	0,686	0,517
B0000076	0,717	0,39	0,466	-0,004	0,470
B0000075	0,794	0,403	0,651	0,276	0,375
B0000082	0,849	0,437	0,899	0,544	0,354

B0000060	1,159	0,387	-1,899	-2,237	0,338
B0000001	1,548	0,555	-1,014	-1,351	0,337
B0000063	0,801	0,344	-0,188	-0,509	0,322
B0000070	0,592	0,264	-0,088	-0,403	0,315
B0000047	1,131	0,41	-1,104	-1,400	0,296
B0000087	0,913	0,475	0,517	0,233	0,283
B0000085	0,981	0,491	-0,057	-0,335	0,278
B0000050	1,000	0,48	-0,345	-0,611	0,267
B0000072	0,900	0,463	0,092	-0,173	0,265
B0000051	1,034	0,5	-0,401	-0,661	0,260
B0000058	1,256	0,56	-0,662	-0,921	0,259
B0000083	1,041	0,527	0,075	-0,183	0,258
B0000057	0,967	0,53	0,800	0,544	0,256
B0000048	1,067	0,543	-0,297	-0,548	0,251
B0000062	1,149	0,543	-0,650	-0,881	0,231
B0000055	1,002	0,527	-0,022	-0,251	0,229
B0000069	1,128	0,536	-0,617	-0,845	0,228
B0000080	1,025	0,516	-0,180	-0,403	0,223
B0000065	1,187	0,587	-0,447	-0,662	0,215
B0000086	1,006	0,45	-0,850	-1,064	0,215
B0000056	0,917	0,403	-0,876	-1,064	0,188
B0000064	1,210	0,64	-0,110	-0,289	0,179
B0000059	1,310	0,683	-0,230	-0,403	0,172
B0000077	1,233	0,649	-0,045	-0,212	0,167
B0000023	1,244	0,491	-0,727	-0,885	0,158
B0000078	1,194	0,648	0,183	0,026	0,157
B0000074	1,105	0,611	0,989	0,834	0,155
B0000073	1,069	0,551	2,071	1,916	0,155
B0000043	0,766	0,39	1,234	1,112	0,122
B0000071	1,281	0,64	0,536	0,425	0,110

B0000066	1,225	0,693	0,781	0,686	0,094
B0000034	1,242	0,615	-0,276	-0,353	0,077
B0000039	1,188	0,579	-0,265	-0,333	0,068
B0000067	1,314	0,745	0,784	0,735	0,049
B0000004	0,928	0,478	-0,239	-0,263	0,023
B0000140	1,492	0,847	1,042	1,097	-0,055
B0000141	1,492	0,847	1,042	1,097	-0,055
B0000142	1,438	0,847	1,596	1,680	-0,084
B0000052	0,707	0,229	-1,710	-1,600	-0,110
B0000081	1,238	0,449	2,605	2,728	-0,124

Bunda ham 1-variantdagi kabi elementlarning ajratish darajasi va nuqtaviy biserial korrelyatsiya koefitsiyenti qiymatlarining mos holda o'zgarayotganligini 2-jadvaldan hamda 6-rasmdan ko'rish mumkin. 2-jadvalda keltirilgan nuqtaviy biserial korrelyatsiya koefitsiyenti qiymatlari B0000061, B0000090, B0000089 ID-raqamli elementlar uchun eng kichik (jadvalda belgilab ko'rsatilgan) qiymatga ega. O'z navbatida ushbu elementlarning ajratish darajasi qiymati ham eng kichik ekanligi aniqlandi. Bu esa ushbu elementlarning qobiliyat darajalarini yaxshi ajratib bermasligini anglatadi.

3-jadvalda test sinovlarida foydalanilgan umumiy elementli 3-variant uchun elementlarning Rash va ikki parametrli modellar bilan hisoblangan qiyinlik darajalari, ularning farqi, ajratish darajasi, nuqtaviy biserial korrelyatsiya koefitsiyentlari berilgan. Jadvaldan Rash va ikki parametrli modellar bilan hisoblangan elementlar qiyinlik darajalarining farqi unchalik katta emasligini, faqat B0000037, B0000109 ID-raqamli elementlar qiyinlik darajalarining farqi katta ekanligini ko'rish mumkin.



6-rasm. 2- variantdagi elementlarning ajratish darajasi (α) va nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlari o'rtasidagi bog'liqlik

3-jadval

3-variantdagi elementlarning Rash va ikki parametrlil model bilan hisoblangan qiyinlik darajalari, ularning farqi, ajratish darajasi va nuqtaviy biserial korrelyatsiya koeffitsiyentlari

Elementlar	Ajratish darajasi (α)	Nuqtaviy biserial korrelyatsiya koeffitsiyenti	Qiyinlik darajasi (2PL)	Qiyinlik darajasi (Rash)	Qiyinli darajalari farqi (2PL-Rash)
B0000037	0,410	0,184	1,740	0,971	0,769
B0000109	0,641	0,27	0,888	0,191	0,697
B0000097	0,750	0,357	1,180	0,605	0,575
B0000128	0,740	0,339	0,959	0,391	0,567
B0000146	0,806	0,347	2,134	1,604	0,530
B0000115	0,776	0,369	0,688	0,191	0,497
B0000130	0,789	0,406	0,569	0,096	0,473
B0000116	0,850	0,452	0,507	0,096	0,411
B0000099	0,682	0,31	-0,333	-0,736	0,404

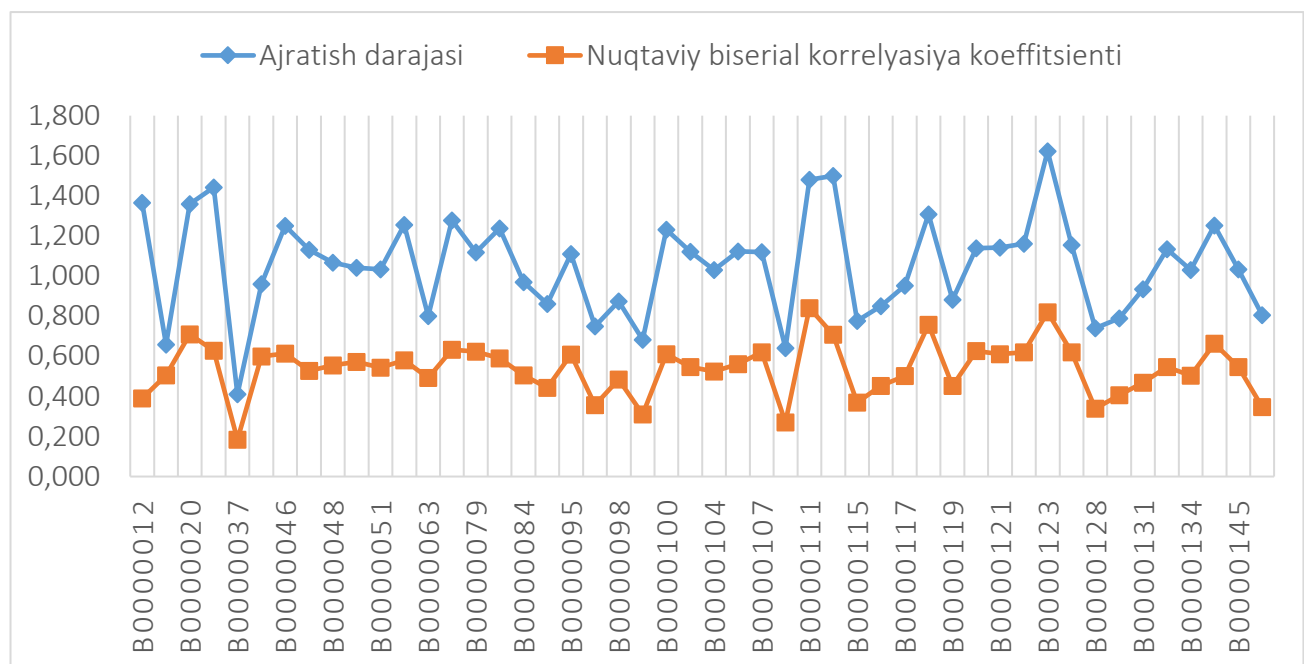
B0000088	0,861	0,443	0,448	0,050	0,398
B0000119	0,883	0,452	-0,021	-0,383	0,362
B0000098	0,874	0,485	-0,203	-0,553	0,351
B0000117	0,953	0,502	0,110	-0,216	0,326
B0000131	0,935	0,468	-0,489	-0,812	0,323
B0000063	0,801	0,493	-0,188	-0,509	0,322
B0000084	0,970	0,505	-0,304	-0,621	0,317
B0000105	1,123	0,561	-0,705	-1,008	0,302
B0000104	1,030	0,524	-0,435	-0,736	0,301
B0000134	1,030	0,503	-0,327	-0,628	0,300
B0000047	1,131	0,528	-1,104	-1,400	0,296
B0000132	1,134	0,547	-0,665	-0,960	0,295
B0000101	1,122	0,547	-0,629	-0,923	0,294
B0000046	1,251	0,614	-0,614	-0,886	0,273
B0000100	1,231	0,611	-0,502	-0,774	0,272
B0000049	1,041	0,572	0,227	-0,041	0,268
B0000051	1,034	0,543	-0,401	-0,661	0,260
B0000058	1,256	0,58	-0,662	-0,921	0,259
B0000068	1,277	0,633	-0,553	-0,812	0,258
B0000048	1,067	0,555	-0,297	-0,548	0,251
B0000122	1,162	0,62	-0,146	-0,383	0,237
B0000121	1,142	0,61	-0,025	-0,259	0,233
B0000095	1,110	0,609	0,188	-0,041	0,230
B0000125	1,154	0,62	-0,031	-0,259	0,228
B0000107	1,121	0,62	0,183	-0,041	0,224
B0000024	1,443	0,628	-0,730	-0,951	0,221
B0000120	1,139	0,627	0,133	-0,086	0,218
B0000145	1,033	0,547	1,227	1,025	0,202
B0000079	1,119	0,623	0,531	0,340	0,191
B0000016	0,658	0,505	0,764	0,602	0,162

B0000020	1,360	0,71	-0,169	-0,234	0,065
B0000012	1,366	0,39	-0,104	-0,152	0,049
B0000045	0,959	0,6	-0,188	-0,234	0,046
B0000111	1,481	0,841	0,349	0,340	0,010
B0000118	1,308	0,757	1,073	1,091	-0,017
B0000123	1,623	0,819	0,552	0,624	-0,071
B0000144	1,252	0,663	1,699	1,772	-0,074
B0000081	1,238	0,59	2,605	2,728	-0,124
B0000112	1,501	0,708	1,507	1,668	-0,161

Yuqorida aytilganidek, bu test variantlari umumiy elementli test variantlari (o'zaro kesishgan elementlar mavjud) bo'lganligi uchun B0000037 ID-raqamli test elementning nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatining 1-variantda ham va 3-variantda ham kichik ekanligini ko'rish mumkin. Shuningdek, bu elementning ajratish darajasi qiymati ham kichik va uni qayta ko'rib chiqish yoki bazadan chiqarish tavsiya etiladi. 7-rasmda elementlarning ajratish darajasi va nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlarining bog'liqligi ko'rsatilgan. Bu yerda ham bog'liqlik 1- va 2-variantlardagidek o'zgarmayotganligini ko'rish mumkin.

4-jadvalda esa test sinovlarida foydalanilgan umumiy elementli 4-variant uchun elementlarning Rash va ikki parametrlil modellar bilan hisoblangan qiyinlik darajalari, ular-

ning farqi, elementlarning ajratish darajasi va nuqtaviy biserial korrelyatsiya koeffitsiyentlari berilgan. Jadvaldan 4-variant uchun ham Rash va ikki parametrlil modellar bilan hisoblangan elementlarning qiyinlik darajalari farqi unchalik katta emasligini, faqat B0000108, B0000182, B0000094 ID-raqamli elementlarning qiyinlik darajalari farqi katta ekanligini, shuningdek bu elementlarning ham nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlari kichik ekanligini ko'rish mumkin. Bunga mos holda ushbu elementlarning qobiliyatlarni ajratish darajasi qiymati ham kichik ekanligi aniqlandi. B0000172 va B0000176 ID-raqamli elementlarning nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlari manfiy bo'lib, mazkur elementlarning qobiliyatlarni ajratish darajasi qiymati juda kichikligi kuzatildi.



7-rasm. 3- variantdagi elementlarning ajratish darajasi (α) va nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlari o'rtasidagi bog'liqlik

4-variant uchun elementlarning ajratish darajasi va nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlarining bog'liqligi 8-rasmda keltirilgan.

Bu rasmdan ham elementlarning qobiliyatlarni ajratish darajasi va nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlarining bir-biriga mos holda o'zgarayotganligi ko'rinadi.

Ushbu maqolada biologiya fanidan tadqiqot maqsadida o'tkazilgan test sinovlarida foydalanilgan umumiy elementli 4 ta variantlar uchun Rash va ikki parametrli modellarning qiyosiy tahlili amalga oshirildi.

Tahlillardan Rash modeli bilan hisoblangan qiyinlik darajalarining ikki

parametrli model bilan hisoblangan qiyinlik darajasiga nisbatan biroz farq qilishi aniqlandi. Bunga sabab ikki parametrli modelda qiyinlik darajasi parametridan tashqari ajratish darajasi parametri bilan ham qo'shimcha moslashtirish amalga oshirilganligidir.

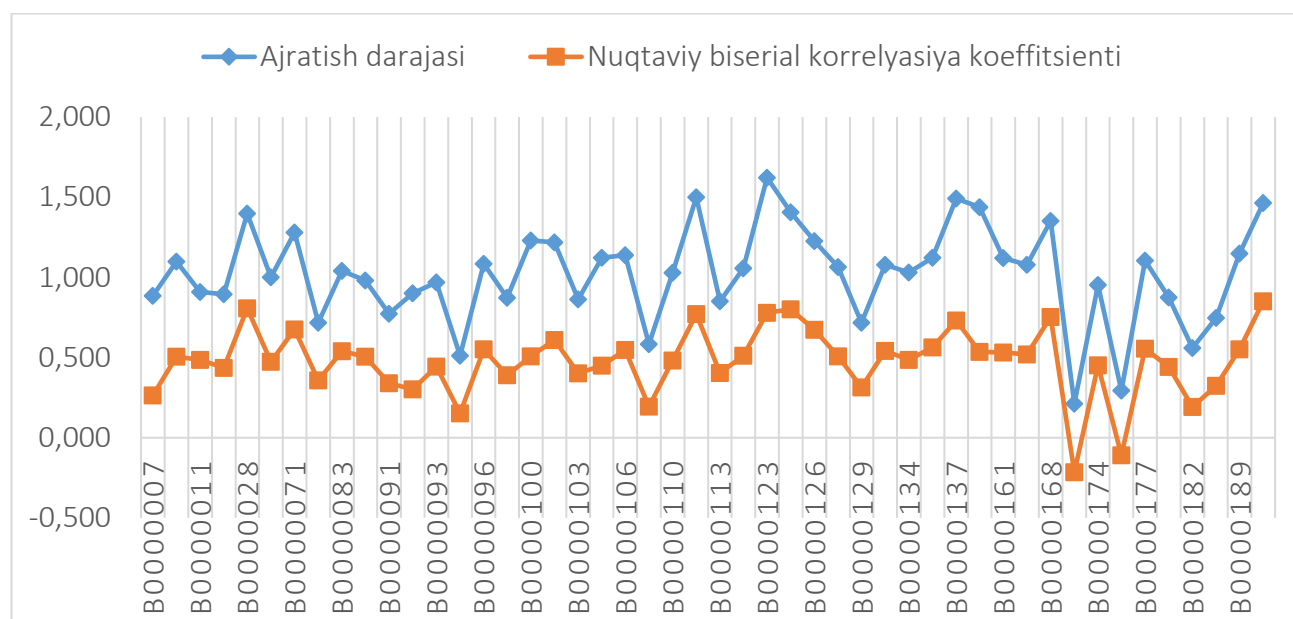
Elementlarning klassik test nazariyasi bilan hisoblangan nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlari va ikki parametrli model bilan hisoblangan ajratish darajasi qiymatlari solishtirildi. Buning natijasida elementlarning ajratish qobiliyati haqida ikkita kattalik ham bir xil xulosalar berishi aniqlandi.

4-jadval

4-variantdagi elementlarning Rash va ikki parametrlil model bilan hisoblangan qiyinlik darajalari, ularning farqi, ajratish darajasi va nuqtaviy biserial korrelyatsiya koeffitsiyentlari

Elementlar	Ajratish darajasi (a)	Nuqtaviy biserial korrelyatsiya koeffitsiyenti	Qiyinlik darajasi (2PL)	Qiyinlik darajasi (Rash)	Qiyinli darajalari farqi (2PL-Rash)
B0000172	0,212	-0,216	7,342	1,353	5,990
B0000176	0,293	-0,109	4,601	1,090	3,510
B0000108	0,583	0,194	1,906	0,800	1,106
B0000182	0,559	0,192	1,391	0,397	0,994
B0000094	0,512	0,152	1,099	0,131	0,968
B0000129	0,717	0,314	0,314	-0,192	0,506
B0000076	0,717	0,358	0,466	-0,004	0,470
B0000113	0,851	0,403	0,770	0,305	0,465
B0000179	0,874	0,442	0,845	0,397	0,448
B0000103	0,862	0,402	0,336	-0,074	0,410
B0000188	0,747	0,324	-0,086	-0,490	0,404
B0000091	0,773	0,339	-0,274	-0,633	0,359
B0000098	0,874	0,39	-0,203	-0,553	0,351
B0000174	0,953	0,451	0,070	-0,268	0,338
B0000093	0,968	0,443	-0,430	-0,739	0,309
B0000110	1,028	0,482	-0,259	-0,562	0,302
B0000105	1,123	0,45	-0,705	-1,008	0,302
B0000134	1,030	0,485	-0,327	-0,628	0,300
B0000114	1,057	0,511	-0,195	-0,490	0,295
B0000127	1,065	0,508	-0,197	-0,490	0,293
B0000164	1,079	0,519	-0,166	-0,454	0,288
B0000161	1,121	0,532	-0,348	-0,633	0,285
B0000133	1,079	0,541	0,089	-0,192	0,281

B0000085	0,981	0,506	-0,057	-0,335	0,278
B0000189	1,149	0,552	-0,287	-0,562	0,275
B0000106	1,138	0,547	-0,217	-0,490	0,273
B0000096	1,086	0,551	0,238	-0,034	0,272
B0000100	1,231	0,507	-0,502	-0,774	0,272
B0000050	1,000	0,474	-0,345	-0,611	0,267
B0000177	1,104	0,556	0,152	-0,114	0,266
B0000135	1,123	0,564	-0,003	-0,268	0,266
B0000083	1,041	0,539	0,075	-0,183	0,258
B0000102	1,219	0,609	-0,002	-0,230	0,228
B0000092	0,902	0,303	-1,541	-1,740	0,199
B0000009	1,098	0,506	-1,008	-1,162	0,154
B0000071	1,281	0,675	0,536	0,425	0,110
B0000126	1,226	0,673	1,510	1,423	0,087
B0000019	0,895	0,435	-0,294	-0,337	0,043
B0000007	0,885	0,265	-0,215	-0,257	0,042
B0000011	0,908	0,486	0,439	0,400	0,040
B0000124	1,407	0,801	1,316	1,353	-0,036
B0000168	1,353	0,753	1,777	1,817	-0,040
B0000123	1,623	0,779	0,552	0,624	-0,071
B0000193	1,463	0,851	1,415	1,496	-0,082
B0000142	1,438	0,535	1,596	1,680	-0,084
B0000112	1,501	0,771	1,507	1,668	-0,161
B0000028	1,398	0,808	0,967	1,181	-0,215
B0000137	1,492	0,731	1,633	2,079	-0,446



8-rasm. 4- variantdagi elementlarning ajratish darajasi (α) va nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlari o'rtasidagi bog'liqlik

Xulosa

Ushbu tadqiqotda 9-sinf o'quvchilari uchun biologiya fanidan o'tkazilgan test sinovlarida foydalanilgan umumiy elementli 4 ta variantlar uchun elementlarning xususiyatlari Rash va ikki parametrli modellar bilan tahlil qilindi.

Ikki parametrli modelda qiyinlik darajasi qiymatlari juda katta bo'lgan elementlarning ajratish darajasi qiymatlari va klassik test nazariyasi bilan hisoblangan nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlari juda kichik ekanligi aniqlandi.

Ajratish darajasi va nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlarining bir-biriga mos holda o'zgarayotganligi elementlarning ajratish darajasi haqida bu ikkita kattalik

ham bir xil ma'lumot berishini ko'rsatdi.

Elementlarining nuqtaviy biserial korrelyatsiya koeffitsiyenti qiymatlari kichik bo'lsa, bunday elementlarning qobiliyatlarni ajratish darajasi qiymatlari ham kichik bo'lishi ko'rsatib berildi.

Ikki parametrli modelda elementlarning qiyinlik darajasi parametridan tashqari ajratish parametri bilan ham ma'lumotlar moslashtirilganligi sababli, model bilan hisoblangan element xarakteristikasi chiziqlari bilan tajribada olingan ma'lumotlar mosligi yaxshi bo'ladi, lekin parametrlar Rash modelidagi parametr kabi barqaror bo'lmasligi mumkin.

Umuman olganda, elementlar sifatini baholashda Rash va ikki parametrli modellar o'zaro to'ldiruvchi yondashuvlar bo'lib, har biri turli

holatlarda foydali bo'ladi. Baholash tizimini takomillashtirishda har ikki model asosidagi natijalardan oqilona foydalanish maqsadga muvofiqdir.

ADABIYOTLAR

1. Rasch G. Probabilistic models for some intelligence and attainment tests, Copenhagen, Danish Institute for Educational research, 1960.
2. R. K. Hambleton and W. J. Van der Linden. Advances in item response theory and applications: An introduction, Advanced Psychological Measurement, 6, 373-378, 1982.
3. Hambleton, R.K., Swaminathan, H.,& Rogers, H.J. Fundamentals of item response theory. Newbury Park, CA: Sage, 1991.
4. Baker, Frank. The Basics of Item Response Theory, ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD, 2001.
5. Ivailo Partchev. A visual guide to item response theory, FriedrichSchiller-Universitat Jena, 2004.
6. 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, Axborotnoma ilmiy-uslubiy jurnali, 1-son, 63-78, 2023.
7. A.B. Normurodov, I.A. Boyxonov. Biologiya fanidan test topshiriqlarining Rayt xaritasi, ichki va tashqi moslik statistikalari, Axborotnoma ilmiy-uslubiy jurnali, 2-son, 56-83, 2023.
8. A.B. Normurodov, M.Dj. Ermamatov, I.A. Boyxonov. Rash modeli bilan moslik: Biologiya fanidan o'tkazilgan test sinovi natijalari, Axborotnoma ilmiy-uslubiy jurnali, 3-son, 4-21, 2023.
9. I.A. Boyxonov, M.Dj. Ermamatov, A.B. Normurodov. "Biologiya fanidan tadqiqot uchun foydalanilgan test topshiriqlari mazmunini statistik tahlil asosida o'rganish", Axborotnoma ilmiy-uslubiy jurnali, 3-son, 4-17, 2024.
10. Lord, F. M. A theory of test scores, Psychometric Monograph. Psychometric Society, No.7, 1952.
11. Birnbaum, A. (1968), Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord M. R. Novick, Statistical theories of mental test scores (chapter 17-20), Reading, M. A.: Addison-Wisley.
12. Gunter Maris, Timo Bechger, Jesse Koops and Ivailo Parchev, Data Management and Analysis of Tests, p. 1-49, 2022.

ITEM CHARACTERISTICS IN CLASSICAL AND ITEM TEST THEORIES

M.J. Ermamatov, A.B. Normurodov, A.R. Sattiyev

Scientific and Educational Practical Center Under the Agency for Assessment of Knowledge and Competences, a.normurodov@uzbmb.uz

Abstract. This paper presents a study on the results of a biology test conducted for 9th-grade students. Within the scope of the research, the difficulty levels of the test items were assessed using modern psychometric approaches — the Rasch model and the two-parameter logistic (2PL) model. The article analyzes the difficulty levels calculated by these two models, the differences between them, the discrimination values of the items, their diagnostic capabilities, and practical applications. Additionally, point-biserial correlation coefficients of the items were determined using classical test theory. These classical indicators were compared with the discrimination parameters obtained from the two-parameter model, and their interrelationships was examined. The fact that the values of the discrimination index and the point-biserial correlation coefficient change in accordance with each other indicates that both of these measures provide the same information about the discrimination level of the items. The research results demonstrate the effectiveness of the Rasch and two-parameter models in determining item quality, as well as revealing the advantages and limitations of each model. The findings of the paper can be useful for developing test items and improving the assessment system.

Keywords: point biserial correlation coefficient, difficulty level, Rasch model, 2PL model.

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'rinishi 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 parametri 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 foydalalanamiz [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 samaradorligi 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 funksionallikka 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.

INTRA-TEXT GAP-FILLING IN MULTILEVEL ENGLISH LANGUAGE TESTS: DESIGN, VALIDITY, AND CEFR ALIGNMENT

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. Gap-filling items are among the most common types of questions in language proficiency tests. They are particularly suitable for multilevel formats because they can target a broad range of difficulty levels and reflect authentic language processing. Intra-text gap-filling items, in which the input text serves as the word bank at the same time, are a type of test item that has been developed relatively recently. The article discusses the principles of designing intra-text gap-filling tasks, examines their validity, and evaluates how they can be aligned with the Common European Framework of Reference for Languages (CEFR).

Keywords: multilevel english language tests, reading comprehension, testlet, gap-filling items, intra-text gap-filling items, validity, reliability, CEFR.

Introduction

The multilevel English language proficiency test has been in use in Uzbekistan since 2022. A multilevel test is designed to assess language proficiency across a range of levels, from beginner to advanced, within a single exam. Developed as an alternative to one-level proficiency exams that were in use before 2022, this test gained popularity among test takers and other stakeholders due to its flexibility and inclusivity for assessing language proficiency across a broad spectrum of abilities. The test consists of four components, corresponding to each language skill. The listening and reading components are further subdivided into smaller

sections, called testlets, each targeting a specific CEFR level. Several different types of questions are used to minimize construct-irrelevant variance – factors unrelated to language proficiency (e.g., test-taking strategy dependence), allowing test-takers to demonstrate their true language abilities. Alderson (2000) argues that high-quality reading tests use a variety of techniques, as this enhances the authenticity of the assessment [1]. Among the various question types employed, one notable example is the use of gap-filling questions, which are included in two of the testlets in the reading section.

1. Types of gap-filling items

Gap-filling tasks are considered to be particularly useful in testing reading [2]. Alderson (2000) states that they are often confused with cloze items [1]. The primary difference between cloze items and gap-filling tasks lies in the deletion process. In cloze items, as originally invented by Taylor (1953), words are removed randomly in the text, i.e., every n-th word is deleted [3]. In contrast, gap-filling tasks involve the removal of words based on a particular rationale, such as targeting content words that are essential to the main idea of the text or the words that carry the coherence of the text [1]. There are a few variations of gap-filling tasks:

- Open gap-filling tasks;
- Banked gap-filling tasks;
- Multiple-choice gap-filling tasks.

In open gap-filling tasks, the test taker is required to deduce the missing words, while in banked and multiple-choice gap-filling tasks, the words are provided before or after the text. A variation of banked gap-filling tasks is used in the Multilevel English language proficiency test. In this version, content words that are essential for maintaining the coherence of the text are deleted, and the remaining words of the text serve as a word bank from which test takers can choose the appropriate words to fill the gaps.

Read the text. Fill in each gap with one word. You must use a word which is somewhere in the rest of the text. Use each word only once. The first gap has been filled in as an example.

My favourite parks

There are no restaurants or cool stores (0) **near** my home. So if there were no parks in my neighborhood I'd (1) like I had nothing to do. But I'm lucky. There are two (2) near my house, and they both feel safe.

El Cariso Park is big with lots of trees, and it has a pool where my sister loves swimming. Every time I've been there, there have been quite a few (3) running or walking and some parents playing with their kids. I often invite friends to go for a run or have picnics with me.

Sometimes I go to Veterans Memorial Park, which is smaller, to read a book or hang out with (4) . There aren't a lot of people around, but it's (5) . I've never felt frightened there.

Fig.1. Intra-text gap-filling task

2. Intra-text gap-filling items

Fig. 1 [4] shows an example of an intra-text banked gap-filling task. As can be seen from the example, gaps are to be filled with the words, which are somewhere in the rest of the text. They can be useful to test the overall understanding of the text as the content words are deleted, as well as to

test the coherence of the text and building a mental model of the text because the deleted words are important to the internal links between the sentences in the text.

Here is another example, from the Multilevel English language tests:

Read the text. Fill in each gap with ONE word. You must use a word which is somewhere in the rest of the text.

Autumn

Autumn is a magical season of change. The leaves on the trees turn beautiful colours of red, orange, and yellow, and fall to the ground. Why do the Q1 _____ change colour? They are green because of a chemical called chlorophyll. This Q2 _____ helps plants to make food from sunlight. In the fall, the days get shorter and the trees don't get as much Q3 _____. Without sunlight, the Q4 _____ stop making chlorophyll, and the leaves change colour.

Why do some animals migrate in the autumn? This is because food becomes scarce in the winter. In warmer climates, there is more Q5 _____ all year round. Some animals, such as bears and bats, sleep in the winter. Sleeping helps these Q6 _____ to save energy during the winter when food is scarce.

(Answer key: 1-leaves, 2-chemical, 3-sunlight, 4-trees, 5-food, 6-animals)

Fig.2. Sample testlet from Multilevel English language proficiency test bank

In this example, the deleted words carry the coherence of the text. It is difficult to guess all the correct words without understanding the text. When test takers read the text, they deduce

the word based on their understanding of the text and further verify it by linking the sentence with the preceding or the following sentence.

3. Item development process

This type of items in the Multilevel English language proficiency test targets lower B1 level. Therefore, texts are selected to suit this level of CEFR in terms of topic, length, lexical and grammatical complexity. After the text is chosen, it needs to be further edited to fit the level and specification requirements. The text should allow the deletion of six words as the testlet contain six items. A number of principles must be followed to maintain the quality of assessment:

- A few sentences at the beginning of the text must remain intact to provide enough context to understanding [1];
- The gaps must be spaced adequately, and the surrounding text

should provide sufficient context and clues for the answers;

- Content words that are essential for maintaining the coherence of the text should be chosen for deletion;
- The selected words must appear elsewhere in the text, ensuring logical flow between sentences;
- Only content words should be deleted;
- Each word must be deleted only once;
- The answer key must include all valid answers, but synonyms or variations in phrasing are not acceptable.

4. CEFR alignment

Input texts as well as tasks in the Multilevel English language proficiency tests are aligned with CEFR according to the following parameters:

- Relevance of topic;
- Nature of information;
- Lexical and grammatical complexity;
- Skill focus.

A Core Inventory for General English [5], developed by British Council, provides core lists of discrete language points, topics and key lexis for each level of the CEFR. This document guides the development of test specifications, according to which the intra-text gap-filling task is aligned

with the B1 level. During the text selection and editing process, item writers make sure that the text is within the capability of an average B1 learner in terms of topical, lexical and grammatical complexity.

According to the CEFR [6, 7], a B1-level reader can:

- understand the main points of clear standard texts on familiar matters regularly encountered in work, school, leisure, etc.;
- understand texts that consist mainly of high frequency everyday or job-related language;
- understand clearly written, straightforward texts on subjects

related to their field and of personal interest.

This means the text should not be overwhelmingly long (around 150 words, according to the Multilevel English language proficiency test specifications [8]) and should not discuss abstract issues. These

descriptors are embedded in the test specifications and monitored closely during the item design and review processes.

Table 1 demonstrates the CEFR alignment mapping in terms of task focus:

Table 1

CEFR alignment mapping of intra-text gap-filling tasks

Aspect	CEFR B1 Relevance	How gap-filling task aligns with CEFR
Reading for gist and detail	Learners must understand the main ideas and key details	Gaps placed at key points (e.g., content words) test understanding.
Contextual inference	Learners are expected to infer meaning from context	Choosing the correct word to fill a gap often requires using surrounding context.
Cohesion and coherence	Learners can follow the thread of a text and basic cohesion	Gaps involving reference words (e.g., “that,” “this”) test this skill.

In summary, intra-text gap-filling tasks at the B1 level strongly align with CEFR expectations if they:

- are based on texts accessible to B1 learners,
- focus on B1-level language forms and meanings,

- require comprehension of both sentence-level and discourse-level information,

- support integrative skills like vocabulary use, grammar, cohesion, and context interpretation.

5. Statistical analysis of intra-text gap-filling items

Table 2 illustrates the analysis of the sample intra-text gap-filling task (Fig. 2) used in the Multilevel English language proficiency tests. The items

are analyzed with classical test theory using SPSS, as well as the item-response theory. Each item in the testlet is given a special code, which

indicates the year (2023) and month (November) the item is administered,

as well as the session number and item's location in the test booklet.

Table 2

Item performance parameters

Item	Mean (difficulty)	Standard deviation	Correlation (discrimination)	Cronbach's alpha if deleted	Cronbach's alpha
23No101	0.691	0.462	0.406	0.808	0.816
23No102	0.776	0.417	0.334	0.811	
23No103	0.626	0.484	0.410	0.808	
23No104	0.799	0.401	0.314	0.811	
23No105	0.657	0.475	0.497	0.804	
23No106	0.861	0.346	0.413	0.809	

The table shows how well the testlet items are performing. The mean difficulty of each item represents its relative difficulty, with values closer to 1 indicating easier items and values closer to 0 indicating more difficult ones.

Correlation value shows whether the item is discriminating more knowledgeable test takers from less knowledgeable ones. The rule of thumb is that this value needs to be greater than 0.2 to consider it an acceptable item [9, 10, 11]. The analysis shows that all the items in the testlet are performing really well in terms of item discrimination. Higher-quality items contribute positively to internal consistency, whereas poor-quality items can decrease it. 'Bad items' are those that fail to discriminate well between higher and lower ability test-takers or fail to contribute to the overall reliability of the test.

Cronbach's alpha is calculated for the whole reading section administered in November 2023. Cronbach's alpha is a measure of the internal consistency of the test, or how reliably the items measure the same underlying construct. A value greater than 0.7 indicates high internal consistency [12]. The 'Cronbach's alpha if deleted' column shows how the internal consistency of the test changes when an item is removed. If the alpha value decreases, it means the item is contributing positively to the test's overall reliability. The table shows that all of the items in the testlet are contributing positively to the internal consistency of the test, which in turn supports the construct validity of the test [9, 10, 11].

While Cronbach's alpha provides insight into the overall internal consistency of the test, item-response theory (IRT) offers a more detailed

analysis of how well the individual items perform, particularly in terms of difficulty and fit. The calculations are done using dexter packet in R, a software environment for statistical computing and graphics [13]. The item

difficulty from the Rasch model (Table 3) is represented by 'beta' values: negative values indicate easier items, while positive values indicate more difficult ones:

Table 3

Item difficulty from the Rasch model

item_id	beta	SE_beta
23No101	-0.780592	0.023149
23No102	-1.224829	0.024546
23No103	-0.363351	0.022502
23No104	-1.375662	0.025204
23No105	-0.434099	0.022569
23No106	-1.833887	0.027857

Item-response theory (IRT) provides a more nuanced view of item performance. Fig. 4 shows that response patterns for items 5 and 6 do not align well with the model's predictions, indicating potential issues with these items' performance or fit within the test structure. Fig. 5 shows the relationship between item difficulty and person ability. IRT describes item difficulty and person ability in the same continuum to show

whether test takers were given the items within their capability or not. In the IRT model of the Multilevel English language proficiency tests, B1-level learners typically fall between -1.5 and 0. The graph in Fig. 5 shows that five out of six items fall within this range, indicating that the test items are appropriately calibrated for B1-level learners' abilities. This can also be considered as evidence of construct validity.

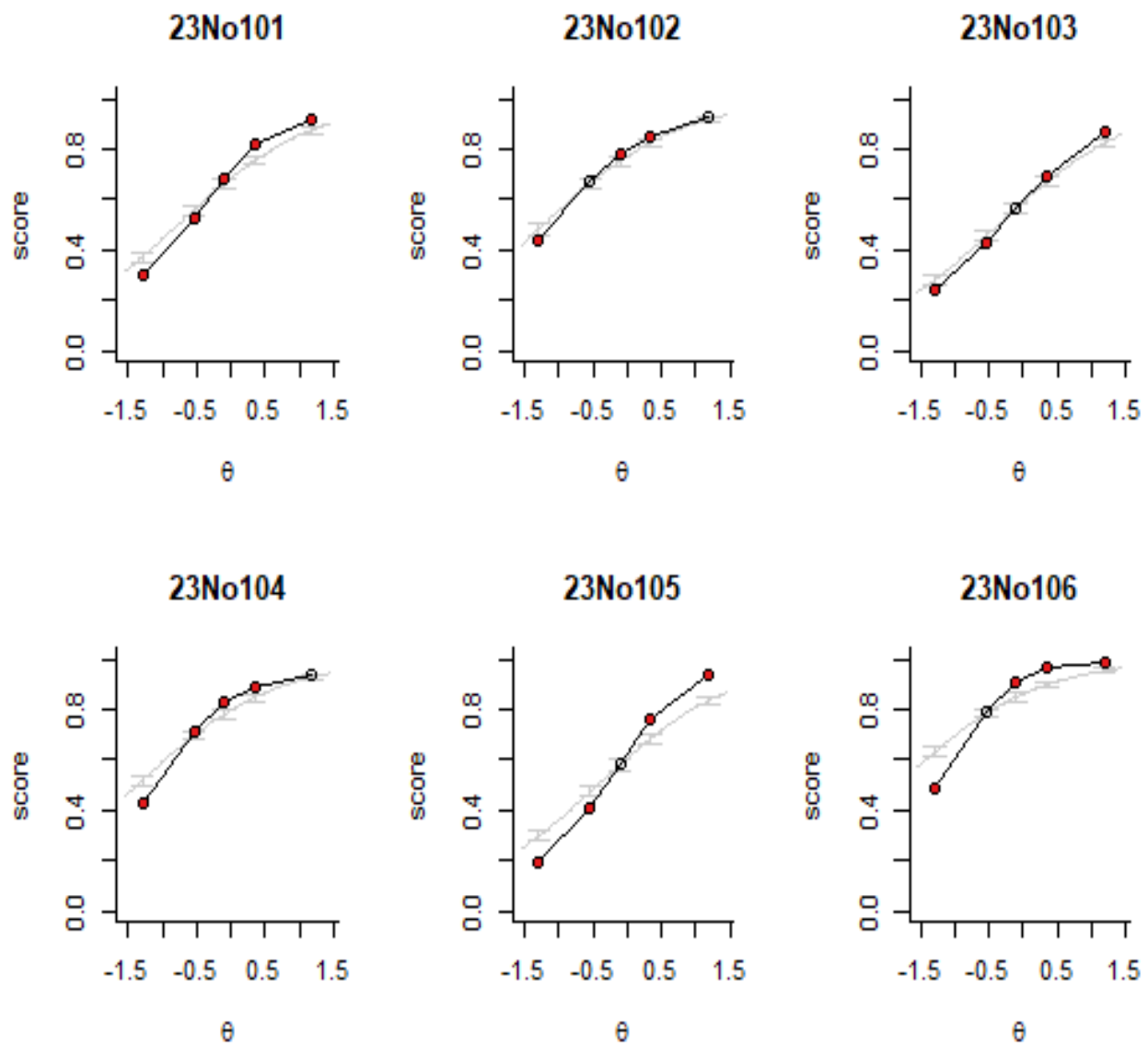


Fig. 4. Test of fit graphs

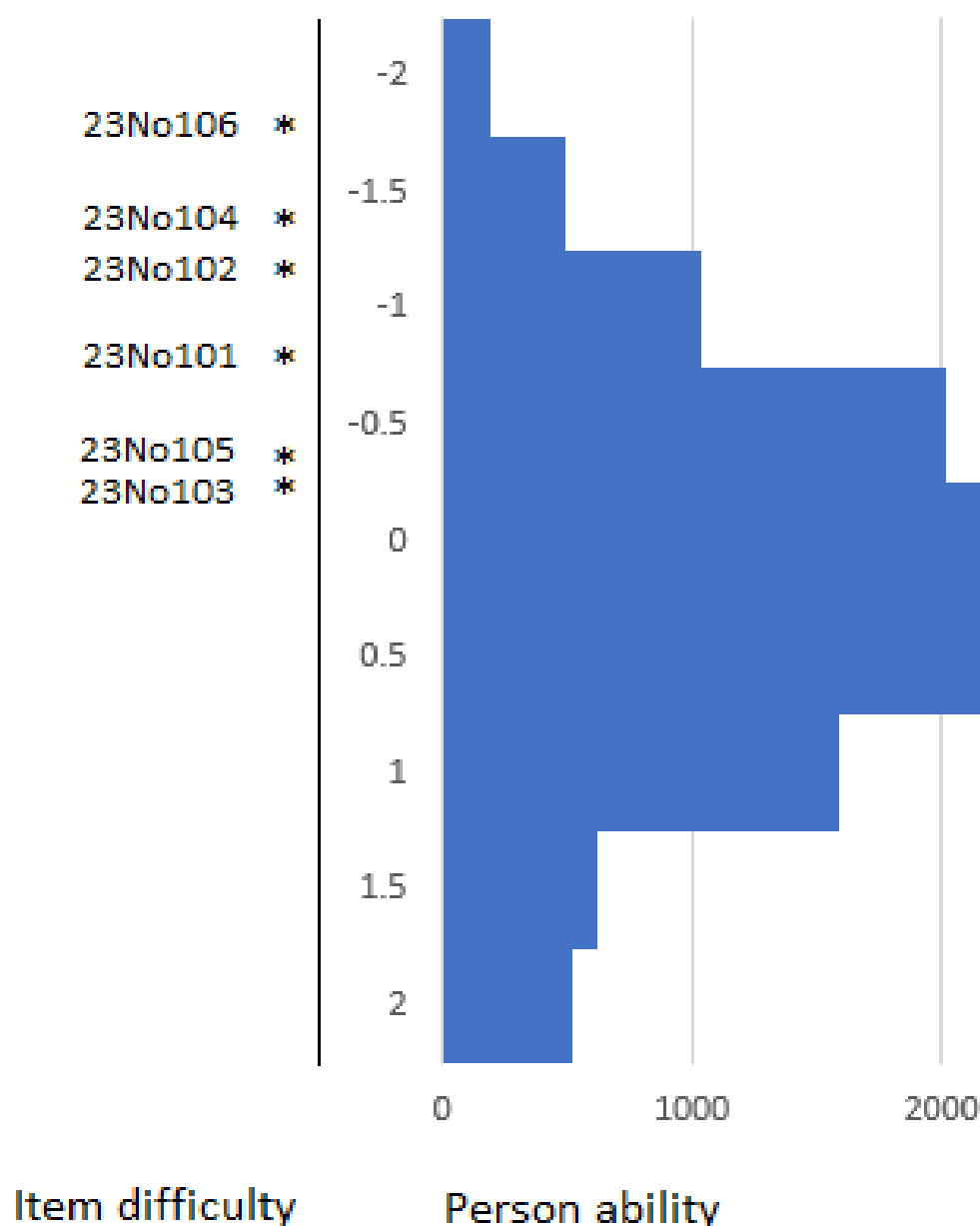


Fig. 5. Item difficulty and person ability map

Conclusion

The study shows the strong empirical evidence of intra-text gap-filling items to test reading comprehension at B1 level. Unlike traditional cloze formats, intra-text gap-filling tasks allow for more controlled and meaningful deletions,

targeting content words that contribute to textual cohesion and comprehension. Carefully designed intra-text gap-filling items can enhance the reading component of language proficiency tests that reflect real-world reading processes.

REFERENCES

1. Alderson, J.C. (2000). *Assessing Reading*. Cambridge University Press, Cambridge.
2. Hughes, A. (2003). *Testing for Language Teachers* (2nd ed.). Cambridge University Press.
3. Taylor, W.L. (1953). Cloze procedure: a new tool for measuring read ability. *Journalism Quarterly*, 30. pp 414-438.
4. <https://h5p.org/h5p/embed/503852>
5. North B., Ortega A., & Sheehan S. (2015). A Core Inventory for General English. British Council
6. Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
7. Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. <https://www.coe.int/lang-cefr>
8. https://uzbmb.uz/page/test_sinovlari_formati
9. Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge University Press.
10. Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment*. McGraw-Hill.
11. Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Prentice Hall.
12. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. doi: 10.1007/BF02310555.
13. Maris, G., Bechger, T., Koops, J., & Partchev, I. (2018) dexter: Data management and analysis of tests. URL: <https://CRAN.Rproject.org/package=dexter>

KO'P DARAJALI INGLIZ TILI TESTLARIDA MATN-BANKLI BO'SHLIQLARNI TO'LDIRISH TOPSHIRIQLARI: DIZAYN, ISHONCHLILIK VA CEFR BILAN MUVOFIQLIK

A.A. Abbosov

*O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi huzuridagi Bilim va malakalarni baholash agentligi,
100084, Toshkent sh., Bog'ishamol k., 12*

Qisqacha mazmuni. Bo'shliqlarni to'ldirish topshiriqlari til bilish darajasini baholash testlarida eng ko'p uchraydigan savol turlaridan biridir. Ular ayniqsa ko'p darajali test formatlari uchun mos keladi, chunki ular turli murakkablik darajalarini qamrab olishi va tildan autentik foydalanishni aks ettirishi mumkin. Matn-bankli bo'shliqlarni to'ldirish topshiriqlari – bunda matnning o'zi bir vaqtning o'zida so'z banki sifatida xizmat qiladi – yaqinda ishlab chiqilgan topshiriq turiga kiradi. Ushbu maqolada matn-bankli bo'shliqni to'ldirish topshiriqlarini ishlab chiqish prinsiplari, ularning ishonchliligi va Umumiy Yevropa Til Mezonlari (CEFR) bilan qanday muvofiqlashtirilishi muhokama qilinadi.

Kalit so'zlar: ko'p darajali ingliz tili testlari, o'qish tushunish, testlet, bo'shliqni to'ldirish topshiriqlari, matn-bankli bo'shliqlarni to'ldirish, ishonchlilik, validlik, CEFR.

MILLIY TEST TIZIMI UCHUN TEST TUZUVCHI MUTAXASSISLARNI TAYYORLASH MASALALARI (O'ZBEK TILINI BILISH DARAJASI SERTIFIKATI MISOLIDA)

K.Q. Jalilov

*Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti, Toshkent shahri,
Yakkasaroy tumani, Yusuf Xos Hojib ko'chasi, 103-uy*

Qisqacha mazmuni. Testning validligini ta'minlashda test topshiriqlarini tuzishga jalb qilinadigan fan ekspertlari baholash va pedagogik o'lchovlarning fundamental tamoyil va talablarini qay darajada tushunishlari hamda test tuzish jarayoniga tatbiq qila olishlari muhim ahamiyat kasb etadi. Shu bilan birga, ta'lim tizimidan baholash bilan shug'ullanuvchi tashkilotlarga test topshiriqlarini tuzish uchun jalb qilinadigan aksar mutaxassislar baholash bo'yicha tizimli bilimga ega emas va bu mazkur tashkilotlar tomonidan o'tkaziladigan testlarning validligiga salbiy ta'sir ko'rsatishi mumkin. Pedagogik dizaynning "orqaga qarab loyihalashtirish" modeli asosida tuzilgan va gibrid tarzda tashkil etilgan qisqa muddatli kurslar muammoni yechishga yordam berishi mumkin.

Kalit so'zlar: test tuzuvchilarni tayyorlash, test topshiriqlari sifati, validlik, "orqaga qarab loyihalashtirish".

I. Kirish

Ma'lumki, validlik har qanday testning sifatini belgilaydigan asosiy va yagona mezon bo'lib, testning maqsadidan kelib chiqqan holda test natijalari talqinlarining (interpretatsiyalarining) qay darajada empirik va nazariy jihatdan asoslanganligini bildiradi [1]. Ya'ni test valid bo'lishi uchun u aynan o'lchanishi lozim bo'lgan xususiyat (konstrukt – ing. *Construct*)larni o'lchashi hamda uning natijalari asosida chiqariladigan xulosalar asoslangan bo'lishi kerak. Ta'limda qo'llaniladigan test (xususan, "yuqori qiymatga ega" – ing. *high-*

stakes" deb qaraladigan testlar) tuzish, ularni o'tkazish va natijalarini e'lon qilish bilan shug'ullanadigan tashkilotlar uchun test tuzuvchi (test topshiriqlarini shakllantiruvchi) mutaxassislarning kompetentlik darajasi testlarning validligiga bevosita ta'sir qiluvchi omillardan biridir [2]. Test validligi dalillari orasida mazmun bilan bog'liq dalillar eng muhim ahamiyat kasb etishi hisobga olinsa, test qamrab oladigan soha (domen – ing. *domain*)ni yaxshi biladigan va test tuzuvchi sifatida jalb qilingan fan mutaxassislarning

validlikni ta'minlashdagi ahamiyati yanada oydinlashadi [3].

Test tuzuvchilar uchun maxsus mashg'ulot (trening)lar test topshiriqlarini tuzishda ularning validligiga salbiy ta'sir qiladigan xatolarni [3, 4] kamaytirishi, shuningdek statistik parametrlarining (qiyinlik darajasi, diskriminatsiya indeksi va boshqalar) baholashda qabul qilingan me'yorlarga mos kelishini ta'minlash orqali testning validligiga ijobiy ta'sir qilishi tadqiqotlarda o'z isbotini topgan. Sirinaj kasalxonasi (Tailand) klinik tibbiyot o'qituvchilari bilan o'tkazilgan tadqiqotda test topshiriqlarini shakllantirish bo'yicha mashg'ulotlar o'tkazilgan va tinglovchilarning mashg'ulotlardan oldingi va keyingi test topshiriqlari, shuningdek, mashg'ulotlarda ishtirok etganlar va ishtirok etmaganlar tomonidan tuzilgan test topshiriqlari solishtirilgan [5]. Natijada, mashg'ulotlar test topshiriqlarining qiyinlik darajasiga (p qiymatiga) ijobiy ta'sir qilgani, ya'ni o'qituvchilar tomonidan tuzilgan o'ta oson yoki o'ta qiyin test topshiriqlari ulushi kamaygani aniqlangan. Shuningdek, mashg'ulot ishtirokchilari tuzgan test topshiriqlarining diskriminatsiya indeksi va nuqtali biseral korrelyatsiyasi xuddi shu

o'qituvchilarning masg'ulotlardan oldin tuzgan test topshiriqlari va mashg'ulotlarda qatnashmagan o'qituvchilar tuzgan test topshiriqlariga nisbatan yuqori ekanligi qayd etilgan. Shunga o'xshash boshqa tadqiqotlarda ham test topshiriqlarini tuzish qoidalari bo'yicha qisqa va uzoq muddatli kurslar va mashg'ulotlar test topshiriqlari sifatiga ijobiy ta'sir qilgani kuzatilgan [6, 7, 8].

Ko'rinadiki, test tuzuvchilarni tayyorlash test tuzish va uning validligini ta'minlash jarayonining muhim va ajralmas qismidir [9]. O'zbekistonda Oliy talim, fan va innovatsiyalar vazirligi huzuridagi Bilim va malakalarni baholash agentligi tomonidan o'tkaziladigan Milliy sertifikat imtihonlarining ham "yuqori qiymatga ega" testlar sirasiga kirishi – oliy ta'lim muassasalariga abituriyentlarni saralab olish, tegishli fanlar pedagoglariga ustamalar tayinlash va boshqa muhim qarorlar qabul qilishda qo'llanilishidan kelib chiqib, mazkur imtihonlar uchun test tuzuvchi mutaxassislarni tayyorlash muhim ahamiyat kasb etadi. Ushbu maqolada joriy etilishi kutilayotgan O'zbek tilini bilish darajasini aniqlash test tizimi uchun test tuzuvchi ekspertlarni tayyorlash kursi tajribasi muhokama qilinadi.

II. O'zbek tilini bilish darajasini aniqlash test tizimi uchun ekspertlar tayyorlash kursi

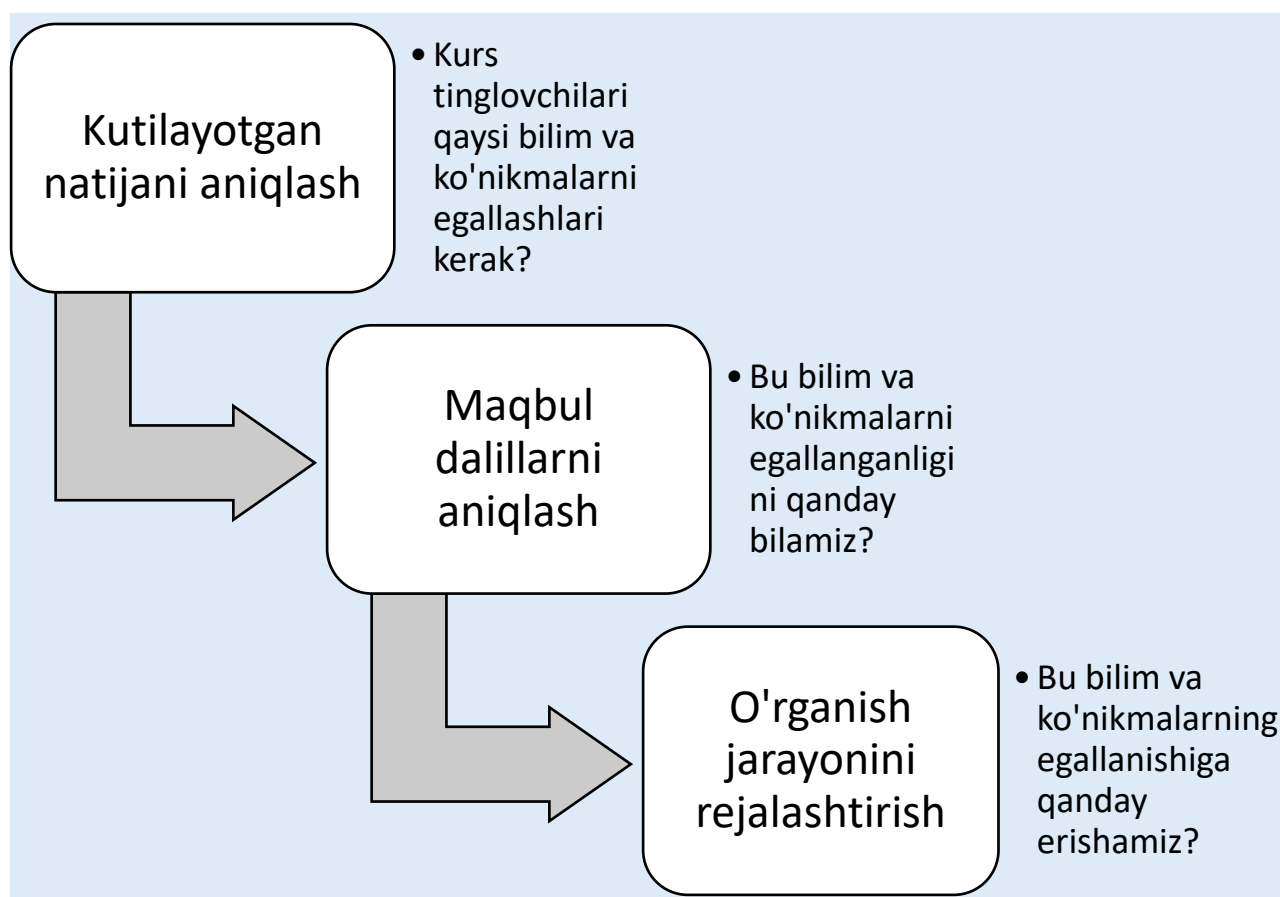
O'zbekiston Respublikasi Prezidentining 2020-yil 20-oktyabrdagi

"Mamlakatimizda o'zbek tilini yanada rivojlantirish va til siyosatini

takomillashtirish chora-tadbirlari to'g'risida"gi PF-6084-farmonida [10] davlat tilining mamlakat ichida va xalqaro miqyosda o'rni va nufuzini oshirish chora-tadbirlaridan biri sifatida Bilim va malakalarni baholash agentligiga xalqaro tajribalar asosida O'zbek tili bo'yicha bilimlarni baholashning zamonaviy tizimini joriy qilish vazifasi yuklatilgan. Mazkur vazifaning ijrosini ta'minlash va xalqaro standartlarga javob beradigan baholash tizimini yo'lga qo'yish qadamlaridan biri sifatida agentlikda o'zbek tili bo'yicha test topshiriqlarini shakllantiruvchi ekspertlar uchun maxsus 120 soatlik kurs tashkil qilindi. Kurs gibrid shaklda tashkil qilindi hamda 60 soat an'anaviy (offlayn) shakldagi ma'ruza va amaliy darslarni, shuningdek, xuddi shuncha soat hajmida Kanvas platformasi orqali o'zlashtiriladigan o'quv materiallari va bajariladigan topshiriqlarni o'z ichiga oldi.

Kursning o'quv dasturi "orqaga qarab loyihalashtirish" ("backward design") yondashuvi asosida ishlab chiqildi. "Orqaga qarab loyihalashtirish" yondashuvi pedagogik loyihalashtirish (pedagogik dizayn – ing. *instructional design*)da qabul qilingan modellardan biri bo'lib, unda dastlab o'rganivchilar

egallashi kerak bo'lgan bilim va ko'nikmalar (kutilayotgan natija) aniqlanadi hamda shundan kelib chiqib ushbu bilim va ko'nikmalarning qay darajada egallanganligini baholash usullari (maqbul dalillar) tanlanadi va o'qitish jarayoni rejalashtiriladi (1-rasm) [11]. Bunday yondashuv o'quv natijalarini aniq tasavvur qilish hamda shu natijalarga olib boradigan o'qitish metodlari va o'quv materiallarini tanlash yoki boshqacha qilib aytganda, "konstruktiv uyg'unlik" (ing. *Constructive alignment*) – o'quv maqsadlari – o'qitish metodlari va materiallari – baholash uchligining o'zaro mosligini [12] ta'minlash imkonini beradi. Shu bilan birga, "orqaga qarab loyihalashtirish" pedagogik loyihalashtirishda keng qo'llaniladigan boshqa yondashuvlar, masalan, ADDIE ("analyze – design – develop – implement – evaluate" – "tahlil qilish – loyihalashtirish – ishlab chiqish – amaliyotga joriy qilish – baholash") yoki 4C-ID ("to'rt qismli pedagogik loyihalashtirish") modellaridan farqli ravishda, uzoq muddatli tayyorgarlikni talab qilmaydi, loyihaga jarayonda o'zgartirishlar kiritish va kutilgan natijalarga tezroq erishish mumkin.



1-rasm. Pedagogik dizaynning “orqaga qarab loyihalashtirish” modeli

Kurs tinglovchilaridan o'zbek tili bo'yicha bilimlarni baholashning zamonaviy tizimini joriy etishda ekspert sifatida faol qatnashish kutilayotgani hisobga olinib, dastlab tinglovchilar kurs yakunida egallashi kerak bo'lgan asosiy kompetensiyalar ro'yxati shakllantirildi:

- o'zbek tilini bilish darajasini baholash tizimi uchun test tafsilotini (spetsifikatsiyasini) ishlab chiqish;
- test tafsilotiga mos ravishda test topshiriqlari bazasini shakllantirish;
- test topshiriqlarining sifatini klassik va zamonaviy test nazariyalari (CTT, IRT) asosida

tahlil qilish hamda test asosida chiqariladigan xulosalarning validligini ta'minlash.

Ushbu asosiy kompetensiyalarni egallash uchun tinglovchilar quyidagi bilim va ko'nikmalarga ega bo'lishi lozimligi aniqlandi:

- baholashning asosiy tamoyillarini tushunish va ularni test jarayoniga tatbiq qilish;
- test topshiriqlarining turlari, ularning maqsadlari va imkoniyatlarini farqlash, maqsadga mos test topshirig'i turini tanlay olish;
- tillarni o'qitish va baholash bo'yicha Umumyevropa ramkalari (CEFR) standartlarini tahlil qilish va ularni

o'zbek tilini bilish darajasini baholash tizimiga tatbiq qilish;

- nutq ko'nikmalari va til bo'yicha bilimlarni baholashning o'ziga xos xususiyatlarini hisobga olish;

- testni yaratish, o'tkazish va natijalarini hisoblashda baholash standartlariga rioya etilishi uchun yo'riqnomalar va talablar ishlab chiqish.

Yuqoridagi – bilim, ko'nikma va kompetensiyalarni aniqlash jarayoni “orqaga qarab loyihalashtirish” jarayonining birinchi bosqichiga (“kutilayotgan natijani aniqlash”) mos keladi. Keyingi qadam – ushbu bilim, ko'nikma va kompetensiyalarning amalda egallanganligini o'lchash (baholash) usullarini aniqlash. Baholashning autentikligi tamoyili baholash maqsadida beriladigan topshiriqlarning o'rganuvchi real hayotda bajaradigan vazifalarga yaqin bo'lishini taqozo qiladi va yuqori darajadagi kognitiv ko'nikmalarni tekshirish imkonini beradi [12]. Shundan kelib chiqib va kurs tinglovchilari keyingi faoliyatida amalga oshirishi lozim bo'ladigan vazifalarni

hisobga olgan holda, kurs tinglovchilarini baholash uchun ikki asosiy va bir-biriga bog'langan kurs ishi rejalashtirildi:

- o'zbek tilini bilish darajasini baholash tizimi uchun test tafsilotini ishlab chiqish;
- ishlab chiqilgan tafsilot asosida namunaviy test variantini shakllantirish.

Nazarimizda, kurs ishlarining bunday shaklda berilishi kurs tinglovchilarining kurs davomida egallagan bilim va ko'nikmalarini real kontekstda baholash imkonini beradi.

“Orqaga qarab loyihalashtirish” modelidagi keyingi qadam – bevosita o'quv jarayonini loyihalashtirish. Kurs mashg'ulotlari modulli struktura asosida rejalashtirildi va modullar “umumiydan xususiyga” tamoyili asosida joylashtirildi: dastlab baholash va pedagogik o'lchovlarning umumiy tamoyil va qonuniyatlari, so'ngra bevosita til ko'nikmalarini baholashga bo'lgan yondashuvlarga bag'ishlangan mavzular rejaga kiritildi (qarang: ilova).

III. Kursning tinlovchilarning baholash haqidagi qarashlariga ta'siri

Kursda jami 10 nafar tinglovchi ishtirok etdi. Tinlovchilarning aksariyati (7 nafar) oliy ta'lim tizimida faoliyat yuritadi, qolganlar umumiy o'rta va professional ta'limda hamda malaka oshirish tizimida ishlashini qayd etishgan (1 nafardan). Tinglovchilarning

aksariyati (60 foizi) asosiy ish faoliyati sifatida dars berishni (pedagogik faoliyat) ko'rsatishdi. Ularning ish tajribasi har xil: 5 yilgacha (60 foizi), 6-10 yil (10 foizi) va 20 yildan oshiq (30 foizi). Tinglovchilar orasida 3 nafar ilmiy darajaga (PhD, fan nomzodi) ega

va 3 nafar ayni paytda PhD darajasini olish uchun ilmiy izlanish olib borayotgan mutaxassislar ham mavjud.

Tinglovchilarning kursda qatnashishdan oldingi baholash va test tuzish bo'yicha tasavvuri va tajribasi har

xilligini kuzatish mumkin (1, 2-jadvallar). Aksar tinglovchilar test tuzish bilan ba'zan yoki muntazam shug'ullanishganini, baholashga oid ilmiy adabiyotlar bilan tanish ekanliklarini qayd qilishgan.

1-jadval

“Kursda qatnashishdan oldin test topshiriqlari tuzish bo'yicha tajribangizni qanday tasvirlaysiz?” savoliga javoblar taqsimoti

Javoblar	Javobni tanlagan tinglovchilar foizi
Ba'zan test topshiriqlari tuzishga to'g'ri kelgan	60
Test topshiriqlari tuzish bilan muntazam shug'ullanaman, bu ishimning bir qismi	30
Test topshiriqlari tuzish va baholash mening asosiy ish faoliyatimga kiradi	10

2-jadval

“Kursda qatnashishdan oldin baholash va test topshiriqlari tuzish bo'yicha ilmiy adabiyotlar (kitoblar, ilmiy maqolalar) bilan qay darajada tanish edingiz?” savoliga javoblar taqsimoti

Javoblar	Javobni tanlagan tinglovchilar foizi
O'qimaganman	30
Ayrim maqolalar va kitoblarni o'qiganman	60
Muntazam o'qib boraman	10

Qizig'i, tinglovchilarning aksariyati kursda qatnashishdan oldin baholash va pedagogik o'lchovlar bo'yicha ilmiy adabiyotlar bilan u yoki bu darajada tanish bo'lganini hamda test topshiriq-

lari tuzish bo'yicha muayyan tajribaga egaligini qayd qilgan bo'lsa-da, kurs ularning baholash, pedagogik o'lchov vositalarini tuzish va qo'llash bo'yicha qarashlarini qisman yoki tubdan

o'zgartirgani ham kuzatilgan (3-jadval). Tinglovchilar kursda olgan bilimlari asosida ish faoliyati davomida o'zlari

oldin tuzgan test topshiriqlariga tanqidiy baho berishgan (4-jadval).

3-jadval

“Kursda berilgan ma'lumotlar va fikrlar sizning baholash va test topshiriqlari tuzish bo'yicha qarashlaringizga qanchalik ta'sir qildi?” savoliga javoblar taqsimoti

Javoblar	Javobni tanlagan tinglovchilar foizi
Qarashlarimga mos keldi	20
Ayrim qarashlarimni o'zgartirdi.	40
Qarashlarimni tubdan o'zgartirdi.	40

4-jadval

“Agar siz ish faoliyatingiz davomida test topshiriqlari tuzgan bo'lsangiz, kursda olgan bilimlaringiz asosida bu test topshiriqlariga qanday baho berasiz?” savoliga javoblar taqsimoti

Javoblar	Javobni tanlagan tinglovchilar foizi
Ularni o'zgartirmagan bo'lardim	0
Ularga ayrim o'zgartirishlar kiritgan bo'lardim	50
Ularni butunlay boshqacha tuzgan bo'lardim.	50

Tinglovchilar test topshiriqlarini yozish va distraktorlarni shakllantirish qoidalari, aprobatsiya natijalari klassik va zamonaviy test nazariyalari asosida tahlil qilish orqali test topshiriqlari

sifatiga hamda testning ishonchliligiga baho berish, validlik dalillari bilan bog'liq mavzular ular uchun yangilik bo'lganini qayd qilishgan.

IV. Xulosa

“Hech kim yaxshi test tuzuvchi ekspert bo'lib tug'ilmaydi, test tuzuvchi ekspertlar tarbiyalanadi” [9]. Test

topshiriqlarini to'g'ri tuzish va testning validligini ta'minlash ekspertdan nafaqat qamrab olinayotgan soha

bo'yicha chuqur bilimni, balki baholash va pedagogik o'lchovlar bo'yicha muayyan kompetensiyalarni talab qiladi. Ta'lim tizimida faoliyat yuritadigan aksar mutaxassislar o'z faoliyati davomida baholash va test topshiriqlari tuzish bilan u yoki bu darajada shug'ullansa-da, baholashning, xususan, "yuqori qiymatga ega" deb tasniflanishi mumkin bo'lgan testlarning validligini ta'minlash uchun bevosita test tuzish jarayoniga jalb qilinadigan ekspert-

larning baholashning fundamental tushunchalari bo'yicha tizimli bilimlarga ega bo'lishi, test topshiriqlarini yozish qoidalaridan xabardorligi va test topshiriqlari aprobatsiyasi natijalari asosida test sifatiga baho bera olishi muhimdir. "Orqaga qarab loyihalashtirish" yondashuvi asosida tuzilgan gibrid formatdagi qisqa muddatli, tizimli kurslar bunday bo'shliqni bartaraf etishda yordam berishi mumkin.

O'zbek tilini bilish darajasini baholash tizimi ekspertlari tayyorlash kursi modullari va mavzulari

1-modul: Baholash nazariyasiga kirish:

- 1.1 Baholash tamoyillari va ularga ta'sir etuvchi omillar
- 1.2 Test turlari va ularning imkoniyatlari
- 1.3 Test yaratish boshqichlari. Test spetsifikatsiyasi va uni yaratish
- 1.4 Aprobatsiya natijalarini o'qish va tahlil qilish

2-modul: Til malakalari va ularni baholash:

- 2.1 Til malakasini aniqlash imtihonlari
- 2.2 CEFR standartlari va o'zbek tili

3-modul: Retseptiv ko'nikmalarni baholash testlari:

- 3.1 Tinglab tushunish ko'nikmalarini baholash
- 3.2. O'qib tushunish ko'nikmalarini baholash
- 3.3 Kontent tanlash tamoyillari

4-modul: Produktiv ko'nikmalarni baholash testlari:

- 4.1 Yozma nutqni baholash
- 4.2. Og'zaki nutqni baholash

ADABIYOTLAR

1. American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). Standards for Educational and Psychological Assessment. Washington, DC: AERA.
2. The International Association for Educational Assessment (2024). International Standards for Educational Assessment Organisations.
3. Haladyna, T. and Rodriguez, M. (2013). Developing and Validating Test Items. New York, NY: Routledge.
4. Popham, W. J. (2000) Modern Educational Measurement: Practical Guidelines for Educational Leaders. Boston: Allyn & Bacon.
5. Iramaneerat, C. (2022). The impact of item writer training on item statistics of multiple-choice items for medical student examination. Siriraj Medical Journal, 64(6), pp. 178-182.
6. Dellenges, M. A., Curtis, D. A. (2017) Will a Short Training Session Improve Multiple-Choice Item-Writing Quality by Dental School Faculty? A Pilot Study. Journal of Dental Education, 81(8), pp. 948-955.
7. Naeem, N., van der Vleuten, C. and Alfari, E.A. (2012) Faculty development on item writing substantially improves item quality. Adv in Health Sci Educ 17, pp. 369–376.
8. Abdulghani H. M., Irshad M., Haque S., Ahmad T., Sattar K., Khalil M.S. (2017) Effectiveness of longitudinal faculty development programs on MCQs items writing skills: A follow-up study. PLoS ONE 12(10).
9. Downing, S. and Haladyna, T. (2006) Handbook of Test Development. Mahwah, NJ: Lawrence Erlbaum Associates.
10. O'zbekiston Respublikasi Prezidentining 2020-yil 20-oktabrdagi PF-6084-son Farmoni. "Mamlakatimizda o'zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to'g'risida". <https://lex.uz/docs/-5058351>.
11. Wiggin, G. and McTighe, J. (2005) Understanding by Design. 2nd edition. Alexandria, VA: Association for Supervision and Curriculum Development.
12. Biggs, J. B. and Tang, C. (2011). *Teaching for Quality Learning at University: What the Student Does*. Maidenhead: McGraw-Hill.
13. Wyatt-Smith, C. , Klenowski, V. and Colbert, P., Eds. (2014) Designing Assessment for Quality Learning. New York: Springer.

ISSUES OF TEST ITEM WRITERS TRAINING FOR THE NATIONAL TESTING SYSTEM (ON THE EXAMPLE OF THE UZBEK LANGUAGE PROFICIENCY TESTS)

K.K. Djalilov

Tashkent State University of Uzbek Language and Literature named after Alisher Navoi, Tashkent city, Yakkasaray district, Yusuf Khos Hojib street, 103

Abstract. In ensuring the validity of the test, it is important that the subject matter experts (SMEs) contracted as test item writers understand the fundamental principles and requirements of test development and can apply these principles to the process of item writing. However, the fact that many SMEs who work in educational institutions and serve as test item writers for organisations that develop and administer tests lack systematic knowledge of assessment principles and test development procedures can negatively affect the validity of tests used by such organisations. Short-term courses, modelled on the “backward design” approach and offered in a hybrid mode, can help to address the issue.

Keywords: test item writer training, quality of test items, validity, "backward design".

KIMYO FANIDAN TEST NATIJALARINING TAFSILOTLAR ASOSIDAGI 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 kimyo fanidan milliy sertifikat uchun o'tkazilgan test sinovi natijalari asosida talabgorlarning hududlar, yosh va tillar kesimida statistik tahlillari amalga oshirildi. 25–35 yosh oralig'idagi ishtirokchilarning natijalari boshqa yosh guruhlariga nisbatan yuqorili, aksincha 20 yoshgacha bo'lgan talabgorlarning natijalari boshqa yosh guruhlariga nisbatan past ekanligi aniqlandi. Qiyinlik darajasi yuqori bo'lgan test topshiriqlari test tafsilotiga asosan qaysi tematik mavzularga mos kelishi o'rganildi.

Kalit so'zlar: anorganik kimyo, organik kimyo, kognitiv ko'nikma, Rash modeli, qiyinlik darajalari, test tafsiloti.

I. Kirish

Bugungi kunda ta'lim tizimi oldida turgan eng dolzarb masalalardan biri – bu o'quvchilarning bilimini adolatli, ishonchli baholash, ta'lim dasturida belgilangan bilim va ko'nikmalarni qanday darajada egallaganligi haqida asosli xulosalar chiqarish zaruratidir [1]. Zamonaviy baholash usullari o'quvchilarning nafaqat bilim darajasini aniqlash, balki ularning tafakkur doirasi, mustaqil fikrlashi va muammoli vaziyatlarda to'g'ri yechim topa olish qobiliyatini baholashga ham qaratilishi lozim [2].

Tabiiy fanlar, jumladan kimyo fanidan o'quvchilarning o'zlashtirish darajasi va bilim sifati – ularning kelgusidagi kasbiy tayyorgarligi, ilmiy mulohaza qilishi va mustaqil qarorlar

qabul qilish qobiliyatiga bevosita ta'sir ko'rsatadi [3]. Kimyo o'z mohiyatiga ko'ra nazariy bilimlar va amaliy kompetensiyalarni uyg'unlashtiruvchi, murakkab tushunchalarga asoslangan, yuqori darajadagi kognitiv faoliyatni talab etuvchi fandır [4]. Shu boisdan ham kimyo fanida qo'llaniladigan pedagogik o'lchov vositalari nafaqat bilimni eslab qolish, balki tahlil qilish, solishtirish, xulosa chiqarish, sabab va oqibat aloqalarini anglash kabi yuqori darajadagi aqliy ko'nikmalarini baholay olishi zarur [5-8]. Mazmuniy jihatdan puxta ishlab chiqilgan, didaktik va psixometrik talablarni qanoatlantiruvchi test topshiriqlari o'quvchilarning o'zlashtirish darajasini ishonchli aniqlash imkonini beradi [9].

Shu nuqtai nazardan qaraganda, test topshiriqlari qanday shakllantirilgani o'quvchilarning kimyo fanini qanday darajada o'zlashtirganligini aniqlashda muhim ahamiyat kasb etadi. Bu esa test topshiriqlarining sifat ko'rsatkichlari bilan o'quvchilarning o'zlashtirish darajasi o'rtasidagi o'zaro bog'liqlikni chuqur ilmiy tahlil qilish zaruratini yuzaga keltiradi. Bunday tahlil o'z navbatida ta'lim sifatini oshirish, o'qituvchilarning bilimni baholash amaliyotini takomillashtirish va ilmiy asoslangan didaktik qarorlar qabul qilishiga xizmat qiladi [10-12].

Bugungi ta'lim tizimida bilimni baholashda muqobil javobli testlar keng qo'llanilayotgan samarali pedagogik o'lchov vositalardan biri hisoblanadi. Testlar yordamida qisqa vaqt ichida katta hajmdagi ma'lumotlarni yig'ish, o'quvchilarning bilim darajasini oldindan belgilangan mezonlar asosida taqqoslash, ularning o'zlashtirishdagi farqlarini aniqlash imkoniyati mavjud. Ayniqsa, muhim ahamiyatga ega baholashlarda test jarayonining sifat ko'rsatkichlari uni boshqa pedagogik o'lchov vositalari shakllaridan ustun qo'yadi [13,14].

Biroq, test natijalarining oddiy tahlili asosida baholash sifati haqida aniq xulosa chiqarish mumkin emas. Baholash vositalarining haqiqiy diagnostik qiymatini aniqlash uchun test topshiriqlarining sifat ko'rsatkichlari chuqur o'rganilishi lozim.

Kimyo fanidan talabgorlarning tayyorgarlik darajasini baholovchi test topshiriqlari majmuasi mazmunan fanning barcha bo'limlarini qamrab olishi lozim. Jumladan, test tuzishda kimyo fanining barcha bo'limlari, shuningdek, ushbu bo'limlarga tegishli kichik tematik mavzular inobatga olinishi kerak [15-17]. Bu yondashuv test mazmunining to'liqlik (komplekslik), mavzulararo muvofiqlik va kompetensiyaviy yondashuv talablari asosida shakllantirilishini ta'minlaydi.

Shu sababli, kimyo fanidan bilimni baholash uchun tuziladigan test topshiriqlarini quyidagi asosiy yo'nalishlar bo'yicha tuzish maqsadga muvofiqdir:

- **Umumiy kimyo:** modda va uning xossalari, kimyoviy bog'lanishlar, reaksiya turlari, modda miqdori, energiya o'zgarishlari, kinetika va muvozanat va hokazo.
- **Anorganik kimyo:** kimyoviy elementlar va ularning birikmalari, davriy qonuniyatlar, gidroksidlar, oksidlar, kislotalar va tuzlar, metall va metallmaslar xossalari va hokazo.
- **Organik kimyo:** uglerod birikmalarining tasnifi, tuzilish nazariyasi, asosiy sinflar (alkanlar, alkenlar, alkinlar, aromatik birikmalar, spirtlar, kislotalar), reaksiyalar va ularning mexanizmlari va hokazo.

Mazkur mazmundagi yo'nalishlarning to'g'ri qamrab olinishi test

topshiriqlarining validlik va ishonchlik darajasini oshiradi. Shuningdek, bu holat ta'lim standartlariga muvofiqlik, milliy baholash mezonlariga asoslanish, va o'quvchilar bilimini diagnostik tahlil qilish imkoniyatini yaratadi.

Ushbu maqolada kimyo fanidan milliy sertifikat uchun o'tkazilgan test sinovi natijalaridan foydalanildi. Kimyo fanidan milliy sertifikat uchun o'tkazilgan test sinovlarida 3 ta test variantidan foydalanildi. Har bir variant 43 ta test topshirig'idan iborat bo'lib, 1-35 – yopiq test topshiriqlari, 36-40-qisqa javobli ochiq test topshiriqlari va 41-43 – kengaytirilgan

javobli ochiq test topshiriqlaridan iborat bo'lib, ajratilgan vaqt javoblar varaqasini bo'yash bilan birgalikda 250 daqiqani tashkil etadi. Ammo statistik tahlillar olib borish uchun (41-43) kengaytirilgan javobli ochiq test topshiriqlari natijalaridan foydalanilmadi. Test sinovlarida jami 26152 nafar talabgor ishtirok etdi.

Mazkur maqolada kimyo fanidan milliy sertifikat uchun o'tkazilgan test sinovi natijalari test tafsiloti asosida talabgorlarning o'quv dasturida belgilangan mavzularni o'zlashtirish darajasi statistik usullar orqali tahlil qilingan va xulosalar chiqarilgan

II. Asosiy qism

Kimyo fanidan milliy sertifikat uchun foydalanilgan test topshiriqlarining qiyinlik darajalarini aniqlash uchun zamonaviy test nazariyasi (Item Response Theory - IRT)dan foydalanildi.

Bu nazariya nafaqat psixometrik baholashda qo'llaniladigan zamonaviy statistik yondashuv, balki u test topshiriqlari (element) bilan talabgorlar o'rtasidagi munosabatni ehtimollik modellariga asoslangan holda ifodalaydi. Mazkur nazariyaga xos bo'lgan Rasch modeli test topshiriqlarining qiyinlik darajalarini va talabgorlarning qobiliyat darajalarini alohida statistik parameter sifatida baholash imkonini beradi [18-

21]. Bu talabgorlarning qobiliyat darajalarini va test topshiriqlarining qiyinlik darajalarini chiziqli o'lchash orqali amalga oshiriladi. Rash modeli test natijalari orqali talabgorlarning shaxsiy qobiliyatiga nisbatan obyektiv baholashga imkon beradi va har bir test topshirig'ining Rash modelga qanchalik mos kelishini INFIT va OUTFIT statistik ko'rsatkichlari orqali aniqlaydi [22]. Bu esa test topshiriqlarining sifatini ilmiy asosda o'rganish, juda oson yoki juda qiyin test topshiriqlarini aniqlash va ularni takomillashtirish imkoniyatini beradi. Rash modeli nafaqat test topshiriqlari xususiyatlarini tahlil qilishda, balki o'quvchilarning bilim darajasini

ishonchli aniqlashda ham yuqori aniqlik va adolatlilikni ta'minlaydi.

Rash modeliga ko'ra, dixotomik elementlarga individual javoblar talabgorlarning qobiliyat darajasi va test topshirig'i qiyinligi bilan

aniqlanadi. Ma'lum bir qobiliyatga ega bo'lgan sinaluvchining ma'lum bir qiyinlikdagi test topshirig'iga to'g'ri javob berish ehtimolligini aniqlaydi. Bu quyidagi matematik formula orqali ifodalanadi [23-22]:

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

Bu yerda $X_{si}=1$ s-sinaluvchining i test topshirig'iga to'g'ri javob berish ehtimolligi, θ_s -qobiliyat o'zgaruvchisi, b_i -topshiriq qiyinlik darajasi, e - natural logarifm asosi ($e=2,718...$).

Rash modeli asosida aniqlangan test topshiriqlarining qiyinlik darajalari 1-jadvalda ko'rsatilgan, 1-, 2- va 3-variantlar uchun test

topshiriqlarining qiyinlik darajalari mos ravishda $(-2,314:5,369)$, $(-196:4,74)$ va $(-1,914:2,75)$ oralig'ida ekanligi aniqlandi. Test topshiriqlarining qiyinlik darajalari $(-3:3)$ logit birligi oralig'ida taqsimlanishi ilmiy va statistik tahlillar ko'ra maqsadga muvofiqdir [22-26].

1-jadval

Kimyo fanidan milliy sertifikat uchun o'tkazilgan 1-test sinovi natijalari bo'yicha test topshiriqlarining aniqlangan qiyinlik darajalari

№	1-variant	2-variant	3-variant
1	-1,723	-0,448	0,351
2	0,310	-1,960	0,319
3	0,204	-1,237	-0,758
4	-1,237	-0,124	0,406
5	0,008	-0,041	-1,367
6	-0,464	-0,038	-1,597
7	-0,491	-0,672	-0,052
8	-0,572	0,591	-1,512
9	-0,920	-0,933	-0,661
10	0,127	-0,713	0,155
11	0,631	0,631	0,349
12	-1,243	-0,955	-0,322

13	0,291	-0,085	-0,149
14	-0,165	-1,914	-1,914
15	-1,451	-1,948	-0,418
16	-1,948	-0,003	-0,299
17	-0,430	0,299	-0,307
18	0,030	-0,330	-0,330
19	1,150	0,368	0,230
20	-0,662	-0,820	-1,295
21	0,556	0,327	-1,098
22	-0,102	0,291	0,229
23	-0,304	-0,859	0,011
24	-1,547	0,003	0,571
25	0,160	0,311	-0,854
26	-0,780	-0,780	-0,601
27	-2,314	-0,867	-0,641
28	0,396	0,952	0,952
29	0,367	-0,665	-0,665
30	0,753	-0,328	-0,585
31	-1,564	-1,564	-0,977
32	0,249	-0,184	-0,184
33	0,655	-0,057	-0,100
34	0,547	0,204	-0,320
35	0,653	1,050	-0,092
36	5,369	2,666	2,735
37	2,123	4,741	1,453
38	3,837	2,407	2,750
39	2,470	1,997	2,374
40	5,278	3,745	2,192

Biroq test topshiriqlari orasida qiyinlik darajasi (-3:3) logit birliklari oralig'idan tashqarida uchraydigan bir yoki ikkita topshiriqning mavjudligi yuqori qobiliyatli test talabgorlarni aniqlash imkoniyatini oshiradi. Shu

bilan birga, bunday topshiriqlarning son jihatdan ko'p bo'lishi pedagogik o'lchovlar nazariyasi tamoyillariga zid hisoblanadi[22, 23].

Qiyinlik darajasi bo'yicha 1- va 2-variantlaridagi (-3:3) logit birligi

oralig'idan tashqarida bo'lgan test topshiriqlarining qiyinlik darajalari, mutaxassislar fikriga ko'ra, juda past yoki juda yuqori deb baholansa, ushbu topshiriqlar o'rniga $(-3:3)$ logit birligiga mos keladigan qiyinlik darajasidagi test topshiriqlaridan foydalanish maqsadga muvofiq bo'ladi. Agar mutaxassislar tomonidan ularning qiyinlik darajasi me'yoriy deb topilsa, test topshiriqlari qiyinlik darajasining $(-3:3)$ logit birligi diapazonidan tashqarida bo'lishi sabablarini aniqlash kerak bo'ladi.

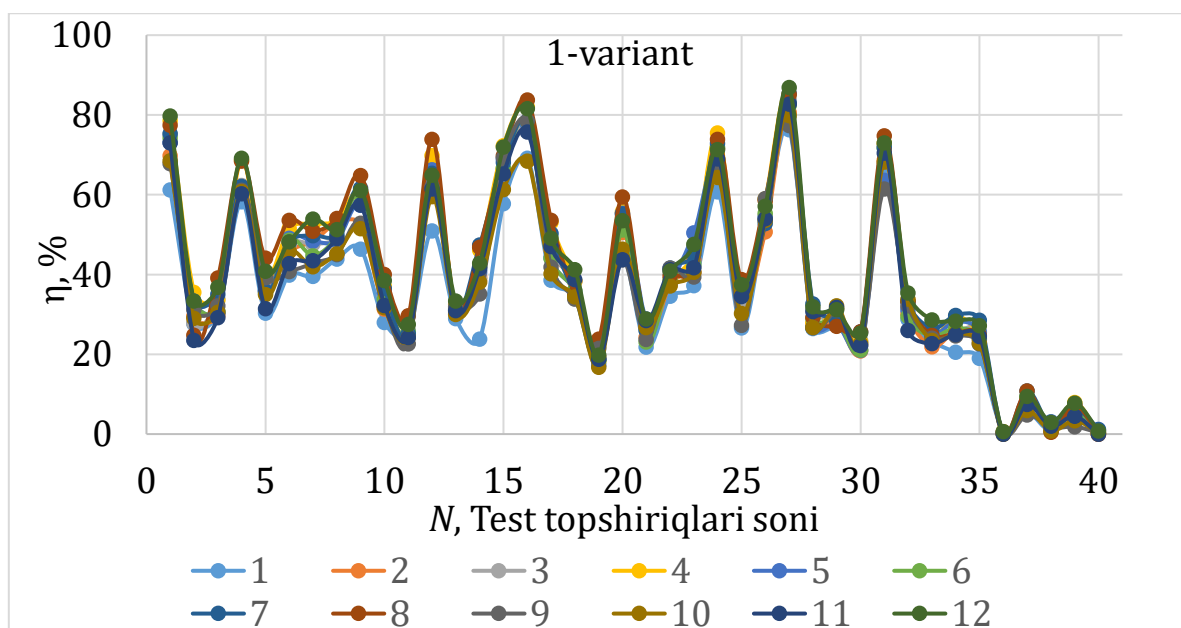
Qobiliyat va qiyinlik darajalari bir xil shkalada bo'lishi ta'minlangan turli xil hududlardagi talabgorlarning kimyo fanidan test sinovi natijalarini solishtirish orqali ularning ta'lim dasturini o'zlashtirish ko'rsatkichi, ta'lim olayotgan ta'lim muassasalarining samaradorligini, unda faoliyat yuritayotgan pedagoglarning bilimlari haqida xulosalar chiqarish mumkin.

1-a, b va c rasmlarda kuzatilgan yuqori cho'qqilar aynan shu test topshiriqlariga talabgorlar tomonidan to'g'ri javob berish ulushining yuqoriligini anglatadi. Bu esa 1-jadvalda keltirilgan Rash modeli asosida aniqlangan test topshiriqlarining qiyinlik darajalari ham

aynan shu cho'qqilarga mos keluvchi test topshiriqlari uchun nisbatan past ekanligini ko'rsatadi. Bu ularni nisbatan oson test topshiriqlari sifatida tasniflash imkonini beradi. Shuni alohida ta'kidlash joizki, ushbu test topshiriqlarning oson deb baholanishi nafaqat ularning tuzilishiga bog'liq, balki ular qamrab olgan tematik mavzular talabgorlar tomonidan yaxshi o'zlashtirilgan bo'lish ehtimolini ham ko'rsatadi.

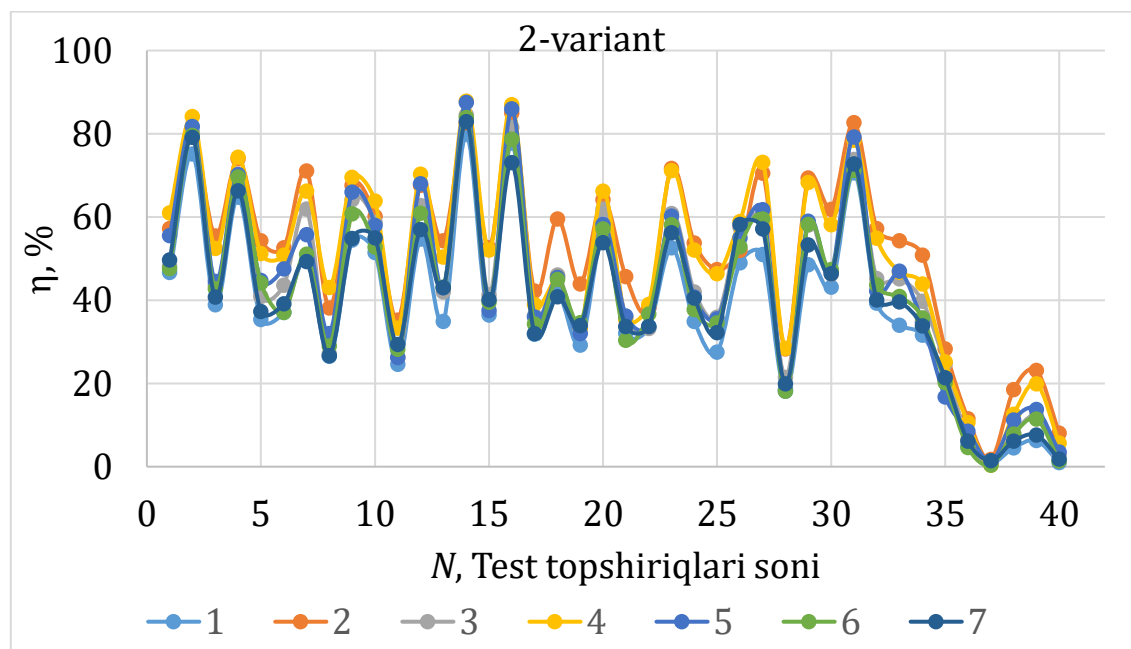
1-variantda ko'zatilayotgan quyi nuqtalar 2-, 3-, 5-, 10-, 11-, 13-, 19-, 21-, 25-, 28-, 29-, 30- va 35-40-tartibli test topshiriqlariga mos keladi. Ushbu test topshiriqlarning Rash modeli asosida aniqlangan qiyinlik darajalari ham o'rtacha qiymatdan yuqori bo'lib, ularning qiymati 1-jadvalda ko'rsatilgan. Ayniqsa 36-, 38- va 40-test topshiriqlarining qiyinlik darajalari mos ravishda 5,369, 3,837 va 5,278 ga tengligi bu test topshiriqlarining o'ta qiyin $(-3:3)$ logit birligidan tashqarida) ekanligini anglatadi.

1-a, b va c rasmlarda kimyo fanidan milliy sertifikat uchun o'tkazilgan test sinovlari natijalari asosida aniqlangan talabgorlarning test topshiriqlariga to'g'ri javob berish ulushining test topshiriqlari soniga bog'liqlik grafigi keltirilgan.



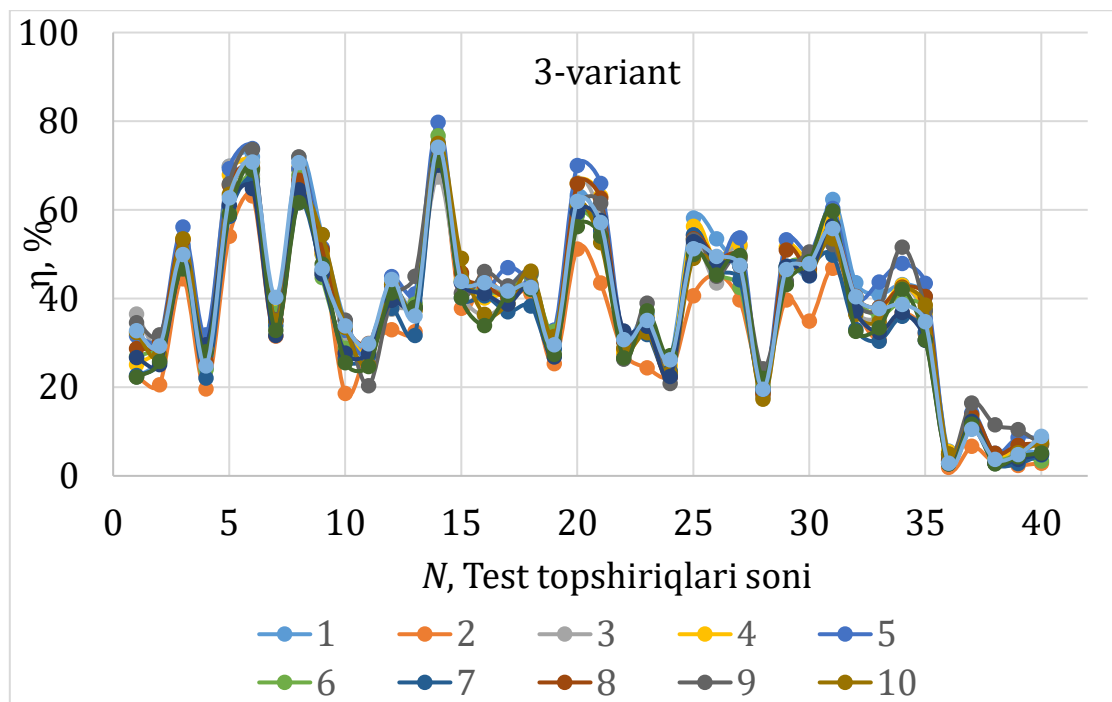
a)

- | | |
|------------------------|---|
| 1 Andijon viloyati | 7 Qashqadaryo viloyati |
| 2 Namangan viloyati | 8 Jizzax viloyati |
| 3 Farg'ona viloyati | 9 Navoiy viloyati |
| 4 Buxoro viloyati | 10 Samarqand viloyati |
| 5 Xorazm viloyati | 11 Sirdaryo viloyati |
| 6 Surxondaryo viloyati | 12 Toshkent viloyati va Toshkent shahri |



b)

- | | |
|---------------------|----------------------|
| 1 Andijon viloyati | 5 Jizzax viloyati |
| 2 Namangan viloyati | 6 Navoiy viloyati |
| 3 Farg'ona viloyati | 7 Samarqand viloyati |
| 4 Buxoro viloyati | |



c)

- | | |
|---------------------------------|---|
| 1 Qoraqalpog'iston Respublikasi | 8 Qashqadaryo viloyati |
| 2 Andijon viloyati | 9 Jizzax viloyati |
| 3 Namangan viloyati | 10 Navoiy viloyati |
| 4 Farg'ona viloyati | 11 Samarqand viloyati |
| 5 Buxoro viloyati | 12 Sirdaryo viloyati |
| 6 Xorazm viloyati | 13 Toshkent viloyati va Toshkent shahri |
| 7 Surxondaryo viloyati | |

1-rasm. test topshiriqlariga to'g'ri javob berish ulushi(%)ning test topshiriqlarining tartib raqamiga bog'liqligi

Hududlar kesimida olib borilgan statistik tahlil natijalari shuni ko'rsatadiki, barcha foydalanilgan variantlarda bir qancha yuqori cho'qqilar va quyi nuqtalar mavjud bo'lib, bu o'zgarish tendensiyasi barcha hududlar kesimida va barcha variantlar uchun o'zgarmasligini saqlab qolgan.

2-variantda kuzatilayotgan quyi nuqtalar ham mos ravishda 1-, 3-, 5-, 7-, 11-, 13-, 15-, 17-, 19-, 21-, 25-, 28- va 34-40-tartibli test topshiriqlariga mos keladi, ushbu test topshiriqlarning

Rash modeli asosida aniqlangan qiyinlik darajalari ham o'rtacha qiymatdan yuqori bo'lib, ularning qiymati 1-jadvalda ham ko'rsatilgan, 37- va 40-test topshiriqlarining qiyinlik darajalari mos ravishda 4,741 va 3,74 ga tengligi, bu test topshiriqlarining ham ((-3:3) logit birligidan tashqarida) qiyin darajasi yuqori ekanligini bildiradi.

Xuddi shu kabi o'zgarish qonuniyati 3-variant uchun ham takrorlangan bo'lib, kuzatilgan quyi nuqtalar 1-, 2-, 4-, 7-, 10-, 11-, 13-, 19-,

22-, 24-, 28-, va 35-40-tartibli test topshiriqlariga mos keladi. Bu test topshiriqlarining ham Rash modeli asosida aniqlangan qiyinlik darajasi o'rtacha qiymatdan yuqori ekanligini ko'rsatadi (1-jadval).

Barcha foydalanilgan test variantlarida 36-40 gacha (qisqa javobli ochiq test topshiriqlari) bo'lgan test topshiriqlariga to'g'ri javob berish ulushi juda past bo'lib o'rtacha 8,5 foizni tashkil qiladi. Bu esa ushbu test topshiriqlarining qiyinlik darajasi yuqori ekanligi, talabgorlar tomonidan taxminiy javob berish ehtimolining mavjud emasligi, ular qamrab olgan tematik mavzular talabgorlar tomonidan yaxshi o'zlashtirilmagan bo'lishi ham mumkin.

Ushbu statistik tahlillar natijalari asosida talabgorlarning kimyo fanining qaysi kichik tematik mavzularini yetarli darajada o'zlashtirmaganliklarini aniqlash, shuningdek, fan dasturi yoki ta'lim jarayonida mavjud bo'lgan metodik kamchiliklar va tizimli muammolar yuzasidan asosli xulosalar chiqarish imkoniyati yuzaga keladi.

Bu holatda kimyo fanining test tafsilotlariga murojaat qiladigan bo'lsak, talabgorlarning ayrim mavzular bo'yicha o'zlashtirish darajasi past ekanligi kuzatildi, bu esa ularning mazkur tematik mavzularni yetarli darajada o'zlashtirmaganligini yoki shu mavzulardan nisbatan qiyin test topshiriqlari mavjudligini ko'rsatadi.

Jumladan:

Kimyoviy elementlar davriy sistemasi, atomlarning davriylik xossalari, atom tuzilishi, izotop, izobar, izoton, izoelektron tushunchalari, elektron konfiguratsiya, pauli prinsipi, klechkovski va gund qoidalari, kvant sonlari, kimyoviy bog'lanish turlari (kovalent, ion, metall, vodorod, donor-akseptor), kovalent bog'lanishning ba'zi bir xususiyatlari, struktura formulalari, kristall panjara turlari, gibridlanish va uning xillari, anorganik birikmalarning eng muhim sinflari, ularning klassifikatsiyasi, ular o'rtasidagi genetik bog'lanishlar, asoslar, aromatik uglevodorodlar, aldegid va ketonlar, karbon kislotalar, ularning molekula tuzilishi, gomologik qatori, izomeriyasi, nomenklaturasi, oddiy va murakkab efirlar, sovunlar, yog'lar, ularning molekula tuzilishi, gomologik qatori, izomeriyasi, nomenklaturasi, tabiatda tarqalishi, nitrobirikmalar, aminlar va romatik aminlar, olinishi va xossalari, aminokislotalar, oqsillar, neft, tabiiy gaz va toshko'mir, tabiiy va sintetik yuqori molekulyar birikmalar (polimerlar, polimerlanish va polikondensatlanish reaksiyalari, polimerlarning olinishi va tuzilishi, kauchuk va tolalar) kabi tematik mavzulardan tuzilgan test topshiriqlariga berilgan javoblar ulushi past ekanligi aniqlandi.

Yoki yuqorida ko'rsatilgan tematik mavzular kimyo fanining o'quv dasturida yetarli akademik dars

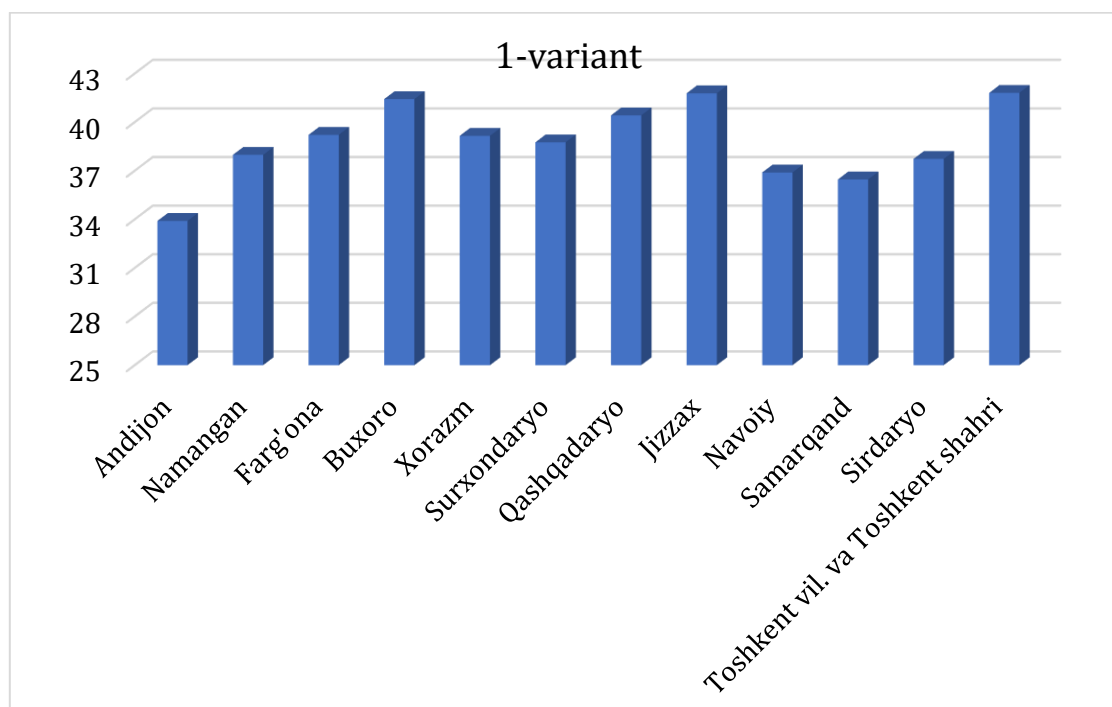
soatlari bilan qamrab olinmagan bo'lishi, shuningdek, ta'lim jarayonida uslubiy kamchiliklar va tizimli muammolar mavjud bo'lishi mumkin.

Talabgorlarning test natijalarini o'zaro solishtirish orqali nafaqat ularning bilim darajasi, balki ta'lim muassasalari samaradorligi va pedagoglar faoliyati to'g'risida ham asosli xulosalar chiqarish mumkin. 2-a, b va c rasmda hududlar kesimida kimyo fanidan o'tkazilgan test sinovlari natijalari asosida o'zlashtirish darajasining o'rtacha qiymati (foizlarda) berilgan.

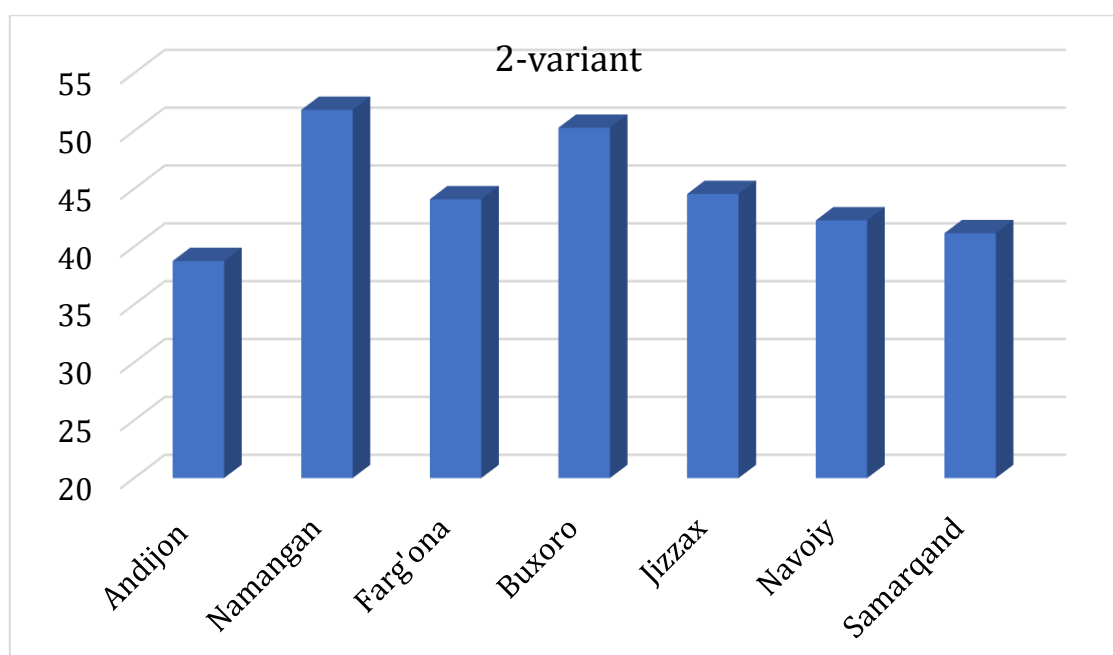
1-variant bo'yicha olingan natijalar tahliliga ko'ra, kimyo fanidan fan dasturida keltirilgan tematik mavzularni o'zlashtirishining eng yuqori o'rtacha ko'rsatkichlari Buxoro

viloyati (41,42 foiz), Jizzax viloyati (41,77 foiz), shuningdek, Toshkent viloyati va Toshkent shahri (41,8 foiz)da kuzatilgan. Aksincha, fan dasturini o'zlashtirish darajasi eng past bo'lgan hududlar esa Andijon viloyati (33,9 foiz) va Samarqand viloyati (36,46 foiz)da kuzatilgan (2-a rasm).

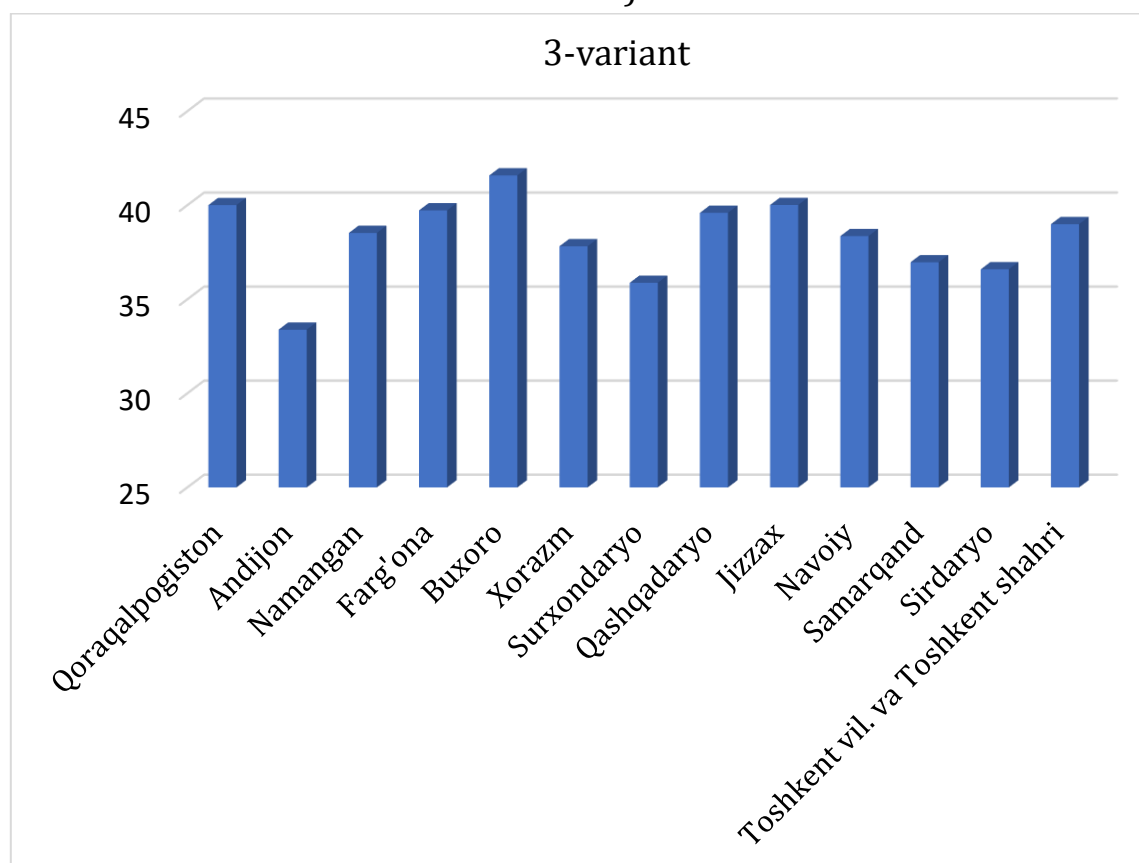
2-variant bo'yicha olingan natijalar tahliliga ko'ra, kimyo fanidan fan dasturida keltirilgan tematik mavzularni o'zlashtirishining eng yuqori o'rtacha ko'rsatkichlari Namangan viloyati (51,82 foiz) va Buxoro viloyati (52,27 foiz)ga to'g'ri keladi. Aksincha, fan dasturini o'zlashtirish darajasi eng past bo'lgan hududlar qatoriga Andijon viloyati (38,76 foiz) va Samarqand viloyatida (41,16 foiz) kuzatilgan (2-b- rasm).



a)



b)



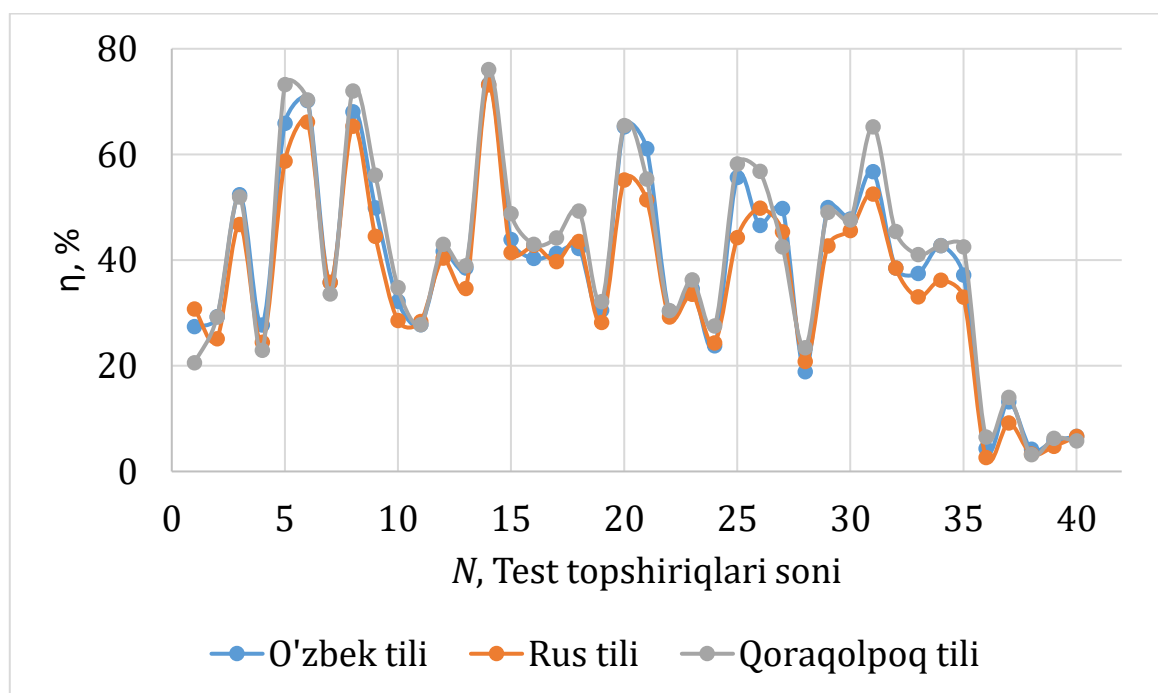
c)

2-rasm. Hududlar kesimida talabgorlarning test topshiriqlariga to'g'ri javob berish o'rtacha qiymatining ulushi (%)

Xuddi shuningdek, 3-variant bo'yicha olingan natijalar tahliliga ko'ra, kimyo fanidan fan dasturida keltirilgan tematik mavzularni

o'zlashtirishining eng yuqori o'rtacha ko'rsatkichlari Qoraqolpog'iston Respublikasi (40 foiz), Buxoro viloyati (41,58 foiz), Jizzax viloyati (40,01 foiz)ga to'g'ri kelgan bo'lsa, fan dasturini o'zlashtirish darajasi eng past bo'lgan hududlar esa Andijon viloyati (33,38 foiz) va Surxondaryo viloyati (35,87 foiz) ga to'g'ri keladi (2-c rasm) Kimyo fanidan milliy sertifikat uchun o'tkazilgan test sinovlarida har bir test variant talabgorlar talabiga asosan o'zbek, rus va qoraqalpoq tillarida taqdim etildi.

Bu esa talabgorlarning test natijalari asosida tillar kesimida statistik tahlil olib borish imkoniyatini yaratadi. Statistik tahlil natijalari ko'ra kimyo fanidan qoraqalpoq tilida test topshirgan talabgorlarning eng yuqori o'rtacha ko'rsatkichlari o'zbek va rus tilida test topshirgan talabgorlarga nisbatan yuqoriroq ekanligi aniqlandi. Aksincha, kimyo fanidan rus tilida test topshirgan talabgorlarning test natijalari nisbatan pastroq ekanligi aniqlandi (3-rasm).



3-rasm. Tillar kesimida talabgorlarning test topshiriqlariga to'g'ri javob berish ulushi (%)

Kimyo fanidan milliy sertifikat uchun o'tkazilgan imtihonlarda turli yoshdagi ishtirokchilarning qatnashgani test natijalarini yosh toifalari bo'yicha ham tahlil qilish imkonini

beradi. Bu esa test topshiriqlari mazmunining yoshga qarab o'zlashtirilish darajasini aniqlash va baholashda ular haqida xulosalar chiqarishga asos bo'ladi.

4-rasmda talabgorlarning yosh guruhlarida test topshiriqlariga to'g'ri javob berish ulushi (foiz) keltirilgan. O'tkazilgan statistik tahlillar natijalariga ko'ra 25–35 yosh oralig'idagi talabgorlarning natijalari boshqa yosh guruhlariga nisbatan yuqoriligi, aksincha 20 yoshgacha bo'lgan talabgorlarning natijalari barcha foydalanilgan test variantlari uchun boshqa yosh guruhlariga nisbatan past ekanligi aniqlandi.

25–35 yoshdagi talabgorlar, odatda, oliy ta'limni yakunlagan yoki faol akademik yoki kasbiy faoliyat bilan shug'ullanayotgan yosh guruhiga kiradi. Bu yoshda xotira, mantiqiy fikrlash, muammoga tez va to'g'ri yechim topish kabi kognitiv ko'nikmalar nisbatan barqaror va yuqori ekanligi ehtimoldan yiroq emas. Shuni ham takidlash kerakki bu yoshdagi ishtirokchilar o'z bilim darajasini diagnostika qilish va professional faoliyatini yuqori darajada saqlab turish maqsad qilgan bo'lishi mumkin,

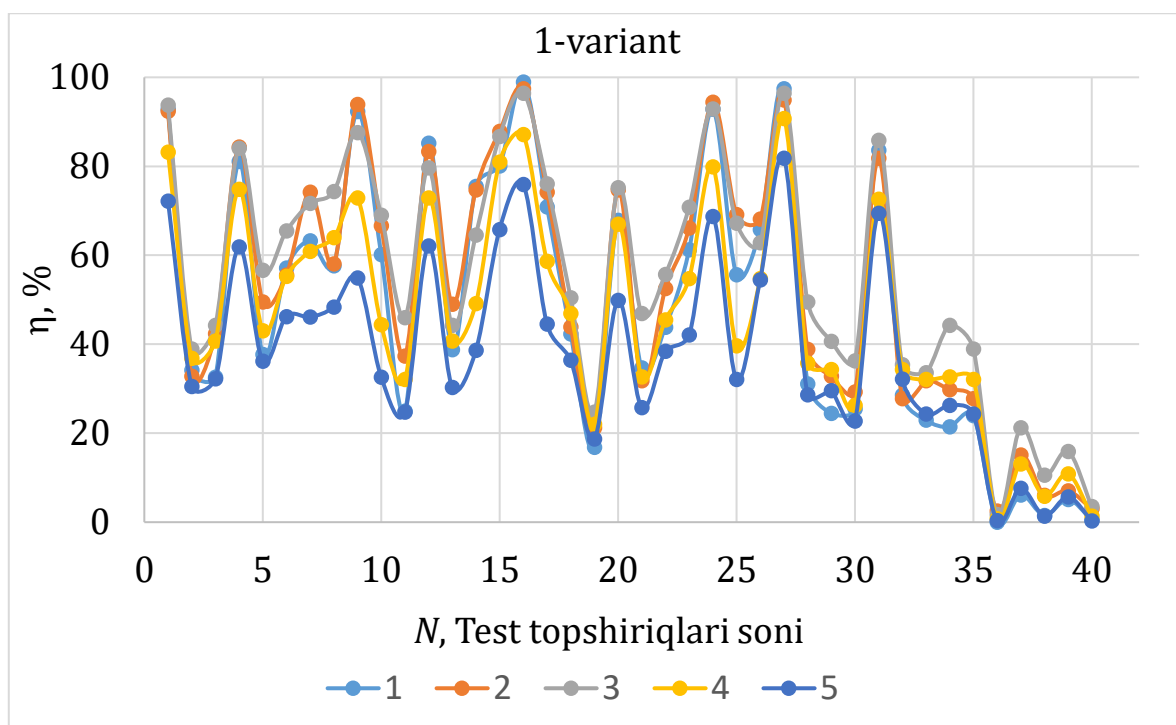
bu esa ularning tayyorgarlik darajasini oshirishga xizmat qilgan.

Aynan 25–35 yosh oralig'i optimal bilimlarni sinov sharoitida namoyon etish uchun qulay davr bo'lishi ham ehtimoldan xoli emas.

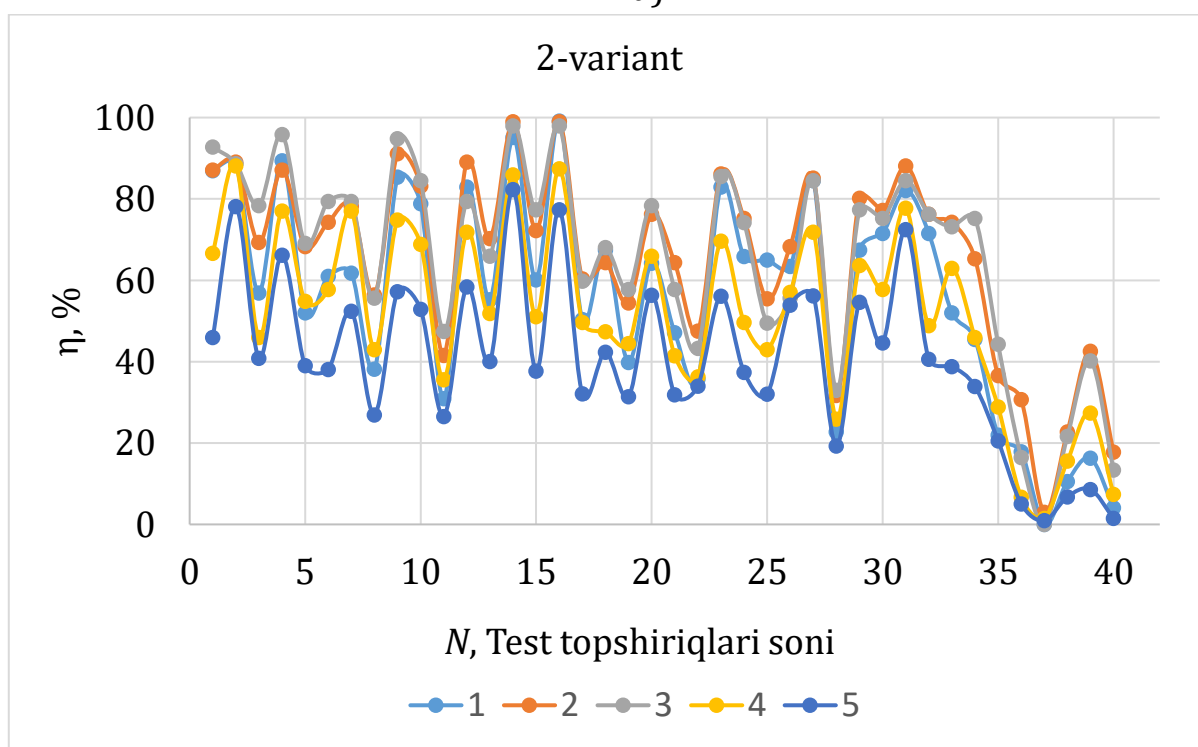
20 yoshgacha bo'lgan ishtirokchilar test topshiriqlarida eng past natijalarni qayd etganligi, ularning fanga oid bilimlari hali yetarli darajada shakllanmagan bo'lishi yoki testga tayyorgarlik darajasi past ekanligini ko'rsatadi.

Barcha variantlarda yoshga xos tafovutlar kuzatilgani test topshiriqlari mazmunining muayyan yosh guruhlariga nisbatan murakkabroq yoki mosroq bo'lganligini ham ko'rsatishi mumkin. Bu esa test topshiriqlarini ishlab chiqishda yoshga moslik masalasiga e'tibor qaratish zarurligini anglatadi.

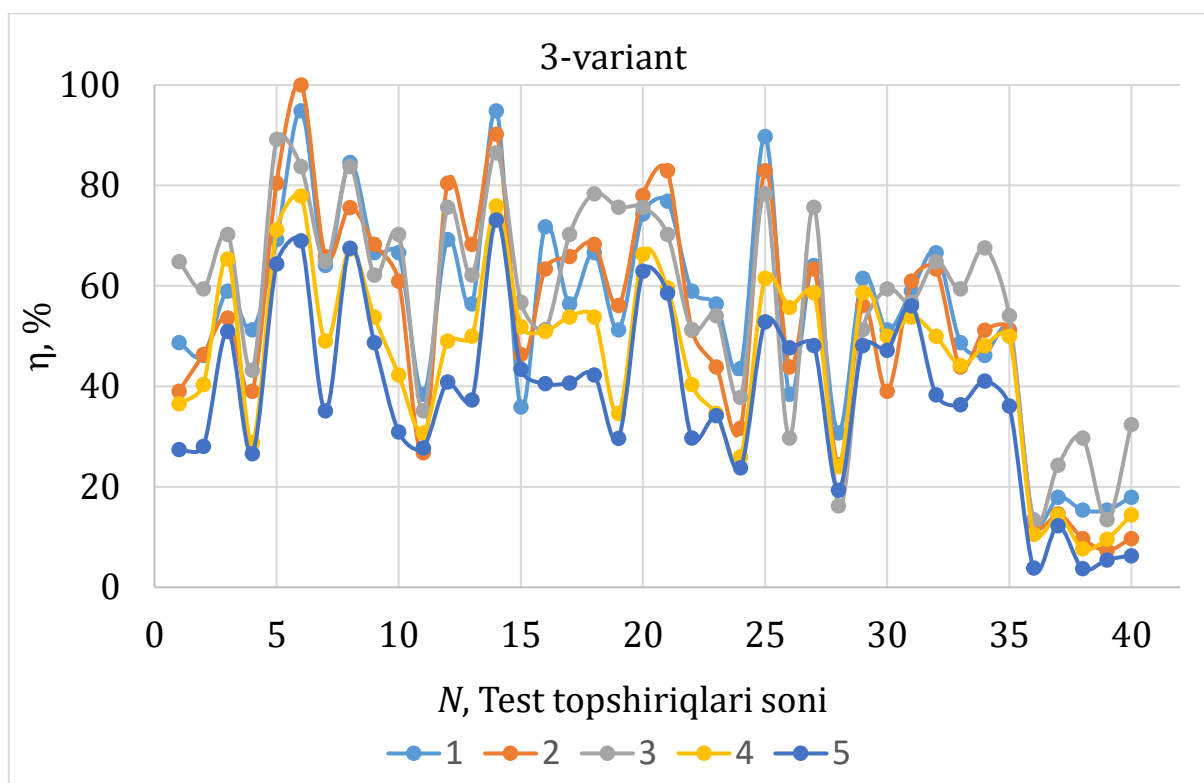
Bunday holat yosh va bilim o'zlashtirish darajasi o'rtasidagi bog'liqlik mavjudligini tasdiqlaydi.



a)



b)



c)

1–35 yoshdan kata, 2–30-35 yosh oralig'i 3–25-30 yosh oralig'i
 4–20-25 yosh oralig'i 5–20 yoshdan kichik

4-rasm. Yosh kesimida talabgorlarning test topshiriqlariga
 to'g'ri javob berish ulushi (foiz)

Xulosa

Ushbu statistik tahlillar natijalari asosida talabgorlarning kimyo fanining fan dasturidagi o'zlashtirilishi past bo'lgan tematik mavzulariga yetarli darajada e'tibor qaratilishi, ta'lim jarayonida mavjud bo'lgan uslubiy kamchiliklar va tizimli muammolar bartaraf etilishi zarurdir.

Kimyo fanidan milliy sertifikat imtihonlarida o'zbek, rus va qoraqalpoq tillarida ishtirok etgan talabgorlarning natijalari asosida olib borilgan statistik tahlil natijalariga

ko'ra, qoraqalpoq tilida test topshirgan talabgorlar natijalarining o'rtacha ko'rsatkichlari yuqori bo'lishi qoraqalpoq tilida ishtirok etgan talabgorlar sonining nisbatan kamligi bilan izohlanishi mumkin.

Hududlar kesimida olib borilgan statistik tahlil natijalariga asosan Qoraqalpog'iston Respublikasi, Buxoro, Namangan, Jizzax va Toshkent viloyatlari va Toshkent shahrida ishtirok etgan talabgorlar natijalarining ko'rsatkichlari yuqori va

Andijon, Surxondaryo va Samarqand viloyatlarida ishtirok etgan talabgorlar natijalrining ko'rsatgichlari pastligi aniqlandi.

O'tkazilgan statistik tahlillar natijalariga ko'ra 25–35 yosh oralig'idagi ishtirokchilarning natijalari boshqa yosh guruhlariga nisbatan yuqoriligi aniqlandi. 25–35 yoshdagi talabgorlar oliy ma'lumotli, faol kasbiy

faoliyat bilan shug'ullanayotgan, yuqori kognitiv ko'nikmaga ega ekanligi bilan izohlash mumkin. Aksincha 20 yoshgacha bo'lgan talabgorlarning natijalari boshqa yosh guruhlariga nisbatan past ekanligi bu ularning fanga oid bilimlari hali yetarli darajada shakllanmaganligi yoki testga tayyorgarlik darajasi past ekanligini ko'rsatadi.

ADABIYOTLAR

1. Black P., Wiliam D. Assessment and classroom learning, Assessment in Education: principles, policy & practice. 1998. T. 5. №. 1. S. 7-74.
2. Pellegrino J. W., Chudowsky N., Glaser R. The nature of assessment and reasoning from evidence, Knowing what students know: The science and design of educational assessment. 2001. S. 37-54.
3. Bybee R. W. et al. The BSCS 5E instructional model: Origins and effectiveness, Colorado Springs, Co: BSCS. 2006. T. 5. №. 88-98.
4. Taber K. Chemical misconceptions: prevention, diagnosis and cure. – Royal Society of Chemistry, 2002. – T. 1.
5. Ermamatov M.Dj., Baratov A.A., Mirvaliyev Z.Z., Normurodov A.B., Sulaymonov A.A. Ta'lim tizimida bilimlarni baholashda Blum taksonomiyasining o'rni, Axborotnoma, 1-2, 2022, 4-10 b.
6. Bloom, B.S.; Engelhart, M.D.; Furst, E.J.; Hill, W.H.; Krathwohl, D.R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitiv domain. New York: David McKay Company.
7. Q.A. Amonov, A.A. Baratov. Blum taksonomiyasi asosida fizikadan test topshiriqlarini shakllantirish, Axborotnoma, 2, 2023, 84-96 b.
8. Anderson, L. W., Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
9. Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.
10. Sattiyev A.R., Ermamatov M.Dj. Rayt xaritasi, ichki va tashqi moslik statistikalari: kimyo fanidan milliy sertifikat uchun o'tkazilgan test sinovi natijalari, Axborotnoma, 1, 2024, 18-96 b.
11. Ermamatov M. Dj. Rash modelida test topshiriqlari qiyinlik darajalarining invariantligi, Axborotnoma, 4, 2023, 4-11 b.
12. Stiggins R., Chappuis J. Using student-involved classroom assessment to close achievement gaps, Theory into practice. 2005. T. 44. №. 1. S. 11-18.
13. Haladyna T. M., Downing S. M. Construct-irrelevant variance in high-stakes testing, Educational Measurement: Issues and Practice. 2004. T. 23. №. 1. S. 17-27.
14. Pellegrino J. W., Chudowsky N., Glaser R. The nature of assessment and reasoning from evidence, Knowing what students know: The science and design of educational assessment. 2001. S. 37-54.

15. Gilbert J. K. Multiple representations in chemical education. – Dordrecht: Springer, 2009. T. 4. S. 1-8.
16. Taber K. Chemical misconceptions: prevention, diagnosis and cure. – Royal Society of Chemistry, 2002. T. 1.
17. <https://uzbmb.uz/page/fanlar>.
18. Baker, Frank (2001). The Basics of Item Response Theory, ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.
19. Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991), Fundamentals of item response theory. Newbury Park, CA: Sage.
20. Rasch G. (1960), Probabilistic models for some intelligence and attainment tests, Copenhagen, Danish Institute for Educational research.
21. Wright, B. D., & Stone, M. H. (1979). Best test design. Chicago, IL: Mesa Press.
22. Ermamatov M.Dj., Sattiyev A.R., Normurodov A.B., Olimbekov Z.O., Baratov A.A. Fizika fanidan o'tkazilgan test sinovi natijalari: rayt xaritasi, ichki va tashqi moslik statistikalari, rash modeli bilan moslik, Axborotnoma, 1, 2023, 4-62 b.
23. Rasch G. "An item analysis which takes individual differences into account." British journal of mathematical and statistical psychology 19.1 1966, 49-57.
24. Ermamatov M.Dj., Alimov M.D., Sulaymonov A.A., Sattiyev A.R. Kalibrovkalangan test topshiriqlari: Sharq tillaridan o'tkazilgan test sinovi natijalarining statistik tahlili, Axborotnoma, №. 3-4, 16-83 b., 2022.
25. Ermamatov M.Dj., Abbosov A., Baratov A.A. Test topshiriqlarini kalibrovkalash va qobiliyatlarini tenglashtirish, Axborotnoma, №. 3-4, 4-16 b., 2022.
26. Ermamatov M.Dj., Sattiyev A.R., Normurodov A.B., Olimbekov Z.O., Baratov A.A. Fizika fanidan o'tkazilgan test sinovi natijalari: Rayt xaritasi, ichki va tashqi moslik statistikalari, Rash modeli bilan moslik, Axborotnoma, №1, 2023, 4-62.

STATISTICAL ANALYSIS OF CHEMISTRY TEST RESULTS ON THE BASIS OF DETAILS

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. This article analyzes the results of the national chemistry test for the Kazakh National Certificate in Chemistry by region, age, and language. It was found that the results of participants aged 25–35 were higher than those of other age groups, while the results of participants under 20 were lower than those of other age groups. It was studied which thematic topics correspond to the test tasks with a high level of difficulty, based on the test details.

Keywords: inorganic chemistry, organic chemistry, cognitive skills, Rash model, difficulty levels, test details.

YOZISH VA GAPIRISH KO'NIKMALARINI BAHOLOVCHI MUTAXASSISLARNI TAYYORLASH TIZIMI

M.P. Abdullaeva

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

Qisqacha mazmuni. Ushbu maqolada chet tillar bo'yicha yozish va gapirish ko'nikmalari uchun ekspertlarni masofaviy tayyorlash va ularni Bilimlarni baholash agentligining Chet tillarni bilish va egallash darajasini baholash boshqarmasi tomonidan tashkil etiladigan test sinovlarining yozish va gapirish ko'nikmalarini baholash jarayoniga jalb etish bo'yicha metodologiya tahlil qilingan. Maqolada chet tillar bo'yicha fan mutaxassislari bo'lgan nomzodlarni tanlab olish bosqichlari, ularning malakasini oshirish mexanizmi va shuningdek umuman bilimni baholash jarayonining sifatini oshirish va xolisligini ta'minlash choralariga e'tibor qaratilgan.

Kalit so'zlar: bilimni baholovchi mutaxassis, mutaxassislarni tayyorlash, baholash mezonlari, masofaviy ta'lim shakli.

Kirish

Baholash valid va ishonchli bo'lishi uchun sifatli test topshiriqlarini ishlab chiqish bilan birga natijalarni to'g'ri baholash ham muhimdir, ayniqsa yozish va gapirish kabi ko'nikmalarni baholash, bilimni baholashning eng muhim bosqichidir. Bilimlarni baholash vositalarining amaliyotda to'g'ri va xolis qo'llanilishi esa, o'z navbatida, malakali ekspertlarning ishtirokini taqozo etadi. Ayniqsa, masofaviy ta'lim texnologiyalar jadal rivojlanib borayotgan bir davrda, bilim va ko'nikmalarni masofadan turib adolatli baholay oladigan mutaxassislarni tayyorlash zarurati ortib bormoqda.

Bilimlarni baholovchi mutaxassislarni tayyorlash jarayoni ishonchli, adolatli va samarali imtihon tizimini shakllantirishda muhim rol o'ynaydi [1] va bu bosqich chet tilini bilish darajasini baholashning ajralmas qismini tashkil etadi. Bilimni baholashning yuqori aniqligi va izchilligini ta'minlash uchun baholovchi mutaxassislarning bu sohadagi bilim va malakalarini oshirish zarur ekanligi keng e'tirof etilgan [2]. Ushbu maqolada, avvalo, bilimni baholaydigan mutaxassislarni tayyorlash bo'yicha tarixiy jarayonlarga qisqacha to'xtalib o'tilgan. Shuningdek, bilimni baholovchi mutaxassislarni tayyorlash jarayoniga hozirgi kunda qanday

yondashilayotgani, ya'ni ekspert malakasi, monitoring ishlarining ishonchliligi va baholashning ijtimoiy

amaliyot sifatida qaralishi nuqtayi nazaridan qanday yondashuvlar mavjudligi yoritib beriladi.

I. Chet tilini bilish darajasini baholash bo'yicha baholovchi mutaxassislarni tayyorlash kursiga nomzodlarni jalb qilish va saralash

Chet tilini bilish darajasini baholash bo'yicha mutaxassislar malakasini oshirish O'zbekiston Respublikasi Vazirlar Mahkamasining 2020-yil 17-sentabrdagi "Test tizimini takomillashtirishga qaratilgan qo'shimcha chora-tadbirlar to'g'risida"gi 562-sonli qarorida belgilangan vazifalarga muvofiq amalga oshirilmoqda. Ushbu qarorda baholash va pedagogik o'lchovlar sohasidagi tadqiqotlarni rivojlantirish, shuningdek chet tillari bo'yicha malaka oshirish kurslarini tashkil etish, yetakchi xalqaro tashkilotlar bilan hamkorlikni yo'lga qo'yish va Markaziy Osiyoda baholash bo'yicha mintaqaviy ilmiy maktab yaratish ko'zda tutilgan.

Mazkur topshiriqlarni amalga oshirish maqsadida, O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi huzuridagi **Bilimlarni baholash agentligi** huzurida "**Ilmiy-o'quv amaliy markazi**" davlat muassasasi tashkil etilgan. **Ilmiy-o'quv amaliy markazi** tomonidan **malaka oshirish kurslari yo'lga qo'yilgan** bo'lib, ularning doirasida chet tillar bo'yicha test topshiriqlarini shakllantirish, pedagogik o'lchovlar va bilimlarni baholash

uchun mutaxassislarni tayyorlash bo'yicha tizimli faoliyat olib borilmoqda.

Chet tilini bilish darajasini baholovchi ekspertlarni tayyorlash jarayoni bosqichma-bosqich va tizimli ravishda amalga oshiriladi. Dastlabki bosqich bu nomzodlarni **ochiq e'lon asosida jalb qilish** va **aniq belgilangan mezonlar asosida saralashdir** [3].

Kursga nomzodlarni jalb qilish bo'yicha e'lonlar va videolavhalar tashkilotning rasmiy axborot kanallari va veb-saytlari orqali keng ommaga taqdim etiladi. E'lon matnida esa quyidagi **asosiy talablar aniq ko'rsatiladi**:

- nomzod o'z sohasida **mutaxassis** bo'lib, tegishli darajadagi diplomga ega bo'lishi (bakalavr, magistr va boshqa);

- C1 darajadagi sertifikat (Milliy sertifikat, IELTS, TOEFL va boshqa xalqaro tan olingan sertifikatlardan biri) orqali chet tilini bilish darajasini tasdiqlay olishi.

Shuningdek, e'londa chet tillarni bilish darajasini aniqlash bo'yicha yozish va gapirish ko'nikmalarini baholash uchun malaka oshirish

kursini muvaffaqiyatli tamomlaganligi to'g'risida sertifikatga ega bo'lgan taqdirda u Bilimlarni baholash agentligida ekspert bo'lib faoliyat yuritishi mumkinligi ma'lum qilinadi [4].

E'lon orqali kurs mashg'ulotlari o'tkaziladigan **sana va davomiyligi** ham ko'rsatiladi. E'lon bo'yicha nomzodlar belgilangan <https://my.uzbmb.uz/ilmiy-service> **elektron tizim orqali ro'yxatdan o'tadi.** Bu yondashuv Respublikamizning barcha hududlaridan bilimli va xolis nomzodlarni saralab olish, manfaatlar to'qnashuvining oldini olish imkonini beradi.

Ro'yxatdan o'tgan barcha nomzodlarga **Zoom platformasi orqali masofaviy mashg'ulotlarda ishtirok etish uchun havola** SMS xabar shaklida yuboriladi. Kurs **36 akademik soatdan iborat** bo'lib, u **asinxron va sinxron** uslublarda olib boriladi. Bu esa tinglovchilarning ta'lim jarayoniga moslashuvchanligini ta'minlab, tinglovchilarga individual o'rganish sur'atini belgilash imkonini beradi.

Mashg'ulotlar quyidagicha tashkil etiladi:

- **nazariy materiallar** jonli efirda yuqori malakali mutaxassislar tomonidan taqdim etiladi. Darslar interaktiv uslubda o'tkazilib, ishtirokchilarning faol muloqotda bo'lishi rag'batlantiriladi.

- amaliy mashg'ulotlar esa Zoom va Canvas platformalari orqali olib boriladi. Canvas platformasi orqali tinglovchilarga mustaqil bajarilishi lozim bo'lgan topshiriqlar, test namunalarini tahlil qilish, yozma va og'zaki javoblarni baholash kabi amaliy topshiriqlar berib boriladi hamda shu platformaning o'zida ularning bajargan ishlari tekshirib boriladi;

- har bir modul yakunida tinglovchilarning fikr-mulohazalari o'rganiladi. Bu esa mashg'ulotlar sifatini baholash va zarur hollarda o'quv strategiyasini takomillashtirish imkonini beradi;

- har bir uy vazifasi va amaliy topshiriq Bilimlarni baholash agentligi Chet tillarni bilish va egallash darajasini baholash boshqarmasining bosh mutaxassislari tomonidan baholanadi, baholash natijalari esa ishtirokchilarning shaxsiy portfoliolari shakllanishiga xizmat qiladi.

Kursning asosiy ustunligi – **zamonaviy pedagogik texnologiyalar, baholash va pedagogik o'lchovlar tamoyillariga asoslanganligidir.** Sinxron va asinxron mashg'ulotlarning kombinatsiyasi egallangan bilimlarni o'zlashtirishda auditoriyalarga moslashishga yordam beradi. Ayniqsa, Canvas **platformasidan** foydalanish orqali:

- **fikr va mulohaza (feedback)** tizimi yo'lga qo'yiladi;
- har bir ishtirokchining o'zlashtirish darajasini yakka tartibda

monitoring qilish imkoniyati yaratiladi;

- vaqtdan foydalanib mustaqil ta'lim olish imkoniyati kengayadi;
- portfolio orqali ishtirokchilarning natijalari va ko'rsatkichlari raqamli platformada saqlanadi.

Kurs doirasidagi mashg'ulotlar interaktiv yondashuv asosida loyihalashtirilgani zamonaviy pedagogik tamoyillarga, xususan, konstruktivistik o'quv modeliga tayanadi. Bu modelga ko'ra, ta'lim jarayonida tinglovchilar faqat tayyor nazariy bilimlarni egallab qolmasdan, balki amaliy ko'nikma orqali chuqur tushunchalarga ega bo'ladi. Bu yondashuv xalqaro ekspertlar tavsiyalariga mos keladi va baholash sohasida **ko'p bosqichli, adolatli va obyektiv tizimni** yaratishga xizmat qiladi. Mazkur tamoyillar Amerika ta'lim tadqiqotchilari assotsiatsiyasi (AERA), Psixologlar assotsiatsiyasi (APA) hamda O'lchov va baholash milliy kengashi (NCME) tomonidan ishlab chiqilgan "Standards for Educational and Psychological Testing" (2014) [5] hujjatida aniq belgilangan.

Bilimlarni baholashda adolat va obyektivlik esa test topshiriqlarining standartlashtirilgan mezonlarga asoslanishi, ekspertlar malakasini doimiy ravishda oshirib borilishi hamda test natijalarini tahlil qilishda statistik vositalardan foydalanilishi orqali ta'minlanadi [6].

Bundan tashqari, interaktiv metodlar xalqaro ekspertlar tomonidan ta'lim samaradorligini oshirish, o'rganilgan bilimlarning real hayotdagi qo'llanilishini ta'minlash vositasi sifatida e'tirof etilgan (OECD, 2013). Bu yondashuv baholash sohasidagi ko'p bosqichli baholash tizimini shakllantirishda muhim ahamiyat kasb etadi. U adolatli, ishonchli va obyektiv baholashni ta'minlash uchun zarur bo'lgan amaliy ko'nikmalarni va tinglovchilarning dars jarayonidagi erkin fikrlashlarini rivojlantiradi (AERA, APA, NCME, 2014).

Kurs yakunida tinglovchilarga real baholash ko'nikmalarini amalda sinab ko'rish uchun **20 ta yozma ish (writing) va 20 ta og'zaki nutq (speaking) audio materiallari** taqdim etiladi. Bu materiallar **Canvas platformasida joylashtiriladi**. Tinglovchilar kurs davomida olgan bilim va ko'nikmalari asosida, **belgilangan baholash mezonlariga** tayangan holda yozma va og'zaki materiallarni mustaqil baholaydilar. Ular tomonidan baholangan natijalar **Bilimlarni baholash agentligi Chet tillarni bilish va egallash darajasini baholash boshqarmasining bosh mutaxassislari** tomonidan tekshirilib, monitoring o'tkaziladi va tahlil qilinadi. Yozma va og'zaki materiallarni baholash mezonlariga to'liq mos kelgan, xolis va aniqlik bilan tekshirgan nomzodlarga **baholovchi ekspert sifatida** faoliyat yuritishi uchun **taklif**

etiladi [7]. Shuningdek, har bir chet tillari imtihonidan so'ng, ekspertlar faoliyatini baholash bo'yicha aniq va izchil monitoring tizimi joriy etilgan [8]. Bunda test sinovida ishtirok etgan talabgorlarga tegishli bo'lgan 20 tadan yozma ishlari va gapirish ko'nikmalari bo'yicha audio materiallari Chet tillarni bilish va egallash darajasini baholash boshqarmasi bosh mutaxassislari tomonidan baholanadi. Ushbu baholanagan ishlar ekspertlarni monitoring qilish maqsadida ularga taqdim qilinadi. Ular baholagan ishlar bilan bosh ekspertlar baholagan ishlar o'rtasida tafovut kuzatilganda bu ishlar chuqur tahlil qilinadi.

Tafvut sabablarini aniqlashda quyidagi holatlar e'tiborga olinadi:

- baholovchining loqaydlik bilan yondashganligi;
- yozma ishni to'liq o'rganmaganligi;

- audio materialni oxirigacha tinglamaganligi;
- baholash mezonlari bo'yicha bilimi yetishmasligi va h.k.

Mazkur omillar asosida har bir xato dalillar bilan asoslab beriladi va zarurat tug'ilganda, ekspert qayta malaka oshirish kursiga yuboriladi. Qayta o'qitish jarayoni Ilmiy markaz tomonidan 24 soatlik o'quv dasturi asosida Zoom va Canvas platformalari orqali tashkil etiladi. Bu kurslar ekspertlarning baholash bo'yicha malakasini oshirish, me'yoriy hujjatlar va mezonlarga asoslangan holda ishlash malakasini mustahkamlashga xizmat qiladi.

Shu tarzda, kursdan keyingi monitoring va takroriy malaka oshirish kurslari orqali ekspertlar tomonidan test topshiruvchilarning natijalarini tizimli va sifatli baholanishi ta'minlanadi.

II. Chet tilini baholash bo'yicha mutaxassislarni tayyorlash kursi samaradorligi

Chet tillarini baholashda ishonchli, adolatli va professional yondashuvni ta'minlash, zamonaviy ta'lim tizimi oldida turgan eng muhim vazifalardan biridir. Shu maqsadda 2021-yildan boshlab chet tilini bilish darajasini baholash bo'yicha mutaxassislarni tayyorlashga ixtisoslashgan malaka oshirish kurslari muntazam yo'lga qo'yildi.

Bugungi kunga qadar ushbu kurslarda 1400 nafarga yaqin chet tili mutaxassislari bo'lgan tinglovchilar ishtirok etgan bo'lib, ularning 700 nafaridan ortig'i yakuniy baholashdan muvaffaqiyatli o'tib, baholovchi ekspert sifatida tanlab olingan.

Kurslar yakunida tinglovchilar o'rtasida baholash mezonlari, kurs mazmuni va amaliy qo'llanilishi bo'yicha doimiy ravishda monitoring

olib boriladi. So'nggi o'tkazilgan so'rovnoma 274 nafar ishtirokchi qatnashib, o'z fikr va takliflarini bildirgan. Quyida aynan shu so'rovnoma natijalari asosida kursning samaradorligi va uning bilimlarni baholash jarayoniga ta'siri tahlil qilindi.

Ilmiy markaz tomonidan tashkil etilgan malaka oshirish kurslarida Respublikaning barcha hududlaridan tajribali mutaxassis ishtiroki ta'minlandi.

Malaka oshirish kurslarini muvaffaqiyatli tamomlagan ishtirokchilar o'rtasida o'tkazilgan so'rovnoma natijalari asosida kursning samaradorligi, ishtirokchilarning malaka oshirish kursidan qoniqish darajasi hamda kelgusidagi faoliyatini rivojlantirish bo'yicha muhim omillarni aniqlashga yordam berdi.

So'rovnoma ishtirok etgan ekspertlarning faoliyat sohasi o'rganilganda ularning qaysi ta'lim turida faoliyat yuritayotgani aniqlashtirildi (1-rasm).

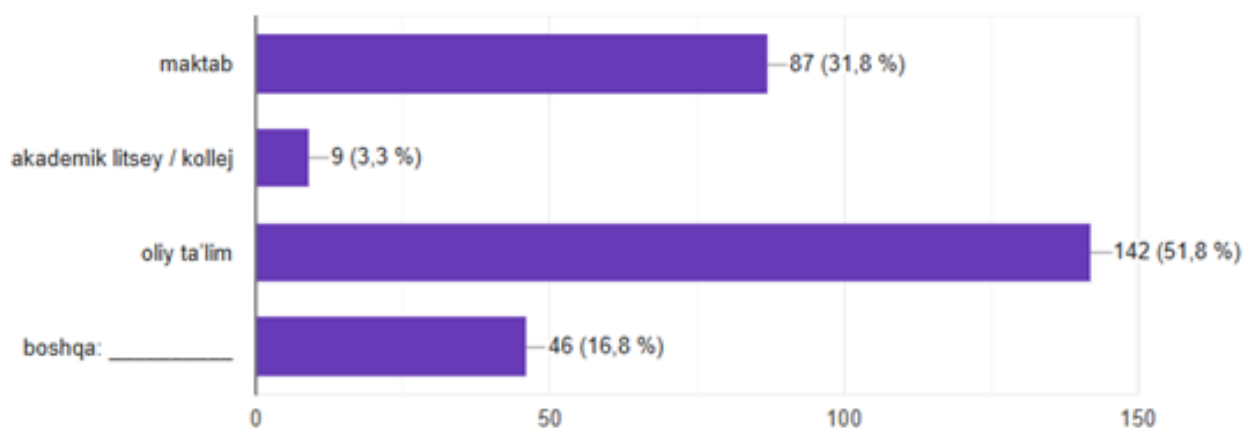
Oliy ta'lim muassasalarida faoliyat olib boruvchi ishtirokchilar soni **142 nafar (51,8 foiz)**ni tashkil etdi. Bu eng katta guruh bo'lib, bundan oliy ta'lim tizimidagi pedagoglarning malaka oshirish kurslariga bo'lgan ehtiyoji yuqoriligi ko'rinadi. Ularning aksariyati test topshiriqlarini shakllantirish, bilimni baholash va

natijalarni tahlil qilish bilan bevosita shug'ullanadilar.

Maktab o'qituvchilari 87 nafar ishtirokchi (31,5 foiz) bilan ikkinchi o'rinda turadi va ular aynan umumiy o'rta ta'lim tizimida ishlashini bildirganlar. Bu holat umumiy o'rta ta'lim tizimidagi bilimni baholash sohasidagi savodxonlikni kuchaytirishga ehtiyoj borligini ko'rsatadi.

Boshqa toifadagi 46 nafar (16,8 foiz) ishtirokchi o'z faoliyat sohasini noan'anaviy yoki aralash shakllarda ko'rsatganlar. Bunga mustaqil ta'lim markazlari, nodavlat o'quv kurslari yoki xalqaro ta'lim loyihalarida ishlovchi mutaxassislar kiradi. **Akademik litsey yoki kollejlarda** ishlovchilar esa **9 nafar (3,3 foiz)** bo'lib, bu guruh so'rovnoma ishtirokchilari orasida eng kichik ulushni tashkil qiladi.

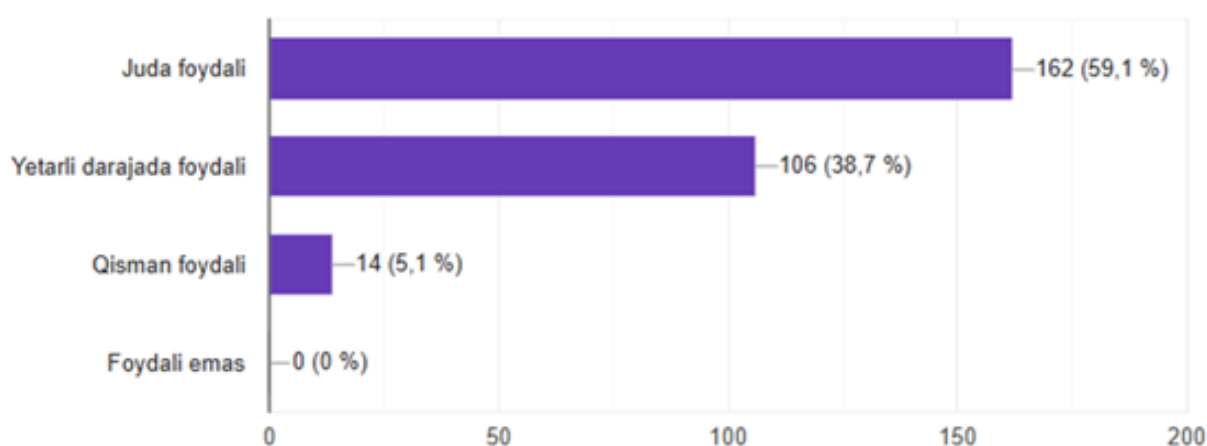
Mazkur tahlil natijalariga ko'ra, ekspertlarning katta qismi oliy ta'lim va umumiy o'rta ta'lim tizimida faoliyat yuritayotganligi aniqlandi. Bu esa malaka oshirish kurslarini rejalashtirishda aynan ushbu ikki yo'nalish, jumladan oliy va umumiy o'rta ta'lim uchun mos modullarni ishlab chiqish zarurligini anglatadi. Shu bilan birga, boshqa sohalarda faoliyat yuritayotgan mutaxassislar ham e'tibordan chetda qolmasligi kerak, ular uchun ham moslashtirilgan o'quv dasturlar ishlab chiqilishi lozim.



1-rasm. So'rovnomada ishtirok etgan ekspertlarning faoliyat sohasi o'rganilgan diagramma

So'rovnomada qatnashganlarning 59,1 foizdan ortig'i (**162 nafar**) kursning nazariy qismini **"juda foydali"**, 38,7 foiz (106 nafar) ishtirokchi esa **"yetarli darajada foydali"** deb baholagan. Faqat 14 nafar ishtirokchi bu qismni **"qisman foydali"** deb topgan. Bu ko'rsatkichlar (2-rasm) nazariy modulning tinglovchilarga aniq va foydali tarzda yetkazilganini anglatadi. Umuman

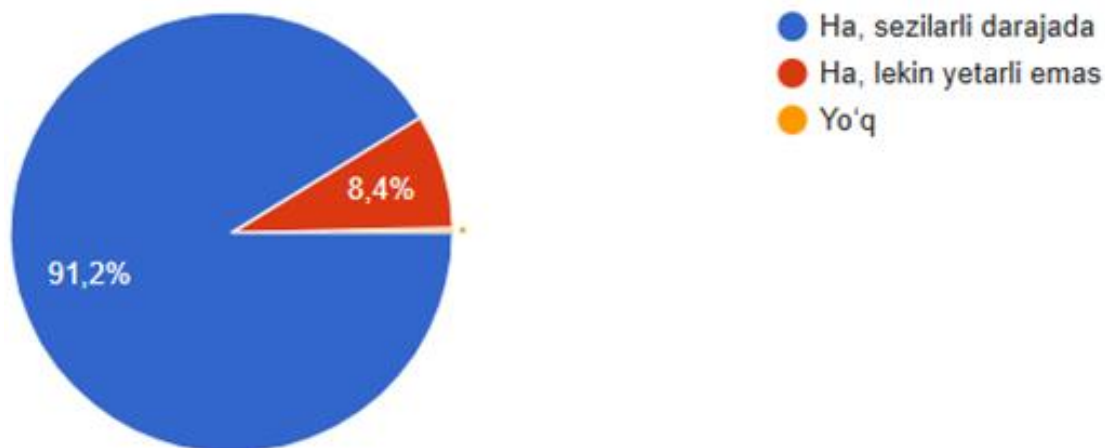
olganda, kursning nazariy qismi yuqori darajada samara berganligi, ko'pchilik ishtirokchilarda ijobiy taassurot qoldirganligi aniq bo'ldi. Bu esa kelgusidagi kurslarni taskil etishda kursning nazariy qismini saqlab qolish, uni yanada boyitish, va zarur hollarda turli darajadagi ishtirokchilar uchun **innovatsion pedagogik texnologiyalarni** qo'llash zarurligini ko'rsatadi.



2-rasm. Malaka oshirish kursining mazmuni va metodikasi haqida ekspertlarning fikrlari aks etgan diagramma

Ishtirokchilarning **91,2 foizdan ortig'i** (274 nafar) amaliy mashg'ulotlar (test topshiriqlarini shakllantirish, baholash, takliflar berish) **sezilarli darajada ko'nikmalarini oshirishga yordam berganligini** ta'kidlaganlar. Atigi 8,4 foiz (23 nafar)

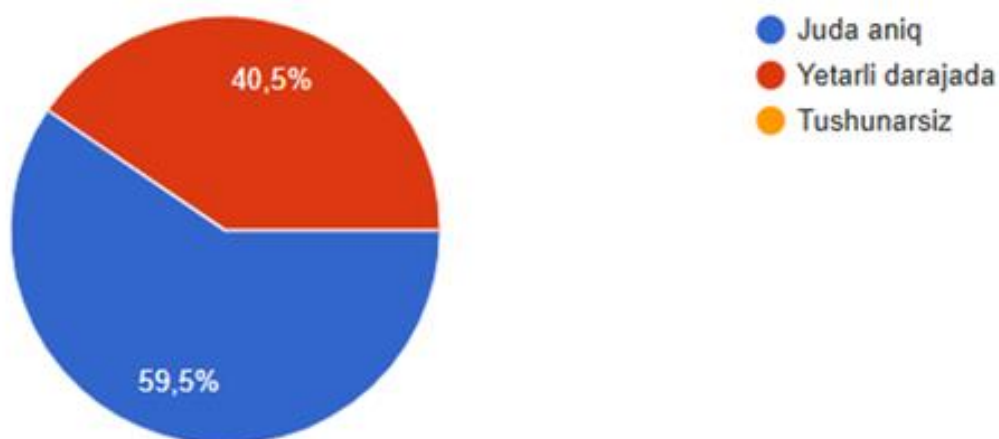
ishtirokchilar amaliy mashg'ulotlar foydali bo'lganini bildirgan, biroq ularning fikricha, bu yetarli darajada bo'lmagan, ya'ni ular yanada ko'proq mashq, real holatlarga asoslangan vazifalar yoki tahliliy muhokamalar bo'lishi kerakligini istaganlar (3-rasm).



3-rasm. Amaliy mashg'ulotlarning (bilimni baholash) ko'nikmalarni rivojlantirishga ta'siri haqidagi ishtirokchilarning fikrlari

Ekspertlarning 59,5 foizi (166 nafar) topshiriqlar **“juda aniq”**, 59,5 foizi (105 nafar) esa **“yetarli darajada”** tushuntirilganligini bildirganlar. Ushbu ko'rsatkichlar ekspertlarning vazifalarni qanday tushunganiga bevosita bog'liq bo'lib, kurs ishtirokchilarining mustaqil ishlash

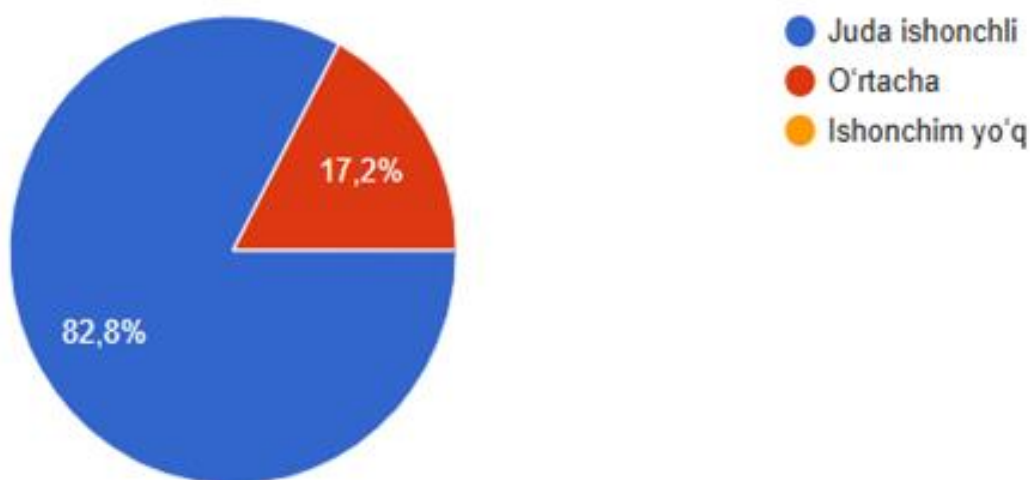
salohiyatini ko'rsatadi. Bu esa ularda mustaqil test tuzish va baholash mezonlarini qo'llash hamda vazifalarni ijodiy yondashuvlar bilan bajarish ko'nikmalari shakllanganligini anglatadi (4-rasm).



4-rasm. Mustaqil vazifalarning (test topshiriqlarini shakllantirish, yozish va gapirish ko'nikmalarini baholash) aniqlik darajasi haqidagi ishtirokchilarning fikrlari

Yozish va gapirish kabi produktiv ko'nikmalarni baholashga doir savollarga javoban 82,8 foiz (227 nafar) ishtirokchilar ushbu ko'nikmalarni **“juda ishonchli”** tarzda qo'llay olishini bildirganlar. Bu holat kurs

davomida baholash mezonlari, baholash jarayonining bosqichlari hamda baholovchi sifatida ishlash bo'yicha tinglovchilarga yetarli nazariy bilim va amaliy tayyorgarlik berilganini yaqqol tasdiqlaydi (5-rasm).



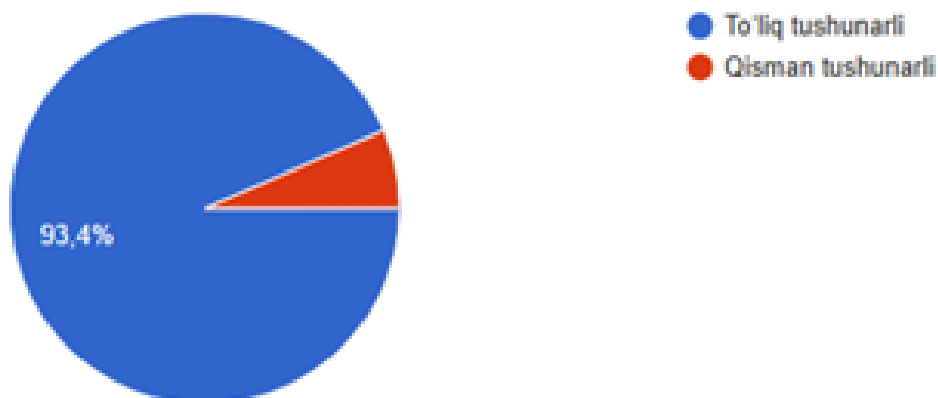
5-rasm. Yozma va og'zaki ko'nikmalarni baholash bo'yicha o'zini ishonchli his qilgan ishtirokchilarning ko'rsatkichi

93,4 foiz (256 nafar) ishtirokchilar baholash mezonlari va xolislikni ta'minlash mexanizmlarini **“to'liq**

tushunganini”, 6,6 foizi (18 nafar) esa **“qisman tushunganini”** bildirganlar. Ushbu natijalar baholashda foyda-

laniladigan topshiriqlar, xolislik va adolat tamoyillariga oid mavzular kurs davomida yetarli darajada yoritilganligini va ishtirokchilarda aniq

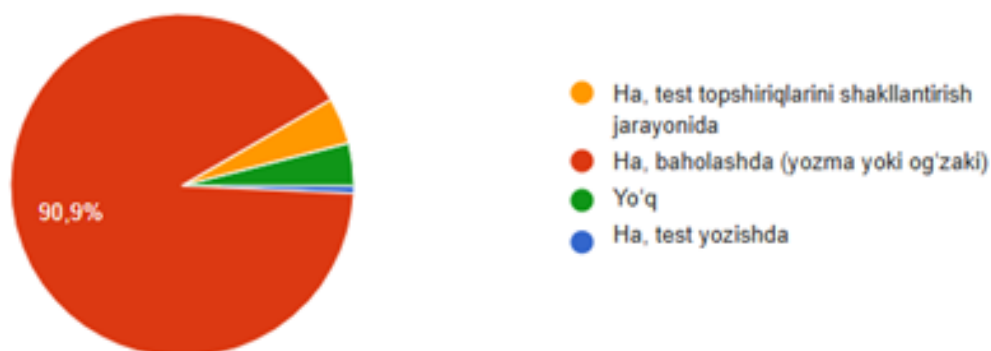
tushuncha shakllanganligini ko'rsatadi. Bu holat kurs materiallari mazmunan samarali bo'lganligini tasdiqlaydi (6-rasm).



6-rasm. Baholash mezonlari va xolislik tamoyillarining tushunilishi

So'rovnoma natijalariga ko'ra, kurs yakunlangach, ishtirokchilarning aksariyati, 90,9 foizi (**249 nafar**) o'z hududlarida yozma va gapirish ko'nikmalarni baholash jarayonlarida **faol ishtirok etganligini** bildirganlar.

Bu esa malaka oshirish kursi davomida o'zlashtirilgan nazariy bilim va amaliy ko'nikmalar tinglovchilar tomonidan **real test jarayonlarida muvaffaqiyatli qo'llanilganligini** ko'rsatadi.



7-rasm. Kursdan so'ng tinglovchilarning bilimni baholash jarayonidagi ishtiroki

Shuningdek, **ishtirokchilarning 4,0 foizi (11 nafar) test topshiriqlarini shakllantirish jarayonlarida** qatnashganligini ta'kidlagan

bo'lsa, ularning 0,7 foizi (**2 nafar**) esa test topshiriqlarini tahlil qilishda ishtirok etganligini bildirganlar. Ishtirokchilarning 4,4 foizi (**12 nafar**) hali

baholash yoki test tuzish bo'yicha ishtirok etmaganligini bildirganlar. Bu guruh uchun alohida kuzatuv yoki qayta ko'rib chiqish maqsadga muvofiqdir.

Umuman olganda, ushbu natijalar kursdan keyingi **real jarayonga ulanish (transfer of training)** yuqori darajada bo'lganini ko'rsatadi, bu esa kursning amaliy yo'naltirilganligini va samaradorligini yana bir bor tasdiqlaydi.

Malaka oshirish kurslari tinglovchilarining fikr va mulohazalari asosida quyidagi asosiy yo'nalishlar bo'yicha quyidagi:

- **amaliy mashg'ulotlar sonini oshirish;**
- tinglovchilarga **monitoring natijalari tahlili bilan tanishish imkoniyatini yaratish;**

- **jonli (zoom) darslar sonini ko'paytirish;**
- yozma va og'zaki ishlarning **izohli namunalarini ko'rish imkoniyatini yaratish;**
- **platformada savol-javob (Q&A)** bo'limlarini tashkil qilish;
- ekspert tanlovida shaffoflikni oshirish
- **takroriy treninglar va davriy malaka oshirish kurslarini tashkil etish**

kabi takliflarini bildirganlar. Bundan tashqari ayrim tinglovchilar tomonidan nazariy qism yetarli bo'lganligi, kelajakda **amaliyotga ko'proq e'tibor berish** lozimligini bildirganlar. Kelgusidagi malaka oshirish kurslarini tashkil qilish jarayonida mazkur takliflarni ham inobatga olinishi belgilab qo'yildi.

Xulosa

Chet tillarini bilish darajasini baholashda, ayniqsa, yozish va gapirish kabi produktiv ko'nikmalarni baholovchi ekspertlarni masofaviy shaklda o'qitish va ularni tayyorlash jarayoni tizimli ravishda o'rganildi.

Ilmiy-o'quv amaliy markazi tomonidan tashkil etilgan malaka oshirish kurslari zamonaviy texnologiyalar va interaktiv yondashuvlar asosida loyihalashtirilganligini afzalliklari ko'rsatib berildi. Bundan tashqari, monitoring va qayta

tayyorlash mexanizmlarining mavjudligi mutaxassislarning doimo yuqori malakada bo'lishini ta'minlashda, shuningdek, ularning faoliyat sifati va samaradorligini uzluksiz ravishda nazorat qilib borishda muhim omil sifatida xizmat qilishi ko'rsatib berildi.

Umuman olganda, ushbu tajriba O'zbekistonda bilimlarni baholash sohasida zamonaviy, xolis va ishonchli baholash tizimini shakllantirishda muhim omil bo'lib xizmat qilishi ko'rsatib berildi.

ADABIYOTLAR

1. Komil Jalilov *Baholash nazariyasi asoslari (testologiya, pedagogik o'lovlar) Toshkent Akademyashr 2020 (6-bet)*
2. A.A. Abbosov Training raters for assessing productive skills: enhancing accuracy and reliability (28-bet) 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
3. <https://www.britishcouncil.org/exam/english/aptis/research/assessment-literacy>, British council english language assessment research group how language assessment works, an a to z of second language assessment: how language teachers understand assessment concepts edited by christine coombe issn 2516-8649 © british council 2018
4. Bilimlarni baholash agentligi. (2024). Ekspertlarni tayyorlash va baholash metodikasi bo'yicha ichki hujjatlar. Toshkent: BBA.
5. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.
6. Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
7. The effect of rater training on rating behaviors in peer assessment among secondary school students *International Journal of Assessment Tools in Education* 2024, Vol. 11, No. 3, 507–523 <https://doi.org/10.21449/ijate.1438798>
8. <https://www.britishcouncil.org/exam/english/aptis/research/assessment-literacy>, Edited by Christine Coombe An a to z of second language assessment: how language teachers understand assessment concepts issn 2516-8649 © british council 2018 (page 42)

RATER TRAINING SYSTEM FOR ASSESSING WRITING AND SPEAKING SKILLS

M.P. Abdullaeva

Scientific-Educational Practical Center under the Agency for Assessment of Knowledge and Competences, 12 Bog'ishamol Street, Tashkent, Uzbekistan

Abstract. This article analyzes the methodology for the remote training of experts in assessing writing and speaking skills in foreign languages, as well as their involvement in the evaluation of these skills during test sessions organized by the Department for Assessment of Foreign Language Proficiency at the Agency for Assessment of Knowledge and Competences.

The article focuses on the stages of selecting candidates who are foreign language teachers, the methods of improving their qualifications, and the measures aimed at enhancing the overall quality and objectivity of the evaluation process.

Keywords: assessment specialist, expert training, assessment criteria, distance learning.