ҒТАМР 14.25.09 https://doi.org/10.63597/UTO3105-4161.2026.5.1.003

К.Т. Туралбаева*[1], Г.С. Джарасова[2], И.Ж. Жуматова[3], Г.Т. Сагнаева[4], Б.К. Асемова[5]

[1,3,4,5]ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана қ., Қазақстан Республикасы

[2]ҚР ҒЖБМ «Жоғары және жоғары оқу орнынан кейінгі білім комитеті», Астана қ., Қазақстан Республикасы

*e-mail: kengebahit@mail.ru

[1]ORCID 0009-0004-6640-0522, [2]ORCID 0000-0001-9736-0499, [3]ORCID 0009-0001-5382-4519, [4]ORCID 0009-0009-7651-508X, [5]ORCID 0009-0003-4857-4679

МАТЕМАТИКА ПӘНІ БОЙЫНША ТЕСТТІҢ ПСИХОМЕТРИЯЛЫҚ ТАЛДАУЫ: РАШ МОДЕЛІ НЕГІЗІНДЕ БАҒАЛАУ

Мақалада Ұлттық бірыңғай тестілеу (ҰБТ) шеңберінде математика пәні бойынша апробациядан өткен тест тапсырмаларына жүргізілген психометриялық талдау нәтижелері ұсынылады. Зерттеудің негізгі мақсаты – тесттің өлшеу сипаттамаларын бағалау, жеке тапсырмалардың жұмыс істеу ерекшеліктерін анықтау және тест нәтижелерінің сенімділігі мен валидтілігін қамтамасыз ету. Талдау барысында бір параметрлі Раш моделі (Rasch model) қолданылды. Бұл тәсіл тапсырмалардың күрделілігі мен қатысушылардың дайындық деңгейін бір шкалада қарастыруға мүмкіндік береді, сонымен қатар тесттің сапасын объективті және дәл бағалауға жағдай жасайды. Раш моделі тесттің әрбір элементінің қатысушылар арасындағы айырмашылықты қаншалықты дұрыс көрсететінін анықтауға көмектеседі. Зерттеу деректері ретінде математика пәні бойынша бір тест нұсқасының апробациясына қатысқан 326 тестіленушінің нәтижелері пайдаланылды. Мәліметтер WINSTEPS бағдарламасында (5.8.3.0 нұсқасы) өңделіп, тапсырмалардың күрделілігі, қатысушылардың қабілет деңгейі, модельге сәйкестік көрсеткіштері (Infit/Outfit MNSQ, ZSTD), дискриминативтілігі және тесттің сенімділігі анықталды. Нәтижелер тесттің жалпы алғанда сенімді екенін көрсетті. Сонымен қатар, жекелеген тапсырмалар бойынша мазмұндық және құрылымдық түзетулер қажет екендігі анықталды. Мұндай жетілдірулер тест нәтижелерінің дәлдігін арттыруға, әртүрлі дайындық деңгейіндегі қатысушыларды неғұрлым дұрыс саралауға және тестті қолдану тиімділігін жоғарылатуға мүмкіндік береді.

Түйін сөздер: ҰБТ, Раш моделі, психометриялық талдау, тапсырмалардың қиындығы, тесттің сенімділігі, валидтілік.

Кіріспе

ҰБТ тапсырмалары білім алушылардың тек пәндік білімін ғана емес, сонымен қатар олардың функционалдық сауаттылығын бағалауға бағытталған. Осыған байланысты тест мазмұны мен құрылымына қойылатын талаптар жыл сайын нақтыланып, жетілдіріліп отырады.

Қазіргі уақытта тестілеу жүйелері педагогикалық өлшеулер теориясына сүйенеді және білім сапасын объективті бағалаудың маңызды құралдарының бірі ретінде қарастырылады [1]. ҰБТ форматында қолданылатын тапсырмалар әртүрлі формада ұсынылады: бір дұрыс жауабы бар сұрақтар, бірнеше дұрыс жауабы бар тапсырмалар, контекстік тапсырмалар және сәйкестікті анықтауға арналған тапсырмалар.

Тапсырмалар базасын әзірлеу көпсатылы үдеріс болып табылады. Ол тапсырмаларды құрастырудан бастап, сараптамадан өткізу, түзету және апробациялау кезеңдерін қамтиды. Алайда тек сарапшылар пікірімен шектелу жеткіліксіз. Тест сапасын жан-жақты бағалау үшін нақты деректерге сүйенген талдау қажет [2].

Соңғы жылдары классикалық тест теориясымен қатар заманауи психометриялық тәсілдер, соның ішінде Раш моделі кеңінен қолданыла бастады. Бұл модель тестіленушінің

қабілеті мен тапсырманың қиындығы арасындағы байланысты бағалауға мүмкіндік береді және тест нәтижелерін интерпретациялаудың дәлдігін арттырады [2, 3].

Зерттеу материалдары мен әдістері

Зерттеу барысында Қазақстанның әртүрлі өңірлерінен қатысқан 326 тестіленушінің нәтижелері қарастырылды. Талдау математика пәні бойынша 40 тапсырмадан тұратын бір тест нұсқасының деректері негізінде жүргізілді.

Бір шкала бойынша сұрақтың қиындығы мен қатысушылардың шеберлік деңгейін бағалауға мүмкіндік беретін негізгі құрал ретінде Rasch моделі пайдаланылды.

Зерттеу объектісі мен контингенті:

- Республиканың әр түрлі өңірлерінен таңдалған 326 қатысушы.

Қолданылған құралдар мен материалдар:

- Математика пәні бойынша 40 тапсырмадан тұратын тест;
- Деректерді өңдеу үшін WINSTEPS бағдарламасы (5.8.3.0 нұсқасы) [3].

Статистикалық және психометриялық әдістер:

- Бір параметрлі Раш моделі (Rasch model) және оның политомды нұсқасы (тапсырмалар 0, 1, 2 балл бойынша бағаланған жағдайда) [4, 5];
- Тест элементтерінің күрделілігін, қатысушылар қабілетін бағалау;
- Модельге сәйкестік индекстері (Infit/Outfit MNSQ және ZSTD);
- Тапсырмалардың дискриминативтілігі (Pt-Measure Corr.);
- Тесттің сенімділігі (Reliability) және жіктелуі (Separation) [6, 7];

Деректерді өңдеу:

- Психометриялық деректерді компьютерлік бағдарлама арқылы талдау;
- Тесттің әр элементінің қалай жұмыс істейтінін бағалау, коррекция қажет тапсырмаларды анықтау [8, 9].

Нәтижелер және талқылау

Контингент пен шкаланың жалпы сипаттамасы. Қатысушылардың қабілет деңгейлерін және тапсырмалардың қиындығын таралуын талдау (1 кестесіндегі «Person-Item Map» картасын талдау) тесттің тепе-теңдігін анықтауға алғашқы маңызды қорытындылар жасауға мүмкіндік береді.

- Тапсырмалардың қиындық диапазоны (Item Measure): Тапсырмалар қиындықтың кең ауқымын қамтиды – (-1,86) мен (+1,71) логит аралығында. Бұл жақсы көрсеткіш, яғни тест өте жеңіл және өте қиын тапсырмаларды қамтиды.

- Қатысушылардың қабілет деңгейінің диапазоны (Person Measure): Сынақтан өткендер (-2,33) мен (+1,84) логит аралығында таралған, бұл топтың дайындық деңгейі бойынша әртүрлілігін айқындайды.

- Шкаланы ортаға келтіру: Қатысушылардың қабілет деңгейінің орташа көрсеткіші (Mean Person) (-0,36) логитті құрайды, ал тапсырмалардың орташа қиындығы (Mean Item) бастапқыда 0,00 деп белгіленген.

Осыған байланысты қатысушылардың таралуы сәл төмен қабілеттер жағына ығысқан. Бұл көрсеткіш берілген контингент үшін тесттің орташа деңгейде қиын болғанын білдіреді. Топ, жалпы алғанда, шкала бағытындағы «орташа гипотетикалық» студентке қарағанда тестті біршама нашар орындады.

- Шкала бойынша қамту: Тапсырмалар жалпы алғанда қатысушылардың көпшілігінің қабілет диапазонын жақсы қамтиды. Дегенмен, картаның визуалды талдауы жоғары қабілетті деңгейлерде (+1,5 логиттен жоғары) тапсырмалардың жеткіліксіз екенін көрсетеді. Ең мықты оқушылар үшін тест олардың нәтижелерін дәл ажыратуға жеткілікті мүмкіндік бермей тұр.

Сенімділік пен жіктелу бойынша талдау. Тесттің сенімділігін бағалау педагогикалық өлшеулердің негізгі талаптарының бірі болып табылады. Сенімділік көрсеткіштері өлшеу нәтижелерінің тұрақтылығын және тесттің әртүрлі топтарда қолданылған жағдайда ұқсас нәтижелер беру қабілетін сипаттайды [10].

Кронбах альфа коэффициенті тесттің ішкі үйлесімділігін бағалау үшін кеңінен қолданылатын статистикалық көрсеткіштердің бірі болып табылады [11].

1 кестесінен алынған тест сапасының негізгі көрсеткіштері төменде 1-кестеде берілген:

1-кесте - Тест сапасының негізгі көрсеткіштері

Көрсеткіш	Мәні	Интерпретация
Қатысушылардың сенімділігі (Person Reliability)	0,83 (Model)	Жақсы. Тест пәндерді білім деңгейі бойынша сенімді түрде ажыратады. Мұнда 0,8-ден жоғары мән топтық және жеке шешімдер қабылдау үшін қолайлы.
Тапсырмалардың сенімділігі (Item Reliability)	0,98 (Model)	Өте жоғары. Тапсырманың күрделілігінің иерархиясы өте тұрақты. Егер бұл тест ұқсас үлгіде қайталанса, тапсырманың күрделілігінің реті сақталады.
Қатысушылардың жіктелуі (Person Separation)	2,19	Орташа. Дайындық деңгейіне қарай шамамен 3-4 статистикалық тұрғыда ерекшеленетін қатысушылар тобын анықтауға мүмкіндік береді.
Тапсырмалардың жіктелуі (Item Separation)	7,74	Жоғары. Тапсырмалар қиындығы бойынша өте жақсы ерекшеленеді. Көрсеткіш тапсырмалардың қиындықтардың кең ауқымын қамтитынын растайды.
Кронбах коэффициенті (Cronbach's Alpha)	0,83	Жақсы ішкі үйлесімділік. Қатысушылардың сенімділік көрсеткішімен корреляцияланады.
Орташа балл (Raw Score)	27 / 50	54%. Бұл қорытынды тесттің таңдалған қатысушылар үшін біршама қиын болғанын растайды (нәтиже жалпы баллдың ортасынан асады)

Тапсырмалар жіктелуінің жоғары көрсеткіші (7,74) және өте жоғары сенімділік көрсеткіші (0,98) бұл құралдың қиындық шкаласы бойынша өте жақсы реттелгенін білдіреді, бұл тесттің күшті жақтарының бірі болып саналады.

Модельдің сәйкестігін тексеру (Infit/Outfit). Модельдің сәйкестігі Infit және Outfit статистикалары (MNSQ және ZSTD) арқылы тексерілді. Сапалы тестте MNSQ орташа мәндері 1,0-ге, ал ZSTD мәндері 0-ге жақын болуы тиіс.

- *Қатысушылар* (Persons): Infit/Outfit MNSQ орташа мәндері 1,01 және 1,02, ал ZSTD мәндері нөлге жақын. Бұл қатысушылардың модельге сәйкес әрекет ететінін көрсетеді. Дегенмен, Outfit MNSQ максималды мәндерінің 2,52-ге жетуі кейбір қатысушылардың болжанбаған жауап бергендігін білдіреді (мысалы, қабілетті оқушылар жеңіл сұрақтарда қателесуі мүмкін немесе әлсіз оқушылар қиын сұрақтарды кездейсоқ дұрыс жауабымен тапсыруы мүмкін).

- *Тапсырмалар* (Items): Орташа көрсеткіштер де идеалға жақын: Infit MNSQ = 1,00, Outfit MNSQ = 1,02. Орташа ZSTD мәндері (-0,23 және -0,09) «underfit» тенденциясының аздап бар екенін көрсетеді (яғни, шу көлемі жоғары, болжамдылығы төмен), бұл нақты деректер үшін жалпы қабылданады. Дегенмен, кейбір тапсырмалардың ZSTD мәндерінің 3,85-ке жетуі олардың жан-жақты талдануын қажет етеді.

Тапсырмалардың жұмыс жасау қабілетін талдау. Тапсырмалардың сапасын бағалау кезінде олардың қиындық деңгейі мен дискриминациялық қабілеті маңызды рөл атқарады. Бұл көрсеткіштер тест тапсырмаларының жоғары және төмен дайындық деңгейіндегі қатысушыларды ажырату қабілетін анықтайды [12].

Тапсырмалардың статистикалық көрсеткіштерін талдау классикалық тест теориясы мен заманауи психометриялық әдістердің үйлесімді қолданылуын талап етеді [13].

Талдаудың ең маңызды бөлігі – қорытынды бағалауды бұрмалайтын, күткеніміздей жұмыс жасамайтын тапсырмаларды анықтау. Тапсырмалар проблемалардың маңызды

деңгейіне қарай Infit/Outfit, ZSTD мәндері және нүктелік-бисериалдық корреляция (PT-measure) көрсеткіштері негізінде топтарға бөлінді. Тапсырмалар проблемалық деңгейіне қарай 2,3-кестелерде ұсынылады.

1) Тапсырмаларды проблемалық дәрежесіне қарай жіктеу

А ТОБЫ: КРИТИКАЛЫҚ МӘСЕЛЕЛЕР (Тез арада тексеруді немесе жоюды қажет етеді).

Бұл тапсырмалар тесттің жалпы тестімен бірдей айнымалыны өлшемейді немесе тапсырма шартында екіойлылық бар болуы мүмкін.

2-кесте - Проблемалық тапсырмалардың А-тобы

Тапсырма	Мәселе	Ұсыныс
MC-9	Outfit MNSQ = 1,53 (жоғары шу), ZSTD = 3,85, Pt-measure corr. = -0,01 (**теріс!**). Тапсырма «босқа» немесе зиянды жұмыс істейді: қабілеті жоғары оқушылар оны әлсіздерге қарағанда нашар орындайды.	Тапсырманы жан-жақты тексеру. Критикалық ақау.
MC-24	Infit ZSTD = -4,08, Outfit ZSTD = -4,11 (күшті overfit). Өте болжамды тапсырма. Немесе мәтінде дұрыс жауапты меңзейтін мәлімет бар.	Тапсырманы тексеру. Күрделендіру немесе ауыстыру мүмкіндігін қарастыру.
SOOTV-31	Infit ZSTD = 3,58, Outfit ZSTD = 3,48 (күшті underfit). Жоғары шу, болжамдылық төмен. Exact match (дәл сәйкес келу көрсеткіші) – 40,2%, күтілген деңгейден төмен (48,0%).	Тапсырма шартының анықтылығын тексеру.

В ТОБЫ: ОРТАША МӘСЕЛЕЛЕР (тексеру ұсынылады)

Тапсырмалар күткендегідей жақсы жұмыс істемейді, бірақ бұл критикалық емес.

3-кесте - Проблемалық тапсырмалардың В-тобы

Тапсырма	Мәселе	Ұсыныс
MC-14	Infit ZSTD = 2,34, Outfit ZSTD = 2,63, Pt-measure corr. = 0,16 (төмен).	Мүмкіндігінше дистракторларды тексеру.
MC-21	Infit ZSTD = 2,79, Pt-measure corr. = 0,19 (төмен).	Мүмкіндігінше тапсырма шарты тексеру.
CONT-30	Infit ZSTD = 2,00, Outfit ZSTD = 2,55 (шектелген жағдай).	Мүмкіндігінше мазмұнына талдау жүргізу.

2) Жоғары психометриялық қасиеттерге ие тапсырмалар

Тест тапсырмаларының көпшілігі (80%-дан астам) тамаша психометриялық қасиеттерді көрсетеді. Олардың ішінде: MC-1, MC-2, MC-3, MC-7, MC-10, MC-12, MC-15, MC-16, MC-17, MC-22, MC-23, MC-25, CONT-27, CONT-28, CONT-29, MC-39, MC-40 және т.б. Олар қабылданатын Infit/Outfit көрсеткіштеріне ие, жалпы баллмен жоғары корреляция көрсетеді (>0.3) және категориялардың шекті мәндері дұрыс реттелген.

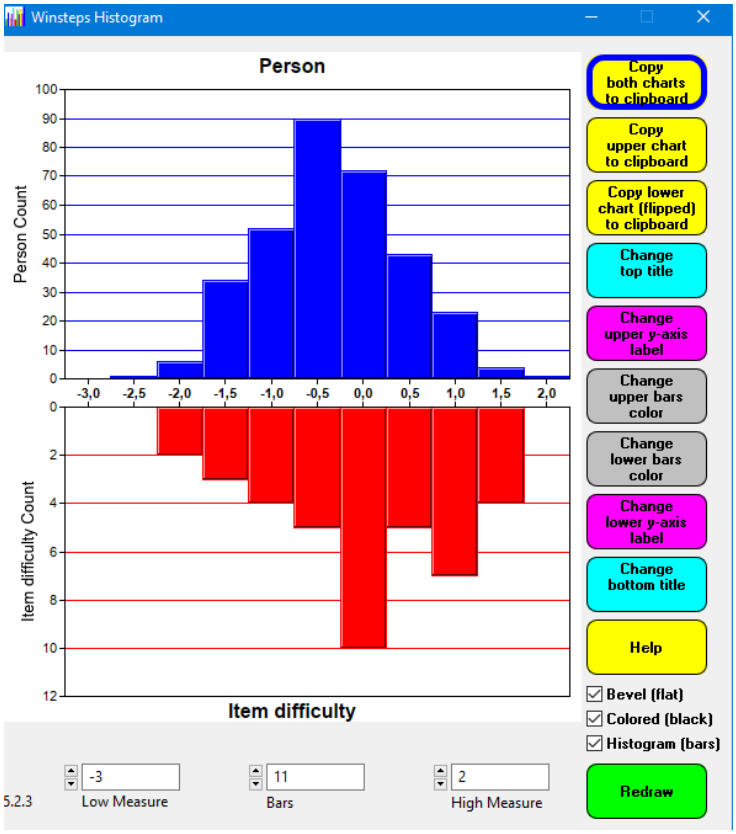
Дифференциалдық функционалдауды талдау (DIF және DPF). DIF (Differential Item Functioning) талдауы тест тапсырмаларының әртүрлі топтар үшін бірдей жұмыс істейтінін анықтауға мүмкіндік береді және бағалаудың әділдігін қамтамасыз етуде маңызды рөл атқарады [14].

DIF талдауы көрсеткендей, көптеген тапсырмаларда белгілі бір топқа артықшылық беретін жүйелі ауытқулар байқалмайды. Яғни, тест осы белгі бойынша әртүрлі оқушы топтары үшін әділ деп қарастыруға болады.

DPF (Differential Person Functioning): жауаптары ерекше экстремалды болған жеке оқушылар анықталды (мысалы, ID 320, 322). Белгілі бір типтегі тапсырмалар бойынша олардың нәтижелері (мысалы, С типі) жалпы модельден айтарлықтай ерекшеленеді, бұл

көшіріп жазу, немқұрайлылық немесе білімдегі нақты олқылықтарға байланысты болуы мүмкін.

Раш моделі бойынша гистограмманы сипаттау (Person-Item Histogram). Раш моделі аясында алынған қатысушылар мен тапсырмалардың таралу гистограммасы (1-сурет) көрсеткендей, негізгі тапсырмалар массасы тестке қатысушылардың дайындық деңгейіне қарағанда төмен қиындық деңгейінде орналасқан. Көптеген қатысушылар латенттік айнымалының орташа мәндері аймағында шоғырланған, ал тапсырмалар айтарлықтай теріс логиттер аймағына ығысқан, бұл жеңіл тапсырмалардың көлемі көбірек екенін көрсетеді.



1-сурет: Раш моделі бойынша гистограмма

Мұндай көрсеткіш тесттің төмен және орта деңгейдегі дайындықты тиімді өлшейтінін білдіреді, дегенмен жоғары дайындық деңгейі бар қатысушыларды айыра білуде шектеулі мүмкіндік бар екендігін анықтайды.

Ұсыныстар. Нәтижелерді түсіндіру кезінде тек статистикалық көрсеткіштерге емес, тапсырмалардың мазмұнына да назар аудару қажет. Бұл тәсіл тесттің валидтілігі мен сенімділігін арттыруға мүмкіндік береді.

Математика пәні бойынша тестті Раш моделі арқылы психометриялық талдау жүргізу нәтижесінде бірқатар маңызды ұсыныстар айқындалды.

- «Проблемалық» тапсырмаларды қайта қарау: МС-9 тапсырмасын қайта өңдеу (теріс корреляция көрсеткен);
- «Жоғары underfit» көрсеткіштері бар тапсырмаларға мазмұндық талдау жүргізу: SOOTV-31, МС-14, МС-21.
- «Күшті overfit» көрсеткіші бар тапсырмаларда (МС-24) болжамдылықтың бар-жоғын тексеру.

Психометриялық талдау нәтижелері бойынша тестті жетілдіру.

Математика пәні бойынша тестті Раш моделі арқылы психометриялық талдау жүргізу нәтижесінде қорытындылар жасалды, сонымен қатар кемшіліктерді қалпына келтіру бойынша жұмыстар жүргізілді.

1. **Тесттің жалпы сапасы**: Тест қатысушылар (0,83) мен тапсырмалардың (0,98) сенімділік көрсеткіштері бойынша жақсы әрі жоғары нәтижелер көрсетеді. Құрал жалпы математика білімін өлшеу және оқушыларды дайындық деңгейі бойынша дифференциациялау үшін жарамды.

2. **Модельге сәйкестік**: Деректер негізінен Раш моделіне сәйкес келеді, бұл Infit/Outfit орташа мәндерінің 1,0-ге жақын болуымен расталады.

3. **Шкала балансы**: Тест қатысушылардың қабілеттерінің орташа ауқымын жақсы қамтиды, бірақ өте жақсы оқитын оқушыларға арналған қиын тапсырмалардың аздығы байқалады және бұл қатысушылар тобының аздап жоғары қиындығы бар тапсырмаларға ығысуы байқалады.

Психометриялық талдау нәтижелеріне сүйене отырып, тесттегі анықталған кемшіліктерді жою және өлшеу сапасын арттыру бойынша кешенді жұмыс жүргізілді. Ең алдымен, МС-9 тапсырмасы бойынша күтілетін психометриялық көрсеткіштерге қатысты ауытқулар тіркелді, сондықтан ол қосымша тексеру мен нақтылауға жататын тапсырмалар тізіміне енгізілді. Айқын underfit белгілері бар бірнеше тапсырмалар (SOOTV-31, МС-14, МС-21) мазмұндық сараптамадан өтті, нәтижесінде тапсырма шартының құрылымы, нұсқаулықтар және дистракторларды құру логикасы нақтыланды, екіойлылықты жою және жауаптардағы кездейсоқ шуды азайту мақсатында өзгерістер енгізілді.

Overfit белгілері бар тапсырмалар (МС-24) үшін жасырын болжам мен артық қарапайымдықкатексеру жүргізілді. Нәтижесінде кейбір тапсырмалар күрделендірілді, ал кейбірі балама шарттармен алмастырылды, бұл өзгерістер өлшенетін құрылымды сақтау арқылы жасалды.

Қорытынды

Жүргізілген талдау нәтижелері қарастырылған тест нұсқасының жалпы алғанда жеткілікті деңгейде сенімді екенін көрсетті. Тапсырмалар мен қатысушылар көрсеткіштерінің үйлесімділігі тесттің құрылымы дұрыс қалыптастырылғанын аңғартады.

Жалпы алғанда, алынған деректер Раш моделін тест сапасын бағалау мен жетілдіруде тиімді қолдануға болатынын айқындайды. Алынған нәтижелер сынақтың жалпы жақсы өлшеу қасиеттеріне ие екенін және үміткерлердің дайындық деңгейін бағалау үшін пайдалануға болатынын көрсетеді.

Дегенмен, одан әрі нақтылауды қажет ететін кейбір міндеттер анықталды. Бұларды реттеу, әсіресе дайындық деңгейі жоғары қатысушылар үшін өлшемдердің дәлдігін жақсартады.

Әдебиеттер тізімі

1. Магомедов А. Г. Применение тестирования как инструмента педагогических измерений: возможности и ограничения // Современное педагогическое образование. - 2022. - № 9. - С. 62-67.

2. Crocker L., Algina J. Introduction to Classical and Modern Test Theory. - Mason, OH : Cengage Learning, 2008. - 540 p.

3. Linacre M. Practical Rasch Measurement. Tutorial 1. Analysis of dichotomous (true-false, multiple-choice) // Winsteps. - 2011. - July. - https://www.winsteps.com/tutorials.htm

4. Bond T. G., Fox C. M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. - 4th Edition. - New York : Routledge, 2021. - 520 p.

5. Wright B. D., Masters G. N. Rating Scale Analysis: Rasch Measurement. - Chicago : MESA Press, 1982. - 240 p.

6. Анализ влияния первичных баллов заданий на уровни подготовленности респондентов в дихотомической модели Раша на основе взвешенного метода максимального правдоподобия / Е. Б. Белов, М. В. Алексеев, Н. П. Китаев, А. И. Кучумов // Вестник Российского экономического университета имени Г. В. Плеханова. - 2019. - № 2 (104). - С. 169-180. - http://dx.doi.org/10.21686/2500-3925-2019-2-89-98

7. Engelhard G. Jr. Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences. - New York : Routledge, 2013. - 480 p.

8. Boone W. J., Staver J. R., Yale M. S. Rasch Analysis in the Human Sciences. - Dordrecht : Springer, 2014. - 270 p.

9. Bond T., Yan Z., Heene M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. - Routledge, 2020. - 540 p.

10. Корнилов С. А. Проблема надежности психологических шкал и ее значение для психодиагностики в научных и практических целях // Экспериментальная психология в России: традиции и перспективы / под ред. В. А. Барабанщикова, А. Л. Журавлева. - М. : Изд-во «Институт психологии РАН», 2010. - С. 101-105.

11. Cronbach L. J. Coefficient alpha and the internal structure of tests // Psychometrika. - 1951. - Vol. 16, № 3. - P. 297-334. - https://doi.org/10.1007/BF02310555

12. Evaluating Item Difficulty and Discrimination in Knowledge Tests // Food Safety Institute. - 2024. - Jan. 4. - https://foodsafety.institute/research-methodology/evaluating-item-difficulty-discrimination-tests/

13. Thompson N. Classical Statistics and their Interpretation // Assess. - https://assesshelp.zendesk.com/hc/en-us/articles/360037560611-Classical-Statistics-and-their-Interpretation

14. Карданова Е. Ю., Иванова А. Е. Психометрические исследования: современные методы и новые возможности для образования // Вопросы образования. - 2023. - № 3. - С. 8-19. - https://doi.org/10.17323/vo-2023-17951

15. Wright B. D., Stone M. H. Best Test Design. - Chicago : MESA Press, 1979. - 224 p.

К.Т. Туралбаева, Г.С. Джарасова, И.Ж. Жуматова, Г.Т. Сагнаева, Б.К. Асемова

ПСИХОМЕТРИЧЕСКИЙ АНАЛИЗ ТЕСТА ПО МАТЕМАТИКЕ: ОЦЕНКА НА ОСНОВЕ МОДЕЛИ РАША

В статье представлены результаты комплексного психометрического анализа заданий по предмету «Математика», прошедших апробацию в рамках Единого национального тестирования (ЕНТ). Цель исследования заключалась в оценке измерительных свойств теста, анализе функционирования 40 заданий, а также проверке степени соответствия эмпирических данных теоретической модели и надежности теста. Для проведения анализа использовалась однопараметрическая модель Раша (Rasch model), которая позволяет учитывать сложность заданий и уровень подготовки участников на единой шкале, обеспечивая объективную оценку качества теста. В исследовании были использованы данные 326 участников, которые проходили апробацию одного варианта теста по математике. Психометрическая обработка данных осуществлялась с помощью специализированной программы WINSTEPS (версия 5.8.3.0), что позволило вычислить основные метрики теста, включая показатели сложности заданий, способности участников, соответствие модели (Infit/Outfit MNSQ и ZSTD), дискриминативность заданий и надежность инструмента. Результаты анализа показали высокий уровень общей надежности теста, что подтверждает его пригодность для оценки функциональной грамотности абитуриентов. Вместе с тем, выявлены отдельные задания, требующие содержательной корректировки и совершенствования, особенно для более подготовленных участников. Улучшение этих заданий позволит повысить точность измерений и эффективность тестирования на всех уровнях подготовки.

Ключевые слова: ЕНТ, модель Раша, психометрический анализ, трудность заданий, надежность теста, валидность.

K. Turalbaeva, G. Jarassova, I. Zhumatova, G.Sagnaeva, B. Assemova

PSYCHOMETRIC ANALYSIS OF A MATHEMATICS TEST: EVALUATION BASED ON THE RASCH MODEL

The article presents the results of a comprehensive psychometric analysis of tasks in the subject «Mathematics», which were piloted within the framework of the Unified National Testing (UNT). The purpose of the study was to evaluate the measurement properties of the test, analyze the functioning of 40 tasks, and assess the degree to which empirical data correspond to the theoretical model, as well as the reliability of the test. The analysis was conducted using the one-parameter Rasch model, which allows for accounting for task difficulty and participant ability on a single scale, providing an objective assessment of test quality. The study utilized data from 326 participants who completed the pilot of a single version of the mathematics test. Psychometric data processing was carried out using the WINSTEPS software (version 5.8.3.0), which enabled calculation of key test metrics, including task difficulty, participant ability, model fit (Infit/Outfit MNSQ and ZSTD), task discriminative power, and instrument reliability. The analysis results demonstrated a high overall level of test reliability, confirming its suitability for assessing the functional literacy of students. At the same time, several tasks were identified that require content refinement and improvement, particularly for participants with higher levels of preparation. Enhancing these tasks will increase measurement precision and improve the effectiveness of testing across all levels of participant ability.

Keywords: UNT, Rasch model, psychometric analysis, item difficulty, test reliability, validity.

References

1. Magomedov, A. G. (2022). Primenenie testirovaniya kak instrumenta pedagogicheskikh izmereniy: vozmozhnosti i ogranicheniya [Application of testing as a tool for pedagogical measurements: opportunities and limitations]. *Sovremennoe pedagogicheskoe obrazovanie*, (9), 62–67.

2. Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.

3. Linacre, M. (2011, July). *Practical Rasch measurement. Tutorial 1. Analysis of dichotomous (true-false, multiple-choice)*. Winsteps. https://www.winsteps.com/tutorials.htm

4. Bond, T. G., Fox, C. M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge.

5. Wright, B. D., Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. MESA Press.

6. Belov, E. B., Alekseev, M. V., Kitaev, N. P., & Kuchumov, A. I. (2019). Analiz vliyaniya pervichnykh ballov zadaniy na uровni podgotovlennosti respondentov v dikhotomicheskoy modeli Rasha na osnove vzveshennogo metoda maksimal'nogo pravdopodobiya [Analysis of the influence of primary item scores on respondents' ability levels in the dichotomous Rasch model based on the weighted maximum likelihood method]. *Vestnik Rossiyskogo ekonomicheskogo universiteta imeni G. V. Plekhanova*, (2), 169–180. http://dx.doi.org/10.21686/2500-3925-2019-2-89-98

7. Engelhard, G., Jr. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. Routledge.

8. Boone, W. J., Staver, J. R., Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.

9. Bond, T., Yan, Z., Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge.

10. Kornilov, S. A. (2010). Problema nadezhnosti psikhologicheskikh shkal i ee znachenie dlya psikhodiagnostiki v nauchnykh i prakticheskikh tselyakh [The problem of reliability of psychological scales and its significance for psychodiagnostics for scientific and practical purposes]. In V. A. Barabanshchikov & A. L. Zhuravlev (Eds.), *Eksperimental'naya psikhologiya v Rossii: traditsii i perspektivy* (pp. 101-105). Institute of Psychology RAS.

11. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. https://doi.org/10.1007/BF02310555

12. Food Safety Institute. (2024, January 4). *Evaluating item difficulty and discrimination in knowledge tests*. https://foodsafety.institute/research-methodology/evaluating-item-difficulty-discrimination-tests/

13. Thompson, N. (n.d.). *Classical statistics and their interpretation*. Assess. https://assesshelp.zendesk.com/hc/en-us/articles/360037560611-Classical-Statistics-and-their-Interpretation

14. Kardanova, E. Y., & Ivanova, A. E. (2023). Psikhometricheskie issledovaniya: sovremennye metody i novye vozmozhnosti dlya obrazovaniya [Psychometric research: Modern methods and new opportunities for education]. *Voprosy obrazovaniya / Educational Studies Moscow*, (3), 8–19. https://doi.org/10.17323/vo-2023-17951

15. Wright, B. D., & Stone, M. H. (1979). *Best test design*. MESA Press.

Сведения об авторах:

Туралбаева Кенжекул Турганбаевна (автор-корреспондент) – заведующий лабораторией, РГП на ПХВ «Национальный центр тестирования» МНВО РК, Астана, Казахстан, e-mail*:* kengebahit@mail.ru

Джарасова Гульжан Сагидуллаевна – кандидат педагогических наук, председатель, Комитет высшего и послевузовского образования, Астана, Казахстан, e-mail: yulzhan@mail.ru

Жуматова Индира Жаналыковна – заведующий лабораторией, РГП на ПХВ «Национальный центр тестирования» МНВО РК, Астана, Казахстан, e-mail*:* i.zh.86@bk.ru

Сагнаева Гульнар Туяковна – главный эксперт, РГП на ПХВ «Национальный центр тестирования» МНВО РК, Астана, Казахстан, e-mail: cgulnar_t@mail.ru

Асемова Бибіфатима Кәкімжанқызы – ведущий эксперт РГП на ПХВ «Национальный центр тестирования» МНВО РК, Астана, Казахстан, e-mail: assemova00@mail.ru

Авторлар туралы мәлімет:

Туралбаева Кенжекул Турганбаевна (автор-корреспондент) – зертхана меңгерушісі, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана, Қазақстан, e-mail: kengebahit@mail.ru

Джарасова Гульжан Сагидуллаевна – педагогика ғылымдарының кандидаты, Жоғары және жоғары оқу орнынан кейінгі білім комитетінің төрағасы, Астана, Қазақстан, e-mail: yulzhan@mail.ru

Жуматова Индира Жаналыковна – зертхана меңгерушісі, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана, Қазақстан, e-mail: i.zh.86@bk.ru

Сагнаева Гульнар Туяковна - бас сарапшы, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана, Қазақстан, e-mail:cgulnar_t@mail.ru

Асемова Бибіфатима Кәкімжанқызы – жетекші сарапшы, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана, Қазақстан, assemova00@mail.ru

Information about authors:

Turalbayeva Kenzhekul Turganbayevna (corresponding author) – Head of laboratory, Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Astana, Kazakhstan, e-mail: kengebahit@mail.ru

Jarassova Gulzhan Sagidullaevna – Candidate of Pedagogical Sciences, Chairman of the Committee of Higher and Postgraduate Education, Astana, Kazakhstan, e-mail: yulzhan@mail.ru

Zhumatova Indira Zhanalykovna – Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Head of laboratory, Astana, Kazakhstan, e-mail: i.zh.86@bk.ru

Sagnaeva Gulnara Tuyakovna – Chief Expert, Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Astana, Kazakhstan,e-mail: cgulnar_t@mail.ru

Assemova Bibifatima Kakimzhanovna – Senior expert, Republican State Enterprise on the Right of Economic Management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Astana, Kazakhstan, e-mail: assemova00@mail.ru