Qabul qilindi: 8-sentabr 2026-yil • Ma’qullandi: 14-sentabr 2026-yil • Nashr qilindi: 17-sentabr 2026-yil

PEDAGOGIK TESTLARNING VALIDLIGI VA ISHONCHLILIGINI TA’MINLASHNING NAZARIY ASOSLARI

Xalikov Faroxidin Abdumannobovich
Farg‘ona davlat universiteti mustaqil izlanuvchisi
Orcid: 0009-0000-8156-2491

Annotatsiya. Maqolada pedagogik testlarning validligi va ishonchliligini asoslash shartlari tahlil qilinadi. Tarixiy-qiyosiy va konseptual tahlil orqali ball talqini, o‘lchash aniqligi va pedagogik qaror farqlanadi. Konstrukt, topshiriq, dalil, talqin va qaror izchilligiga tayangan metodik model taklif etiladi.

Kalit so‘zlar: pedagogik test; validlik; ishonchlilik; konstrukt; o‘lchash xatosi; validatsiya; adolatlilik; pedagogik qaror.

KIRISH

Pedagogik testning vazifasi faqat javoblarni ballga aylantirish emas, balki shu ball asosida o‘quv natijasi haqida asoslangan xulosa chiqarishga xizmat qilishdir. Muammo «test to‘g‘ri tuzilganmi?» savolidan kengroq: undan kim uchun, qaysi mazmun doirasida va qanday qaror uchun foydalanish mumkinligi aniqlashtirilishi kerak. AERA, APA va NCME standartlarida validlik muayyan maqsad uchun taklif etilgan ball talqinini nazariya hamda dalillar qay darajada qo‘llab-quvvatlashi bilan bog‘lanadi [1, 1-bob]. Shuning uchun «valid test» iborasi qo‘llanish sohasi ko‘rsatilmaganda yetarlicha aniq emas.

Masalaning nazariy genezisida L.J. Cronbach va P.E. Meehlning konstrukt validligiga oid yondashuvi muhim: kuzatiladigan javob bilan nazariy tushuncha o‘rtasidagi bog‘lanish tekshiriladigan munosabatlar tizimida asoslanadi [2]. M.T. Kanening argumentga asoslangan yondashuvi esa ballardan chiqariladigan xulosalar va foydalanish da’volarini, ularning taxminlari hamda dalillarini ochiq tekshirishni taklif etadi [3]. Bu yo‘nalishlar testni mustaqil texnik buyum sifatida emas, nazariya, o‘lchash va qaror o‘rtasidagi vosita

sifatida o‘rganish uchun tayanchdir.

O‘zbekiston kontekstida K.Q. Jalilovning o‘zbek tilini bilish darajasi sertifikati uchun test tuzuvchilar tayyorgarligiga oid tadqiqoti fan ekspertizasi bilan baholash kompetentligini farqlash zaruratini ko‘rsatadi [4, 54–61-b.]. Mazkur xulosani barcha o‘qituvchilarga umumlashtirish mumkin emas; biroq u pedagogik test yaratishni maxsus metodologik tayyorgarlik bilan bog‘lash uchun mahalliy ilmiy asos beradi. Ushbu maqolada shu masala alohida fan metodikasi doirasidan pedagogika nazariyasi darajasiga olib chiqiladi.

Tadqiqotning maqsadi validlik va ishonchlilikni ta’minlashning nazariy shartlarini tizimlashtirib, ularni pedagogik testni loyihalash, tekshirish va qo‘llash bosqichlari bilan bog‘lashdir. Asosiy savol: mazmuniy jihatdan asosli talqinni yetarli o‘lchash aniqligi va mas’uliyatli pedagogik qaror bilan qanday muvofiqlashtirish mumkin? Nazariy-metodik hissa mavjud tushunchalarni yangidan nomlash emas, ularning o‘qituvchi faoliyatidagi vazifalarini «konstrukt, topshiriq, dalil, talqin, qaror» zanjirida farqlash va tekshiruv mezonlariga aylantirishdan iborat.

MATERIALLAR VA METODLAR

Tadqiqot nazariy-konseptual va metodik loyihalash xarakteriga ega. Manbalar maqsadli tanlandi: konstrukt validligi haqidagi birlamchi nazariy matn, argumentga asoslangan validatsiya maqolasi, xalqaro professional standartlar, ishonchlilik bo‘yicha metodologik nashrlar va O‘zbekistondagi test tuzuvchilar tayyorgarligiga oid tadqiqot. 2014-yilgi standartlar tarixiy nashr sanasi bilan, raqamli baholashning 2025-yilgi qo‘llanmasi va 2026-yilgi metodologik bahs bilan birgalikda o‘rganildi. Bu barcha nashrlarni qamrab oluvchi tizimli sharh yoki meta-tahlil emas.

Tarixiy-qiyosiy usul tushunchalarning rivojlanishidagi urg‘ularni, konseptual tahlil validlik, ishonchlilik, adolatlilik va standartlashtirish chegaralarini aniqlashga xizmat qildi. Argumentativ tahlilda har bir talqin uchun da’vo, tayanch dalil, muqobil izoh va qo‘llanish chegarasi ajratildi. Metodik modellashtirish orqali ushbu farqlar test yaratishning ketma-ket ishlariga aylantirildi. Pedagogik vaziyatlar real qatnashchilarni tavsiflamaydi, taklif etilgan mezonlarni tushuntiruvchi shartli misollardir.

Mustaqil tanlanma, eksperiment yoki empirik ballar bazasi mavjud emas; shuning uchun maqolada ishonchlilik koeffitsiyenti hisoblab topilgani yoxud model samaradorligi tasdiqlangani da’vo qilinmaydi. Standart xatoga oid raqamlar faqat hisoblash namunasidir. «Natijalar» bo‘limida empirik ta’sir emas, tushunchalar chegarasi, dalillar matritsasi va amaliy tekshiruv tartibi beriladi. Manbalardagi professional tavsiyalar O‘zbekiston uchun avtomatik huquqiy majburiyat sifatida talqin qilinmaydi.

NATIJALAR

Validlik va ishonchlilikning vazifaviy chegaralari. Konstrukt test yordamida xulosa chiqarilishi ko‘zda tutilgan nazariy belgi yoki bilim sohasi. «Pedagogik tushunchalarni eslab aytish», «vaziyatni tahlil qilish» va «sinfda asoslangan harakat qilish» bir xil konstrukt emas. Mazkur farqdan kelib chiqib, maqolada topshiriqning mazmuni bilan xulosaning qamrovini moslashtirish talabi qo‘yiladi: ta’rifni tanlash topshirig‘i bo‘yicha olingan natija o‘z-o‘zidan amaliy pedagogik mahoratning to‘liq dalili bo‘la olmaydi. Bu taklif etilayotgan modelning boshlang‘ich metodik chegarasi.

Ishonchlilik ballarning takroriy o‘lchash sharoitlaridagi izchilligini, jumladan, vaqt, test shakli yoki baholovchi o‘zgarishiga sezgirligini ifodalaydi [5]. Validlik esa bu ballar nimani anglatishi haqidagi asoslashdir. Demak, barqaror natija noto‘g‘ri mazmunni ham izchil aks ettirishi mumkin. Aksincha, mazmuni o‘rinli topshiriqlar haddan tashqari noaniq baholansa, individual qarorni yetarli aniqlikda asoslay olmaydi. Standartlashtirish tartiblarning bir xilligini ta’minlashga xizmat qiladi, lekin tanlangan mazmunning to‘g‘riligini alohida isbotlamaydi.

Yagona validlik konsepsiyasida mazmun, javob berish jarayoni, ichki tuzilma, boshqa o‘zgaruvchilar bilan aloqalar va qo‘llash oqibatlari haqidagi dalillar bir-biridan mustaqil «validlik turlari» sifatida olinmaydi [1, 1-bob]. Ularni tekshirish hajmi maqsadga bog‘liq. Quyidagi matritsa standartlarning aynan tarjimasi emas, shu yondashuvni pedagogik test yaratishga moslashtiruvchi metodik taklifdir.

1-jadval

Ball talqinini asoslash uchun dalillar matritsasi

Dalil yo‘nalishi	Tekshiriladigan savol	Tavsiya etiladigan material	Asosiy cheklov
Mazmun	Muhim natijalar va fikrlash amallari qamrab olinganmi?	O‘quv natijalari xaritasi, spetsifikatsiya, ekspert izohi.	Ekspert roziligi boshqa dalillarni almashtirmaydi.
Javob jarayoni	Talaba ko‘zlangan bilim amalidan foydalanadimi?	Kognitiv suhbat, yechim izohi, xato tahlili.	To‘g‘ri javobning o‘zi to‘g‘ri strategiyani isbotlamaydi.
Ichki tuzilma	Ball birligi nazariy tuzilishga muvofiqmi?	Topshiriqlararo aloqalar, mos faktor yoki IRT modeli.	Modelga moslik barcha talqinlarni oqlamaydi.
Tashqi aloqalar	Natija tegishli faoliyat bilan bog‘lanadimi?	Mustaqil amaliy vazifa, mezon ko‘rsatkichi.	Bitta korrelatsiya sababiylikni ko‘rsatmaydi.
Qo‘llash oqibatlari	Qaror nimaga olib keladi; xato kimga ta’sir qiladi?	Qaror auditi, qayta baholash va apellatsiya qaydlari.	Noqulay natija o‘z-o‘zidan invalidlik dalili emas.

Jadvaldagi yo‘nalishlar bir-birini ballar yig‘indisi kabi qoplamaydi. Masalan, mazmuniy qamrovdagi nuqsonni yuqori ishonchlilik koeffitsiyenti bilan «yopish» mumkin emas. Shu bilan birga, har qanday salbiy ijtimoiy natijani o‘lchovning yaroqsizligi deb e’lon qilish ham asossiz: muammo test mazmunidami, o‘qitish imkoniyatidami yoki undan foydalanish siyosatidami — alohida tekshiriladi. Mualliflik taklifida dalil yetishmasligi bilan talqinga qarshi dalil mavjudligi ham ajratiladi: birinchisi qo‘shimcha tekshiruvni,

ikkinchisi talqinni qayta ko‘rib chiqishni talab qiladi.

Ishonchlilikni faqat bitta koeffitsiyentga tenglashtirish tekshiruvni toraytiradi. Klassik test nazariyasida kuzatilgan ball kutiladigan haqiqiy ball va xato komponentiga ajratiladi; «haqiqiy ball» insonning o‘zgarmas mohiyati emas, belgilangan o‘lchash sharoitlaridagi nazariy kutilmadir [5]. Umumlashtirish nazariyasi topshiriq, baholovchi va vaziyat kabi xato manbalarini ajratishga; topshiriqqa javob nazariyasi esa model shartlari bajarilganda turli qobiliyat darajalaridagi aniqlikni tekshirishga imkon beradi [6].

Alfa topshiriqlarning bir o‘lchamliligini yoki validlikni isbotlamaydi. 2026-yilgi metodologik bahsda S. Zitzmann va G.A. Orona mazmunning xilma-xilligi past alfaga avtomatik izoh emasligini, koeffitsiyent ballning nazariy modeli bilan bog‘liq talqin qilinishini ta’kidlaydi [7]. Demak, mazmunni sun’iy toraytirib alfaga erishish ham, past ko‘rsatkichni «fan murakkab» deb tekshiruvsiz oqlash ham maqbul yo‘l emas. Omega ham avtomatik yechim bo‘lmaydi: uning turi va hisoblash modeli ko‘zlangan umumiy yoki ko‘p komponentli ballga mos bo‘lishi kerak.

Pedagogik qaror uchun ball birligida ifodalangan o‘lchashning standart xatosi, SEM ham ahamiyatli. Bir xil guruh va xato manbalari uchun klassik bog‘lanish quyidagicha yoziladi [5, 22-b.]:

$$SEM = SD \times \sqrt{(1 - r)}.$$

Bu yerda SD — kuzatilgan ballarning standart og‘ishi, r - shu sharoitdagi ishonchlilik bahosi. Faqat tushuntirish uchun SD = 10 ball va r = 0,84 deb olinsa, $SEM = 10 \times \sqrt{0,16} = 4$ ball bo‘ladi. Bu haqiqiy tadqiqot natijasi emas. Hisob olingan ball mutlaq aniqlikdagi nuqta emasligini ko‘rsatadi; 4 ball barcha talabalar uchun bir xil individual xato degani emas. Chegara atrofida xato o‘zgarishi mumkinligi sababli yuqori oqibatli qarorlarda shartli aniqlik alohida baholanadi [6].

Model uchun amaliy qoida sifatida «yagona universal minimal koeffitsiyent» o‘rniga uch savol taklif etiladi: qaysi xato manbasi hisobga olindi, qanday qaror chiqariladi va xato qaror oqibati nimadan iborat? Guruhlararo tadqiqot xulosasi, individual rivojlantiruvchi tavsiya va attestatsiya bir xil aniqlik talabiga ega deb olinmaydi. Koeffitsiyent bilan birga

tanlanma tavsifi, hisoblash usuli va uning noaniqligi berilishi kerak; faqat «ishonchlilik yuqori» jumlasi tekshiriladigan hisobot o‘rnini bosa olmaydi.

Normaga yo‘naltirilgan talqinda shaxsning guruhdagi nisbiy o‘rni, mezonga yo‘naltirilgan talqinda esa belgilangan o‘quv talabini egallash darajasi markazda turadi. Shu farq modelda alohida qayd qilinadi. O‘tish balini faqat guruh o‘rtachasi yoki istalgan foizdan kelib chiqib belgilash o‘rniga, minimal kutiladigan faoliyat tavsifi, ekspert hukmi va qaror oqibatlarini asoslash taklif etiladi. Chegarani belgilash bilan o‘sha chegara atrofidagi o‘lchash aniqligini tekshirish bir-birini almashtirmaydigan ishlardir [1, 5-bob; 6].

Validatsiyani testning hayotiy sikliga integratsiyalash. Taklif etilayotgan modelda ish topshiriq yozishdan emas, ball asosida bildiriladigan da’vodan boshlanadi. Avval maqsad, auditoriya, konstrukt va foydalanish chegarasi yoziladi; keyin mazmun hamda kognitiv talablar spetsifikatsiyasi tuziladi. Har bir topshiriq «qaysi faoliyat dalilini beradi?» savoliga javob berishi lozim. Ekspert faqat kalitni tekshirmaydi: noaniq ibora, mavzuga aloqasiz til yuklamasi va konstruktning tushib qolgan qismini ham qayd etadi. Bunday ish o‘qituvchining fan bilimini baholash savodxonligi bilan bog‘laydi [4].

Keyingi qadamda kichik kognitiv sinov topshiriqning qanday tushunilganini aniqlash uchun taklif etiladi; keyingi aprobatsiya esa o‘quvchilar tayyorgarligi, til va o‘qish sharoitlaridagi xilma-xillikni hisobga oladi. Tanlanma hajmi bitta tayyor son bilan belgilanmaydi: model murakkabligi, guruhlar bo‘yicha tekshiruv va talab qilinadigan aniqlik asoslanadi. Topshiriq statistikasi mazmuniy ekspertiza bilan birga talqin qilinadi. Past ko‘rsatkichni ko‘rib avtomatik chiqarish o‘rniga kalit, noaniqlik, mazmuniy zarurat va topshiriqning umumiy konstruktga aloqasi tekshiriladi.

Baholash rubrikasi, topshiriq yoki til varianti o‘zgarsa, avvalgi dalillarni o‘zgarishsiz ko‘chirish emas, ta’sirlangan xulosani qayta tekshirish taklif etiladi. Yakuniy hujjatda dalil, cheklov, foydalanish sharti va qayta ko‘rib chiqish sababi qayd etiladi. O‘qituvchi, test tuzuvchi va natijadan foydalanuvchining javobgarligi shu tariqa kuzatiladigan ishlar bilan belgilanadi. Maqsad barcha testlarni murakkab statistik tizimga aylantirish emas, qaror xavfiga mutanosib, ammo dalili ochiq jarayon yaratishdir.

MUHOKAMA

Pedagogik vaziyatda modelning qo‘llanishi. Shartli misolda «Pedagogika nazariyasi» fanining yakuniy testi tarbiyaviy vaziyatni tahlil qilish kompetentligini baholash uchun yaratiladi, biroq topshiriqlar asosan ta’riflarni eslashga yo‘naltirilgan, deb faraz qilinadi. Talabalar ularni barqaror bajargan taqdirda ham, shu natijadan «murakkab tarbiyaviy vaziyatda mustaqil qaror qabul qiladi» degan xulosa chiqarish uchun dalil yetarli emas. Bu yerda muammo statistik tartibning yomonligida emas, topshiriq bilan xulosa orasidagi uzilishdadir.

Model asosida spetsifikatsiyani qayta tuzish taklif etiladi: talaba vaziyatdagi faktni taxmindan ajratishi, kamida muqobil izohlarni ko‘rib chiqishi va pedagogik harakatini asoslay olishi kerak. Bir vaziyatda o‘quvchining darsda sukut saqlashi tavsiflanadi; ishtirokchidan «sust o‘quvchi» yorlig‘ini tanlash emas, qo‘shimcha dalil ehtiyojini va yordam usulini belgilash so‘raladi. Bir nechta mazmuniy to‘g‘ri javob bo‘lishi mumkinligi rubrikada oldindan tan olinadi. Baholash o‘qituvchi fikriga moslikni emas, pedagogik dalillash sifatini tekshirishi ko‘zda tutiladi.

Ochiq javoblar mustaqil baholanib, tafovutlar rubrika bandlari bo‘yicha muhokama qilinadi. Qo‘sh baholash dastlabki mustaqil ballarni saqlashi kerak; yakuniy kelishilgan ballning o‘zi baholovchilar ishonchliligiga dalil emas. Vaziyatli javobni sinfdagi real xulqqa umumlashtirish uchun alohida amaliy kuzatuv zarur. Shunday qilib, testning mazmunini kengaytirish bilan baholash aniqligini ta’minlash birgalikda loyihalanadi; misol texnologiya samaradorligining empirik isboti sifatida berilmaydi.

Til, raqamli muhit va adolatlilik shartlari. O‘zbek, rus yoki boshqa tillardagi variantlar o‘rtasida so‘zma-so‘z tarjima teng o‘lchashni kafolatlamaydi. ITC qo‘llanmasi konstruktning muvofiqligi, ko‘rsatmalar mazmuni, mos tarjima tartibi va maqsadli guruhdagi aprobatsiyani tekshirishni nazarda tutadi [8]. Mahalliy qo‘llash uchun fan mutaxassisi, tilshunos va o‘lchovchi hamkorligi taklif etiladi. Bir xil tayyorgarlik darajasida guruhlar topshiriqqa turlicha javob bersa, differensial ishlash sabablari o‘rganiladi; statistik farqning o‘zi kamsitish yoki tarjima xatosi haqida yakuniy hukm emas.

Raqamli testda qurilma, interfeys va texnik resurslarning ta’siri maqsadli bilimdan ajratilishi kerak. ITC va ATPning 2025-yilgi qo‘llanmasi texnologik baholashda validlik, aniqlik, foydalanish imkoniyati, xavfsizlik va maxfiylikni o‘zaro bog‘laydi [9]. Shu asosda sinovdan oldingi tanishtirish, uzilishlarni qayd etish va zarur moslashtirishlarni hujjatlashtirish taklif etiladi. Agar raqamli tezkorlik konstruktga kirmasa, uni fan bilimining bilvosita mezoniga aylantirmaslik lozim. Avtomatik baholash ham kalit, rubrika va insoniy qayta tekshiruv o‘rnini mutlaq egallamaydi.

Cheklovlar va empirik tekshiruv istiqbollari. Taklif etilgan model test sifatining sertifikati yoki tayyor psixometrik shkala emas. Uning nazariy mantiqi turli ta’lim bosqichlarida bir xil samaradorlikni kafolatlamaydi. Keyingi tadqiqot uchun spetsifikatsiya bo‘yicha mazmuniy qamrov, javob jarayonlari, ko‘zlangan tuzilma, baholovchilar mosligi va qarorlarning aniqligini alohida tekshirish rejalashtirilishi mumkin. O‘qitish ta’siridan kelgan haqiqiy o‘zgarishni qayta testdagi xato bilan aynanlashtirmaslik, yangi tanlanma yoki ajratilgan tekshiruv ma’lumotida xulosalarni baholash muhim metodik vazifadir.

Pedagoglar tayyorgarligida «testni tanqidiy ekspertizalash» mashg‘ulotini joriy etish taklif etiladi. Tinglovchi bir xil topshiriqdan kelib chiqadigan asosli va asossiz xulosalarni ajratadi, xato manbalarini belgilaydi, tekshiruv usulini tanlaydi va ball haqidagi hisobotni qayta yozadi. Uning faoliyati koeffitsiyentni yoddan aytish bilan emas, mazmun–dalil–qaror bog‘lanishini tushuntira olishi bilan baholanadi. Bu tavsiyaning natijadorligi nazoratli metodik sinov orqali aniqlanishi kerak.

XULOSA

Pedagogik testlarning sifatini ta’minlash validlik va ishonchlilikni bir ko‘rsatkichga birlashtirish emas, ularning turlicha vazifalarini izchil muvofiqlashtirishni talab etadi. Validlik muayyan ball talqini va foydalanish maqsadining daliliy asosini, ishonchlilik esa belgilangan sharoitlardagi o‘lchash aniqligini ifodalaydi. Mazmuniy moslik, aniqlik va adolatli qo‘llashdan birining kuchliligi boshqasidagi nuqsonni avtomatik bartaraf etmaydi.

Maqolaning natijasi «konstrukt–topshiriq–dalil–talqin–qaror» izchilligi, validlik dalillari matritsasi va xato manbasiga mos ishonchlilik profilining ishlab chiqilishidir.

Pedagogik tatbiq uchun spetsifikatsiyadan boshlangan, aprobatsiya va qaror monitoringi bilan davom etadigan jarayon taklif qilindi. Test natijasining ilmiy qiymati raqamning mavjudligida emas, qaysi xulosaga, qanday aniqlikda va qaysi chegaralarda xizmat qilishi ochiq asoslanishidadir. Modelni empirik sinash va mahalliy kontekstga moslashtirish keyingi ishning zarur bosqichidir.

FOYDALANILGAN ADABIYOTLAR

1. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: AERA, 2014.

2. Cronbach L.J., Meehl P.E. Construct validity in psychological tests // Psychological Bulletin. 1955. Vol. 52, No. 4. P. 281–302. DOI: 10.1037/h0040957

3. Kane M.T. Validating the interpretations and uses of test scores // Journal of Educational Measurement. 2013. Vol. 50, No. 1. P. 1–73. DOI: 10.1111/jedm.12000

4. Jalilov K.Q. Milliy test tizimi uchun test tuzuvchi mutaxassislarni tayyorlash masalalari (o‘zbek tilini bilish darajasi sertifikati misolida) // Axborotnoma. 2025. № 2. B. 54–64.

5. Livingston S.A. Test Reliability—Basic Concepts. ETS Research Memorandum RM-18-01. Princeton, NJ: Educational Testing Service, 2018. 6. Lee W., Harris D.J. Reliability in educational measurement // Cook L.L., Pitoniak M.J. (eds.). Educational Measurement. 5th ed. Oxford University Press, 2025. P. 277–381. DOI: 10.1093/oso/9780197654965.003.0005

7. Zitzmann S., Orona G.A. When Cronbach’s alpha does (not) indicate the reliability of domain-specific knowledge tests and why // Frontiers in Psychology. 2026. Vol. 17. Article 1796702. DOI: 10.3389/fpsyg.2026.1796702

8. International Test Commission. The ITC Guidelines for Translating and Adapting Tests. 2nd ed., version 2.4. 2017.

9. International Test Commission, Association of Test Publishers. Guidelines for Technology-Based Assessment. Version 1.1. 2025.