PEDAGOGIK TEST NATIJALARINI STATISTIK TAHLIL QILISHDA EHTIMOLLAR NAZARIYASI VA MATEMATIK STATISTIKA USULLARIDAN FOYDALANISH

Mannonova Nilufar Shomurod qizi
Ilmiy rahbar: ***Boqiyev.X***
Shahrisabz Davlat Pedagogika Instituti
1-kurs matematika yo‘nalishi Magistiratura talabasi
mannonovanilufar40@gmail.com
https://orcid.org/0009-0000-6832-7454

Annotatsiya: *Mazkur maqolada pedagogik test natijalarini statistik tahlil qilishda ehtimollar nazariyasi va matematik statistika usullaridan foydalanish masalalari yoritilgan. Test ballari tasodifiy miqdor sifatida talqin qilinib, tavsifiy statistika, normal taqsimotga moslikni tekshirish, ishonch intervallari, statistik gipotezalarni tekshirish va ta’sir kattaligini baholash usullari tizimlashtirilgan. Testning sifati topshiriqlar qiyinligi, ajratish qobiliyati va ishonchlilik ko‘rsatkichlari orqali baholangan. Usullarning qo‘llanilishi 60 nafar o‘quvchi va 20 ta topshiriqdan iborat model (simulyatsiya qilingan) ma’lumotlar misolida ko‘rsatilgan. Maqola natijalaridan o‘qituvchilar va tadqiqotchilar test natijalarini ob’ektiv talqin qilishda foydalanishi mumkin.*

Kalit so‘zlar: *pedagogik test, ehtimollar nazariyasi, matematik statistika, normal taqsimot, ishonch intervali, statistik gipoteza, t-mezon, KR-20, Rasch modeli, ta’sir kattaligi.*

Аннотация: *статье рассматривается применение методов теории вероятностей и математической статистики при статистическом анализе результатов педагогического тестирования. Тестовые баллы трактуются как случайная величина; систематизированы методы описательной статистики, проверки согласия с нормальным распределением, построения доверительных интервалов, проверки статистических гипотез и оценки величины эффекта. Качество теста оценено по показателям трудности, дифференцирующей способности и надёжности. Применение методов продемонстрировано на модельных (смоделированных) данных: 60 обучающихся и 20 заданий. Результаты могут быть использованы педагогами и исследователями для объективной интерпретации результатов тестирования.*

Ключевые слова: *педагогический тест, теория вероятностей, математическая статистика, нормальное распределение, доверительный интервал, статистическая гипотеза, t-критерий, KR-20, модель Раша, величина эффекта.*

Abstract: *This article examines the use of probability theory and mathematical statistics in the statistical analysis of pedagogical test results. Test scores are treated as random variables, and the methods of descriptive statistics, goodness-of-fit testing for normality, confidence intervals, statistical hypothesis testing, and effect-size estimation are systematized. Test quality is evaluated through item difficulty, discrimination, and reliability indices. The application of these methods is demonstrated on model (simulated) data from 60 students and 20 test items. The findings can help teachers and researchers interpret test results objectively.*

Keywords: *pedagogical test, probability theory, mathematical statistics, normal distribution, confidence interval, statistical hypothesis, t-test, KR-20, Rasch model, effect size.*

KIRISH

Ta’lim sifatini baholashda test usuli eng keng tarqalgan vositalardan biriga aylangan. Test natijalari o‘quvchilarni saralash, o‘quv dasturlarining samaradorligini aniqlash va ta’lim jarayonini takomillashtirish uchun asos bo‘lib xizmat qiladi. Biroq test balli o‘quvchining haqiqiy bilim darajasini to‘liq aks ettirmaydi: unga topshiriqlarning o‘ziga xosligi, o‘quvchining o‘sha paytdagi holati, javobni taxmin qilish imkoniyati kabi tasodifiy omillar ta’sir qiladi. Shu sababli test natijalarini faqat o‘rtacha ball yoki foiz ko‘rsatkichi bilan talqin qilish noto‘g‘ri xulosalarga olib kelishi mumkin.

Test natijalarini tahlil qilishda uch savolga javob topish zarur: birinchidan, olingan ball qanchalik aniq o‘lchangan; ikkinchidan, guruhlar o‘rtasidagi farq tasodifiymi yoki qonuniy; uchinchidan, testning o‘zi sifatlimi. Bu savollarga ehtimollar nazariyasi va matematik statistika usullari ilmiy asoslangan javob beradi: ehtimollar nazariyasi tasodifiylikni modellashtirishga, matematik statistika esa tanlanma ma’lumotlari asosida umumiy to‘plam haqida ishonchli xulosa chiqarishga imkon beradi.[1]

Ta’lim sifatini nazorat qilish va baholashning zamonaviy, shu jumladan raqamli vositalarini joriy etish O‘zbekiston ta’lim siyosatining ustuvor yo‘nalishlaridan biri sifatida me’yoriy hujjatlarda o‘z aksini topgan.[2,3] Amaliyotda esa test natijalari ko‘pincha o‘rtacha ball va foiz ko‘rsatkichlari bilan cheklanishi kuzatiladi. Mazkur maqolaning maqsadi pedagogik test natijalarini statistik tahlil qilishning asosiy ehtimoliy-statistik usullarini tizimlashtirish va ularni model ma’lumotlar misolida ko‘rsatishdan iborat.

[1] Гмурман, В. Е. (2003). Теория вероятностей и математическая статистика. Москва: Высшая школа.

[2] O‘zbekiston Respublikasining “Ta’lim to‘g‘risida”gi Qonuni, O‘RQ-637-son, 2020-yil 23-sentabr.

[3] O‘zbekiston Respublikasi Prezidentining 2020-yil 5-oktyabrdagi “«Raqamli O‘zbekiston – 2030» strategiyasini tasdiqlash va uni samarali amalga oshirish chora-tadbirlari to‘g‘risida”gi PF-6079-son Farmoni.

MAVZUGA OID ADABIYOTLAR TAHLILI

Test natijalarini ehtimoliy-statistik tahlil qilishning nazariy asoslari xorijiy tadqiqotlarda keng yoritilgan. F.Lord va M.Novik klassik test nazariyasini ishlab chiqib, kuzatilgan ballni haqiqiy ball va tasodifiy xatolikning yig‘indisi sifatida ifodalagan hamda ishonchlilikni dispersiyalar nisbati orqali aniqlagan.[4] L.Kronbax test topshiriqlarining ichki izchilligini baholovchi alfa koeffitsiyentini taklif etgan,[5] G.Kyuder va M.Richardson esa dixotomik topshiriqlar uchun ishonchlilik formulalarini asoslab bergan.[6] L.Kroker va J.Algina klassik va zamonaviy test nazariyasining statistik apparatini tizimli bayon etgan.[7] G.Rasch topshiriq qiyinligi va o‘quvchi qobiliyati o‘rtasidagi ehtimoliy bog‘liqlikni modellashtirgan,[8] bu yondashuv keyinchalik R.Hembleton va hamkasblari tomonidan topshiriq javobi nazariyasi (Item Response Theory) doirasida rivojlantirilgan.[9] Statistik mezonlarni tanlash va natijalarni talqin qilishda Dj.Koenning ta’sir kattaligi va statistik quvvat haqidagi qarashlari muhim o‘rin tutadi.[10]

MDH mamlakatlari olimlari pedagogik testlarni loyihalash va natijalarini qayta ishlash masalalariga alohida e’tibor qaratganlar. Xususan, V.S.Avanesov test topshiriqlarini tuzish va pedagogik o‘lchash nazariyasini tizimlashtirgan,[11] M.B.Chelishkova pedagogik testlarni konstruksiyalash nazariyasi va amaliyotini, jumladan test sifatini statistik baholash usullarini yoritgan.[12] V.E.Gmurmanning ehtimollar nazariyasi va matematik statistika bo‘yicha asarlari statistik usullarni o‘rganishda keng qo‘llaniladigan asosiy manbalardan biridir.[13] Shunga

[4]Lord, F. M., & Novick, M. R. (1968). Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley.

[5]Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297–334.

[6]Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. Psychometrika, 2(3), 151–160.

[7]Crocker, L., & Algina, J. (1986). Introduction to Classical and Modern Test Theory. New York: Holt, Rinehart and Winston.

[8] Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Danmarks Paedagogiske Institut.

[9]Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of Item Response Theory. Newbury Park, CA: Sage Publications.

[10]Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

[11]Аванесов, В. С. (2002). Композиция тестовых заданий. Москва: Центр тестирования.

[12]Челышкова, М. Б. (2002). Теория и практика конструирования педагогических тестов. Москва: Логос.

[13]Гмурман, В. Е. (2003). Теория вероятностей и математическая статистика. Москва: Высшая школа.

qaramay, bu manbalarda statistik usullar ko‘pincha alohida bo‘limlar sifatida bayon etilgan bo‘lib, ularni bitta test natijalari misolida bosqichma-bosqich qo‘llash tartibi kamroq ko‘rsatilgan.

O‘zbekistonda ta’lim sifatini baholash va ta’lim jarayonini raqamlashtirish masalalari, birinchi navbatda, me’yoriy-huquqiy hujjatlarda belgilangan.[14, 15] Biroq test natijalarini ehtimollar nazariyasi va matematik statistika vositalari bilan tahlil qilish tartibini yagona izchil sxema sifatida bayon etish zarurati saqlanib qolmoqda. Mazkur maqola shu ehtiyojni qisman qondirishga qaratilgan.

TADQIQOT METODOLOGIYASI

Tadqiqotda tizimli tahlil, qiyosiy tahlil, matematik modellashtirish va statistik tahlil usullaridan foydalanildi. Statistik usullarning qo‘llanilishini ko‘rsatish uchun model ma’lumotlar hosil qilindi: ikkita guruh (tajriba va nazorat, har biri 30 nafar o‘quvchi) 20 ta dixotomik (to‘g‘ri – 1, noto‘g‘ri – 0) topshiriqdan iborat testni bajargan deb faraz qilindi. Javoblar Rasch modeliga muvofiq simulyatsiya qilindi: o‘quvchilar qobiliyati normal taqsimotdan (tajriba guruhida o‘rtacha 0,6, nazorat guruhida 0,0; standart og‘ish 1,0), topshiriqlar qiyinligi esa −1,8 dan +1,8 gacha teng oraliqda tanlandi. Test balli to‘g‘ri javoblar soni sifatida (0–20) hisoblandi.

Hisob-kitoblar Python muhitida (SciPy kutubxonasi) bajarildi; ahamiyatlilik darajasi $\alpha = 0{,}05$ deb olindi. Shuni alohida ta’kidlash kerakki, ma’lumotlar haqiqiy o‘quvchilardan yig‘ilmagan, shu sababli keltirilgan sonli natijalar usullarning ishlash tartibini ko‘rsatish uchun mo‘ljallangan va biror o‘qitish metodikasining samaradorligi haqida empirik xulosa hisoblanmaydi.

NATIJALAR

Test balining ehtimoliy modeli. O‘quvchining test balini X bilan belgilaymiz. Bir xil bilim darajasidagi o‘quvchi turli sharoitda turlicha ball olishi mumkinligi sababli X tasodifiy miqdor hisoblanadi. Klassik test nazariyasida u quyidagicha ifodalanadi:[16]

$$X = T + E \qquad (1)$$

bu yerda T — haqiqiy ball, E — tasodifiy xatolik bo‘lib, $M(E) = 0$ va E bilan T o‘zaro bog‘liq emas deb qabul qilinadi. Natijada dispersiyalar uchun quyidagi tenglik o‘rinli bo‘ladi, testning ishonchliligi esa haqiqiy ball dispersiyasining umumiy dispersiyadagi ulushi sifatida aniqlanadi:

$$\sigma^2_X = \sigma^2_T + \sigma^2_E, \quad \rho = \sigma^2_T / \sigma^2_X \qquad (2)$$

[14] O‘zbekiston Respublikasining “Ta’lim to‘g‘risida”gi Qonuni, O‘RQ-637-son, 2020-yil 23-sentabr.

[15] O‘zbekiston Respublikasi Prezidentining 2020-yil 5-oktyabrdagi “«Raqamli O‘zbekiston – 2030» strategiyasini tasdiqlash va uni samarali amalga oshirish chora-tadbirlari to‘g‘risida”gi PF-6079-son Farmoni.

[16] Lord, F. M., & Novick, M. R. (1968). Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley.

Alohida topshiriqning natijasi Bernulli tasodifiy miqdoridir: to‘g‘ri javob 1 qiymatni p ehtimollik bilan, noto‘g‘ri javob 0 qiymatni $1 - p$ ehtimollik bilan qabul qiladi. n ta bir xil sharoitdagi topshiriqdan to‘g‘ri bajarilganlari soni binomial taqsimotga bo‘ysunadi:

$$P(X = k) = C^k_n \cdot p^k \cdot (1 - p)^{n-k} \quad (3)$$

Bu formula tasodifiy taxmin qilish ehtimolini baholashga imkon beradi. Masalan, har bir savolda 4 ta javob variantidan faqat bittasi to‘g‘ri bo‘lgan 10 ta savolda faqat taxmin qilib 6 yoki undan ko‘p to‘g‘ri javob berish ehtimoli (p = 0,25) taxminan 0,020 ga teng. Demak, bunday natijani tasodif bilan izohlash qiyin, shu sababli test natijalarini talqin qilishda tasodifiy to‘g‘ri javoblarning ulushini hisobga olish zarur.

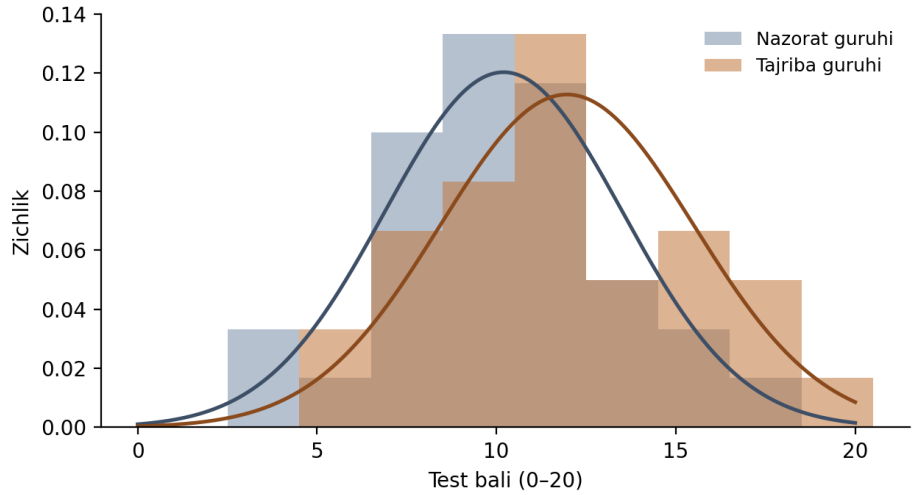
Tavsifiy statistika va normallikni tekshirish. Tahlilning birinchi bosqichida har bir guruh uchun markaziy tendensiya va tarqoqlik ko‘rsatkichlari hisoblanadi (1-jadval).

1-jadval. Test ballarining tavsifiy statistikasi

Ko‘rsatkich	Tajriba guruhi	Nazorat guruhi
O‘quvchilar soni (n)	30	30
O‘rtacha ball ($\bar{x}$)	11,97	10,20
Standart og‘ish (s)	3,54	3,32
Standart xato ($s/\sqrt{n}$)	0,65	0,61
Mediana	12	10
Minimum–maksimum	6–19	3–17
Asimmetriya koeffitsiyenti	0,18	−0,01
Ekssess koeffitsiyenti	−0,77	−0,07
O‘rtachaning 95% ishonch intervali	(10,65; 13,29)	(8,96; 11,44)
Shapiro–Uilk mezoni, p-qiymat	0,509	0,624

Tajriba guruhida o‘rtacha ball 11,97, nazorat guruhida 10,20 ni tashkil etdi; standart og‘ishlar bir-biriga yaqin (3,54 va 3,32). Shapiro–Uilk mezoni[17] har ikki guruh uchun normallik gipotezasini rad etmadi (p = 0,509 va p = 0,624), asimmetriya koeffitsiyentlari nolga yaqin, ya’ni taqsimotlar deyarli simmetrik (1-rasm). Bu parametrik mezonlarni qo‘llash uchun asos bo‘ladi.

[17] Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3–4), 591–611.



1-rasm. Test ballarining taqsimoti va moslashtirilgan normal egri chiziqlar[18]

Normal taqsimotda natijalarning taxminan 68,3% i (μ ± σ), 95,4% i (μ ± 2σ) va 99,7% i (μ ± 3σ) oralig‘ida joylashadi. Ballarni taqqoslash uchun standartlashtirilgan shkalalardan foydalaniladi:

$$z = (x - \bar{x}) / s, \quad T = 50 + 10z \tag{4}$$

Masalan, butun tanlanma (n = 60) bo‘yicha $\bar{x}$ = 11,08 va s = 3,51 bo‘lganda 15 ball olgan o‘quvchining z-balli taxminan 1,12, T-balli esa taxminan 61 ga teng, ya’ni u o‘rtachadan taxminan bir standart og‘ishdan ko‘proq yuqori natija ko‘rsatgan.

Ishonch intervallari. Tanlanma o‘rtachasi umumiy to‘plam o‘rtachasi μ ning faqat baholovchisi bo‘lgani uchun uning aniqligi ishonch intervali bilan ifodalanadi. Standart og‘ish noma’lum bo‘lganda (1 − α) ishonchlilik darajasidagi interval:

$$\bar{x} \pm t_{\alpha/2;\, n-1} \cdot s / \sqrt{n} \tag{5}$$

Tajriba guruhi uchun $t_{0,025;\, 29}$ = 2,045, standart xato 0,65 bo‘lgani uchun 95% ishonch intervali (10,65; 13,29) ni, nazorat guruhi uchun (8,96; 11,44) ni tashkil etdi (1-jadval). Intervallar qisman ustma-ust tushadi, bu guruhlar o‘rtasidagi farq katta emasligidan dalolat beradi, biroq xulosani gipotezani tekshirish bilan tasdiqlash kerak.

Statistik gipotezalarni tekshirish. Guruhlar o‘rtasidagi farqni baholash uchun nol gipoteza H_0: $\mu_1 = \mu_2$ (farq yo‘q) va muqobil gipoteza H_1: $\mu_1 \neq \mu_2$ ilgari suriladi. Dispersiyalar teng deb olinmaganda Uelch t-mezoni ishlatiladi:[19]

$$t = (\bar{x}_1 - \bar{x}_2) / \sqrt{(s^2_1/n_1 + s^2_2/n_2)} \tag{6}$$

[18] Muallif tomonidan model (simulyatsiya qilingan) ma’lumotlar asosida tayyorlangan.

[19] Welch, B. L. (1947). The generalization of “Student’s” problem when several different population variances are involved. Biometrika, 34(1–2), 28–35.

Farqning amaliy ahamiyatini baholash uchun Koen ko‘rsatkichi hisoblanadi:[20]

$$d = (\bar{x}_1 - \bar{x}_2) / s_p \quad (7)$$

bu yerda s_p — birlashtirilgan standart og‘ish. Natijalar 2-jadvalda keltirilgan.

2-jadval. Guruhlarni taqqoslash natijalari

Ko‘rsatkich	**Qiymat**	***p*-qiymat**
Levin mezoni (dispersiyalar tengligi)	$F = 0{,}18$	0,672
Uelch *t*-mezoni	$t = 2{,}00$; $df \approx 57{,}8$	0,051
Mann–Uitni mezoni	$U = 570{,}5$	0,074
O‘rtachalar farqi	1,77 ball; 95% IO: (−0,01; 3,54)	—
Koen ko‘rsatkichi (*d*)	0,52	—

Tajriba guruhining o‘rtacha balli nazorat guruhidan 1,77 ballga yuqori. Biroq Uelch mezoni bo‘yicha $p = 0{,}051$ bo‘lib, $\alpha = 0{,}05$ darajada nol gipoteza formal ravishda rad etilmaydi; farqning 95% ishonch intervali (−0,01; 3,54) nolni deyarli chetlab o‘tadi. Noparametrik Mann–Uitni mezoni[21] ham o‘xshash natija berdi ($p = 0{,}074$). Shu bilan birga Koen ko‘rsatkichi $d = 0{,}52$ o‘rtacha darajadagi ta’sirni bildiradi. Bunday chegaraviy holatda “farq bor” yoki “farq yo‘q” degan qat’iy xulosa chiqarish o‘rniga tanlanma hajmi yetarli emasligini hisobga olish kerak: $d \approx 0{,}5$ ta’sirni 80% quvvat bilan aniqlash uchun har bir guruhda taxminan 64 nafar o‘quvchi talab etiladi.[22]

Testning sifat ko‘rsatkichlari. Test natijalarini talqin qilishdan oldin testning o‘zi sifatli ekaniga ishonch hosil qilish kerak. Buning uchun topshiriqlar qiyinligi indeksi (p_i), ajratish qobiliyati (D_i) va topshiriq bilan umumiy ball o‘rtasidagi nuqtali-biseriyal korrelyatsiya (r_{pb}) hisoblanadi:[23]

$$p_i = m_i / N, \quad D_i = p_i^{yuqori} - p_i^{past} \quad (8)$$

bu yerda m_i — *i*-topshiriqni to‘g‘ri bajarganlar soni, N — o‘quvchilar soni, p^{yuqori} va p^{past} — umumiy bali bo‘yicha eng yuqori va eng past 27% o‘quvchilardagi to‘g‘ri javoblar

[20] Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

[21] Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics, 18(1), 50–60.

[22] Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

[23] Crocker, L., & Algina, J. (1986). Introduction to Classical and Modern Test Theory. New York: Holt, Rinehart and Winston.

ulushi. Odatda $0{,}3 \le p_i \le 0{,}7$ bo‘lgan, $D_i \ge 0{,}30$ va $r_{pb} \ge 0{,}20$ topshiriqlar yaxshi hisoblanadi (3-jadval).

3-jadval. Topshiriqlarni tahlil qilish natijalari

№	*p*	*D*	r_{pb}	№	*p*	*D*	r_{pb}
1	0,83	0,19	0,19	11	0,48	0,56	0,24
2	0,83	0,50	0,40	12	0,45	0,44	0,22
3	0,70	0,50	0,36	13	0,42	0,50	0,36
4	0,73	0,38	0,17	14	0,40	0,81	0,44
5	0,77	0,25	0,17	15	0,52	0,44	0,22
6	0,75	0,38	0,23	16	0,43	0,62	0,40
7	0,67	0,25	0,17	17	0,35	0,56	0,38
8	0,68	0,19	0,10	18	0,28	0,06	−0,07
9	0,70	0,44	0,23	19	0,27	0,44	0,31
10	0,60	0,56	0,32	20	0,22	0,56	0,39

Tahlil shuni ko‘rsatdiki, 20 ta topshiriqdan 12 tasi qiyinlik bo‘yicha maqbul oraliqda (0,30–0,70), 5 tasi (1, 2, 4, 5, 6-topshiriqlar) nisbatan oson, 3 tasi (18, 19, 20-topshiriqlar) nisbatan qiyin. Ajratish qobiliyati 5 ta topshiriqda (1, 5, 7, 8, 18) 0,30 dan past; r_{pb} esa 6 ta topshiriqda (1, 4, 5, 7, 8, 18) 0,20 dan past. Ayniqsa 18-topshiriq ($p = 0{,}28$; $D = 0{,}06$; $r_{pb} = -0{,}07$) umumiy ball bilan deyarli bog‘lanmagan, uni mazmun jihatidan qayta ko‘rib chiqish yoki testdan chiqarish lozim.

Testning ichki izchilligi dixotomik topshiriqlar uchun Kyuder–Richardson formulasi (KR-20) bilan baholanadi,[24] o‘lchashning standart xatosi esa ishonchlilik koeffitsiyenti orqali topiladi:

$$KR\text{-}20 = (k/(k-1)) \cdot (1 - \Sigma p_i q_i / s^2), \quad SEM = s\sqrt{(1-r)} \qquad (9)$$

bu yerda k — topshiriqlar soni, $q_i = 1 - p_i$. Model ma’lumotlar uchun KR-20 = 0,69, SEM ≈ 1,94 ball. Juft va toq topshiriqlar bo‘yicha ikkiga bo‘lish usulida korrelyatsiya 0,49, Spirmen–Braun tuzatishidan keyin esa 0,66 ni tashkil etdi. Ishonchlilik 0,70 ga yaqin: guruh darajasidagi tadqiqotlar uchun qabul qilinishi mumkin, ammo alohida o‘quvchi haqida qaror qabul qilish uchun yetarli emas. Test uzunligini k barobar oshirgandagi ishonchlilik Spirmen–Braun formulasi bilan baholanadi:

$$r^* = k \cdot r / (1 + (k-1)r) \qquad (10)$$

[24] Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. Psychometrika, 2(3), 151–160.

Unga ko‘ra topshiriqlar soni ikki baravar oshirilsa (40 tagacha), bir xil sifatdagi topshiriqlar qo‘shilgan holda ishonchlilik taxminan 0,82 ga yetadi.

Zamonaviy test nazariyasi elementlari. Klassik nazariyaning kamchiligi shundaki, topshiriq ko‘rsatkichlari muayyan tanlanmaga bog‘liq bo‘ladi. Rasch modeli esa *i*-topshiriqni to‘g‘ri bajarish ehtimolini o‘quvchi qobiliyati θ va topshiriq qiyinligi b_i orqali ifodalaydi:[25]

$$P(X_i = 1 \mid \theta, b_i) = e^{\theta - b_i} / (1 + e^{\theta - b_i}) \qquad (11)$$

Agar $\theta = b_i$ bo‘lsa, to‘g‘ri javob ehtimoli 0,5 ga teng; $\theta - b_i$ farqi ortgan sari u birga intiladi. Model parametrlari eng katta ishonchlilik usuli bilan baholanadi va o‘quvchilar qobiliyati hamda topshiriqlar qiyinligini bitta shkalada ifodalash, turli variantdagi testlar natijalarini solishtirish imkonini beradi.[26] Mazkur maqolada model ma’lumotlar aynan shu modeldan hosil qilingani sababli, kuzatilgan qiyinlik ko‘rsatkichlarining tartibi (1-topshiriqdan 20-topshiriqqa qarab p_i ning kamayishi) berilgan qiyinlik parametrlari tartibini qayta tikladi.

MUHOKAMA

Tadqiqot natijalari pedagogik test natijalarini tahlil qilish ketma-ket bosqichlardan iborat jarayon ekanligini ko‘rsatdi: avval ma’lumotlar taqsimoti va test sifati tekshiriladi, keyin gipotezalar sinaladi va natijaning amaliy ahamiyati baholanadi. Misoldagi chegaraviy natija ($p = 0{,}051$) statistik xulosa faqat *p*-qiymatga tayanmasligi kerakligini yaqqol ko‘rsatadi. *p*-qiymat nol gipoteza to‘g‘ri bo‘lish ehtimolini bildirmaydi; u faqat kuzatilgan farq nol gipoteza to‘g‘ri bo‘lganda qanchalik kutilmagan ekanini ifodalaydi. Shu sababli ishonch intervali va ta’sir kattaligini birga keltirish, tanlanma hajmini oldindan asoslash maqsadga muvofiq.

Shu bilan birga, mazkur tadqiqotning bir qator cheklovlari mavjud. Birinchidan, ma’lumotlar simulyatsiya yo‘li bilan hosil qilingan, shuning uchun natijalarni real ta’lim jarayoniga bevosita ko‘chirib bo‘lmaydi. Ikkinchidan, parametrik mezonlar normallik va kuzatuvlar mustaqilligi shartlariga tayanadi; bu shartlar buzilganda noparametrik usullarga murojaat qilish kerak. Uchinchidan, bir nechta gipotezani bir vaqtda tekshirishda birinchi tur xato ehtimoli oshadi va Bonferroni kabi tuzatishlar talab etiladi. To‘rtinchidan, test ishonchliligi 0,70 atrofida bo‘lganda yakka natijalar bo‘yicha qat’iy qarorlar qabul qilmaslik lozim.

Demak, test natijalarini ob’ektiv talqin qilish uchun faqat o‘rtacha ballni hisoblash yetarli emas; buning uchun taqsimot shaklini tekshirish, o‘lchash xatosini hisobga olish,

[25] Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Danmarks Paedagogiske Institut.

[26] Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of Item Response Theory. Newbury Park, CA: Sage Publications.

statistik mezonni to‘g‘ri tanlash, ta’sir kattaligini baholash va testning sifat ko‘rsatkichlarini nazorat qilishni o‘z ichiga olgan kompleks yondashuv zarur.

XULOSA

Pedagogik test natijalarini statistik tahlil qilishda ehtimollar nazariyasi va matematik statistika usullaridan foydalanish baholashni subyektiv fikrlardan ob’ektiv, miqdoriy asoslangan xulosalarga o‘tkazadi. Test balini tasodifiy miqdor sifatida qarash, taqsimot qonuniyatlarini aniqlash, ishonch intervallari va statistik mezonlar orqali xulosalarning ishonchliligini baholash, topshiriqlar qiyinligi, ajratish qobiliyati va ishonchlilikni nazorat qilish ta’lim sifatini boshqarishda muhim ahamiyatga ega. Model misol shuni ko‘rsatdiki, o‘rtacha ballardagi farq chegaraviy ahamiyatga ega bo‘lishi, test ishonchliligi esa 0,70 atrofida bo‘lishi mumkin va bunday holatda natijalarni ehtiyotkorlik bilan talqin qilish, tanlanma hajmini oshirish hamda sifatsiz topshiriqlarni qayta ko‘rib chiqish zarur.

Yuqoridagi tahlillar asosida quyidagi takliflar beriladi:

Test natijalarini tahlil qilishda o‘rtacha ball bilan birga standart og‘ish, ishonch intervali va ta’sir kattaligini keltirish, guruhlarni taqqoslashdan oldin taqsimotning normalligini tekshirish va mos (parametrik yoki noparametrik) mezonni tanlash tavsiya etiladi.

Test tuzilgandan so‘ng uni amaliyotga joriy etishdan oldin topshiriqlarning qiyinligi, ajratish qobiliyati va ishonchliligini (KR-20 yoki Kronbax alfa) hisoblash, sifatsiz topshiriqlarni almashtirish, o‘qituvchilarning statistik savodxonligini muntazam oshirib borish va kelgusida real o‘quvchilar ma’lumotlari asosida usullarni sinovdan o‘tkazish maqsadga muvofiq.

FOYDALANILGAN ADABIYOTLAR RO‘YXATI:

1. Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

2. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297–334.

3. Crocker, L., & Algina, J. (1986). Introduction to Classical and Modern Test Theory. New York: Holt, Rinehart and Winston.

4. Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of Item Response Theory. Newbury Park, CA: Sage Publications.

5. Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. Psychometrika, 2(3), 151–160.

6. Lord, F. M., & Novick, M. R. (1968). Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley.

7. Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics, 18(1), 50–60.

8. Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Danmarks Paedagogiske Institut.

9. Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52(3–4), 591–611.

10. Welch, B. L. (1947). The generalization of “Student’s” problem when several different population variances are involved. Biometrika, 34(1–2), 28–35.

11. Аванесов, В. С. (2002). Композиция тестовых заданий. Москва: Центр тестирования.

12. Гмурман, В. Е. (2003). Теория вероятностей и математическая статистика. Москва: Высшая школа.

13. Челышкова, М. Б. (2002). Теория и практика конструирования педагогических тестов. Москва: Логос.

14. O‘zbekiston Respublikasining “Ta’lim to‘g‘risida”gi Qonuni, O‘RQ-637-son, 2020-yil 23-sentabr.

15. O‘zbekiston Respublikasi Prezidentining 2020-yil 5-oktyabrdagi “«Raqamli O‘zbekiston – 2030» strategiyasini tasdiqlash va uni samarali amalga oshirish chora-tadbirlari to‘g‘risida”gi PF-6079-son Farmoni.