ULTTYQ TESTILEÝ ORTALYǴY

ISSN: 3105-4161 (Online)

ПЕДАГОГИКАЛЫҚ ӨЛШЕМДЕР

ҒЫЛЫМИ-ПРАКТИКАЛЫҚ ЖУРНАЛЫ

№3/2025

ПЕДАГОГИЧЕСКИЕ ИЗМЕРЕНИЯ

НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ

PEDAGOGICAL MEASUREMENTS

JOURNAL OF RESEARCH AND PRACTICE

Қазақстан Республикасының Ғылым және жоғары білім министрлігі «Ұлттық тестілеу орталығы» шаруашылық жүргізу құқығындағы республикалық мемлекеттік кәсіпорны

«ПЕДАГОГИКАЛЫҚ ӨЛШЕМДЕР»

ҒЫЛЫМИ-ПРАКТИКАЛЫҚ ЖУРНАЛЫ

Республиканское государственное предприятие на праве хозяйственного ведения «Национальный центр тестирования» Министерства науки и высшего образования Республики Казахстан

НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ

«ПЕДАГОГИЧЕСКИЕ ИЗМЕРЕНИЯ»

Republican state enterprise under the right of economic management «National Testing Center» of the Ministry of Science and Higher Education of the Republic of Kazakhstan

JOURNAL OF RESEARCH AND PRACTICE

«PEDAGOGICAL MEASUREMENTS»

№3(3) / 2025

Жылына 4 рет шығарылады
2025 жылдан бастап шығады

Астана
2025

ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК

«ПЕДАГОГИКАЛЫҚ ӨЛШЕМДЕР»

ғылыми-практикалық журналы

Бас редакторы:
Абдрасилов Б.С.
ҚР ҒЖБМ Ұлттық тестілеу орталығының директоры, ҚР ҰҒА академигі

Жауапты редактор:
Шинетова Л.Е.

Редакция алқасы:
Ниязов Т.Ж. - э.ғ.д., Ұлттық тестілеу орталығы
Байжанов Н.А. - з.ғ.к., ETS стажері, АҚШ
Дидарбекова Н.А. - ф.ғ.к., Ұлттық тестілеу орталығы
Танирбергенова А.Ш. - п.ғ.к., Ы.Алтынсарин атындағы Ұлттық білім академиясы
Калкеева К.Р. - п.ғ.к., профессор, Л.Н. Гумилев атындағы Еуразия ұлттық университеті
Абаева Г.А. - п.ғ.к., доцент, Абай атындағы Қазақ ұлттық педагогикалық университеті
Абсатова М.А. - п.ғ.к., процессор, Абай атындағы Қазақ ұлттық педагогикалық университеті
Карданова Е. Ю. - ф.-м.ғ.к., доцент, «Экономика жоғары мектебі» Ғылыми-зерттеу университеті, Ресей
Ширинов Е. - Әзірбайжан Республикасының мемлекеттік емтихан орталығы
Patrick Kyllonen - PhD, ETS, АҚШ
Nico Dieteren - Dutchtestologist, Нидерланды
Huseyn Vali - AQA Global Assessment Services, Ұлыбритания
Temur Nezir - PhD, профессор, Гази университеті, Түркия
Matthew Courtney - PhD, Назарбаев Университет

Техникалық редактор:
Бейсенова Ж.Ж.

«Педагогикалық өлшемдер» ғылыми-практикалық журналы. – 2025. – 3 т. – 96 б.

Редакцияның мекенжайы: 010000, Қазақстан Республикасы, Астана қ., Родниковая к-сі, 1/1, тел. 8(7172)799-889, і.н.157, e-mail: naukauto24@gmail.com, сайт: https://testcenter-zhurnal.kz/.

Қазақстан Республикасы Мәдениет және ақпарат министрлігімен тіркелген. №KZ54VPY00116912 есепке қою туралы куәлігі.

© ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК

МАЗМҰНЫ / СОДЕРЖАНИЕ / CONTENT

БІЛІМ БЕРУ НӘТИЖЕЛЕРІН БАҒАЛАУДЫҢ ЗАМАНАУИ ТӘСІЛДЕРІ ЖӘНЕ БІЛІМ БЕРУ САПАСЫН БАСҚАРУ / СОВРЕМЕННЫЕ ПОДХОДЫ К ОЦЕНКЕ ОБРАЗОВАТЕЛЬНЫХ РЕЗУЛЬТАТОВ И УПРАВЛЕНИЕ КАЧЕСТВОМ ОБРАЗОВАНИЯ / MODERN APPROACHES TO EDUCATIONAL OUTCOMES ASSESSMENT AND EDUCATION QUALITY MANAGEMENT

К.Т.Туралбаева, Т.Ж. Ниязов, Б.К. Асемова
ҰЛТТЫҚ БІРЫҢҒАЙ ТЕСТІЛЕУДЕГІ МАТЕМАТИКА ПӘНІ БОЙЫНША СӘЙКЕСТІКТІ АНЫҚТАУҒА АРНАЛҒАН ТЕСТ ТАПСЫРМАЛАРЫ…….…….. 6

Ж.Ж. Бейсенова, Л.Е. Шинетова, Г.К. Жабаева
ПСИХОМЕТРИЧЕСКИЙ АНАЛИЗ ТЕСТОВЫХ ЗАДАНИЙ ПО ОЦЕНКЕ ЗНАНИЙ ПЕДАГОГОВ В РЕСПУБЛИКЕ КАЗАХСТАН (ПО РЕЗУЛЬТАТАМ АПРОБАЦИИ)…….. 20

Ұ. Садық, А. К. Макыбаева, Н. Б. Нуримбетов
ҚАЗТЕСТ ЖҮЙЕСІНДЕГІ ДИАГНОСТИКАЛЫҚ ТЕСТІЛЕУДІҢ ӨТКІЗІЛУ БАРЫСЫ МЕН ТӘЖІРИБЕСІ..... 30

И.У. Сагиндиков, А.Ш. Танирбергенова, А. Мағауия
БІЛІМ БЕРУ ҰЙЫМДАРЫНДАҒЫ ТӘРБИЕ ПРОЦЕСІН БАҒАЛАУДЫҢ ЭКОЖҮЙЕЛІК ТӘСІЛДЕМЕСІ..... 40

Ш.Б. Алтыбаева, А.К. Каркенова, М.Д. Каргалдаева
ПСИХОМЕТРИЧЕСКАЯ ВЕРИФИКАЦИЯ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ УПРАВЛЕНИЯ БАЗОЙ ТЕСТОВЫХ ЗАДАНИЙ……..... 55

D. Chuteneva
GENDER, MULTILINGUALISM, AND CAREER CHOICE IN KAZAKHSTAN: AN INTERSECTIONAL VIEW OF UNDERGRADUATE ASPIRATIONS……... 68

Джафарзода С.Ф., Шодиниё А.Ш., Мирзоаминзода Х.М.
РАЗВИТИЕ НАЦИОНАЛЬНОЙ СИСТЕМЫ ТЕСТИРОВАНИЯ В РЕСПУБЛИКЕ ТАДЖИКИСТАН…… 76

ЖАСАНДЫ ИНТЕЛЛЕКТ ЖӘНЕ БІЛІМ БЕРУДІ БАҒАЛАУДАҒЫ ИННОВАЦИЯ / ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ИННОВАЦИИ В ОЦЕНИВАНИИ ОБРАЗОВАНИЯ / ARTIFICIAL INTELLIGENCE AND INNOVATIONS IN EDUCATIONAL ASSESSMENT

N. Baizhanov, A. Sabyrov, A. Makhmutova, S. Kadyrov
LARGE LANGUAGE MODELS IN EDUCATIONAL MEASUREMENT OF KAZAKH LANGUAGE PROFICIENCY……..... 84

БІЛІМ БЕРУ НӘТИЖЕЛЕРІН БАҒАЛАУДЫҢ ЗАМАНАУИ ТӘСІЛДЕРІ ЖӘНЕ БІЛІМ БЕРУ САПАСЫН БАСҚАРУ

СОВРЕМЕННЫЕ ПОДХОДЫ К ОЦЕНКЕ ОБРАЗОВАТЕЛЬНЫХ РЕЗУЛЬТАТОВ И УПРАВЛЕНИЕ КАЧЕСТВОМ ОБРАЗОВАНИЯ

MODERN APPROACHES TO EDUCATIONAL OUTCOMES ASSESSMENT AND EDUCATION QUALITY MANAGEMENT

ҒТАМР 14.15.25 https://doi.org/10.63597/UTO3105-4161.2025.3.3.001

К.Т. Туралбаева[1], Т.Ж. Ниязов[2], Б.К. Асемова[3]
ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана қ, Қазақстан Республикасы
*e-mail: kengebahit@mail.ru

[1]ORCID 0009-0004-6640-0522, [2]ORCID 0009-0003-3046-3532, [3]ORCID 0009-0003-4857-4679

ҰЛТТЫҚ БІРЫҢҒАЙ ТЕСТІЛЕУДЕГІ МАТЕМАТИКА ПӘНІ БОЙЫНША СӘЙКЕСТІКТІ АНЫҚТАУҒА АРНАЛҒАН ТЕСТ ТАПСЫРМАЛАРЫ

Зерттеуде Ұлттық бірыңғай тестілеу шеңберінде тест тапсырмаларын енгізу мәселелері қарастырылады. Негізгі назар тестілердің құрылымы мен мазмұнын жақсартуға аударылған, бұл студенттердің білімдерін тереңірек бағалауға және олардың логикалық ойлауын дамытуға бағытталған. Зерттеу барысында қолданыстағы тестілеу әдістері талданып, арнайы пәндер бойынша білім тереңдігін бағалауға бағытталған жаңа тапсырмалар әзірленіп, енгізілді. Мақалада мультимедиялық элементтері бар тест тапсырмалары сәйкестікті анықтауға арналған жаңа тест тапсырмалары қарастырылады. Сәйкестікті анықтауға арналған тест тапсырмалары ақпаратты талдау және салыстыру дағдыларын тексеру, сыни ойлауды дамыту және элементтер арасындағы логикалық байланыстарды орнату қабілетін бағалау мақсатында қосылған. Сәйкестікті анықтауға арналған тест тапсырмалары фактілерді білуді ғана емес, олардың арасындағы өзара байланысты түсінуді де тексеруге мүмкіндік береді, бұл оқушылардың білімін кешенді бағалауға және ақпаратты әртүрлі контекстерде қолдана алу қабілетін дамытуға ықпал етеді. Авторлар математика пәні бойынша сәйкестікті анықтауға арналған тест тапсырмаларын сипаттап, олардың білім деңгейін бағалау жүйесіндегі нақты рөлін ашады. Сонымен қатар, мақалада бұл тапсырмалардың заманауи білім беру жүйесіндегі маңызы мен қажеттілігіне ерекше назар аударылады, олардың оқыту мен бағалаудың жаңа тәсілдеріне ықпал ететіндігі көрсетіледі.

Жаңа форматтағы тест тапсырмаларын апробациялау нәтижелері тестіленушілердің білім тереңдігі мен логикалық дағдыларын бағалауда жоғары тиімділікті көрсетті. Сонымен қатар, тестілердің сенімділігі, күрделілігі, корреляциялық сипаттамалары және басқа да көрсеткіштерді қамтитын статистикалық деректер талданды. Алынған деректер негізінде тестілеуді әрі қарай жақсарту бойынша ұсыныстар жасалды, бұл Ұлттық бірыңғай тестілеуде бағалаудың объективтілігі мен тиімділігін арттыруға көмектеседі.

Түйін сөздер: ҰБТ, тест тапсырмасы, мультимедиялық элементтер, тестіленуші, бағалау, математика, спецификация, тест тапсырмасын әзірлеу, сәйкестікті анықтау.

Кіріспе

Білім беру саласында соңғы жылдары оқыту үдерісін жаңғыртуға бағытталған реформалар қарқынды жүргізілуде. Бұл өзгерістердің маңызды құрамдас бөлігі ретінде білім алушылардың дайындық деңгейін бағалаудың тиімді әдістерін енгізу мен жетілдіруге ерекше назар аударылуда. Осы тұрғыда тестілеу технологиясы мен тестілеу жүйесі оқу жетістіктерін бағалаудың маңызды құралы ретінде кеңінен қолданылып келеді. Оның ішінде ҰБТ – Қазақстан Республикасында жоғары оқу орындарына түсу кезінде талапкерлердің білімін кешенді бағалайтын негізгі механизмдердің бірі [1].

Ұлттық бірыңғай тестілеуді өткізу қағидаларына сәйкес, талапкерлер Қазақстан тарихы, математикалық сауаттылық, оқу сауаттылығы (оқыту тілі бойынша) және шығармашылық дайындықты қажет ететін білім беру бағдарламаларына түсушілерден басқа барлық талапкерлер екі бейіндік пән бойынша ҰБТ тапсырады [2].

ҰБТ форматы халықаралық тәжірибені ескере отырып үнемі жетілдіріліп отырады. Соның ішінде, Дүниежүзілік банк жобасының шеңберінде халықаралық консультанттардың ұсынылған реформалар негізінде тест құрылымын жаңарту, оның ішінде мультимедиялық элементтері бар жаңа тапсырмалар енгізу көзделген.

Жүргізілген зерттеу аясында ҰБТ-ға сәйкестікті анықтауға арналған тест тапсырмаларын енгізу бойынша талдау жүргізілді. Аталған жобаның шеңберінде халықаралық консультанттармен бірлесіп жүргізілген зерттеулер мен олар негізінде әзірленген ұсыныс есепке алынып, сараптама жұмыстары орындалды. Бұл зерттеу мен талдау нәтижесінде тест тапсырмаларының құрылымы мен мазмұнын жетілдіруге қатысты нақты ұсыныстар әзірленіп, тестілеудің тиімділігін арттыру мақсатында қажетті өзгерістер енгізілді. Бұл өзгерістер білім алушылардың пәндік білімін тереңірек бағалауға және олардың логикалық ойлау қабілеттерін дамытуға бағытталған. Бұл ұсыныстарды сатылы түрде жүзеге асыру мақсатында 2024 жылғы ҰБТ форматында біршама өзгерістер орын алды. ҰБТ-ның бейінді пәндері бойынша жаңа формадағы тест тапсырмалары, нақты айтқанда, сәйкестікті анықтауға арналған 5 тест тапсырмасы енгізілді [3].

ҰБТ-ға сәйкестікті анықтауға арналған тест тапсырмаларын енгізу алдында кешенді ақпараттық және түсіндірме іс-шаралары жүргізілді. Осы форматтағы тест тапсырмаларын қамтитын байқау сынақтары кең аудиторияға ұсынылып, білім беру ұйымдарында білім алушыларды қамти отырып, тест тапсырмаларының апробациясы бірнеше рет жүзеге асырылды. Жүргізілген зерттеулер мен талдаулардың нәтижесінде сәйкестікті анықтауға арналған тест тапсырмаларының құрылымы мен мазмұнын жетілдіруге бағытталған ұсыныстар мен тәсілдер айқындалды. Жаңа форматтағы тест тапсырмалары түсуішілердің бейіндік пән бойынша білімінің тереңдігі мен толықтығын бағалауға мүмкіндік береді, бұл зерттеу нәтижелерінде ұсынылған тәжірибелік деректермен дәлелденген.

Зерттеу әдістері

Зерттеу жүргізу барысында сәйкестікті анықтауға арналған тест тапсырмаларының ерекшеліктері мен қасиеттерін толық ашып көрсету мақсатында зерттеудің талдау әдісі қолданылады. Талдаудың негізгі міндеті – тапсырмалардың оң және сол жақ жиындарын салыстырып, олардың ортақ қасиеттерін және арасындағы ерекшеліктерді анықтау. Осылайша, жаңа форматтағы тест тапсырмаларының әртүрлі аспектілерін тереңірек зерттеу мүмкіндігі туындайды, бұл өз кезегінде ғылыми тұрғыдан құнды ақпаратты алуға мүмкіндік береді.

Зерттеу барысында талдау әдісінен басқа да негізгі ғылыми әдістер қолданылды, атап айтқанда статистикалық талдау, жалпылау және бағалау әдістері. Бұл әдістер тест тапсырмаларын қарастыру кезінде кеңінен қолданылып, олардың тиімділігін арттыруға септігін тигізді.

Ағымдағы зерттеуде сандық және сапалық деректер негізгі және қосымша ақпарат ретінде қолданылады. Зерттеу Ұлттық тестілеу орталығының базасында ҰБТ-ға арналған жаңа форматта қалыптастырылған тест тапсырмалары қолданылған тест нәтижесі бойынша жүргізілді. Сандық деректерге математика пәні бойынша тест тапсырмаларының және тестілеуге қатысқан түлектердің саны кіреді. Сапалық мәліметтер тест нәтижелерін математикалық және статистикалық өңдеу әдістерімен талдау барысында алынған тест тапсырмаларының сапасын бағалау нәтижелерімен қамтылады.

Зерттеу барысында апробация нәтижесі бойынша математикадан сәйкестікті анықтауға арналған тапсырмаларының сапасы зерттеледі. Апробация нәтижесінде тест тапсырмаларының жалпы статистикалық көрсеткіштері (сенімділік, тапсырмалардың қиындық деңгейі, корреляциялық сипаттамалар және басқа да көрсеткіштер) және жеке тапсырмалардың

ерекшеліктері (қиындық деңгейі, дифференциациялық қабілет және т.б.) талданды. Сонымен қатар, сараптамалық бағалау арқылы тапсырмалардың дұрыстығы мен ҰБТ талаптарына сәйкестігі тексерілді. Тесттің валидтілігі міндетті түрде зерттеліп, оның нақты талаптарға сәйкес нәтиже беруі бағаланды. Талдау нәтижелері бойынша тест тапсырмалары қайта өңделіп, кейбір тапсырмалар жойылуы мүмкін. Осы зерттеудің қорытындысы бойынша математика пәні бойынша тестілеуді ҰБТ жүйесінде қолдану мүмкіндігіне қатысты шешімдер қабылданады.

Сәйкестікті анықтауға арналған тест тапсырмалары: құрылымы, мақсаттары және бағалау әдістемелері

Сәйкестікті анықтауға арналған тест тапсырмалары – бұл бір жиынның элементтеріне басқа жиынның элементтерін сәйкестендіріп қойып шығуды талап ететін тапсырмалар [4]. Сәйкестікті анықтауға арналған тест тапсырмалары өзін-өзі бақылау жұмысында және ағымдағы білім деңгейін бақылау кезінде де қолайлы. Мұндай түрдегі тапсырмалар әрбір оқу пәнінде болатын жинақты білімді тексеруге мүмкіндік береді.

Сәйкестікті анықтауға арналған тест тапсырмаларының көптеген параллель концепциялары бар, мысалы:

- Терминдер және анықтамалар
- Нысандар және бейнелер
- Себептер және салдарлар
- Оқиғалар және даталар
- Заңдылықтар және формулалар
- Болмыстар және құбылыстар
- Заттар және қасиеттер
- Шамалар және өлшемдер
- Қағидалар және сипаттамалар
- Авторлар және шығармалар т.б.

Сәйкестікті анықтауға арналған тест тапсырмалары арқылы түсушінің пәндік білімі ғана емес, жоғарыда тізімде ұсынылған концепциялар арасындағы қарым-қатынасты логикалық тұрғыда байланыстыра білу икемділігі тексеріледі.

Сәйкестікті анықтауға арналған тест тапсырмалары түсушілердің ұқсас заттар/категориялар/түсініктер арасындағы байланысты немесе бірлестіктерін анықтау қабілетін, сонымен қатар басқа да кең көлемдегі дағдылары мен біліктерін бағалайды. Осыған орай сәйкестікті анықтауға арналған тест тапсырмаларын қолданудағы басты мақсат – оқушының түрлі жиынтықтар арасындағы байланыстар мен бірлестіктерді анықтай алу деңгейін бағалау болып табылады. Мектеп оқушыларына арналған жаңа бағдарлама мазмұнына сәйкес бөлім бойынша жиынтық бағалау (БЖБ) және тоқсандық жиынтық бағалау (ТЖБ) кезінде сәйкестікті анықтауға арналған тест тапсырмалары жиі қолданылады [5]. Педагогикалық тұрғыдан мұндай тапсырмаларды қолданудың мәні – оқушылардың оқу әрекетін белсендіру, сәйкестендірілетін элементтердің ассоциацияларын нығайту және құбылыстардың өзара байланысын түсінуді жақсарту.

Сәйкестікті анықтауға арналған тест тапсырмалары физикалық, математикалық және оларды іс жүзінде қолдану аясындағы басқа да формулаларды жинақтау үшін өте пайдалы. Кез-келген оқу пәнінде зерттелетін объектілердің (ұғымдар, шамалар) түрге, топқа, типке немесе т.б. бөлінетіндігі немесе жіктелетіндігі туралы мәліметтер кездеседі. Осы топтардың әрқайсысы үшін көптеген қасиеттер мен сипаттамалар, қағидалар, қолдану ережелері мен мөлшерлері бар, сондықтан осы терминдер мен олардың сипаттамаларын пайдалана отырып сәйкестендіруге арналған тест тапсырмаларын әзірлеуге мүмкіндік бар. Нақты бір білім саласындағы білімді, икемділік пен машықтануды тез бағалау мүмкіндігі және тапсырма элементтерінің үнемді орналасуы сәйкестікті анықтауға арналған тест тапсырмаларының басты артықшылығы болып

саналады. Сәйкестікті анықтауға арналған тест тапсырмалары бір дұрыс жауапты тест тапсырмаларына қарағанда барынша ұтымды болады. Мысалы, бір дұрыс жауапты таңдауға арналған біртекті тест тапсырмаларының бірнешеуін алып, олардан сәйкестікті анықтауға арналған бір ғана тест тапсырмасын жасау ұтымдырақ [6].

Сәйкестікті анықтауға арналған тест тапсырмаларында екі жиынға да ортақ шарт ұсынылады. Сондай-ақ әр жиындағы элементтердің біркелкілігін сақтау міндеттеледі. Мұндай тапсырмаларды бағалаудың әртүрлі нұсқалары бар. Оның ішінде ең тиімдісі: барлық дұрыс сәйкестендірілген тапсырма үшін – 2 балл, ал бір қате жіберілген жағдайда – 1 балл, екі немесе одан да көп сәйкессіздік орын алған жағдайда 0 балл беріледі.

ҰБТ форматы бойынша сәйкестікті анықтауға арналған тест тапсырмаларын бағалау келесі түрде жүзеге асырылады [1]:

- Бір дұрыс сәйкестік анықталса – 1 балл;
- Екі дұрыс сәйкестік анықталса – 2 балл;
- Дұрыс сәйкестік анықталмаса – 0 балл.

ҰБТ электронды форматта өтетіндіктен, сәйкестікті анықтауға арналған тест тапсырмалары мультимедиялық элементі бар тапсырмалар ретінде енгізілген. Тестіленушілерге тапсырманың сол жақ элементтері толық көрініп тұрады, ал оң жақ элементтер ашылмалы тізім ретінде беріледі. Нұсқаулық «сәйкестендіріңіз» немесе «сәйкестікті анықтаңыз» деген сөздерден тұрады. Көп жағдайда «сәйкестікті анықтаңыз» деген нұсқаулықты элементтердің атауымен толықтырып ұсынатын тапсырмалар да кездеседі. Алайда бағандардың атауларынан екі жиын элементтері арасындағы сәйкестікті анықтау қажет екендігі жақсы көрініп тұрғандықтан, жиындар атауымен толықтырудың қажеті жоқ [7].

Сәйкестікті анықтауға арналған тест тапсырмаларына қойылатын **талаптар**:

- оң жақтағы элементтер саны сол жақ жиындағы элементтер санынан 2 есе артық болуы тиіс;
- әр жиындағы элементтер мазмұны мен ұзындығы жағынан біркелкі болуы тиіс;
- элементтер нақты бір тақырып (тақырыпша) аясында таңдалып алынуы тиіс;
- элементтер ішінде дұрыс емес жауаптар дұрыс жауаптарға ұқсайтындай, тартымды болуы тиіс;
- сол жақ жиындағы 2 элемент А), В) латын әріптері арқылы берілуі тиіс [8].

Жоғарыда аталған талаптарды орындай отырып, сәйкестікті анықтауға арналған тест тапсырмаларын толықтай танып білу үшін мысалдар қарастырамыз.

1-мысал. (a_n) арифметикалық прогрессияда $S_n = 155,\ a_1 \cdot a_{10} = 58$ екені белгілі болса, онда сәйкестікті анықтаңыз:

А) егер прогрессия кемімелі болса, онда оның төртінші мүшесінің мәні	10
	11
	20
	29
В) егер прогрессия өспелі болса, онда оның төртінші мүшесінің мәні	10
	11
	20
	29

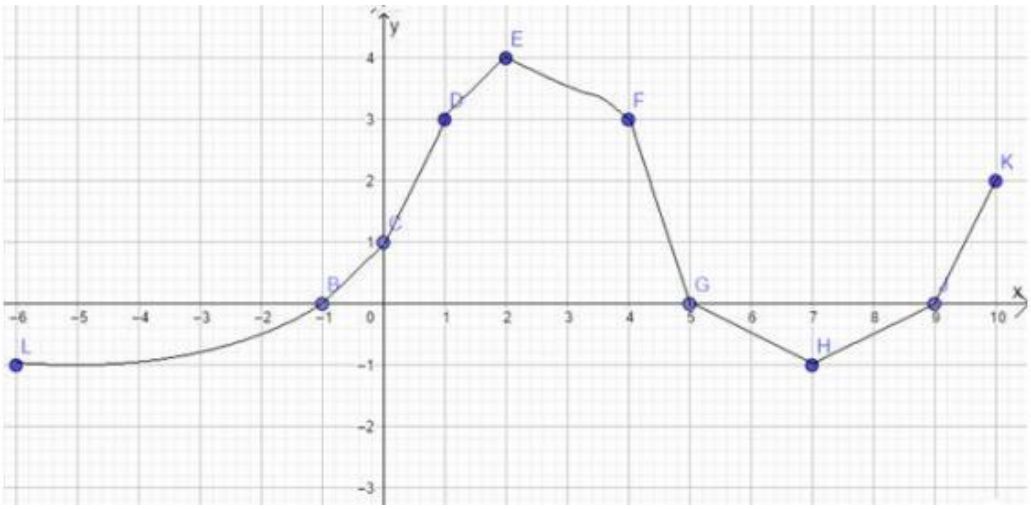
Бұл есеп ҰБТ-ға арналған тест спецификациясы бойынша «Тізбектер» бөлімінің «Арифметикалық және геометриялық прогрессиялар» тақырыбына жатады [9]. Тапсырма толығымен талаптарға сай құрастырылған.

Тапсырманың бастапқы шарты кестеде берілген сәйкестендіруге тиісті жиындар үшін ортақ түрде беріледі, яғни сәйкестіктерді анықтау сол бастапқы шартқа сүйеніп іске асырылады. Сол жақ жиын элементтері прогрессияның екі түрлі жағдайдағы (өспелі, кемімелі) бірдей реттегі мүшелерін сұрау мақсатында, ал оң жақтағы элементтер сол мүшелердің мәндерін сәйкестендіріп анықтау үшін берілген. Тестілеу барысында сол жақтағы екі элемент үшін де оң жақтағы элементтер бірдей және бір ретте ұсынылады.

Бұл мысал арифметикалық прогрессияның өспелі және кемімелі болу жағдайларына байланысты әзірленген. Сол жақтағы екі элементке оң жақтағы 4 элементтің ішінен екі элемент сәйкес келеді. Әрбір дұрыс сәйкестік үшін тестіленушіге бір баллдан беріледі. Тапсырма үшін максималды балл саны – 2 балл.

Сол жақ жиындағы екі элементке оң жақтағы элементтердің ішінен біреуі ғана сәйкес келетін жағдайлар кездесуі мүмкін. Осы жағдайды толық сипаттай алатындай тағы бір мысалды қарастырамыз.

2-мысал. $y = f(x)$ функциясының графигі берілген, сәйкестікті анықтаңыз:



A) $f(1)$ болғандағы функцияның мәні	**3**
	4
	1
	0
B) $f(4)$ болғандағы функцияның мәні	**3**
	4
	1
	0

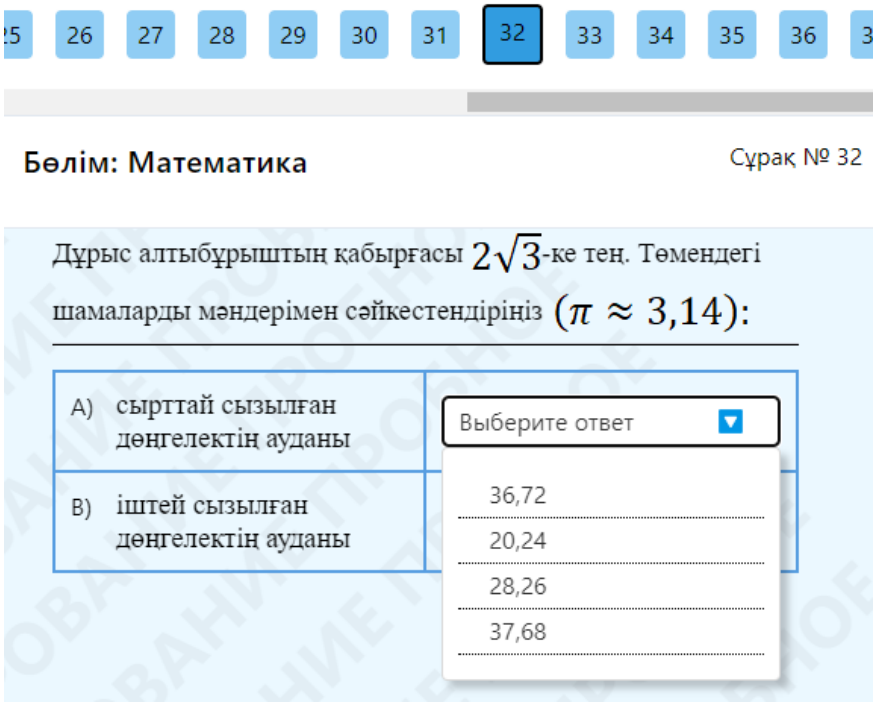
Бұл есеп Тест спецификациясы бойынша «Математикалық модельдеу мен талдау» бөлімінің «Математикалық анализ бастамалары» тақырыбы үшін әзірленген. Тапсырма толығымен талаптарға сай құрастырылған.

Берілген функция графигін пайдаланып, сол жақтағы элементтер үшін бір ғана дұрыс жауап бар екендігін анықтай аламыз. Бұл екі жағдайда да функцияның мәні 3-ке тең болады. Сол

жақтағы әрбір элемент үшін тестіленуші 3-ті таңдаған кезде, бұл тапсырма үшін ол толықтай екі балл иеленеді.

ҰБТ-да Математика пәні бойынша сәйкестікті анықтауға арналған тест тапсырмаларын математиканың барлық бөлімдері үшін құрастыруға болады. Тестіленушілерге келетін тест нұсқасының параллельдігін сақтау мақсатында 2024 жылғы ҰБТ-да сәйкестікті анықтауға арналған тест тапсырмалары математика пәні бойынша тест спецификациясының 5 тақырыбы бойынша әзірленді.

Сәйкестікті анықтауға арналған тест тапсырмаларын тапсырмалар базасына енгізу және тестілеу платформасында тестіленушіге дұрыс көрсету мақсатында Орталықтың программалық жабдықтауларға жауапты мамандары жаңа функционал әзірледі. Аталған функционалдың көмегімен сәйкестікті анықтауға арналған тест тапсырмалары тестіленушілерге платформада төмендегідей түрде ұсынылды.



Сәйкестікті анықтауға арналған тест тапсырмалары 2024 жылғы ҰБТ-да барлық бейіндік пәндер бойынша тест құрамына енгізілді [10]. Мұндай түрдегі тапсырмалардың енгізілгендігіне талапкерлер мен қоғам тарапынан оң көзқарас танытылғандығын ерекше атап өтуге болады. Алдағы уақытта сәйкестікті анықтауға арналған тест тапсырмалары қолданылатын тақырыптар аясы кеңейтілетін болады.

Мәліметтерді жинау құралдары

Зерттеуде деректерді жинаудың негізгі құралы ретінде ҰБТ-ға арналған тест тапсырмаларының апробация кезінде математика пәні бойынша алынған тест нәтижелері болды. Таңдалған нұсқаға сәйкес, тест нәтижелерінің матрицалары 40 тапсырма аясында талданды.

Тестілеу сапасын бағалау тестіленушілердің білім деңгейі мен тапсырмалардың қиындық деңгейін бір өлшемді интервалдық шкалада өлшеу тәсіліне негізделді. Деректерді өңдеу үшін Winsteps 5.5.2.0 бағдарламасы қолданылды [11]. Зерттеу объектісі ретінде сенімді нәтижелер алу мақсатында тестіленушілер саны көп болған нұсқалар қарастырылды [12]. Тестке қатысушылардың жалпы саны бір нұсқа бойынша – 471.

Нәтиже және талқылау

Зерттеу барысында математика пәні бойынша жаңа форматтағы тесттің сенімділігі мен валидтілігін бағалау, сондай-ақ тапсырмалардың қиындық деңгейі мен модельге сәйкестігін анықтау мақсатында апробацияның нәтижелері талданып, сараптама жүргізілді. Зерттеу нәтижесінде альфа Кронбах мәні (0, 1,0) қолданылады. Альфа Кронбах 1,0-ге жақын болғанда, ол сенімді көрсеткіш болып табылады [13].

Апробация нәтижесін талдау барысында Кронбах альфасы 0,78 көрсеткішін көрсетті. Бұл коэффициент тесттің ішкі тұрақтылығын бағалайды. 0,78 мәні тест элементтерінің жоғары ішкі үйлесімділігін көрсетіп, оның сенімділігін дәлелдейді. Мұндай көрсеткіштер тесттің жоғары деңгейдегі сенімділігі мен әртүрлі қабілет деңгейлері бар тестіленушілерді тиімді ажырата алу мүмкіндігін көрсетеді. Статистикалық талдау нәтижелері тесттің өлшеу қасиеттерінің негізділігін растап, оны тестіленушілер арасындағы айырмашылықтарды бағалау құралы ретінде қолдануға болатынын айқындайды.

Тест тапсырмаларының статистикалық сипаттамаларын қарастырып, барлық тапсырмалардың модельмен сәйкестігін бағалайық.

1-кестеде математика пәні бойынша тест тапсырмаларының статистикалық сипаттамалары берілген. Сәйкестік статистикасы кестенің соңғы екі бағанында көрсетілген.

1-кесте – Келісім статистикасы

Тапсырма нөмірі	Қиындықты бағалау (Measure)	Өлшеу қателігі (SE)	Корреляция коэффициенті	Келісім статистикасы	
				Взвешенная (OUTFIT MNSQ)	Простая (INFIT MNSQ)
Item1	-1.12	0.12	0.49	1.09	1.04
Item2	-1.90	0.14	0.53	0.74	0.90
Item3	-0.81	0.12	0.65	0.66	0.78
Item4	-1.49	0.13	0.56	0.90	0.88
Item5	-0.54	0.12	0.65	0.71	0.76
Item6	2.07	0.15	0.27	2.02	1.24
Item7	-0.58	0.12	0.63	0.70	0.81
Item8	0.53	0.12	0.43	1.40	1.14
Item9	-1.83	0.13	0.47	0.97	1.02
Item10	3.08	0.20	0.29	1.38	1.10
...					

Қарапайым сәйкестік статистикасы шектен тыс күтпеген жауаптарға, яғни, мықты тестіленуші жеңіл тапсырмаға қате жауап берген немесе әлсіз тестіленуші қиын тапсырмаға дұрыс жауап берген жағдайларға сезімтал болып келеді. Салмақталған статистика экстремалды күтпеген жауаптардың әсерін азайтуға мүмкіндік береді. Сәйкестік статистикасы үшін рұқсат етілген интервал (0,5; 1,5) аралығы болып табылады [14].

1-кесте деректері бойынша, осы тесттегі ең оңай тапсырма 2-тапсырма, оның қиындығы -1,90 логитті құрайды, ең қиын тапсырмалар 6 және 10-тапсырмалар, олардың қиындығы сәйкесінше 2,07 және 3,08 логитке тең. Кестеде берілген статистикалық талдау нәтижелері тест тапсырмаларының 90%-ының модельмен жақсы үйлесетінін көрсетеді.

Кесте деректеріне сүйене отырып, 6-тапсырманың келісім статистикасы белгіленген интервалға сәйкес келмейтіні анықталды. Сонымен қатар, 6 және 10-тапсырмалардың корреляция коэффициенттері салыстырмалы түрде төмен (сәйкесінше 0,27 және 0,29) екені байқалады. Корреляция коэффициенті -1 мен +1 аралығында өзгеретін көрсеткіш болып табылады, және оның мәні неғұрлым төмен болса, айнымалылар арасындағы өзара байланыс соғұрлым әлсіз болатынын білдіреді [15]. Төмен корреляция мәні бұл тапсырмалардың қатысушылардың жалпы білім деңгейін дәл өлшеуде шектеулі екенін көрсетеді.

6-тапсырма жоғары деңгейдегі күрделілікке ие (Measure = 2.07) және сенімді бағалауға ие (SE = 0.15). Оны негізінен жақсы дайындалған қатысушылар ғана сәтті орындай алады.

OUTFIT және InFit MNSQ коэффициенттерінің жоғары болуы (сәйкесінше 2.02 және 1.24) тапсырманың берілген қатысушылар тобы үшін тым күрделі болуы мүмкін екенін көрсетеді.

10-тапсырма жоғары деңгейдегі күрделілікке ие, оның өлшеу мәні (Measure = 3.08) төмен стандартты қатемен (SE = 0.20) белгіленген, бұл оның жоғары сенімділігін растайды. Outfit MNSQ коэффициентінің жоғары болуы (1.38) күтпеген жауаптардың бар екенін білдіреді, бұл қосымша деректерді талдауды қажет етеді. InFit MNSQ мәні (1.10) жалпы деректердің жеткілікті деңгейде сәйкестігін көрсетеді.

Логистикалық модель негізінде құрылған 6 және 10 тест тапсырмаларының сипаттамалық қисығы (Item Characteristic Curve, ICC) қарастырылады. Логистикалық модельдің негізгі аспектілері:

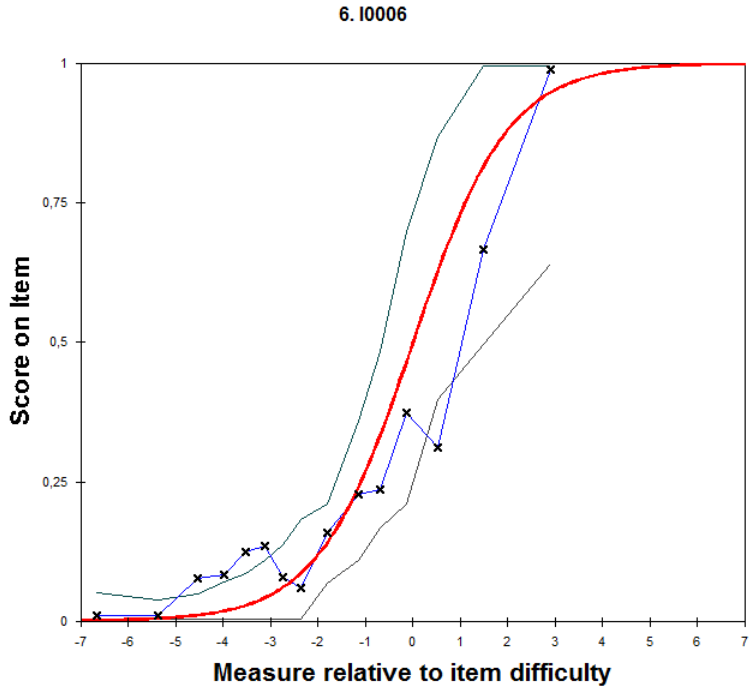
X oci (Measure relative to item difficulty) – Тест тапсырмасының қиындық деңгейіне қатысты тестіленушілердің қабілет деңгейін көрсетеді. Теріс мәндер – төмен қабілет деңгейін, оң мәндер – жоғары қабілет деңгейін білдіреді.

Y oci (Score on Item) – Тапсырманы дұрыс орындау ықтималдығы. 0 – дұрыс емес жауап, 1 – дұрыс жауап беру ықтималдығы 100%.

Қызыл түспен берілген қисық сызық – Логистикалық қисық, яғни модель бойынша дұрыс жауап беру ықтималдығының қабілет деңгейіне тәуелділігін көрсетеді.

Қара түспен берілген нүктелер – Нақты тестілеу нәтижелерінен алынған эмпирикалық деректер, яғни тестіленушілердің шынайы жауаптары.

Көк және қара түспен берілген қисықтар – Бұлар сенімділік интервалдары немесе дұрыс жауап беру ықтималдығының альтернативті апроксимациясы.

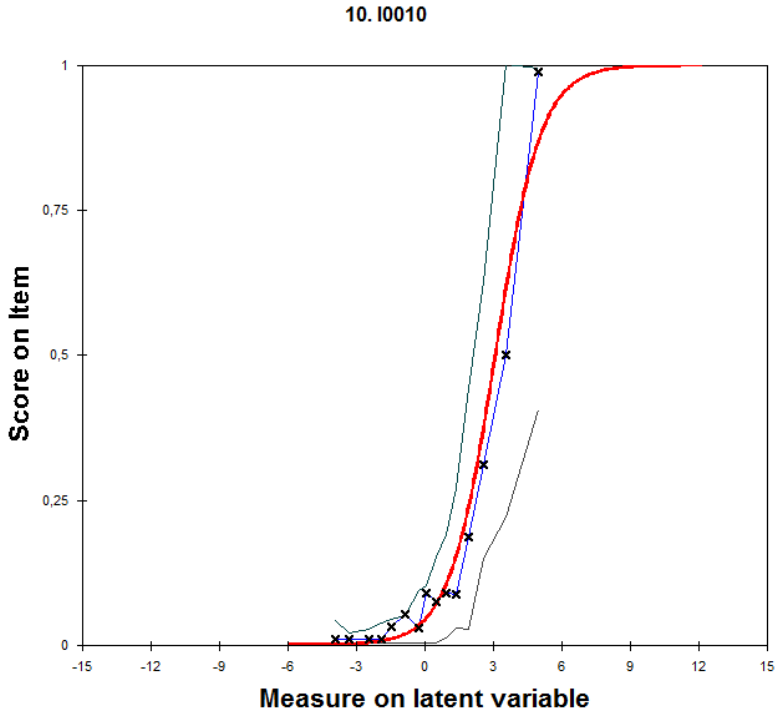


1-сурет - 6-тапсырманың қиындық деңгейіне қатысты тестіленушілердің қабілет деңгейінің қисығы

1-суретте 6-тапсырма бойынша қабілеті жоғары қатысушылардың дұрыс жауап беру ықтималдығы жоғары екені анық байқалады, ал қисықтың логикалық түрде өсуі бұл заңдылықты растайды. Графикте қара түспен белгіленген нақты деректер шамамен қызыл түспен көрсетілген теориялық қисыққа сәйкес келеді, бұл модельдің деректермен жақсы үйлесетінін көрсетеді. Дегенмен, кейбір нүктелер қисықтан сәл ауытқыған, бұл күтпеген жауаптардың болуын білдіруі мүмкін (мысалы, төмен қабілетті қатысушылар күрделі тапсырманы дұрыс орындауы немесе керісінше).

2-суретте 10-тапсырма бойынша қабілеті жоғары қатысушылардың дұрыс жауап беру ықтималдығы жоғары екені анық байқалады, ал қисықтың логикалық түрде өсуі бұл заңдылықты растайды.

Көптеген нүктелер (қара түспен белгіленген нақты деректер) күтілетін S-тәрізді логистикалық қисыққа сәйкес келеді, бұл модельдің деректермен жақсы үйлесетінін көрсетеді. Графиктің төменгі жағында (-6-дан 0-ге дейін) нүктелердің шашырауы байқалады, бұл қабілеттілігі төмен тестіленушілердің күтпеген жауаптардың болуын білдіреді. Демек қабілеті төмен субъектілердің болжамды жауаптары болғанын немесе немесе тестілеу кезінде болған қателіктерді білдіреді.



2-сурет - 10-тапсырманың қиындық деңгейіне қатысты тестіленушілердің қабілет деңгейінің графигі

Тест тапсырмаларының сапасын талдау нәтижелері олардың күрделілік деңгейін, модельге сәйкестігін және сенімділігін бағалауға мүмкіндік береді. 6 және 10-тапсырмалардың жоғары күрделілікке ие екендігі анықталды, алайда олардың корреляция коэффициенттері төмен, бұл олардың тесттің жалпы бағалау дәлдігіне ықпалын шектеуі мүмкін екенін көрсетеді. OUTFIT және InFit MNSQ көрсеткіштерінің жоғары мәндері кейбір тапсырмалардың белгілі бір қатысушылар тобы үшін тым қиын болуы мүмкін екенін білдіреді. Осы нәтижелер негізінде тест тапсырмаларын жетілдіру және олардың бағалау сапасын арттыру бойынша ұсынымдар берілуі қажет.

Қорытынды

Білім беру саласындағы реформалар аясында Ұлттық бірыңғай тестілеу жүйесін жетілдіру мақсатында сәйкестікті анықтауға арналған тест тапсырмалары енгізілді. Бұл тапсырмалар түсушілердің пәндік білімін тереңірек бағалауға, логикалық ойлау қабілеттерін дамытуға және білім деңгейін объективті түрде өлшеуге мүмкіндік береді. Зерттеу барысында математика пәні бойынша жаңа форматтағы тест тапсырмаларының құрылымы, мазмұны және тиімділігі талданды. Апробация нәтижелері бойынша тест тапсырмаларының сенімділігі мен валидтілігі бағаланды, сонымен қатар олардың қиындық деңгейі мен модельге сәйкестігі анықталды.

Зерттеу нәтижелері бойынша сәйкестікті анықтауға арналған тест тапсырмаларының көпшілігі жоғары сенімділікке ие болып шықты. Алайда, кейбір тапсырмалар жоғары қиындық деңгейіне байланысты қатысушылардың білім деңгейін дәл өлшеуде шектеулерге ие болды. Осыған байланысты, бұл тапсырмаларды одан әрі жетілдіру және олардың қиындық деңгейін теңшеу қажеттілігі айқындалды.

Сәйкестікті анықтауға арналған тест тапсырмаларын енгізу ҰБТ жүйесіндегі бағалау тиімділігін арттыруға және түсушілердің білім деңгейін тереңірек бағалауға мүмкіндік береді. Бұл тапсырмалардың қолданылуы білім беру сапасын арттыруға және оқу үдерісін жаңғыртуға үлес қосады. Алдағы уақытта бұл форматтағы тапсырмалардың қолданылатын тақырыптар аясын кеңейту және олардың сапасын одан әрі жетілдіру мақсатында жұмыстар жалғасуы тиіс.

Зерттеу нәтижелері білім беру саласындағы сарапшыларға, мұғалімдерге және тестілеу жүйесін жетілдіруге қатысушыларға пайдалы ақпарат пен ұсыныстар ұсынады. Сәйкестікті анықтауға арналған тест тапсырмаларын қолдану білім алушылардың білім деңгейін объективті және тиімді түрде бағалауға мүмкіндік береді, бұл білім беру сапасын арттыруға және оқу үдерісін жаңғыртуға үлес қосады.

Әдебиеттер

1. Всемирный Банк. (World Bank). Реформы в образовании и тестировании. Стратегия и проекты для стран с переходной экономикой. [Электронный ресурс] URL: https://www.worldbank.org/en/topic/education.
2. Ұлттық бірыңғай тестілеуді өткізу қағидалары. Қазақстан Республикасы Білім және ғылым министрінің 2017 жылғы 2 мамырдағы №204 бұйрығы.
3. www.gov.kz. ҰБТ-2024: өзгерістер туралы. [Электронный ресурс] URL: https://www.gov.kz/memleket/entities/zhetysu-edu/press/news/details/6957933.
4. Сидорик, В.В., Чичко, О.И. (2010). Теория и практика разработки тестовых заданий. Учебно-методическое пособие для слушателей системы повышения квалификации. Учебное издание. Минск.
5. Жалпы орта білім беру деңгейі пәндерінен жиынтық бағалау бойынша тапсырмалар жинағы (2020). Әдістемелік ұсынымдар. Нұр-Сұлтан: Ы.Алтынсарин атындағы Ұлттық білім академиясы.
6. Аванесов, В.С. (2005) Форма тестовых заданий. Учебное пособие для учителей школ, лицеев, преподавателей вузов и колледжей. 2 изд, переработанное и расширенное. М.: Центр тестирования.
7. Аванесов, В.С. (1998) Композиция тестовых заданий. Учебная книга для преподавателей вузов, учителей школ, аспирантов и студентов педвузов. М.: Адепт.
8. Клайн, П. (1994) Справочное руководство по конструированию тестов. Киев.
9. testcenter.kz. ҰБТ тест спецификациялары. [Электронный ресурс] URL: https://testcenter.kz/upload/iblock/fc4/07_.pdf.
10. sputnik.kz. ҰБТ-2024: министр жаңа өзгерістер болатынын айтты. [Электронный ресурс] URL:https://sputnik.kz/20230913/ubt-2024-ministr-zhana-ozgerister-bolatynyn-aytty-38416460.html?utm_source.
11. Trevor G. Bond, Christine M. Fox. Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Second Edition. - Psychology Press, 2013 г. [Электронный ресурс] URL: https://books.google.ru/books?id=Yk7yi47RYIkC&hl=ru&source=gbs_navlinks_s.
12. Маслак, А.А. (2006) Измерение латентных переменных в социально-экономических системах. Монография. Славянск-на-Кубани: Издательский центр СГПИ.
13. sciencedirect.com Альфа-коэффициент Кронбаха [Электронный ресурс] URL: https://www.sciencedirect.com/topics/nursing-and-health-professions/cronbach-alpha-coefficient.
14. Анисимова, Т.С., Маслак, А.А. Осипов, С.А. (2005) Анализ качества заданий с выбором одного правильного ответа. Педагогические Измерения.
15. Антонов, В.Н. (2003). Основы тестологии: теория и практика разработки тестов. М.:Издательство «Гуманитариум».

К.Т. Туралбаева, Т.Ж. Ниязов, Б.К. Асемова

ТЕСТОВЫЕ ЗАДАНИЯ НА УСТАНОВЛЕНИЕ СООТВЕТСТВИЯ ПО ПРЕДМЕТУ МАТЕМАТИКА В ЕДИНОМ НАЦИОНАЛЬНОМ ТЕСТИРОВАНИИ

В данном исследовании рассматриваются проблемы внедрения тестовых заданий в рамках Единого национального тестирования (ЕНТ). Основное внимание уделяется совершенствованию структуры и содержания тестов, что направлено на более глубокую оценку знаний студентов и развитие их логического мышления. В ходе исследования были проанализированы существующие методы тестирования, разработаны и внедрены новые типы заданий, ориентированные на оценку глубины знаний по конкретным предметам. В статье рассматриваются тестовые задания с мультимедийными элементами и новые задания на установление соответствия. Задания на установление соответствия были включены с целью проверки навыков анализа и сравнения информации, развития критического мышления и оценки способности устанавливать логические связи между элементами. Эти задания позволяют проверять не только знание фактов, но и понимание взаимосвязей между ними, что способствует комплексной оценке знаний учащихся и их способности применять информацию в различных контекстах. Авторы описывают тестовые задания на установление соответствия по предмету «Математика» и раскрывают их конкретную роль в системе оценки уровня знаний. Кроме того, в статье особое внимание уделяется значимости и необходимости этих заданий в современной системе образования, показывая, как они способствуют внедрению новых подходов к обучению и оцениванию.

Результаты апробации тестовых заданий нового формата показали высокую эффективность в оценке глубины знаний и логических навыков тестируемых. Также были проанализированы статистические данные, включающие надежность, сложность, корреляционные характеристики и другие показатели тестов. На основе полученных данных были разработаны рекомендации по дальнейшему совершенствованию тестирования, что поможет повысить объективность и эффективность оценки в рамках Единого национального тестирования.

Ключевые слова: ЕНТ, тестовое задание, мультимедийные элементы, тестируемый, оценивание, математика, спецификация, разработка тестовых заданий, установление соответствия.

K. Turalbaeva, T. Niyazov, B. Assemova

TEST TASKS FOR DETERMINING CORRESPONDENCE IN THE SUBJECT OF MATHEMATICS IN THE NATIONAL UNIFIED TESTING

The research examines the challenges of introducing new test items within the Unified National Testing (UNT) framework. The primary focus is on improving the structure and content of the tests to provide a deeper assessment of students' knowledge and to develop their logical thinking skills. The study analyzed existing testing methods and then developed and implemented new types of questions designed to evaluate the depth of knowledge in specific subjects. The article discusses test items with multimedia elements and new matching questions. The matching questions were included to check students' skills in analyzing and comparing information, to foster critical thinking, and to assess their ability to establish logical connections between elements. These questions allow for testing not only the knowledge of facts but also the understanding of the interrelationships between them. This contributes to a comprehensive evaluation of students' knowledge and their ability to apply information in different contexts. The authors describe the matching questions for the subject «Mathematics» and reveal their specific role in the knowledge assessment system. Furthermore, the article highlights the importance and necessity of these questions in the modern education system, showing how they contribute to new approaches in teaching and evaluation.

The results of the new test item format's trial run demonstrated high effectiveness in assessing the depth of knowledge and logical skills of the test-takers. Statistical data, including reliability, difficulty, correlation characteristics, and other test indicators, were also analyzed. Based on the findings, recommendations for further

improving the testing were developed, which will help increase the objectivity and effectiveness of the assessment within the Unified National Testing.

Keywords: UNT, test item, multimedia elements, test-taker, evaluation, mathematics, specification, test item development, matching.

References

1. Vsemirnyi Bank (World Bank). Reformy v obrazovanii i testirovanii. Strategiya i proekty dlya stran s perekhodnoi ekonomikoi. [Elektronnyi resurs] URL: https://www.worldbank.org/en/topic/education.
2. Ŭlttyq biryñğai testileudi ótkizu qağıdalar. Qazaqstan Respublikasy Bilim jəne ğylym ministriniñ 2017 jylğy 2 mamyrdağy №204 bŭiryğy.
3. www.gov.kz. ŬBT-2024: ózgerister turaly. [Elektronnyi resurs] URL: https://www.gov.kz/memleket/entities/zhetysu-edu/press/news/details/6957933.
4. Sidorik, V.V., Chichko, O.I. (2010). Teoriya i praktika razrabotki testovykh zadanii. Uchebno-metodicheskoe posobie dlya slushatelei sistemy povysheniya kvalifikatsii. Uchebnoe izdanie. Minsk.
5. Zhalpy orta bilim beru deñgeii pənderinen zhiyntyq bağalau boyynsha tapsyrmalar zhinagy (2020). Ədistemelik ŭsynymdar. Nur-Sultan: Y. Altynsarin atyndağy Ŭlttyq bilim akademiyasy.
6. Avanesov, V.S. (2005) Forma testovykh zadanii. Uchebnoe posobie dlya uchitelei shkol, litseev, prepodavatelei vuzov i kolledzhei. 2 izd., pererabotannoe i rasshirennoe. M.: Tsentr testirovaniya.
7. Avanesov, V.S. (1998) Kompozitsiya testovykh zadanii. Uchebnaya kniga dlya prepodavatelei vuzov, uchitelei shkol, aspirantov i studentov pedvuzov. M.: Adept.
8. Klain, P. (1994) Spravochnoe rukovodstvo po konstruirovaniyu testov. Kiev.
9. testcenter.kz. ŬBT test spetsifikatsiyalary. [Elektronnyi resurs] URL: https://testcenter.kz/upload/iblock/fc4/07_.pdf.
10. sputnik.kz. ŬBT-2024: ministr zhaña ózgerister bolatynyn aitty. [Elektronnyi resurs] URL: https://sputnik.kz/20230913/ubt-2024-ministr-zhana-ozgerister-bolatynyn-aytty-38416460.html?utm_source.
11. Trevor G. Bond, Christine M. Fox. Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Second Edition. Psychology Press, 2013 g. [Elektronnyi resurs] URL: https://books.google.ru/books?id=Yk7yi47RYIkC&hl=ru&source=gbs_navlinks_s.
12. Maslak, A.A. (2006) Izmerenie latentnykh peremennykh v sotsialno-ekonomicheskikh sistemakh. Monografiya. Slavyansk-na-Kubani: Izdatelskii tsentr SGPI.
13. sciencedirect.com. Alfa-koeffitsient Kronbakha [Elektronnyi resurs] URL: https://www.sciencedirect.com/topics/nursing-and-health-professions/cronbach-alpha-coefficient.
14. Anisimova, T.S., Maslak, A.A. Osipov, S.A. (2005) Analiz kachestva zadanii s vyborom odnogo pravilnogo otveta. Pedagogicheskie Izmereniya.
15. Antonov, V.N. (2003). Osnovy testologii: teoriya i praktika razrabotki testov. M.: Izdatelstvo «Gumanitarium».

Сведения об авторах:

Туралбаева Кенжекул Турганбаевна – РГП на ПХВ «Национальный центр тестирования» МНВО РК, заведующий лабораторией, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: kengebahit@mail.ru.

Ниязов Талгат Жаксыбаевич – РГП на ПХВ «Национальный центр тестирования» МНВО РК, заместитель директора, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: n.talgin@bk.ru.

Асемова Бибіфатима Кәкімжанқызы – РГП на ПХВ «Национальный центр тестирования» МНВО РК, старший эксперт, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: assemova00@mail.ru.

Авторлар туралы мәлімет:

Туралбаева Кенжекул Турганбаевна – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, зертхана меңгерушісі, Родниковая көш., 1/1, 010008, Астана, Қазақстан, e-mail: kengebahit@mail.ru.

Ниязов Талгат Жаксыбаевич – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, директордың орынбасары, Родниковая көшесі, 1/1, 010008, Астана, Қазақстан, e-mail: n.talgin@bk.ru.

Асемова Бибіфатима Кәкімжанқызы – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, аға сарапшы, Родниковая көш., 1/1, 010008, Астана, Қазақстан, assemova00@mail.ru.

Information about authors:

Turalbayeva Kenzhekyl Turganbayevna – 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, Rodnikovaya St., 1/1, 010008, Astana, Kazakhstan, e-mail: kengebahit@mail.ru.

Niyazov Talgat Zhaksybayevich – 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, Deputy Director, Rodnikovaya St., 1/1, 010008, Astana, Kazakhstan, e-mail: n.talgin@bk.ru.

Assemova Bibifatima Kakimzhanovna – 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, Senior expert, Rodnikovaya St., 1/1, 010008, Astana, Kazakhstan, assemova00@mail.ru.

МРНТИ 14.35.09 https://doi.org/10.63597/UTO3105-4161.2025.3.3.002

Ж.Ж. Бейсенова[1], Л.Е. Шинетова[2], Г.К. Жабаева[3]
[1,2,3]РГП на ПХВ «Национальный центр тестирования» МНВО РК, г. Астана, Республика Казахстан

[1]ORCID 0000-0002-7347-1538, [2]ORCID 0000-0003-4280-7999, [3]ORCID 0009-0006-1003-6316

ПСИХОМЕТРИЧЕСКИЙ АНАЛИЗ ТЕСТОВЫХ ЗАДАНИЙ ПО ОЦЕНКЕ ЗНАНИЙ ПЕДАГОГОВ В РЕСПУБЛИКЕ КАЗАХСТАН (ПО РЕЗУЛЬТАТАМ АПРОБАЦИИ)

В статье представлены детальные результаты психометрического анализа тестовых заданий по математике, которые использовались в рамках апробации процедуры оценки знаний педагогов (ОЗП) в Казахстане. Исследование основано на сравнительном анализе трех различных вариантов теста, подготовленных для аттестации. Для оценки качества заданий и выявления их сильных и слабых сторон были задействованы методы как классической теории тестов (КТТ), так и современной теории тестов (IRT), в частности, модель Раша. Целью работы являлась не только оценка существующих оценочных инструментов, но и разработка практических рекомендаций по их усовершенствованию. Анализ апробации показал, что, несмотря на общую высокую надежность тестов (коэффициент альфа-Кронбаха превышает 0,84) и их хорошую способность дифференцировать испытуемых по уровню подготовленности, в каждом варианте были обнаружены отдельные проблемные задания. Было выявлено наличие заданий с низкой или даже отрицательной дискриминативностью, а также дистракторов, которые не работали должным образом. Все замечания, выявленные в ходе апробации, были учтены и устранены в последующей работе по корректировке банка заданий. Это позволило повысить точность и валидность оценочных инструментов в целом.

Ключевые слова: оценка знаний педагогов, психометрический анализ, апробация, надежность, валидность, классическая теория тестов, модель Раша.

Введение

В современной образовательной системе Казахстана процедура оценки знаний педагогов (ОЗП) представляет собой важный инструмент, который направлен на повышение качества обучения и профессионального статуса учителей. В соответствии с приказом Министра образования и науки Республики Казахстан от 27 января 2016 года №83 «Об утверждении Правил и условий проведения аттестации педагогов», ОЗП проводится для определения уровня профессиональной компетентности педагога с помощью тестов, разработанных уполномоченным органом в области образования [1, 2]. Этот процесс распространяется на педагогов дошкольных организаций, предшкольных классов, учителей начального, среднего и общего среднего образования, а также на работников дополнительного, специального и технического образования. ОЗП также затрагивает руководителей организаций образования, методистов, лиц без стажа и тех, кто желает возобновить педагогическую деятельность [3].

В отличие от прошлых форматов, ОЗП больше сфокусирована на практических навыках и методических компетенциях [4]. Такой подход позволяет получать объективные данные об уровне подготовки учителей, выявлять системные пробелы и находить лучшие способы для развития всей системы образования [5]. Ключевым компонентом ОЗП является тестирование, которое должно обладать высокой надежностью и валидностью. Успешно пройденный тест – это не просто подтверждение квалификации. Это возможность для карьерного роста и увеличения заработной платы, что, несомненно, мотивирует учителей постоянно совершенствоваться. Использование электронного формата делает оценку еще более объективной и прозрачной [1].

В этом исследовании мы провели всесторонний анализ тестовых заданий по математике, чтобы оценить их качество и соответствие современным требованиям. Результаты, которыми мы делимся, были получены в ходе апробации этого инструмента.

Материалы и методы исследования

Для анализа были выбраны три варианта теста по математике, использованные при апробации аттестации педагогов: 369168, 369169 и 369694. Выборка составила 190, 185 и 182 участника соответственно. Психометрический анализ проводился с использованием двух ключевых теоретических подходов. В рамках классической теории тестов (КТТ) анализ включает расчет показателей надежности (коэффициенты *KR-20* и *альфа-Кронбаха*), средней трудности и распределения заданий, а также дискриминативности (*pBis*) и эффективности дистракторов [6]. В рамках современной теории тестов (IRT) в рамках модели Раша оценивались надежность, разделяющая способность (*Person Separation*) и соответствие модели (*Infit MNSQ и Outfit MNSQ*), где допустимый диапазон значений установлен в пределах 0,7-1,3. Для расчетов использовались специализированные программы, такие как TAP, Jamovi и Winsteps [7].

Результаты и обсуждение

Результаты анализа по КТТ

Анализ данных был проведен нами с помощью специальных программ TAP и Jamovi. В ходе анализа было выяснено следующее. Все три варианта теста продемонстрировали высокий уровень надежности, что подтверждается значениями коэффициентов KR-20 и альфа-Кронбаха, превышающими 0,84. Это отличный показатель, который говорит о том, что результаты тестов стабильны и внутренне согласованы [8]. Проще говоря, если бы участники прошли этот тест еще раз, их результаты, скорее всего, были бы очень похожими (таблица 1).

Таблица 1 - Основные результаты анализа (КТТ)

Показатель	Вариант 369168	Вариант 369169	Вариант 369694
Число участников	190	185	182
Средний балл	27,53	28,62	31,1
Стандартное отклонение	7,70	8,38	8,64
Минимум / Максимум	9 / 47	9 / 46	10/ 47
Асимметрия	0,033	-0,095	-0,413
Эксцесс	-0,557	-0,677	-0,764
Надежность (KR-20 и альфа-Кронбаха)	0,84	0,87	0,89
Средняя трудность	0,551	0,572	0,622
Число заданий трудности < 0,2	3	0	0
Число заданий трудности > 0,8	4	4	11
Проблемные задания с дискриминативностью < 0	3 (№2, 22, 34)	1 (№44)	1 (№34)
Задания с дискриминативностью < 0,1	5 (№2, 9, 17, 22, 34)	3 (№ 1, 44, 45)	2 (№34, 49)

Таблица представляет собой сводку основных результатов психометрического анализа, выполненного в рамках КТТ с использованием программного обеспечения TAP.

Распределение баллов по всем вариантам в целом соответствует нормальному, с приемлемыми показателями асимметрии и эксцесса. Вариант 369694 показал небольшое смещение в сторону более высоких баллов (отрицательная асимметрия), что может быть обусловлено его чуть меньшей сложностью (рисунок 1). В целом, такое распределение говорит о том, что тесты успешно дифференцируют испытуемых по уровню подготовки.

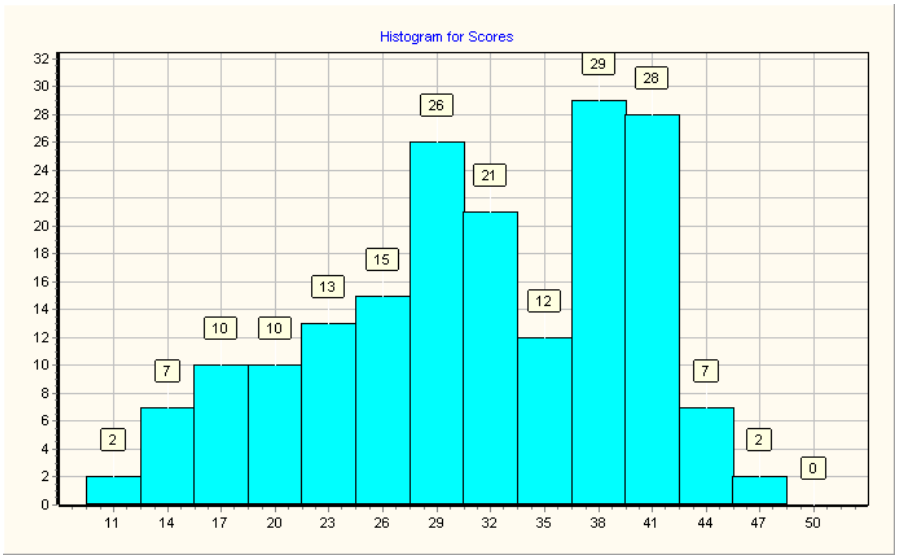


Рисунок 1 - Гистограмма распределения баллов по варианту 369694 (данные ТАР)

Несмотря на высокие общие показатели, углубленный анализ выявил отдельные задания, требующие дополнительного внимания. Так, было обнаружено, что некоторые вопросы имеют низкую или даже отрицательную дискриминативность (таблица 1), что может свидетельствовать о слабой корреляции с общим уровнем подготовленности испытуемых. Иными словами, такие задания не всегда эффективно дифференцируют сильных и слабых участников, что может быть обусловлено их формулировкой или возможностью случайного угадывания. Эти особенности могут оказывать влияние на точность измерения и, как следствие, на валидность оценочного инструмента. Кроме того, анализ дистракторов показал, что функционирование некоторых из них не всегда соответствовало ожиданиям: отдельные из них были слишком очевидными, а другие, напротив, вводили в заблуждение даже высококвалифицированных испытуемых.

Дистракторный анализ проблемных заданий в КТТ

Дистракторный анализ является ключевым этапом в процессе психометрической оценки качества заданий. Он позволил нам выявить и идентифицировать проблемные задания, которые не соответствуют установленным критериям эффективности (таблицы 2) [9]. В отличие от традиционного анализа, фокусирующегося только на общем проценте правильных ответов, дистракторный анализ дает возможность детально изучить привлекательность каждого дистрактора. Это имеет огромное значение для повышения надежности и валидности теста [10].

Анализ проблемных заданий в КТТ начинается с изучения коэффициентов дискриминации и точечно-бисериальной корреляции (pBis) для каждого варианта ответа. Высокий положительный pBis для правильного ответа (например, задание №9, вариант A: pBis=0,10386) указывает на то, что этот вариант ответа чаще выбирают студенты с высокими общими баллами. В то же время, эффективные дистракторы должны иметь отрицательный pBis (например, задание №9, варианты C и D). Это свидетельствует о том, что они привлекают испытуемых с низкими результатами, что является желаемой характеристикой.

Проблемные задания, напротив, демонстрируют аномальные результаты. Например, в задании №22, правильный ответ C имеет низкий pBis (-0,0681), а его дискриминация (0,0126) практически равна нулю, что указывает на отсутствие связи с общим уровнем знаний испытуемых. Более того, дистрактор D в этом же задании имеет положительный pBis (0,0828), что свидетельствует о его привлекательности для более сильных студентов. Такие результаты требуют пересмотра или исключения задания из теста, так как оно не выполняет своей основной функции по дифференциации испытуемых.

Таблица 2 – Результаты дистракторного анализа тестовых заданий

Вариант 369168

Задание № 2

	correct	key	N	rspP	pBis	discrim	lower	mid66	upper
A		A	51	0,26842	0,0832	0,15441	0,2206	0,2273	0,375
B	*	B	35	0,18421	-0,025	0,0021	0,1765	0,197	0,179
C		C	8	0,04211	-0,1219	-0,0588	0,0588	0,0606	0
D		D	95	0,5	-0,1667	-0,0976	0,5441	0,5	0,446

Задание №9

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A	*	A	33	0,1737	0,10386	0,1176	0,132	0,152	0,25
B		B	44	0,2316	0,00942	0,0798	0,206	0,212	0,2857
C		C	95	0,5	-0,1707	-0,0977	0,544	0,5	0,4464
D		D	18	0,0947	-0,1437	-0,0998	0,118	0,136	0,0179

Задание №17

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A		A	53	0,27895	-0,2903	-0,2574	0,382	0,303	0,125
B	*	B	42	0,22105	0,033	0,062	0,206	0,197	0,268
C		C	47	0,24737	0,2017	0,2279	0,147	0,2424	0,375
D		D	47	0,24737	-0,1454	-0,0326	0,265	0,2424	0,232

Задание №22

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A		A	16	0,0842	-0,1985	-0,1082	0,1618	0,0303	0,0536
B		B	4	0,0211	-0,1522	-0,0263	0,0441	0	0,0179
C	*	C	109	0,5737	-0,0681	0,0126	0,5588	0,5909	0,5714
D		D	61	0,3211	0,0828	0,1218	0,2353	0,3788	0,3571

Задание №34

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A		A	32	0,16842	-0,0413	0,0494	0,147	0,1667	0,196
B		B	27	0,14211	-0,0908	-0,0546	0,162	0,1515	0,107
C	*	C	45	0,23684	-0,0278	0,0861	0,235	0,1667	0,321
D		D	85	0,44737	-0,0661	-0,0809	0,456	0,5	0,375

Вариант 369169

Задание №1

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A		A	34	0,1838	-0,3332	-0,2051	0,286	0,1833	0,0806
B		B	16	0,0865	-0,07	-0,0305	0,111	0,0667	0,0806
C	*	C	45	0,2432	0,0846	0,0684	0,238	0,1833	0,3065
D		D	90	0,4865	0,0712	0,1672	0,365	0,5667	0,5323

Задание №44

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A	*	A	61	0,33	-0,045	0,0371	0,302	0,35	0,3387
B		B	73	0,395	0,1419	0,2783	0,254	0,4	0,5323
C		C	28	0,151	-0,2305	-0,1895	0,254	0,133	0,0645
D		D	23	0,124	-0,1429	-0,126	0,19	0,117	0,0645

Задание №45

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A		A	33	0,178	-0,2907	-0,2048	0,302	0,1333	0,0968
B		B	19	0,103	-0,112	-0,0622	0,143	0,0833	0,0806
C	*	C	69	0,373	0,0333	0,1019	0,317	0,3833	0,4194
D		D	64	0,346	0,1002	0,1651	0,238	0,4	0,4032

Вариант 369694

Задание №34

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A		A	17	0,0934	-0,32895	-0,19201	0,209	0,0357	0,0169
B		B	43	0,2363	0,20532	0,29825	0,0746	0,2857	0,3729
C		C	17	0,0934	-0,1850	-0,1004	0,1343	0,1071	0,0339
D	*	D	105	0,5769	-0,0082	-0,0058	0,5821	0,5714	0,5763

Задание №49

	correct	key	n	rspP	pBis	discrim	lower	mid66	upper
A	*	A	95	0,522	0,0841	0,258	0,403	0,5179	0,661
B		B	56	0,3077	-0,0408	-0,0402	0,3284	0,3036	0,2881
C		C	12	0,0659	-0,0217	-0,0238	0,0746	0,0714	0,0508
D		D	19	0,1044	-0,2883	-0,194	0,194	0,1071	0

Таблица представляет собой результаты анализа ответов на задания, выполненного с помощью программного обеспечения Jamovi.

Таким образом, в ходе дистракторного анализа были выявлены задания, которые требуют пересмотра для повышения валидности и надежности теста (таблица 2) [12]. На основе полученных данных, были выделены три основные проблемы, влияющие на качество теста.

Низкая и отрицательная дискриминативность ключей. Основной задачей тестового вопроса является дифференциация испытуемых по уровню их знаний. В таких заданиях показатель pBis, либо близок к нулю, либо имеет отрицательное значение. Это означает, что правильный ответ на такой вопрос не позволяет надежно отличить сильных участников от слабых, что ставит под сомнение валидность задания.

Чрезмерно легкие задания. Ряд вопросов показал аномально высокую долю правильных ответов (показатель трудности p>0,9). Это свидетельствует о том, что задания не выполняют своей диагностической функции, так как их дистракторы оказались нерабочими – их практически никто не выбрал. Такие вопросы не вносят существенного вклада в общую оценку.

Ключ выбран реже, чем дистрактор. В ряде случаев правильный ответ оказался менее популярным, чем один из неправильных. Это серьезный сигнал о том, что в задании может быть либо ошибка, либо его формулировка слишком сложна и сбивает с толку даже хорошо подготовленных испытуемых.

Для устранения выявленных проблем и оптимизации качества оценочного инструмента нами были предложены следующие рекомендации:

Обязательная переэкспертиза заданий с отрицательной дискриминативностью (pBis < 0). Эти вопросы необходимо подвергнуть глубокому содержательному анализу, чтобы выявить причины их низкой эффективности;

Пересмотр или удаление заданий, где ключ не позволяет эффективно дифференцировать испытуемых по группам или выбирается реже, чем дистрактор;

Повышение качества дистракторов. Для улучшения диагностической ценности вопросов необходимо усилить правдоподобность неправильных вариантов ответа. Это позволит уменьшить вероятность случайного угадывания и повысить точность измерения знаний.

Для более глубокой оценки качества заданий и выявления их соответствия единой шкале способностей после КТТ проведен анализ в IRT (модели Раша) с использованием программы Winsteps [13]. Основные результаты анализа IRT приведены в таблице 3.

Таблица 3 - Основные результаты анализа по IRT (модель Раша)

Показатель	Вариант 369168	Вариант 369169	Вариант 369694
Количество респондентов	190	185	182
Надежность (альфа-Кронбаха)	0,84	0,87	0,88
Надежность Person Reliability	0,83	0,86	0,86
Разделяющая способность Person Separation	2,20	2,45	2,50
Реальная стандартная ошибка (Real RMSE)	0,34	0,35	0,36
Средняя способность участников (Mean Measure Persons)	0,26	0,40	0,67
Стандартное отклонение SD Measure (Persons)	0,83	0,92	0,97
Число заданий и номера с Corr<0	1 (№2)	0	0
Число заданий и номера с Corr< 0,1	1 (№2)	1 (№44, 45)	1 (№34)
Число заданий и номера с Infit MNSQ <0,7	0	0	0
Число заданий и номера с Infit MNSQ > 1,3	2 (№2, 34)	1 (№44)	1 (№34)
Число заданий и номера Outfit MNSQ<0,7	0	2 (№25, 39)	2 (№32, 43)
Число заданий и номера Outfit MNSQ > 1,3	2 (№2, 34)	4 (№1, 38, 44, 45)	4 (№34, 49, 41, 33)
Таблица представляет собой сводку основных результатов психометрического анализа, выполненного в рамках IRT с использованием программного обеспечения Winsteps.			

Полученные результаты демонстрируют высокие показатели надежности и разделяющей способности для всех трех вариантов тестов. Коэффициент альфа-Кронбаха для каждого варианта превышает 0,84, а показатель Person Separation стабильно находится выше 2,0. Это свидетельствует о высокой согласованности заданий и их способности эффективно дифференцировать испытуемых по уровню подготовленности [14].

Однако, несмотря на высокие интегральные показатели, детальный анализ выявил ряд заданий, не соответствующих требованиям модели Раша. В частности, были идентифицированы вопросы с низкими и даже отрицательными значениями корреляции (Corr), а также с нежелательными значениями Infit MNSQ и Outfit MNSQ, выходящими за допустимые пределы (0,7-1,3). Данные показатели свидетельствуют о несоответствии заданий модели и указывают на наличие нестандартных паттернов ответов, что может быть вызвано некорректной формулировкой вопроса, ошибкой в ключе или неоднозначностью содержания.

Наиболее проблемным является Вариант 369168, где задание №2 демонстрирует отрицательную корреляцию, что является критическим индикатором его неработоспособности. В целом, во всех трех вариантах выявлены вопросы с высокими значениями Outfit MNSQ > 1,3, что указывает на непредсказуемость ответов, вероятно, из-за случайного угадывания или невнимательности. Например, в варианте 369694 таких заданий было обнаружено четыре.

Таким образом, несмотря на общую высокую надежность инструмента, наличие отдельных заданий, не соответствующих модели, ставит под вопрос валидность и справедливость оценки, поскольку они искажают результаты и не позволяют получить точное измерение латентной переменной [15].

На основе проведенного анализа, для повышения качества оценочного инструмента были предложены следующие рекомендации.

Проведение экспертной верификации. Все задания, демонстрирующие значительные отклонения от модели Раша, должны быть подвергнуты тщательной экспертной проверке. Это касается вопросов с низкими или отрицательными корреляциями и нестандартными значениями Infit/Outfit MNSQ. По итогам проверки они могут быть переформулированы или исключены из банка заданий.

Усиление контроля при разработке заданий. Необходимо усилить процедуру разработки и содержательной экспертизы тестовых заданий. Это позволит минимизировать ошибки в формулировках и исключить неоднозначность, что в свою очередь повысит их соответствие психометрическим требованиям.

Повторный анализ. После внесения всех корректировок необходимо провести повторную апробацию и повторный психометрический анализ. Это позволит объективно оценить степень улучшения качества теста и подтвердить его соответствие требованиям валидности и надежности.

Заключение

Проведенный психометрический анализ апробации подтверждает общее приемлемое качество тестовых заданий по математике, используемых в рамках ОЗП. Высокие показатели надежности и разделяющей способности тестов свидетельствуют об их пригодности для массовой аттестации. Замечания, выявленные в ходе апробации, были приняты во внимание, и на их основе была проведена необходимая работа по улучшению качества банка заданий. В частности, была проведена дополнительная экспертная проверка заданий, показавших отрицательную дискриминативность, и были пересмотрены или исключены задания, где дистракторы вводили в заблуждение сильных испытуемых. Также был усилен контроль за содержательной экспертизой на этапе разработки тестовых заданий и проведена обязательная

апробация новых заданий. Проведенная работа по корректировке банка тестовых заданий и повторный психометрический анализ позволили обеспечить высокое качество оценочных инструментов, что, в свою очередь, способствует повышению эффективности всей системы аттестации педагогических кадров в Казахстане.

Литература

1. Приказ Министра образования и науки Республики Казахстан от 27 января 2016 года №83 «Об утверждении Правил и условий проведения аттестации педагогов».
2. Закон Республики Казахстан от 27 июля 2007 года № 319-III. Об образовании.
3. Оценка знаний педагогов. Ұлттық тестілеу орталығы. http://new.testcenter.kz/?page_id=15731&lang=ru
4. Методические рекомендации по развитию профессиональной компетенции педагогов. – г. Астана: Национальная академия образования имени И. Алтынсарина, 2024. – 201 стр.
5. Магомедов, А.Г. (2022) Применение тестирования как инструмента педагогических измерений: возможности и ограничения. Современное педагогическое образование, №9, 62-67.
6. Cronbach, L. J. (1951) Coefficient alpha and the internal structure of tests // Psychometrika. 1951. № 16. P. 297-334.
7. Карданова Е.Ю., Иванова А.Е. (2023) Психометрические исследования: современные методы и новые возможности для образования. Вопросы образования, № 3, 8-19.
8. Корнилов, С.А. Проблема надежности психологических шкал и ее значение для психодиагностики в научных и практических целях. Экспериментальная психология в России: традиции и перспективы, 101-105.
9. Crocker, L., & Algina, J. (2008). Introduction to Classical and Modern Test Theory. Cengage Learning.
10. Статистика обучения с jamovi. https://lsj.readthedocs.io/ru/latest/.
11. Thompson, N. Classical Statistics and their Interpretation. https://assesshelp.zendesk.com/hc/en-us/articles/360037560611-Classical-Statistics-and-their-Interpretation.
12. Evaluating Item Difficulty and Discrimination in Knowledge Tests, January 4, 2024. https://foodsafety.institute/research-methodology/evaluating-item-difficulty-discrimination-tests/.
13. Practical Rasch Measurement. Tutorial 1. Analysis of dichotomous (true-false, multiple-choice), Mike Linacre, instructor - July 2011, www.winsteps.com.
14. Е.Б. Белов, М.В. Алексеев, Н.П. Китаев, А.И. Кучумов. Анализ влияния первичных баллов заданий на уровни подготовленности респондентов в дихотомической модели Раша на основе взвешенного метода максимального правдоподобия. Статистика и математические методы в экономике, 2019, №2.
15. Математико-статистические методы в эмпирических социально-экономических исследованиях: учеб. пособие / И.Н. Дубина. – М.: Финансы и статистика; ИНФРА-М. – 2010. – 416с.: ил.

Ж.Ж. Бейсенова, Л.Е. Шинетова, Г.К. Жабаева

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНДАҒЫ ПЕДАГОГТЕРДІҢ БІЛІМІН БАҒАЛАУ БОЙЫНША ТЕСТ ТАПСЫРМАЛАРЫНЫҢ ПСИХОМЕТРИЯЛЫҚ ТАЛДАУЫ (АПРОБАЦИЯ НӘТИЖЕЛЕРІ БОЙЫНША)

Бұл мақалада Қазақстанда педагогтердің білімін бағалау рәсімінде (ОББ) қолданылатын математика пәні бойынша тест тапсырмаларының апробациясы барысындағы толық психометриялық талдауы ұсынылған. Зерттеу аттестаттау кезінде қолданылған тесттің үш түрлі нұсқасын салыстырмалы талдауға негізделген. Тапсырмалардың сапасын бағалау және олардың күшті, әлсіз жақтарын анықтау үшін классикалық тест теориясы (КТТ) және қазіргі тест теориясы (IRT), атап айтқанда, Раш моделінің әдістері қолданылды. Жұмыстың мақсаты қолданыстағы бағалау құралдарын бағалаумен қатар, оларды жетілдіру бойынша нақты ұсыныстар әзірлеу болды. Апробация талдауы тесттердің жалпы сенімділігінің жоғары болуына (Cronbach's Alpha коэффициенті 0,84-тен жоғары) және олардың

тестіленушілерді дайындық деңгейі бойынша жақсы ажырата білуіне қарамастан, әр нұсқада бірқатар проблемалық тапсырмалар анықталғанын көрсетті. Атап айтқанда, дискриминативтілігі төмен немесе тіпті теріс тапсырмалар, сондай-ақ тиімсіз дистракторлар (қате жауап нұсқалары) анықталды. Апробация барысында анықталған барлық ескертулер ескеріліп, тест тапсырмаларының қорын түзету бойынша кейінгі жұмыстарда жойылды. Бұл бағалау құралдарының жалпы дәлдігі мен валидтілігін арттыруға ықпал етті.

Түйін сөздер: педагогтердің білімін бағалау, психометриялық талдау, апробация, сенімділік, валидтілік, классикалық тест теориясы, Раш моделі.

Zh.Zh. Beisenova, L.Y. Shinetova, G.K. Zhabayeva

PSYCHOMETRIC ANALYSIS OF TEST ITEMS FOR ASSESSING TEACHER KNOWLEDGE IN THE REPUBLIC OF KAZAKHSTAN (BASED ON APPROBATION RESULTS)

This article provides a detailed psychometric analysis of mathematics test items used during the approbation of the teacher knowledge assessment procedure (OZP) in Kazakhstan. The research is based on a comparative analysis of three different test versions administered during this piloting phase. To evaluate the quality of the items and identify their strengths and weaknesses, the study employed methods from both classical test theory (CTT) and modern item response theory (IRT), specifically the Rasch model. The purpose of the work was not only to assess the existing evaluation instruments but also to develop practical recommendations for their improvement. The approbation analysis showed that, despite the overall high reliability of the tests (Cronbach's Alpha coefficient exceeding 0,84) and their good ability to differentiate test-takers by proficiency level, a number of problematic items were found in each version. Specifically, items with low or even negative discrimination and ineffective distractors were identified. All issues discovered during the approbation have since been taken into account and rectified in the subsequent work on refining the test item bank. This has contributed to improving the overall accuracy and validity of the assessment instruments.

Keywords: teacher knowledge assessment, psychometric analysis, approbation, reliability, validity, classical test theory, Rasch model.

References

1. Prikaz Ministra obrazovaniya i nauki Respubliki Kazakhstan ot 27 yanvarya 2016 goda №83 «Ob utverzhdenii Pravil i usloviy provedeniya attestatsii pedagogov».
2. Zakon Respubliki Kazakhstan ot 27 iyulya 2007 goda № 319-III. Ob obrazovanii.
3. Otsenka znaniy pedagogov. Ұlttyq testileu ortalygy.
4. Metodicheskiye rekomendatsii po razvitiyu professionalnoy kompetentsii pedagogov. – g. Astana: Natsionalnaya akademiya obrazovaniya imeni I. Altynsarina, 2024. – 201 str.
5. Magomedov, A.G. (2022) Primeneniye testirovaniya kak instrumenta pedagogicheskikh izmereniy: vozmozhnosti i ogranicheniya. Sovremennoye pedagogicheskoye obrazovaniye, №9, 62-67.
6. Cronbach, L. J. (1951) Coefficient alpha and the internal structure of tests // Psychometrika. 1951. № 16. P. 297-334.
7. Kardanova, Ye.Yu., Ivanova A.Ye. (2023) Psikhometricheskiye issledovaniya: sovremennyye metody i novyye vozmozhnosti dlya obrazovaniya. Voprosy obrazovaniya, № 3, 8-19.
8. Kornilov, S.A. Problema nadezhnosti psikhologicheskikh shkal i yeye znacheniye dlya psikhodiagnostiki v nauchnykh i prakticheskikh tselyakh. Eksperimentalnaya psikhologiya v Rossii: traditsii i perspektivy, 101-105.
9. Crocker, L., & Algina, J. (2008). Introduction to Classical and Modern Test Theory. Cengage Learning.
10. Statistika obucheniya s jamovi. https://lsj.readthedocs.io/ru/latest/.
11. Thompson, N. Classical Statistics and their Interpretation. https://assesshelp.zendesk.com/hc/en-us/articles/360037560611-Classical-Statistics-and-their-Interpretation.
12. Evaluating Item Difficulty and Discrimination in Knowledge Tests, January 4, 2024. https://foodsafety.institute/research-methodology/evaluating-item-difficulty-discrimination-tests/.
13. Practical Rasch Measurement. Tutorial 1. Analysis of dichotomous (true-false, multiple-choice), Mike Linacre, instructor - July 2011, www.winsteps.com.

14. Belov, Ye.B., Alekseyev, M.V., Kitayev, N.P., & Kuchumov, A.I. (2019). Analiz vliyaniya pervichnykh ballov zadaniy na urovni podgotovlennosti respondentov v dikhotomicheskoy modeli Rasha na osnove vzveshennogo metoda maksimalnogo pravdopodobiya. Statistika i matematicheskiye metody v ekonomike, №2.

15. Matematichesko-statisticheskiye metody v empiricheskikh sotsialno-ekonomicheskikh issledovaniyakh: ucheb. posobiye / I.N. Dubina. – M.: Finansy i statistika; INFRA-M. – 2010. – 416s.: il.

Сведения об авторах:

Бейсенова Жанар Жумагазиновна – РГП на ПХВ «Национальный центр тестирования» МНВО РК, ведущий научный сотрудник, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: zhanar-beisenova@mail.ru.

Шинетова Ляззат Ермековна – РГП на ПХВ «Национальный центр тестирования» МНВО РК, заведующий лабораторией, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: shinetovalyazzat24@gmail.com.

Жабаева Гульдана Куатовна – РГП на ПХВ «Национальный центр тестирования» МНВО РК, ведущий научный сотрудник, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: danka1919@mail.ru.

Авторлар туралы мәлімет:

Бейсенова Жанар Жумагазиновна – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, жетекші ғылыми қызметкер, Родниковая көш., 1/1, 010008, Астана, Қазақстан, e-mail: zhanar-beisenova@mail.ru

Шинетова Ляззат Ермековна – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, зертхана меңгерушісі, Родниковая көшесі, 1/1, 010008, Астана, Қазақстан, e-mail: shinetovalyazzat24@gmail.com.

Жабаева Гульдана Куатовна – ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, жетекші ғылыми қызметкер, Родниковая көш., 1/1, 010008, Астана, Қазақстан, danka1919@mail.ru.

Information about authors:

Beisenova Zhanar Zhumagazinovna – 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, Leading Researcher, Rodnikovaya St., 1/1, 010008, Astana, Kazakhstan, e-mail: zhanar-beisenova@mail.ru.

Shinetova Lyazzat Yermekovna – 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, Rodnikovaya St., 1/1, 010008, Astana, Kazakhstan, e-mail: shinetovalyazzat24@gmail.com.

Zhabaeva Guldana Kuatovna – 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, Leading Researcher, Rodnikovaya St., 1/1, 010008, Astana, Kazakhstan, e-mail: danka1919@mail.ru.

ҒТАМР 14.01 21 https://doi.org/10.63597/UTO3105-4161.2025.3.3.003

Ұ. Садық[1], А.К. Макыбаева[2], Н.Б. Нуримбетов[3]
ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Астана қ, Қазақстан Республикасы
e-mail: u.sadykova@mail.ru, a_kaskataeva07@mail.ru, nurbekrm7@gmail.com

[1]ORCID 0009-0009-4633-1773, [2]ORCID 0009-0004-1310-9272, [3]ORCID 0009-0006-6910-2410

ҚАЗТЕСТ ЖҮЙЕСІНДЕГІ ДИАГНОСТИКАЛЫҚ ТЕСТІЛЕУДІҢ ӨТКІЗІЛУ БАРЫСЫ МЕН ТӘЖІРИБЕСІ

Мақалада ҚАЗТЕСТ жүйесі аясында жүргізілетін диагностикалық тестілеудің мазмұны, құрылымы, өткізілу тәртібі мен практикалық маңызы жан-жақты қарастырылады. Диагностикалық тест – мемлекеттік қызметшілердің, ұлттық компаниялар мен холдинг қызметкерлерінің, білім беру мекемелері мамандарының және басқа да әлеуметтік топ өкілдерінің қазақ тілін меңгеру деңгейін объективті түрде анықтауға бағытталған құрал ретінде сипатталады. Бұл тестілеудің негізгі мақсаты – қатысушылардың тыңдалым, оқылым дағдылары бойынша тілдік деңгейін бағалау арқылы олардың қазақ тілін меңгерудегі нақты орнын айқындау, әрі қарайғы оқыту бағдарламаларын тиімді жоспарлау болып табылады.

Мақалада тест мазмұнына қойылатын талаптар, тапсырмалар типтері, бағалау критерийлері, қолданылатын әдістемелік құралдар және тестілеуді ұйымдастыру тәртібі нақты мысалдармен түсіндіріледі. Сонымен қатар, ҚАЗТЕСТ жүйесінің даму кезеңдері, оны құқықтық-нормативтік тұрғыдан қамтамасыз ету жолдары, мемлекеттік тіл саясатын іске асырудағы орны және халықаралық тілдік тест жүйелерімен салыстырмалы талдауы берілген. Авторлар диагностикалық тестілеудің артықшылықтарын, атап айтқанда, тіл меңгерушінің қажеттіліктерін нақты анықтауға, оқыту әдістерін бейімдеуге және мемлекеттік тілдің қолдану аясын кеңейтуге қосатын үлесін атап көрсетеді.

Зерттеу нәтижелері диагностикалық тестілеудің қазақ тілін оқыту мен бағалау жүйесіндегі тиімділігін дәлелдейді және оның мемлекеттік тіл саясатын қолдаудағы стратегиялық маңызын негіздейді.

Түйін сөздер: ҚАЗТЕСТ, диагностикалық тестілеу, қазақ тілі, тілдік құзыреттілік, мемлекеттік тіл саясаты, бағалау жүйесі.

Кіріспе

Мемлекеттік тілдің мәртебесін арттыру мен оны дамыту – қазіргі уақытқа дейін күн тәртібіндегі, қоғамның басты назарында тұрған өзекті мәселелердің бірі. Қазақстан тәуелсіздік алғаннан бері тіл саясаты саласында бірқатар заңнамалық және бағдарламалық құжаттар қабылданып, нақты іс-шаралар жүзеге асырылды. 1997 жылы «Қазақстан Республикасындағы тілдер туралы» Заң қабылданды [1]. Одан кейін тілдерді дамытудың бірнеше мемлекеттік бағдарламасы әзірленіп, бекітілді: 1998 жылғы алғашқы мемлекеттік бағдарлама, 2001–2010 жылдарға арналған екінші бағдарлама, 2011–2019 жылдарға арналған үшінші бағдарлама және 2020–2025 жылдарға арналған мемлекеттік бағдарлама [2]. Бүгінде 2023–2029 жылдарға арналған тіл саясатын дамыту тұжырымдамасы жұмыс істеуде [16]. Бұл құжаттардың барлығы мемлекеттік тілдің қолданылу аясын кеңейтуге, оның қоғамдық өмірдің барлық саласында орнығуына бағытталды.

Осы бағдарламалар аясында қазақ тілін меңгергендердің үлесін арттыру, мемлекеттік органдарда іс қағаздарын мемлекеттік тілде жүргізуді қамтамасыз ету негізгі міндеттердің бірі ретінде айқындалды [4]. Әсіресе мемлекеттік қызметшілердің тілдік құзыреттілігін бағалау және оны жетілдіру мақсатында ұлттық тестілеу жүйесін енгізу қажеттілігі туындады. Халықаралық тәжірибеде (TOEFL, IELTS, DALF/DELF, DSH, CELI, TÖMER, ТРКИ және т.б.) бар тіл меңгеру деңгейін анықтайтын ұлттық жүйелер үлгісінде Қазақстанда да осындай жүйе құруға шешім қабылданды. Соның нәтижесінде, 2006 жылы ҚР Білім және ғылым министрлігі Ұлттық

тестілеу орталығының базасында ҚАЗТЕСТ жүйесі құрылды. Бұл жүйе азаматтардың қазақ тілін меңгеру деңгейін объективті бағалау, мемлекеттік тіл саясатын жүзеге асыру және тіл үйренуді ынталандыру мақсатында жасалды [5].

Зерттеу материалдары мен әдістері

Зерттеудің негізгі материалдары – ҚАЗТЕСТ жүйесінің 2006 жылдан бері қалыптасқан нормативтік құжаттары, тест спецификациялары, ұлттық стандарттар (ҚР СТ 1925 – 1929, ҚР СТ 1926 – 1929, ҚР СТ 1927 – 1929, ҚР СТ 1928 – 1929, ҚР СТ 1929 – 1929) [10-15], диагностикалық тесттің өткізілу тәртібі мен нәтижелері туралы ресми деректер. [9] Сонымен қатар Қазақстан Республикасының Конституциясы, «Тіл туралы» Заңы, 2008 жылғы № 835 Үкімет қаулысы, мемлекеттік бағдарламалар (1998–2000, 2001–2010, 2011–2019, 2020–2025, 2023–2029) әдіснамалық негіз ретінде қарастырылды [16].

Зерттеу барысында салыстырмалы талдау, педагогикалық өлшеу әдістері, статистикалық сараптама қолданылды. TOEFL, IELTS, TORFL сияқты халықаралық жүйелермен салыстырмалы талдау жүргізіліп, ҚАЗТЕСТ диагностикасының ерекшеліктері айқындалды [3].

Нәтижелер және талқылау

ҚАЗТЕСТ жүйесінің диагностикалық тестілеуі он жылдан астам уақыт бойы тұрақты түрде жүргізіліп келеді және осы кезеңде жиналған деректер қазақ тілін меңгерудің қоғамдық деңгейін жан-жақты сипаттауға мүмкіндік берді. Талдаулар көрсеткендей, жылдар бойынша нәтижелерде бірқатар оң өзгерістер байқалады, бұл мемлекеттік тіл саясатының тиімділігін айқындай түседі [6].

Диагностикалық тестілеудің негізгі мақсаттары:

- мемлекеттік қызметшілердің қазақ тілін меңгеру деңгейін анықтау;
- әртүрлі сала мамандарын кәсіби міндеттерін мемлекеттік тілде орындауға ынталандыру;
- тіл үйрету курстары мен бағдарламаларының тиімділігін бағалау.

Диагностикалық тест құрылымы тыңдалым, оқылым бөліктерін қамтиды. Тест тапсырмалары деңгейлік принципке негізделіп әзірленген (A1–C1). 2008–2021 жылдары үміткерлерге 150 тапсырма ұсынылса, 2022 жылдан бастап жаңа форматта 100 тапсырма енгізіліп, құрылымы ықшамдалды. Қазіргі уақытта тестілеу 60 тапсырманы құрайды.

Диагностикалық тесттің құрылымы:

Тыңдалым – 20 тапсырма (20 минут)

Оқылым – 40 тапсырма (50 минут)

Барлығы – 60 тапсырма, жалпы ұзақтығы – 70 минут.

Нәтижесі әр блок бойынша пайыздық көрсеткішпен анықталады, қатысушыларға қорытынды анықтама беріледі.

ҚАЗТЕСТ жүйесі мемлекеттік тілдің қоғамдық өмірдегі мәртебесін арттыруға ықпал ететін маңызды құралға айналды [4]. Диагностикалық тестке қатысушылар контингенті жыл сайын кеңеюде. 2008 жылдан бері мемлекеттік қызметшілер, ұлттық компаниялар қызметкерлері, әскери қызметшілер қатысып келеді [17].

2008 жылдан бастап ҚАЗТЕСТ жүйесі бойынша қазақ тілінің белгілі бір деңгейін меңгергендігін растау үшін Болашақ бағдарламасының үміткерлерінен диагностикалық тестілеу тапсыру жеткілікті болса, 2011 жылдан бастап оларға сертификаттық тестілеуден өту міндеттелді.

2011 жылдан Назарбаев зияткерлік мектептерінің мұғалімдері мен қызметкерлеріне диагностикалық тестілеу нәтижесі жеткілікті болса, 2012 жылдан бастап белігілі бір деңгейдің сертификатын иемдену шарт бола бастады.

Осылайша ҚАЗТЕСТ жүйесі бойынша диагностикалық тестілеу сертификаттық тестілеуге дайындықты ұйымдастыру үшін талассыз әсері оң екені, яғни негіз қалыптастыруда ерекше орны бар екені дәлелденді.

1 сурет – Диагностикалық тестілеу тапсыру үрдісі

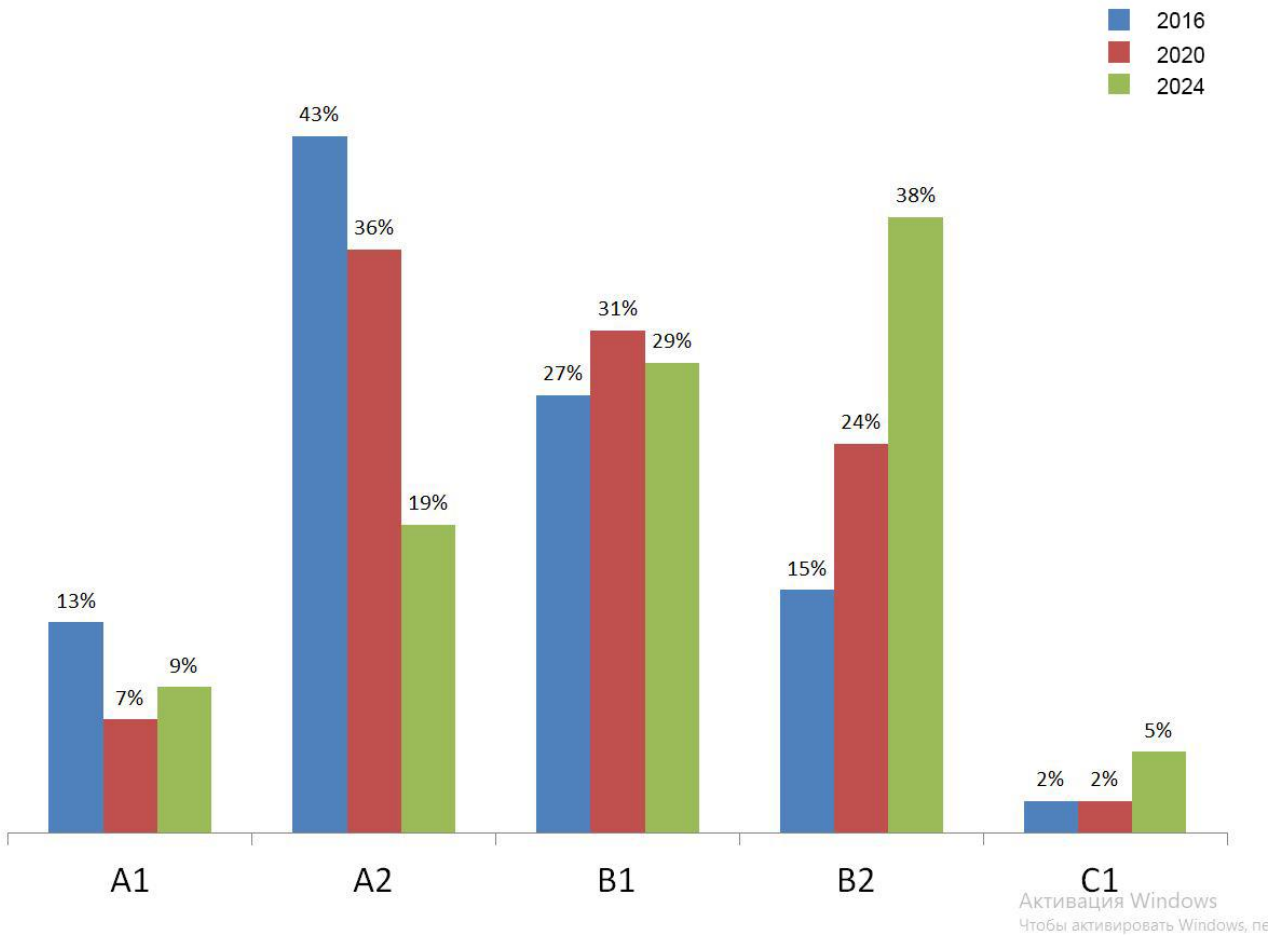
Мысалы, 2011-жылдан бастап бүгінгі күнге дейін диагностикалық тестілеуге қатысқан азаматтардың саны 500 мыңға жуық. Бұл деректер мемлекеттік тіл саясатын жүзеге асырудағы ҚАЗТЕСТ жүйесінің ауқымын көрсетеді.

2016 жылдан 2024 жылға дейінгі диагностикалық тест нәтижелері азаматтардың қазақ тілін меңгеру деңгейінде оң өзгерістер бар екенін көрсетті. Бұл динамика мемлекеттік бағдарламалардың кезең-кезеңімен іске асырылуының және әдістемелік базаның жаңартылуының нақты дәлелі болып табылады [21].

ҚАЗТЕСТ жүйесінің құрамдас бөліктерінің бірі – диагностикалық тестілеу. Ол тест тапсырушының бастапқы тілдік деңгейін анықтау үшін, қызметкердің тілдік дайындығын аралық бағалап, әрі қарайғы оқу процесін тиімді жоспарлауға мүмкіндік беру мақсатында қолданылады [8].

ҚАЗТЕСТ жүйесі бойынша диагностикалық тестілеу – тілді жүйелі әрі нәтижелі меңгерудің алғашқы баспалдағы, тіл меңгеру процесінің стратегиялық бастамасы. Ол тест тапсырушының тілдік мүмкіндіктерін саралап, әрі қарайғы оқу траекториясын дұрыс құруға ықпал етеді. Тіл

үйренушінің бастапқы деңгейін нақтылап, оқыту траекториясын тиімді бағыттауға мүмкіндік береді. Диагностикалық тест нәтижелері негізінде құрастырылған оқыту бағдарламалары жоғары нәтижеге қол жеткізуге септігін тигізеді [7].

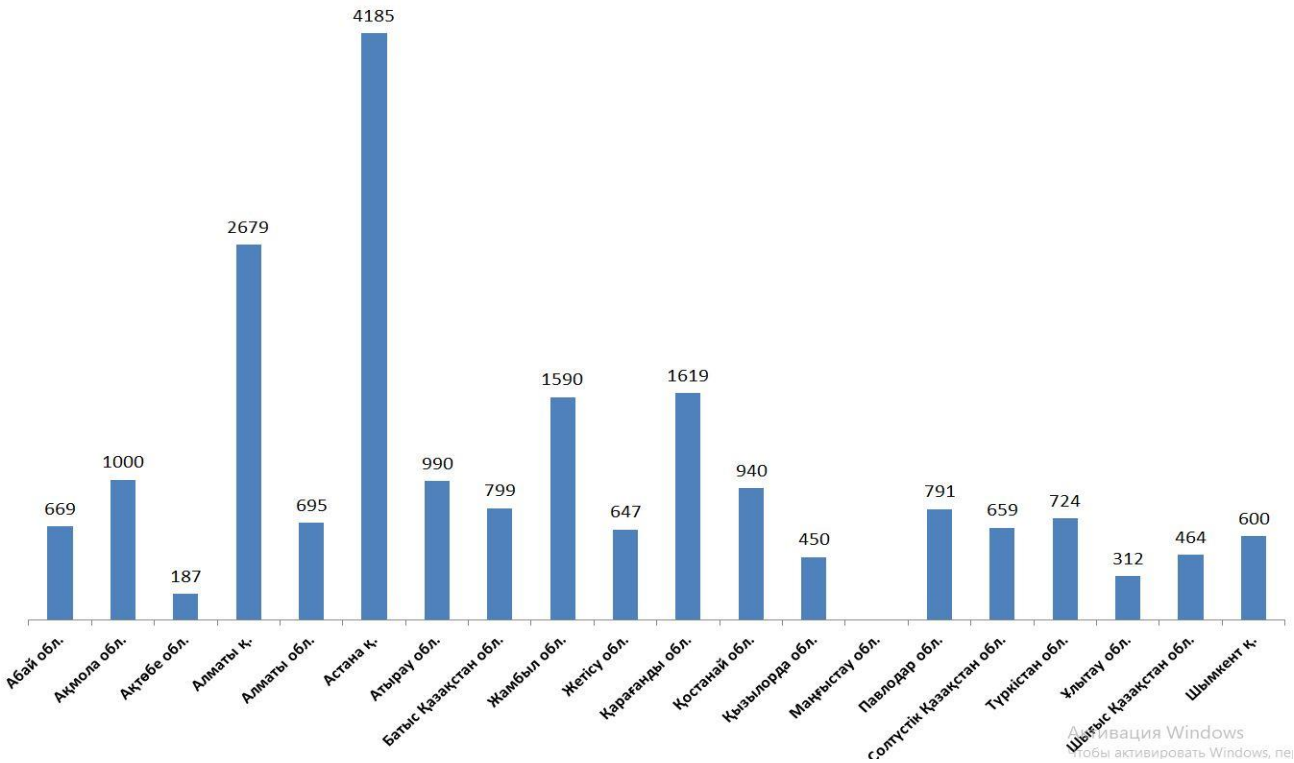


2 сурет – Динамика көрсеткіші (2016 – 2024 жылдар аралығы)

- 2016 жылы тест тапсырушылардың басым бөлігі А1 және А2 деңгейлерінде шоғырланған болатын, яғни тілді бастапқы деңгейде меңгергендер үлесі жоғары болды;
- 2020 жылы жағдай айтарлықтай өзгеріп, тест тапсырушылар арасында В1 және В2 деңгейлері басымдық танытты. Бұл кезеңде тілді орта деңгейде меңгергендер саны артқаны байқалды;
- 2024 жылғы көрсеткіштер оң динамикаға ие болды: тест тапсырушылардың елеулі бөлігі В1-В2 және С1 деңгейінде қазақ тілін меңгергенін көрсетті.

Бұл көрсеткіштер қазақ тілін меңгерудің бастапқы деңгейінен орта және жоғары деңгейге көшу қарқынының бар екенін байқатады. Демек, диагностикалық тест тек білімді өлшеу құралы ғана емес, сонымен қатар азаматтарды тіл үйренуге ынталандыратын маңызды тетікке айналып отыр [18].

2024 жылы ҚАЗТЕСТ жүйесі бойынша 20 мың адам диагностикалық тесттен өтті. Бұл сандық көрсеткіш жүйенің тұрақты түрде сұранысқа ие екенін дәлелдейді. Өңірлер бойынша тестілеуге қатысқандардың үлесі төмендегі диаграммада көрсетілген.



3 сурет - Өңірлер бойынша диагностикалық тестілеуге қатысқандардың үлесі

ҚАЗТЕСТ жүйесінің әдістемелік ерекшеліктеріне тоқталсақ, тест тапсырмаларын әзірлеу үдерісінде деңгейлік оқулықтар, оқу-әдістемелік кешендер («Тіл-құрал»), лексикалық минимумдар кеңінен қолданылады. Тест тапсырмалары халықаралық тәжірибеге бейімделіп, үнемі жаңартылып отырады. Сонымен қатар, тест әзірлеушілер мен сарапшылардың кәсіби біліктілігін арттыру үшін семинарлар, халықаралық тренингтер өткізіледі [17].

ҚАЗТЕСТ жүйесі халықаралық ынтымақтастық саласында да танымал. ETS Global (TOEFL), TOMER (Түркия), ТРКИ (Ресей), ҚЫРҒЫЗТЕСТ (Қырғызстан) орталықтарымен әріптестік орнатылған. Бұл қадамдар ҚАЗТЕСТ нәтижелерін халықаралық стандарттарға сәйкестендіруге және болашақта шетелде мойындатуға мүмкіндік береді [19].

Аталған жүйелердің сол елдегі тілдік саясатқа бірден-бір ықпал ететін құралдар екендігі күмән туғызбайды. АҚШ, Канада, Австралия елдеріне барып белгілі бір оқу орнында оқу үшін міндетті түрде TOEFL нәтижесін талап етеді, дәл сондай Ұлыбританияда IELTS сертификаты шетелден келген азаматтың елде өзін еркін сезініп, тілдік ортада қатысымға бөгетсіз араласуына мүмкіндік береді. Олай болса тілдік білімді бағалау жүйелері тілдік тұлға қалыптастыруға септігін тигізетіні анық және тәжірибе көрсеткендей тілді меңгерудің тетігі. Осы ретте, нәтижеге бағытталған тіл үйренудің, оқытудың әдістемесі қалыптасады. Сонымен, тілдік саясат жоғарыда тілге тиек етілген жүйелер, яғни тетіктер арқылы шешімін тауып отыр. Елге келгенде елдің тілін үйреніп, содан кейін ғана қандай-да бір әрекет түрімен айналыса бастауға болады.

Қазіргі жағдайдағы тіл құрылысының қисыны еліміздегі бірқатар проблемалы аспектілермен байланысты болып отыр. Оның бірі – қоғамдағы мемлекеттік тілді меңгерудің әркелкі деңгейі. Бұл проблемалы аспект мемлекеттік тілді оқытудың бірыңғай әдіснамасы мен стандарттарының енгізілмеуімен, қазақ тілі оқытушылары мен мамандарының кәсіби

дәрежелерінің сәйкессіздігімен, қазақ тілін оқыту инфрақұрылымы қызметінің бірыңғай стандарттарының болмауымен тікелей байланысты болып отыр [20].

Қазақстандағы әрбір азаматтың белгілі бір деңгейде мемлекеттік тілді меңгеруі – көп ұлтты Қазақстан халқын топтастырудың аса маңызды факторы, ал қазақ тілін меңгеру – қазақстандық патриотизмнің құрамдас бөлігі болатыны барлық насихаттық және құқықтық құжаттарда көрініс тапқан. Қазақ тілінің мемлекеттік тіл ретінде бәсекелестікке қабылеттілігі артып, біртіндеп түзеліп келеді десек, бұл тілдік саясаттың жүргізілуінен болар. Кейбір саяси, экономикалық жағдайларға байланысты тілдік ахуалдың толыққанды орындалмай жатқан тұстары бар екені рас. Дей тұрғанмен, ғаламды өз иіріміне салып, өз заңымен жүргізуге бағындырып отырған жаһандану дәуірінде ұлттың келбетін сақтап қалуда қазақ тілі едәуір дамып келеді. Қазақ тілінің мемлекеттік мәртебесін күшейту мақсатында елімізде әлеуметтік-экономикалық өзгерістер жасалуда, айтарлықтай іс-шаралар атқарылуда.

ҚАЗТЕСТ жүйесі бойынша диагностикалық тестілеу – тілді жүйелі әрі нәтижелі меңгерудің алғашқы баспалдағы, тіл меңгеру процесінің стратегиялық бастамасы. Ол тест тапсырушының тілдік мүмкіндіктерін саралап, әрі қарайғы оқу траекториясын дұрыс құруға ықпал етеді. Тіл үйренушінің бастапқы деңгейін нақтылап, оқыту траекториясын тиімді бағыттауға мүмкіндік береді. Диагностикалық тест нәтижелері негізінде құрастырылған оқыту бағдарламалары жоғары нәтижеге қол жеткізуге септігін тигізеді.

Қорытынды

ҚАЗТЕСТ жүйесіндегі диагностикалық тестілеу мемлекеттік тіл саясатының стратегиялық құралы болып табылады. 2011 жылдан бері бұл тестілеуге шамамен 500 мың азамат қатысып, олардың қазақ тілін меңгеру деңгейі жан-жақты бағаланды. Бұл көрсеткіш ҚАЗТЕСТ жүйесінің ауқымын, мемлекеттік тілдің қоғамдық өмірдегі мәртебесін арттыруға қосқан үлесін айқындайды [4, 5, 17].

Диагностикалық тест азаматтардың тіл меңгеру деңгейін анықтап қана қоймай, әрі қарай оқыту бағытын айқындауға, оқу бағдарламаларын жетілдіруге мүмкіндік берді. Жинақталған тәжірибе тесттің тілді меңгерудің объективті көрсеткішін беріп, мемлекеттік тілдің кәсіби ортада қолданылуын күшейтуде маңызды рөл атқарғанын көрсетті [18, 19].

Жүйенің жаңартылған форматы нәтижелердің әділдігін, халықаралық стандарттарға сәйкестігін қамтамасыз етті. Алдағы уақытта ҚАЗТЕСТ тестілеуін онлайн форматта кеңейту, жасанды интеллект негізінде талдау енгізу, халықаралық деңгейде аккредиттеу жүйенің тиімділігін арттырып, қазақ тілінің жаһандық мәртебесін көтеруге ықпал етпек [3, 21].

Әдебиеттер

1. Қазақстан Республикасының «Тіл туралы» Заңы. – Астана, 1997.
2. Қазақстан Республикасының Білім туралы Заңы. – Астана, 2007.
3. CEFR – Common European Framework of Reference for Languages: Learning, Teaching, Assessment. – Council of Europe, 2001.
4. Әбілдаева, Ж.Б. Мемлекеттік тілді меңгеру деңгейін анықтаудың ҚАЗТЕСТ жүйесі // Тілтаным. – 2019. –No3(77). – Б. 45–52.
5. Ерсултанова, Г.С. Тілді меңгеру деңгейін бағалаудағы ҚАЗТЕСТ жүйесінің рөлі: жаңа форматтағы тиімділік // Қазақстан Республикасында мемлекеттік тілді оқытудың жүйесі: деңгейлері мен сатылары. Конференция материалдары. – Алматы: Мемлекеттік тілді дамыту институты, 2024. – 362 б.

6. Кульманова, Ш.Т. Қазақ тілін екінші тіл ретінде оқыту: оқыту технологиясы және бағалау жүйесі. –Алматы: Қазақ университеті, 2021. – 168 б.

7. Ахметова, Г.А. Мемлекеттік тілді оқытуда ҚАЗТЕСТ жүйесінің рөлі // ҚазҰУ хабаршысы. Филология сериясы. – 2020. – No2(180). – Б. 110-115.

8. Нұрғали, Ә. ҚАЗТЕСТ жүйесінің тіл меңгерудегі маңыздылығы // Қазақ тілі мен әдебиеті. – 2021. – No4. – Б. 75–80.

9. KAZTEST.kz – Ресми ақпараттық сайты. – Қолжетімді: http://www.kaztest.kz (қол жеткізу күні: 16.06.2025).

10. Қазақстан Республикасы Ғылым және жоғары білім министрінің 2024 жылғы 23 қыркүйектегі №458 бұйрығы «ҚАЗТЕСТ» қазақ тілін меңгеру деңгейін бағалау жүйесі бойынша қағидаларды бекіту туралы.

11. ҚР СТ 1926-2009. Қазақ тілін меңгерудің коммуникативтік тілдік құзыреттіліктері. Қарапайым деңгей. Жалпы меңгеру.

12. ҚР СТ 1928-2009. Қазақ тілін меңгерудің коммуникативтік тілдік құзыреттіліктері. Базалық деңгей. Жалпы меңгеру.

13. ҚР СТ 1929-2009. Қазақ тілін меңгерудің коммуникативтік тілдік құзыреттіліктері. Орта деңгей. Жалпы меңгеру.

14. ҚР СТ 1925-2009. Қазақ тілін меңгерудің коммуникативтік тілдік құзыреттіліктері. Ортадан жоғары деңгей. Жалпы меңгеру.

15. ҚР СТ 1927-2009. Қазақ тілін меңгерудің коммуникативтік тілдік құзыреттіліктері. Жоғары деңгей. Жалпы меңгеру.

16. Қазақстан Республикасында тіл саясатын дамытудың 2023-2029 жылдарға арналған тұжырымдамасы.

17. Ә.Ж. Болатқызы. Қазақ тілін меңгеру деңгейін бағалаудың ұлттық жүйесі – ҚАЗТЕСТ-тің қазіргі жағдайы мен даму перспективалары // «Білім мен ғылымды дамытудың халықаралық тәжірибесі» халықаралық онлайн конференция материалдары. – 2025.

18. Әділбекова, Ж.Б. ҚАЗТЕСТ жүйесінің тілдік білімді бағалаудағы маңызы мен қызметі // «Bilim. Ǵylym. Innovatsıa» республикалық ғылыми-әдістемелік интернет журналы. – 2022.

19. Әділбекова, Ж.Б. «ǴYLYM JÁNE BILIM – 2024» студенттер мен жас ғалымдардың XIX Халықаралық конференция материалдары. – 2024.

20. Әділбекова, Ж.Б. Мемлекеттік тілді дамытудағы Қазтесттің орны мен мемлекеттік бағдарламаға сәйкес өткізілген тестілеу нәтижелері. – 2023.

21. Yersultanova, G., Serdalina, S. KAZTEST: Test Administration Procedures and Experience. – 2024.

U. Sadyk, A. K. Makybayeva, N. B. Nurimbetov

THE PROCESS AND EXPERIENCE OF DIAGNOSTIC TESTING IN THE KAZTEST SYSTEM

This article provides a comprehensive analysis of the content, structure, administration procedures, and practical significance of diagnostic testing conducted within the KAZTEST system. The diagnostic test is presented as a tool aimed at objectively determining the level of Kazakh language proficiency among civil servants, employees of national companies and holdings, specialists in educational institutions, as well as representatives of various social groups. The primary goal of the testing is to assess participants' proficiency in listening and reading, identify their actual language level, and effectively plan subsequent educational programs.

The article describes in detail the test content requirements, task types, assessment criteria, methodological materials used, and the organization procedures, supported by specific examples. In addition, it outlines the stages of development of the KAZTEST system, its legal and regulatory foundations, and its role in implementing the state language policy, and a comparative analysis with international language testing systems. The authors highlight the advantages of diagnostic testing; particularly its ability to accurately identify learners' needs, adapt teaching methods accordingly, and contribute to the expansion of the use of the Kazakh language in various spheres.

The research findings demonstrate the effectiveness of diagnostic testing in the Kazakh language teaching and assessment system and substantiate its strategic importance in supporting national language policy.

Keywords: KAZTEST, diagnostic testing, Kazakh language, language competence, state language policy, assessment system.

Ұ. Садық, А. К. Макыбаева, Н. Б. Нуримбетов

ПРОЦЕСС ПРОВЕДЕНИЯ И ОПЫТ ДИАГНОСТИЧЕСКОГО ТЕСТИРОВАНИЯ В СИСТЕМЕ КАЗТЕСТ

В статье всесторонне рассматриваются содержание, структура, порядок проведения и практическая значимость диагностического тестирования, проводимого в рамках системы КАЗТЕСТ. Диагностический тест представлен как инструмент, направленный на объективное определение уровня владения казахским языком государственными служащими, сотрудниками национальных компаний и холдингов, специалистами образовательных учреждений, а также представителями различных социальных групп. Основной целью данного тестирования является оценка языкового уровня участников по таким навыкам, как аудирование и чтение, а также определение их реального уровня владения языком для последующего эффективного планирования образовательных программ.

В статье подробно описаны требования к содержанию теста, типы заданий, критерии оценки, используемые методические материалы и порядок организации тестирования с конкретными примерами. Также рассматриваются этапы развития системы КАЗТЕСТ, правовые и нормативные основы ее функционирования, роль в реализации государственной языковой политики и сравнительный анализ с международными системами языкового тестирования. Авторы подчёркивают преимущества диагностического тестирования, в частности его способность точно определять потребности изучающих язык, адаптировать методы обучения и способствовать расширению сферы использования государственного языка.

Результаты исследования доказывают эффективность диагностического тестирования в системе преподавания и оценки казахского языка, а также обосновывают его стратегическое значение в поддержке государственной языковой политики.

Ключевые слова: КАЗТЕСТ, диагностическое тестирование, казахский язык, языковая компетенция, государственная языковая политика, система оценивания.

References

1. Qazaqstan Respublikasynyñ «Tıl turaly» Zañy. – Astana, 1997.
2. Qazaqstan Respublikasynyñ Bılım turaly Zañy. – Astana, 2007.
3. CEFR – Common European Framework of Reference for Languages: Learning, Teaching, Assessment. – Council of Europe, 2001.
4. Äbıldaeva, J.B. Memlekettık tıldı meñgeru deñgeiın anyqtaudyñ QAZTEST jüiesı // Tıltanym. – 2019. – №3(77). – B. 45–52.
5. Ersultanova, G.S. Tıldı meñgeru deñgeiın bağalaudağy QAZTEST jüiesınıñ rölı: jaña formattağy tıımdılık // Qazaqstan Respublikasynda memlekettık tıldı oqytudyñ jüiesı: deñgeilerı men satylary. Konferensia, seminar-keñes materialdarynyñ jinağy. – Almaty: Memlekettık tıldı damytu instituty, 2024. – 362 b.
6. Külmanova, Ş.T. Qazaq tılın ekınşı tıl retında oqytu: oqytu tehnologiasy jäne bağalau jüiesı. – Almaty: Qazaq universitetı, 2021. – 168 b.
7. Ahmetova, G.A. Memlekettık tıldı oqytuda QAZTEST jüiesınıñ rölı // QazŪU habarşysy. Filologia seriasy. – 2020. – №2(180). – B. 110–115.
8. Nūrğali, Ä. QAZTEST jüiesınıñ tıl meñgerudegı mañyzdylyğy // Qazaq tılı men ädebietı. – 2021. – №4. – B. 75–80.

9. KAZTEST.kz – Resmi aqparattyq saity. – Qoljetımdı: http://www.kaztest.kz (qol jetkızu künı: 16.06.2025).

10. Qazaqstan Respublikasy Ğylym jäne joğary bılım ministrınıñ 2024 jylğy 23 qyrküiektegı №458 būiryğy «QAZTEST» qazaq tılın meñgeru deñgeiın bağalau jüiesı boiynşa qağidalardy bekıtu turaly.

11. QR ST 1926–2009. Qazaq tılın meñgerudıñ komunikativtık tıldık qūzyrettılıkterı. Qarapaiym deñgei. Jalpy meñgeru.

12. QR ST 1928–2009. Qazaq tılın meñgerudıñ komunikativtık tıldık qūzyrettılıkterı. Bazalyq deñgei. Jalpy meñgeru.

13. QR ST 1929–2009. Qazaq tılın meñgerudıñ komunikativtık tıldık qūzyrettılıkterı. Orta deñgei. Jalpy meñgeru.

14. QR ST 1925–2009. Qazaq tılın meñgerudıñ komunikativtık tıldık qūzyrettılıkterı. Ortadan joğary deñgei. Jalpy meñgeru.

15. QR ST 1927–2009. Qazaq tılın meñgerudıñ komunikativtık tıldık qūzyrettılıkterı. Joğary deñgei. Jalpy meñgeru. 16. Qazaqstan Respublikasynda tıl saiasatyn damytudyñ 2023–2029 jyldarğa arnalğan tūjyrymdamasy.

17. Adilbekova, Zh.B. The Current State and Development Prospects of the National Kazakh Language Proficiency Assessment System – KAZTEST // Proceedings of the International Online Conference “International Experience in the Development of Education and Science”. – 2025.

18. Adilbekova, Zh.B. The Importance and Function of the KAZTEST System in Language Knowledge Assessment // Bilim. Gylym. Innovatsiya, National Scientific and Methodological Online Journal. – 2022.

19. Adilbekova, Zh.B. Proceedings of the XIX International Conference for Students and Young Scientists “GYLYM JÁNE BILIM – 2024”. – 2024.

20. Adilbekova, Zh.B. The Role of KAZTEST in the Development of the State Language and the Results of Testing Conducted According to the State Program. – 2023.

21. Yersultanova, G., Serdalina, S. KAZTEST: Test Administration Procedures and Experience. – 2024.

Авторлар туралы мәлімет:

Садық Ұлжан Расылханқызы – филология ғылымдарының магистрі, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК ҚАЗТЕСТ жүйесін дамыту жөніндегі басқарма, Тестілеуді ұйымдастыру бөлімінің басшысы, Родниковая көш., 1/1, 010008, Астана, Қазақстан, e-mail: u.sadykova@mail.ru

Макыбаева Асель Кенжебековна – тарих бакалавры, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК ҚАЗТЕСТ жүйесін дамыту жөніндегі басқарма, Тестілеуді ұйымдастыру бөлімінің жетекші сарапшысы, Родниковая көш., 1/1, 010008, Астана, Қазақстан, e-mail: a_kaskataeva07@mail.ru

Нуримбетов Нурбек Батирханович – Астана халықаралық университетінің магистранты, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК ҚАЗТЕСТ жүйесін дамыту жөніндегі басқарма, Тестілеуді ұйымдастыру бөлімінің жетекші сарапшысы, Родниковая көш., 1/1, 010008, Астана, Қазақстан, e-mail: a_kaskataeva07@mail.ru

Сведения об авторах:

Садық Ұлжан Расылханқызы – магистр филологических наук, начальник отдела организации тестирования Управления по развитию системы КАЗТЕСТ РГП на ПХВ «Национальный центр тестирования» МНВО РК, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: u.sadykova@mail.ru

Макыбаева Асель Кенжебековна – бакалавр истории, ведущий эксперт отдела организации тестирования Управления по развитию системы КАЗТЕСТ РГП на ПХВ «Национальный центр тестирования» МНВО РК, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: a_kaskataeva07@mail.ru

Нуримбетов Нурбек Батирханович – магистрант Международного университета Астана, ведущий эксперт отдела организации тестирования Управления по развитию системы КАЗТЕСТ РГП на ПХВ «Национальный центр тестирования» МНВО РК, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: nurbekrm7@gmail.com

Information about Authors:

Ulzhan Rasylkhankyzy Sadyk – Master of Philological Sciences, Head of the Testing Organization Department, Department for the Development of the KAZTEST System, National Testing Center RSE on REM, Ministry of Science and Higher Education of the Republic of Kazakhstan, 1/1 Rodnikovaya St., 010008, Astana, Kazakhstan, e-mail: u.sadykova@mail.ru

Assel Kenzhebekovna Makybayeva – Bachelor of History, Leading Expert of the Testing Organization Department, Department for the Development of the KAZTEST System, National Testing Center RSE on REM, Ministry of Science and Higher Education of the Republic of Kazakhstan, 1/1 Rodnikovaya St., 010008, Astana, Kazakhstan, e-mail: a_kaskataeva07@mail.ru

Nurbek Batirkhanovich Nurimbetov – Master's Student, Astana International University, Leading Expert of the Testing Organization Department, Department for the Development of the KAZTEST System, National Testing Center RSE on REM, Ministry of Science and Higher Education of the Republic of Kazakhstan, 1/1 Rodnikovaya St., 010008, Astana, Kazakhstan, e-mail: nurbekrm7@gmail.com.

ҒТАМР 14.25.05 https://doi.org/10.63597/UTO3105-4161.2025.3.3.004

И.У. Сагиндиков*[1], А.Ш. Танирбергенова[2], А. Мағауия[3]
[1,2,3]Ы. Алтынсарин атындағы Ұлттық білім академиясы Астана қ, Қазақстан Республикасы
*e-mail: Isatay60@mail.ru

[1]ORCID 0009-0004-5522-0649, [2]ORCID 0000-0002-8002-7910, [3]ORCID 0009-0008-5462-6734

БІЛІМ БЕРУ ҰЙЫМДАРЫНДАҒЫ ТӘРБИЕ ПРОЦЕСІН БАҒАЛАУДЫҢ ЭКОЖҮЙЕЛІК ТӘСІЛДЕМЕСІ

Мақалада білім беру ұйымдарындағы тәрбие процесін бағалаудың экожүйелік тәсілдемесі қарастырылады. Қазіргі таңда тәрбие нәтижелерін өлшеу көбінесе формалды көрсеткіштермен шектеліп, білім алушылардың тұлғалық дамуы мен құндылықтық бағдарларын толық ашып көрсетпейді. Осыған байланысты экожүйелік тәсілдеме тәрбие процесін кешенді бағалаудың тиімді тетігі ретінде ұсынылады. Зерттеу барысында макро, мезо, микроорта деңгейлерінің өзара байланысы талданып, білім беру ұйымының тәрбие кеңістігін әлеуметтік серіктестік, отбасы мен мектептің ынтымақтастығы, мектеп мәдениеті тұрғысынан қарастырудың маңызы негізделеді.

Мақалада «Адал азамат» біртұтас тәрбие бағдарламасының негізгі құндылықтарына сүйене отырып, оқушы тұлғасын қалыптастырудың экожүйелік моделін сипаттау ұсынылады. «Адал азамат» моделін енгізу арқылы тәрбие нәтижелерін айқындау, білім алушылардың адалдық, әділдік, жауапкершілік, құрмет сияқты адамгершілік қасиеттерін бағалаудың экожүйелік тәсілдемесі көрсетіледі. Құндылыққа бағдарланған «Адал азамат» моделін жүзеге асыру мектеп экожүйесін дамытуға мүмкіндік береді. Экожүйелік тәсілдеме білім беру ұйымдарының тәрбие жұмысының экожүйесін әлеуметтік ортадағы өзара әрекеттестікті ескере отырып бағалауды қамтамасыз етеді. Бұл тәсілдеме тек іс-шаралардың сандық нәтижелерін ғана емес, сонымен бірге білім алушының құндылық бағдарларын, әлеуметтік дағдыларын, отбасымен және қоғаммен байланысын қамтиды.

Зерттеу нәтижесінде экожүйелік тәсілге негізделген тәрбие процесін бағалау критерийлері мен индикаторлары айқындалды. Сондай-ақ тәрбие экожүйесіне сәйкес құндылыққа бағытталған «Адал азамат» моделі тәрбие процесінің стратегиялық бағыты ретінде ұсынылды.

Түйін сөздер: тәрбиенің экожүйелік тәсілдемесі, «Адал азамат» біртұтас тәрбие бағдарламасы, құндылық, патриотизм, адалдық

Кіріспе

Білім беру жүйесінің басты құндылығы – адам капиталын дамыту. Қазақстанның қазіргі даму кезеңінде қоғамның барлық саласы әлеуметтік әділеттілік пен әрбір адамның әл-ауқатын қамтамасыз ету мақсатында жаңғыртылуда.

Заманауи білім беру тәрбиені біртұтас педагогикалық процестің ажырамас бөлігі ретінде қарастырады.

Тәрбие мен оқыту – бірін-бірі үйлесімді толықтыратын тұтас процесс. Оқыту барысында білім алушылардың интеллектуалдық дамуымен қатар, олардың құндылықтарға негізделген адами қасиеттері де қалыптасуы тиіс. Демек, білім мазмұны мен тәрбие жүйесінің үйлесімділігі қазіргі педагогика ғылымының өзекті мәселелерінің бірі екендігі дау туғызбайды.

Алайда әлемдік жаһандану процесі, ақпараттық технологиялардың қарқынды дамуы және әлеуметтік ортаның өзгеруі өскелең ұрпақтың құндылық бағдарларына әсер етуде. Осы секілді сын тегеуріндер ұрпақ тәрбиесінде келесідей қарама-қайшылықтарды тудырып отыр:

– жаһандық мәдениеттегі үрдістер ықпалына төтеп беру үшін ұлттық бірегейлікті сақтау қажеттілігі;

– ақпараттық ашықтықтың әсері құндылықтар тұтастығына ықпалы нәтижесінде дәстүрлі тәрбие әдістерінің әлсіреуі мен жаңа әлеуметтік құбылыстардың пайда болуы.

Осы орайда қайшылықтарды шешу білім беру жүйесінде тәрбиені құндылықтар мен құзыреттілік тәсілдері арқылы жүзеге асыру тиімді болады.

Құндылыққа негізделген білім беру арқылы жас ұрпақтың ұлттық бірегейлігін нығайтып, жаһандық сын-қатерлерге төтеп беретін, саналы, отаншыл азаматтарды тәрбиелеу басты мақсат болуы тиіс.

Мемлекет басшысы Қ. Тоқаев Ұлттық құрылтайда балалар мен жасөспірімдер тәрбиесіне қатысты өзекті мәселелерді ашық көтеріп, қоғамның назарын аударады [1;2]. Тұлға дамуына ықпал етуші заманауи технологиялық, экономикалық, әлеуметтік және білім саласының ішкі сын-қатерлеріне жауап беру мақсатында әзірленген «Адал азамат» біртұтас тәрбие бағдарламасы – тәлім-тәрбиені жаһандық өзгерістермен үйлесімді ұштастыруға бағытталған маңызды қадам.

2023 жылдан бері барлық орта білім беру ұйымдарында жүзеге асырылып жатқан «Біртұтас тәрбие» бағдарламасында [3;4] құндылықтар жүйесі 2026 жылдан бастап жүзеге асырылатын орта білім берудің жалпыға міндетті стандартында да енгізілген [5;6]. Осылайша тәрбие жұмысында және оқу бағдарламаларында ортақ құндылықтар жүйесі өзара біріздендірілген. Бұл оқыту мен тәрбиенің тұтастығын қамтамасыз етеді.

Ал, 2026 жылғы 1 қыркүйектен бастап қолданысқа енетін Бастауыш және Негізгі орта білім берудің жалпыға міндетті стандарттарының 2-тарауында білім беру мазмұны келесі базалық құндылықтарға бағытталған:

1) тәуелсіздік және патриотизм;
2) бірлік пен ынтымақ;
3) әділдік пен жауапкершілік;
4) заң және тәртіп;
5) еңбекқорлық және кәсіби біліктілік;
6) жасампаздық және жаңашылдық [5;6].

Оқыту мен тәрбие процесінің тұтастығын нығайтуға, құндылықтарды білім беру мен тәрбие жұмысының біртұтас жүйесінде қарастыруға мүмкіндік береді. Осылайша, құндылықтар тек білім мазмұны және тәрбие бағдарламасы арқылы оқушылардың тұлғалық дамуына жан-жақты ықпал етіп, құндылық бағдарлар бекітілуге қол жеткізуді мақсат етеді.

Мемлекет басшысы Бурабайда өткен Ұлттық құрылтайдың төртінші отырысында: «Былтыр мектептерде «Біртұтас тәрбие» бағдарламасы қолға алынды. Бағдарламаны қабылдау – істің басы ғана. Енді жаңарған құжатты еліміздегі барлық мемлекеттік білім ошағына енгізу керек. Оны алдағы уақытта жекеменшік мектептер де басшылыққа алуға тиіс. Себебі тәрбие жұмысы балаларды түгел қамтуы қажет. Сонда ғана бұл бастама жалпыұлттық ауқымға ие болады» деген болатын. Мемлекет басшысы біліммен қатар тәрбиеге айрықша назар аудару өте маңызды деп атап өтіп, құрылтай мүшелері алдағы уақытта осы бағдарламаны «Адал азамат» деп атауды ұсынғанын және бұл ұсыныспен толық келісетінін атады [7].

Зерттеу мәселесі:

Қазіргі білім беру ұйымдарында тәрбие процесін бағалау көбінесе формалды көрсеткіштерге сүйенеді (іс-шаралар саны, қатысушылар үлесі және т.б.). Бұл тәсіл баланың тұлғалық дамуының, әлеуметтік орта ықпалының, отбасы мен мектеп арасындағы өзара әрекеттестіктің толық бейнесін ашып көрсетпейді.

Сондықтан тәрбие процесін бағалауда экожүйелік тәсілдемені қолдану қажеттілігі туындайды. Экожүйелік тәсілдеме баланы дамытудың макро, мезо, микро ортадағы

байланыстарын ескере отырып, тәрбие процесінің тиімділігін анықтауға мүмкіндік береді. Мұндай тәсілдеменің негізінде құндылықтарға сүйенген мектеп экожүйесін қалыптастыру мәдениеті жатыр. Бұл мәдениет мектептің, отбасының және қоғамның біртұтас әрекетінде оқушы бойында құндылықтарды орнықтыруға бағытталады.

Мақсаты: Білім беру ұйымдарындағы тәрбие процесін бағалаудың экожүйелік тәсілдемесінің теориялық негіздерін айқындау және тәрбие экожүйесіне сәйкес «Адал азамат» қалыптастыру моделін ұсыну.

Міндеттері:

- Тәрбие процесін бағалау теорияларына шолу жасап, олардың шектеулері мен мүмкіндіктерін анықтау;
- Экожүйелік тәсілдеменің негізгі ұстанымдарын білім беру ұйымдарының тәрбие процесіне бейімдеу;
- Тәрбие процесін бағалауда макро, мезо, микро орталардың өзара байланысын сипаттау;
- Экожүйелік тәсілге сүйенген тәрбие процесін бағалау критерийлері мен индикаторларын ұсыну;
- Тәрбие экожүйесіне сәйкес «Адал азамат» қалыптастыру моделін ұсыну.

Қазіргі жағдайда ұйымдардың кооперацияға, ынтымақтастыққа және желілік өзара әрекеттесуге негізделген жаңа типтегі өзара тиімді қатынастар құру қабілеті барған сайын маңызды бола түсуде. Осыған орай бүгінгі күнде жаһандану, цифрландыру және мәдени әртүрлілік жағдайында тәрбие жұмысының тиімділігін бағалау әдістерін қайта қарау қажеттілігі туындап отыр. Олай болса, тәрбие процесін бағалаудың экожүйелік тәсілдемесін қалыптастыру білім беру жүйесінің стратегиялық басымдықтарының бірі болмақ.

Бастапқыда бизнесте пайда болған экожүйелер басқа салаларға, соның ішінде білім беру саласына белсенді түрде таралуда. Білім беру экожүйесіне деген қызығушылықтың артуы білім беру ұйымдарында біртіндеп инновациялық процестер мен желілік серіктестіктерді «өзегіне» айналуымен анықталады. Бұл білім беру жүйесіндегі қатынастардың, байланыстардың және функциялардың өзгеруінен көрінеді.

Экожүйе – тірі организмдер қауымдастығынан (биоценоз), олардың тіршілік ету ортасынан (биотоп) және олардың арасындағы зат алмасу мен белсенділікті қамтамасыз ететін байланыс жүйесінен тұратын биологиялық жүйе. Бұл терминді 1935 жылы А.Тенсли ұсынған және белгілі бір аймақтағы барлық денелердің бірлігін және олардың физикалық анықтамамен өзара әрекеттесуін болжайды.

Экожүйе табиғи (мысалы, орман, көл) немесе жасанды (ауылшаруашылық жерлері, саябақтар) болуы мүмкін. Ол түрлердің құрамын сипаттайды, организмдерді, биомассаны және трофикалық құрылымды біріктіреді, оған өндірушілер (өндірушілер), тұтынушылар (тұтынушылар) және органикалық заттарды бұзушылар (редукциялар) кіреді [8].

Экожүйедегі негізгі процестер – бұл заттардың айналымы және оның тұрақтылығы мен жұмысына ықпал ететін энергия ағыны. Яғни экожүйенің басты ерекшілігі – оның тұрақтылыққа, өзін-өзі реттеуге және дамуға қабілеттілігі болып табылады.

XX ғасырдың екінші жартысында экожүйе ұғымы тек табиғи құбылыстарды ғана емес, сонымен қатар әлеуметтік жүйелерді сипаттау үшін де қолданыла бастады. Әсіресе, Ури Бронфенбреннердің (1977) экологиялық жүйе теориясы бұл ұғымды адам дамуы мен тәрбие процесін зерттеуде кеңінен қолдануға жол ашты. Ол тұлғаны дамытушы ортаны төрт деңгейде сипаттады: микро, мезо, экзо және макрожүйе. Бұл тәсіл бойынша, бала тек мектеп пен отбасынан ғана емес, сонымен қатар бұқаралық ақпарат құралдары, мәдени орта, экономикалық жағдайлар секілді сыртқы факторлардан да әсер алады [9].

Педагогикадағы тәрбие процесі дәстүрлі түрде жеке тұлғаны, оның құндылық көзқарасын, азаматтық ұстанымын және әлеуметтік жауапкершілігін қалыптастыруға бағытталған

мұғалімдердің, білім алушылардың, отбасы мен қоғамның мақсатты өзара әрекеттесу жүйесі ретінде анықталады.

Тәрбие процесін бағалау дегеніміз оқушылардың құндылықтарының, құзыреттіліктерінің, мінез-құлық нормаларының қалыптасу деңгейін, сондай-ақ әлеуметтік институттардың тәрбиеге қатысу дәрежесін анықтау.

Нәтижелі тәсілдеме – түпкілікті нәтиже бойынша бағалау (тәрбие, тәртіп, адамгершілік қасиеттер).

Белсенділік тәсілдеме – баланың әртүрлі қызмет түрлеріне (еңбек, шығармашылық, қоғамдық жұмыс) қатысуына баса назар аудару.

Құзыреттілік тәсілдеме – дағдылар мен жағымды әдеттерінің қалыптасуын өлшеу (өзін-өзі басқару, сыни ойлау, коммуникация).

Тәрбие процесін бағалау критерийлері мен индикаторлары:

Экожүйелік тәсілдеме тәрбиенің нәтижелерін тек мектеп ішінде ғана емес, сонымен бірге отбасымен, қоғаммен және мәдени ортамен өзара әрекеттесу арқылы бағалау керек деп болжайды.

Негізгі критерийлері:

Білім алушылардың құндылық бағдарлары – мораль, этика, азаматтық ұстаным нормаларын қабылдау.

Әлеуметтік құзыреттілік – командада жұмыс істей білу, қарым-қатынас мәдениеті, көшбасшылық.

Эмоционалды-психологиялық әл-ауқат – мотивация деңгейі, өзін-өзі бағалау, психологиялық жайлылық.

Әлеуметтік әріптестік – отбасының, қоғамдық ұйымдардың, жергілікті қоғамдастықтың тәрбие процесіне қатысу дәрежесі.

Мектеп климаты мен мәдениеті – сенім, ынтымақтастық және өзара құрмет атмосферасы.

Экожүйелік тәсілдеме осы позицияларды кеңейтеді, өйткені ол мыналарды ескереді:

Микрожүйе – баланың қалыптасуына тікелей ықпал ететін орта. Оған ата-ана, отбасы мүшелері, құрдастары, достары, ұстазы т.б. жақын араласатын адамдарды қамтитын орта.

Мезожүйе – бала өміріндегі микрожүйелердің арасындағы қатынасының жүйесі. Мезожүйе баланың дамуына тікелей емес, жанама түрде әсер етеді. Мезожүйеге мысалы, отбасы мен мектеп арасындағы өзара әрекет (мысалы, ата-ананың мектеппен байланысы, ата-аналар жиналысына қатысуы, мұғалімдермен қарым-қатынасы), отбасы мен балабақша арасындағы байланыс (ата-аналар мен тәрбиешілердің бірлесе жұмыс істеуі), отбасы мен аула/көрші орта арасындағы қатынас (мысалы, ата-аналардың баланың достарымен немесе көрші балаларымен байланысы); мектеп пен қосымша үйірмелер (мысалы, драма үйірмесі, спорт секциясы) арасындағы өзара ықпал. Мезожүйе баланың өміріндегі маңызды адамдар арасындағы байланыстар арқылы оның дамуына әсер етеді. Мезожүйенің әсері баланың жеке ерекшеліктеріне байланысты өзгеріп отырады: бір бала үшін сыныптағы достар маңызды болса, екінші бала үшін драма үйірмесі немесе отбасылық жағдай маңыздырақ болуы мүмкін. Бұл қарым-қатынастардың сапасы – баланың дамуы үшін шешуші фактор. Жақсы үйлесімділік пен ынтымақтастық болса – баланың дамуы да үйлесімді болады.

Экзожүйе – бұл баланың тікелей қатыспайтын, бірақ оған әсер ететін жанама әлеуметтік орта. Баланың дамуына ықпал ететін, бірақ оған тікелей араласпайтын әлеуметтік жүйелер мен құрылымдар. Экзожүйеге: ата-ананың жұмыс орны (ата-ананың жұмыстағы режимі (икемді график, демалыс), ата-анаға көрсетілетін қолдау (әлеуметтік қолдау); әлеуметтік қызметтер мен денсаулық сақтау ұйымдары; ата-ананың әлеуметтік ортасы (достары, туыстары, көршілері); мәдени және қоғамдық ұйымдар; интернет, әлеуметтік желі, БАҚ, фильмдер, мультфильмдер, т.б. медиакеңістік. Мысалы, ата-ананың тұрақты жұмысы мен қолайлы еңбек жағдайы –

отбасының әл-ауқатын жақсартып, баланың қауіпсіз әрі дамытушы ортада өсуіне септігін тигізеді.

Экзожүйе – баланың айналасындағы кең әлеуметтік құрылымдар мен байланыстар жүйесін қамтиды. Бұл жүйелердің ата-анаға көрсететін қолдауы немесе жетіспеушілігі – бала дамуының жанама, бірақ маңызды факторы. Экзожүйе неғұрлым байланысты және қолдаушы болса, бала соғұрлым тұрақты және қауіпсіз ортада дамиды.

Макрожүйе бала өсіп-жетілетін – баланың өміріне әсер ететін ең кең, жалпықоғамдық деңгейдегі мәдени, әлеуметтік, идеологиялық және құқықтық орта. Бұл – қоғамдағы құндылықтар, дәстүрлер, заңдар, мемлекеттік саясат, дін және идеологиялар жиынтығы. Тәрбиелік экожүйені қалыптастыратын стратегиялық деңгей, баланың өмірлік бағытын, құндылық әлемін, әлеуметтену жолын айқындайды.

Ури Бронфенбреннер теориясына сәйкес мысалы, мейірімді отбасында, бірақ азаматтық соғысқа шырмалған елде өмір сүріп жатқан бала. Баланың дамуына микроортасы жағымды үлес қосқанымен, ол өте қауіпті макроорта өмір сүріп жатыр. Сол себепті балада үрей, тұрақсыздық, болашағына сенімсіздік секілді ойлар оның толыққанды өсіп-жетілуіне кедергі келтіреді. Бұл мысалдан баланың толыққанды өсіп-жетілуі үшін барлық ортаның өзара бір бірімен байланысты екенін көруге болады. Сондықтан экожүйелік тәсіл – тәрбиенің кешенділігін мойындап, барлық деңгейдегі орталарды бір бағытта үйлестіруді көздейді. Бүгінгі таңда экожүйелік тәсіл – білім беру ұйымы мен отбасы арасындағы байланысты күшейтуде, әлеуметтік орта мен цифрлық ортада баланың дамуына қолайлы жағдай жасауды қамтиды.

Соңғы жылдары бұл тәсіл білім беру жүйесіне де белсенді енгізіліп отыр. Білім беру экожүйесі деп түрлі деңгейдегі субъектілер (мектеп, отбасы, қосымша білім беру ұйымдары, мұғалім, оқушы, жергілікті қауым, медиа және цифрлық орта) мен олардың өзара байланысынан құралған, оқушының тұлғалық дамуын қамтамасыз ететін ашық жүйені айтады.

Білім берудегі экожүйе – бұл ортақ мақсаттар мен ресурстар төңірегінде әртүрлі мүдделі тараптарды біріктіретін өзін-өзі ұйымдастыратын қауымдастық. Ол мектептер, университеттер, ҮЕҰ және жұмыс берушілер ұсынатын ресми және бейресми оқыту мүмкіндіктерін қамтитын, көбінесе бұл тараптар цифрлық платформалар немесе сертификаттау жүйелері арқылы байланысады. Экожүйелік тәсілдеме отбасы, мектептер, қауымдастықтар және әлеуметтік медиа сияқты әртүрлі білім беру компоненттері арасында тиімді ынтымақтастықты нығайтады (Arthur, J., & Kristjánsson, K, 2022) [10].

Білім берудегі экожүйе термині алғаш рет мектептер мен университеттерді ғана емес сонымен қатар басқа да агенттерді (қауымдастықтар, ата-аналар, цифрлық технологиялар) қамтитын білім беру процесінің дәстүрлі шеңберін кеңейту аясында да ұсынылды. Зерттеулер білім беру экожүйесі үздіксіз оқыту мен тәрбие ортасын құруға ықпал ететін өзара әрекеттесетін элементтердің жиынтығы екендігін көрсетті [11]. Зерттеушілер білім берудегі экожүйе шеңберінде формальді және бейресми білім беруді интеграциялаудың маңыздылығына назар аударады, бұл білім алушылардың жан-жақты дамуына ықпал етеді [12].

Экожүйе бірнеше негізгі принциптермен сипатталады: тұтынушыға бағдарлану, қатысушылардың ынтымақтастығы, реттеу және басқару, қоршаған ортаға әсер ету, цифрлық технологиялар және онлайн платформалар, қатысушылардың бір-біріне пайдасын тигізуі, адам мен табиғаттың бірлескен эволюциясы. Бұл принциптер жаһандық әлеуметтік-экономикалық технологиялық сын-тегеуріндерді шешу үшін ұлттық білім беру жүйелерін өзгерту мүмкіндігіне ие білім берудің экожүйелерін тұжырымдаудың негізін құрайды.

Университеттер бұл экожүйелерде цифрлық трансформацияға бейімделу, білімді басқару және технологиялар трансферті орталықтары ретінде әрекет ету арқылы шешуші рөл атқарады (Mukesh & Pillai K., 2020) [13]. Экожүйелік тәсілдеме инновациялар мен бәсекеге қабілеттілікке

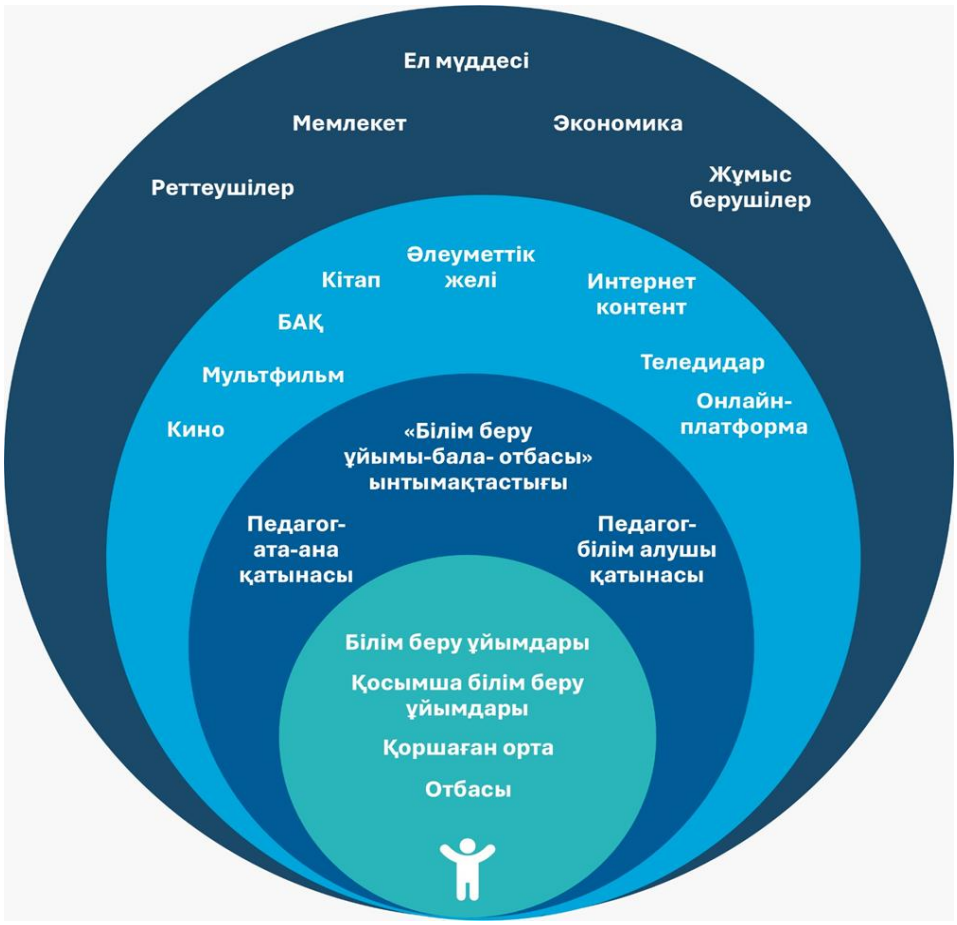
ықпал ететін дәстүрлі иерархиялық модельдерге қарағанда тиімдірек деп саналады (Kovaliuk & Kobets, 2021) [14].

1-сурет - Ури Бронфенбреннер бойынша тәрбие экожүйесі

Құндылыққа бағдарланған білім беруді тәрбие экожүйесі шеңберінде интеграциялай отырып, білім беру ұйымдары тәрбие бағдарламасы аясында «адал азаматты» тәрбиелеуді дамытуды күшейте алады. Бұл интеграция әртүрлі мүдделі тараптардың бірлескен күш-жігеріне сүйенеді және тиімді білім беру ортасын құру үшін цифрлық технологияларды қолданады (Abdul Rokhim et al., 2022) [15]. Ол сондай-ақ білім беру ұйымдарына үздіксіз дамуды ілгерілету және білім алушыларды қазіргі қоғамның талаптарына сай тәрбиелеу арқылы жаһандық әлеуметтік-экономикалық және технологиялық сын-тегеуріндерге төтеп беруге мүмкіндік береді (Kicherova & Trifonova, 2023) [16].

Құндылыққа бағдарланған тәрбие білім беру экожүйелерінің негізгі құрамдас бөліктерінің бірі. Бірқатар зерттеулер білім беру ортасында азаматтық және адамгершілік құндылықтарды қалыптастырудың маңыздылығын атап көрсетеді, бұл білім алушылардың жауапкершілік пен патриотизм сезімін қалыптастыруға ықпал етеді (Lickona, 1996,) [17]. Осы тәсілдеме тұлғаның үйлесімді дамуына әсер етіп, оның қоғамда табысты әлеуметтенуінің алғышарттарын жасайды.

Білім беру экожүйелерін зерттеу табысты тәрбие мен білім беру әртүрлі субъектілердің: мектептердің, отбасылардың қоғамдық ұйымдар мен мемлекеттік органдардың арасындағы өзара белсенді әрекеттесуді қажет ететіндігін көрсетеді (Senge et al., 2000) [18]. Бұл түрлі ресурстарға қол жеткізуді қамтамасыз ететін және білім беру процесінің барлық қатысушылары арасындағы байланысты қолдауға мүмкіндік беретін цифрлық технологиялар маңызды рөл атқарады (Groff, 2013) [19].



2 – сурет. Тұлғаны қалыптастырушы экожүйе

Білім беру экожүйелерін құру озық ойлы ұлтты қалыптастыру арқылы қоғамның дамуына тікелей әсер етеді. Зерттеулер көрсеткендей, тәрбиеге экожүйелік көзқарас әлеуметтік шиеленіс қауіпін азайтады және жастардың табысты әлеуметтенуіне ықпал етеді. Тәрбиенің экожүйесі ұғымы қазақтың ойшылдарында, ұлт зиялыларында тікелей қарастырылмаса да бала тәрбиесінде, оның тұлғалық қалыптасуына отбасы, ұстазы, ортасы ықпал ететіндігін дәлелдейтін ойлар жеткілікті. Сол себепті де қазіргі тәрбие процесін тиімді іске асыру үшін экожүйелік тәсіл тиімді болмақ.

Сондықтан экожүйе ұғымын педагогикалық кеңістікте білім алушының дамуына ықпал ететін көпқырлы, өзара байланысты, динамикалық құрылым деп қарастыруға негіз бар. Себебі ол бір мезгілде тұлғаға әсер ететін тікелей (отбасы, мектеп) және жанама (цифрлық орта, БАҚ) ықпал ететін институттарды қамтиды.

Демек тәрбие экожүйесі заманауи экономиканың қиындықтарына төтеп бере алатын және қоғамның өркендеуіне үлес қоса алатын азаматтарды қалыптастыруға көмектеседі. Себебі қазіргі жаһандану жағдайындағы білім беру жүйесінің негізгі бағыттарының бірі – тәрбие процесін кешенді, жүйелі түрде ұйымдастыру және оны білім беру кеңістігімен бірлікте қарастыру маңызды. Осы ретте тәрбие экожүйесін құру – уақыт талабы екендігі мәлім. Экожүйелік көзқарас тәрбиенің тек мектеп ішінде ғана емес, ол білім алушыны қоршаған әлеуметтік, мәдени, цифрлық және рухани ортада да қалыптасатынын айғақтайды.

Экожүйе тәрбиені оқшау емес, тұтас әлеуметтік феномен ретінде қарастыруға мүмкіндік береді. Себебі тәрбиенің экожүйелік моделінде білім алушы – басты тұлға ретінде танылады.

Оның тұлғалық қалыптасуы мен даму процесі түрлі орталар мен әлеуметтік институттардың өзара байланысы мен ықпалы негізінде жүзеге асады. Бұл байланыстар білім алушының танымдық, эмоционалдық және практикалық аспектілерінің үйлесімді дамуын қамтамасыз етеді. осы ретте айта кетуге болады, біз ұсынған модельдегі тәрбиелік ықпал білім беру ұйымы деңгейімен байланыста қарастырылды.

Тәрбие экожүйесіне сәйкес модель келесі микрожүйелерді қамтиды:

1. Отбасы – тәрбиенің алғашқы және іргелі институты. Бала тұлғасының құндылық негіздері дәл осы ортада қаланады. Модельде отбасы мен мектептің өзара әрекеттестігі тәрбиенің тиімділігін арттырудың негізгі тетігі ретінде қарастырылады.

2. Білім беру ұйымындағы (балабақша, мектеп, колледж) орта – тәрбиенің басты ұйымдастырушылық кеңістігі. Бұл жерде білім алушының азаматтық, отаншылдық, адамгершілік, құқықтық, еңбек және экологиялық мәдениетін қалыптастыру бағытында жүйелі жұмыс жүргізіледі. Білім беру ұйымының тәрбие жұмысын ұйымдастыруда пәннің әлеуеті, педагог тұлғасын, қосымша білім беруді, өзін-өзі басқаруды, мектеп ережелері мен дәстүрлерін тиімді пайдаланудың маңызы зор.

3. Әлеуметтік орта – білім алушы өмір сүретін және қарым-қатынас жасайтын кеңістік. Мәдениет мекемелері, спорт клубтары, жастар ұйымдары, жеке ауласы, қоғамдық бірлестіктер тәрбие экожүйесінің табиғи бөлігі болып табылады.

4. Қосымша білім беру ортасы – үйірмелер, секциялар, өнер мектептері білім алушының шығармашылық әлеуетін дамытып, әлеуметтенуіне мүмкіндік береді. Бұл ортада тұлғаның өзін-өзі танытуына жағдай жасалады.

5. Бұқаралық ақпарат құралдары мен цифрлық орта – қазіргі таңда оқушының сана-сезіміне ең үлкен әсер ететін факторлардың бірі. Экожүйелік модельде цифрлық мәдениет пен медиа сауаттылық тәрбиенің құрамдас бөлігі ретінде қарастырылады.

6. Педагогтер қауымы – мұғалім, тәрбиеші, мектеп әкімшілігі, педагог-психолог және әлеуметтік педагогтер экожүйедегі негізгі бағыттаушы күш. Олардың кәсіби құзыреттілігі мен тұлғалық ұстанымы оқушыға үлгі болуымен қатар, оның дамуын педагогикалық тұрғыда қолдайды.

Ал, ұсынылған модель тәрбие экожүйесінің іргетасы – құндылықтар жүйесі. Модельге сәйкес құндылықтарды білім алушылардың бойына дарыту, құзыреттілік ретінде қалыптастырып, «Адал азамат» бойындағы қасиеттердің сипаттамасын анықтауға болады. Себебі құндылықтар тек жеке адам деңгейінде ғана емес, мектептің бүкіл өмір сүру кеңістігінде, оның мәдениетінде, нормаларында, өзара қарым-қатынасында көрініс табуы тиіс.

Экожүйеге сәйкес тәрбиенің басты мақсаты – білім алушының үйлесімді дамуын, әлеуметтік бейімделуін, тұлғалық және азаматтық қалыптасуын қамтамасыз ету. Бұл мақсатқа жету үшін модель келесі қызметтерді жүзеге асырады:

тәрбиелік әсер ету орталарын біртұтас жүйеге біріктіру;

мектеп пен қоғам арасындағы ынтымақтастықты нығайту;

тұлғалық даму мен әлеуметтенуге жағдай жасау;

тәрбиенің үздіксіздігі мен сабақтастығын қамтамасыз ету;

құндылыққа бағдарланған тәрбиені жүйелендіру.

Демек зерттеулерге сүйене отырып, «Адал азамат» қалыптастыратын тәрбиенің білім беру ұйымы деңгейіндегі экожүйені ұсынуға болады.

Ұсынылған модель тәрбие экожүйесіне сәйкес тәрбиелік ықпалдардың бірізділігі мен сабақтастығын қамтамасыз ететін, білім алушы тұлғасының жан-жақты дамуына бағытталған көпдеңгейлі, құзыреттілікке және құндылыққа бағдарланған кешенді жүйе.

3-сурет - Тәрбие экожүйесіне сәйкес білім беру ұйымы деңгейінде «Адал азамат» қалыптастыру модел

Модель тәрбие экожүйесіне сәйкес құндылыққа бағдарланған, құзыреттілік және іс-әрекеттік тәсілдемелер арқылы іске асырылады.

Тәрбие экожүйесіне сәйкес білім беру ұйымы деңгейінде «Адал азамат» қалыптастыру моделі білім беру мазмұны мен сынып сағаттары және сыныптан тыс іс-шараларды жүйелі түрде ұштастыру арқылы тұлға бойында азаматтық жауапкершілік пен қоғамдық белсенділікті қалыптастырудың тиімді тетігін ұсынады. Тәрбие рөлдік модельдерге (әрекет арқылы үлгі көрсететін тұлғалар), сындарлы кері байланысқа және әлеуметтік серіктестікке негізделген жағдайда ғана, білім алушы өзін қоғам алдындағы жауапты субъект ретінде сезіну мүмкіндігіне ие болады және қауіпсіз, қолайлы ортада дамиды.

Модельге сәйкес таңдалған әдіс-тәсілдер (мысалы, көркем шығармаларды талдау, әңгіме жүргізу, дебат, пікірталас) білім алушының танымдық қызығушылығын ынталандырып, сыни ойлауын жетілдіруге жағдай жасайды. Әлеуметтік жобаларды әзірлеу мен волонтерлік қызметке қатысу білім алушының әлеуметтік жауапкершілік сезімін күшейтіп, ұжымдық іс-әрекеттегі белсенділігін арттырады. Рефлексия жүргізу мен рөлдік ойындар білім алушыларға өз әрекеттері мен тәжірибелерін талдау арқылы мінез-құлықтық стратегияларды тиімді реттеуге мүмкіндік береді. Театрландырылған қойылымдар мен жеке күнделік жүргізу – білім алушылардың өзін-өзі тану, эмпатия және шығармашылық дағдыларын дамытудың бірегей құралы.

Осылайша, оқу процесінің мазмұны мен оған байланысты әдіс-тәсілдердің біртұтас қолданылуы білім алушылардың ұлттық құндылықтарға құрметпен қарап, мәдени мұраны насихаттау, тиімді коммуникация орнату, қоғамдық бастамаларға белсенді араласу және этикалық нормаларды сақтау сияқты құзыреттерін кешенді түрде қалыптастыруға ықпал етеді. Нәтижесінде, өзгерістерге бейім, инновациялық шешімдер ұсынуға қабілетті және мемлекет мүддесін ілгерілетуге дайын «Адал азамат» тұлғасы қалыптасады.

Қорытынды

Қорытындылай келе, тәрбие экожүйесіне сәйкес білім беру ұйымы деңгейінде «Адал азамат» қалыптастыру моделі білім беру ұйымының ресурстары мен сыртқы әлеуметтік ортаның мүмкіндіктерін біріктіре отырып, тәрбиелік процесті жүйелі әрі тиімді ұйымдастырудың заманауи формасы ретінде қарастырылады. Модельге сәйкес тәрбие процесін жаңаша ұйымдастырып, білім беру ұйымы, отбасы, қоғаммен тығыз байланысын дамытуды көздейді. Тәрбие экожүйесіне сәйкес білім беру ұйымы деңгейінде «Адал азамат» құндылықтарын тәрбие процесін жоспарлау, талдау, мониторинг жүйесіне және білім беру ұйымы мәдениетінің артефактілеріне кіріктіруді қарастырады.

Ұсынылған модель білім беру ұйымында тәрбие процесінің кеңістіктік және мазмұндық шегарасын кеңейтіп, білім алушыны субъект ретінде қарастыратын алдын ала жоспарланған, мақсатты, заңдылықтары мен іске асыру механизмдері бар тұтас процесс ретінде жүзеге асыруға мүмкіндік береді.

Қазіргі білім беру парадигмасының тәрбиені оқытумен тең дәрежеде қарастыруы және жеке тұлғаның жан-жақты дамуына бағытталған біртұтас тәрбие жүйесін қалыптастыруға ұмтылуы зерттеудің өзектілігін айқындады.

Формалды көрсеткіштерге сүйену арқылы тәрбиені бағалау баланың тұлғалық және құндылықтық дамуын толық қамти алмайтынын көрсетті. Осыған байланысты экожүйелік тәсілдеме тәрбие процесін бағалауда кешенді көзқарас ұсынады.

Экожүйелік тәсілдеме білім беру ұйымдарының тәрбие жұмысының экожүйесі әлеуметтік ортадағы өзара әрекеттестікті ескере отырып бағалауға мүмкіндік береді. Бұл тәсілдеме тек іс-

шаралардың сандық нәтижелерін ғана емес, сонымен бірге білім алушының құндылық бағдарларын, әлеуметтік дағдыларын, отбасымен және қоғаммен байланысын қамтиды.

Зерттеу нәтижесінде экожүйелік тәсілге негізделген тәрбие процесін бағалау критерийлері мен индикаторлары айқындалды. Сонымен қатар, тәрбие экожүйесіне сәйкес құндылыққа бағытталған «Адал азамат» моделін қалыптастыру тәрбие процесінің стратегиялық бағыты ретінде ұсынылады.

Зерттеу барысында мектеп мәдениетінің құрылымдық, мазмұндық және қатынастық аспектілері талданды. Нәтижесінде мектептегі тәрбие жұмыстарының жүйелілігі мен құндылықтарға сәйкестігін бағалаудың өлшемдері айқындалды. «Адал азамат» моделінің педагогикалық шарттары мен тетіктері нақтыланып, оның білім алушылардың бойында жоғары адами құндылықтарды қалыптастырудағы тиімділігі дәлелденді.

Зерттеу нәтижелері мектеп мәдениеті мен психологиялық ахуалдың тәрбиелік ықпалдың орталық буыны және білім алушының тұлғалық дамуына тікелей әсер ететін орта екендігін нақты көрсетті.

Қорытындылай келе, мектеп мәдениетін тәрбие экожүйесіне сай ұйымдастыру және «Адал азамат» моделін енгізу, сондай-ақ «Адал азамат» біртұтас тәрбие бағдарламасы мен жаңа стандарттағы құндылықтарды білім беру процесіне енгізу болашақ азаматтың адамгершілік келбетін қалыптастыруға бағытталған кешенді әрі ғылыми негізделген процесс. Бұл тәсілдеме білім беру ұйымдарының тәрбие жұмысының сапасын арттырып, білім алушылардың өмірлік ұстанымдарына ықпал ететін негізгі құралдардың біріне айналады.

Әдебиеттер

1. Мемлекет басшысы Қасым-Жомарт Тоқаевтың Ұлттық құрылтайдың «Адал адам – Адал еңбек – Адал табыс» атты үшінші отырысында сөйлеген сөзі 2024 жылғы 15 наурыз [Электронды ресурс] https://www.akorda.kz/kz/memleket-basshysy-kasym-zhomart-tokaevtyn-ulttyk-kuryltaydyn-adal-adam-adal-enbek-adal-tabys-atty-ushinshi-otyrysynda-soylegen-sozi-1525727 (қаралған уақыты 05.09.2024 ж.).

2. Мемлекет басшысы Қасым-Жомарт Тоқаевтың Ұлттық құрылтайдың «Әділетті Қазақстан – Адал азамат» атты екінші отырысында сөйлеген сөзі. 2023 жылғы 17 маусым. – Түркістан. [Электронды ресурс] https://surl.li/tpkuos (қаралған уақыты 05.09.2024 ж.).

3. Қазақстан Республикасы Оқу-ағарту министрінің Жоғары оқу орындарынан басқа білім беру ұйымдарына арналған «Біртұтас тәрбие бағдарламасын» бекіту туралы 2023 жылғы 19 қыргүйектегі №294 бұйрығының 1-қосымшасы.

4. Қазақстан Республикасы Оқу-ағарту министрінің «Біртұтас тәрбие» бағдарламасын бекіту туралы 2024 жылғы 31 шілдедегі №194 бұйрығына 1-қосымша.

5. Бастауыш білім берудің мемлекеттік жалпыға міндетті стандарты. Қазақстан Республикасы Оқу-ағарту министрінің 2025 жылғы 23 қаңтардағы № 12 бұйрығына 1-қосымша. Қазақстан Республикасының Әділет министрлігінде 2025 жылғы 24 қаңтарда № 35670 болып тіркелді [Электронды ресурс] https://adilet.zan.kz/kaz/docs/V2500035670#z35 (қаралған уақыты 05.09.2025 жылы).

6. Негізгі білім берудің мемлекеттік жалпыға міндетті стандарты. Қазақстан Республикасы Оқу-ағарту министрінің 2025 жылғы 23 қаңтардағы № 12 бұйрығына 2-қосымша. Қазақстан Республикасының Әділет министрлігінде 2025 жылғы 24 қаңтарда № 35670 болып тіркелді. [Электронды ресурс] https://adilet.zan.kz/kaz/docs/V2500035670#z211 (қаралған уақыты 05.09.2025).

7. Мемлекет басшысы Қасым-Жомарт Тоқаевтың Ұлттық құрылтайдың төртінші отырысында сөйлеген сөзі 2025 жылғы 14 наурыз. [Электронды ресурс] https://www.akorda.kz/kz/memleket-basshysy-kasym-zhomart-tokaevtyn-ulttyk-kuryltaydyn-tortinshi-otyrysynda-soylegen-sozi-14242 (қаралған уақыты 15.09.2025 жылы).

8. Экосистеме [Экожүйе] Википедия ашық энциклопедиядан алынды. [Электронды ресурс] https://surl.li/zjcvfo (қаралған уақыты 15.03.2024 жылы).

9. Bronfenbrenner U. (1977) Toward an experimental ecology of human development. American Psychologist, 32 (7), 513-531.

10. Arthur, J., & Kristjánsson, K. (2022). The Jubilee Centre framework for character education in schools. University of Birmingham.

11. Van den Akker, J. (2013). _Curriculum Perspectives and Design in Educational Ecosystems_. DOI: 10.1007/978-94-6209-143-6_1.

12. Care, E., Kim, H., Vista, A., & Anderson, K. (2018, November). Education system alignment for 21st century skills: Focus on assessment. Optimizing Assessment for All (OAA). Center for Universal Education at the Brookings Institution.

13. Mukesh H.V. & Rajasekharan Pillai K., 2020. «Role of Institutional Ecosystem in Entrepreneurship Education: An Empirical Reiteration» Journal of Entrepreneurship and Innovation in Emerging Economies, Entrepreneurship Development Institute of India, vol. 29(1), pages 176-205, March.

14. Kovaliuk, Tetiana & Kobets, Nataliya. (2021). The Concept of an Educational Ecosystem for the Digital Transformation of the Ukrainian Economy.

15. Rokhim, Abdul & Suryadi, Suryadi & Supadi, Supadi. (2021). Evaluation of The Implementation of The Inclusion Program. International Journal of Elementary Education. 5, 675. 10.23887/ijee.v5i4.37217.

16. Kicherova M. N., Trifonova I. S. Educación no formal: Una revisión de los estudios modernos. Obrazovanie i nauka = Educación y Ciencia. 2023; 25 (2): 46–67. DOI: 10.17853/1994-5639-2023-2-46-67.

17. Lickona, T. (1996). _Character Education in the Educational Ecosystem: A Value-based Approach_. DOI: 10.1080/0305724920190104.

18. Senge, P., et al. (2000). _Schools as Learning Organizations: A Systems Thinking Approach_. DOI: 10.1016/j.sbspro.2011.11.256.

19. Groff, J. (2013). _Technology and the Learning Ecosystem: An Interconnected Approach_. DOI: 10.2139/ssrn.2386422.

I. Sagindikov, A.Tanirbergenova, A. Magauiya

ECOSYSTEM APPROACH TO THE ASSESSMENT OF THE EDUCATIONAL PROCESS IN EDUCATIONAL ORGANIZATIONS

The article considers an ecosystem approach to the assessment of the educational process in educational organizations. Currently, the measurement of educational outcomes is often limited to formal indicators and does not fully reveal the personal development and value orientations of students. In this regard, the ecosystem approach seems to be an effective mechanism for a comprehensive assessment of the educational process. The study analyzes the relationship between the levels of the macro, meso, and microenvironment, and substantiates the importance of considering the educational space of an educational organization from the point of view of social partnership, family-school cooperation, and school culture.

The article proposes to describe an ecosystem model of student personality formation based on the core values of the Unified Educational Program «Adal Azamat» The article demonstrates an ecosystem approach to determining the results of education through the implementation of the «Adal Azamat» model, assessing the moral qualities of students, such as honesty, fairness, responsibility, respect. The implementation of the «Adal Azamat» value-based model will help develop the school ecosystem. The ecosystem approach enables the evaluation of the ecosystem of educational institutions' upbringing work, taking into account interactions within the social environment. This approach encompasses not only the quantitative results of activities but also the value orientations of learners, their social skills, and their connections with family and society.

As a result of the research, criteria and indicators for evaluating the educational process based on the ecosystem approach were identified. In addition, a value-oriented model titled «Adal Azamat» was proposed as a strategic direction of the educational process aligned with the educational ecosystem.

Keywords: ecosystem approach of education, unified educational program «Adal Azamat», value, patriotism, honesty

И.У. Сагиндиков, А.Ш. Танирбергенова, А. Мағауия

ЭКОСИСТЕМНЫЙ ПОДХОД К ОЦЕНКЕ ВОСПИТАТЕЛЬНОГО ПРОЦЕССА В ОРГАНИЗАЦИЯХ ОБРАЗОВАНИЯ

В статье рассматривается экосистемный подход к оценке воспитательного процесса в организациях образования. В настоящее время измерение результатов воспитания часто ограничивается формальными показателями и не полностью раскрывает личностное развитие и ценностные ориентации обучающихся. В этой связи экосистемный подход представляется эффективным механизмом комплексной оценки воспитательного процесса. В ходе исследования анализируется взаимосвязь уровней макро, мезо, микросреды, обосновывается важность рассмотрения воспитательного пространства образовательной организации с точки зрения социального партнерства, сотрудничества семьи и школы, школьной культуры.

В статье предлагается описать экосистемную модель формирования личности учащегося, опираясь на основные ценности Единой воспитательной программы «Адал азамат». Демонстрируется экосистемный подход к определению результатов воспитания через внедрение модели «Адал азамат», оценке нравственных качеств обучающихся, таких как честность, справедливость, ответственность, уважение. Реализация ценностно-ориентированной модели «Адал азамат» позволит развивать школьную экосистему. Экосистемный подход обеспечивает оценку экосистемы воспитательной работы организации образования с учётом взаимодействия в социальной среде. Этот подход охватывает не только количественные результаты мероприятий, но и ценностные ориентации обучающихся, их социальные навыки, а также связь с семьёй и обществом.

В результате исследования были определены критерии и индикаторы оценки воспитательного процесса на основе экосистемного подхода. Также предложена ценностно-ориентированная модель «Адал азамат» как стратегическое направление воспитательного процесса, соответствующее воспитательной экосистеме.

Ключевые слова: экосистемный подход воспитания, единая воспитательная программа «Адал азамат», ценность, патриотизм, честность.

References

1. Tokayev, K.-Zh. (2024, March 15). Memleket basshysy Kasym-Zhomart Tokayevtin «Adal adam – Adal enbek – Adal tabys» atty uchinshi otyrysynda soylegen sozi. Akorda. [Электронды ресурс] URL: https://www.akorda.kz/kz/memleket-basshysy-kasym-zhomart-tokaevtyn-ulttyk-kuryltaydyn-adal-adam-adal-enbek-adal-tabys-atty-ushinshi-otyrysynda-soylegen-sozi-1525727 (Accessed: 05.09.2024).
2. Tokayev, K.-Zh. (2023, June 17). Memleket basshysy Kasym-Zhomart Tokayevtin Ulttық kuryltaydyn «Adiletti Kazakhstan – Adal azamat» atty ekinshi otyrysynda soylegen sozi. [Электронды ресурс] URL: https://surl.li/tpkuos (Accessed: 05.09.2024).
3. Kazakhstan Respublikasy Oqu-agartu ministirinin Zhoғary oqu oryndarynan basqa bilim beru ұyymdaryna arnalgan «Birtұtas tarbiye barғdarlamasyn» bekitu turaly 2023 zhylғy 19 қyrguyektegi №294 buyrygynyn 1-kosymshasy. (2023).
4. Kazakhstan Respublikasy Oqu-agartu ministirinin «Birtұtas tarbiye» barғdarlamasyn bekitu turaly 2024 zhylғy 31 shildedegi №194 buyrygyna 1-kosymsha. (2024).
5. Bastauysh bilim berudin memlekettik zhalpyga mindetti standarty. Kazakhstan Respublikasy Oqu-agartu ministrinin 2025 zhylғy 23 қantaradagy № 12 buyrygyna 1-kosymsha. Kazakhstan Respublikasynyn Adilet ministrliginde 2025 zhylғy 24 қantarda № 35670 bolyp tirkeldi. (2025). [Электронды ресурс] URL: https://adilet.zan.kz/kaz/docs/V2500035670#z35 (Accessed: 05.09.2025).
6. Negizgi bilim berudin memlekettik zhalpyga mindetti standarty. Kazakhstan Respublikasy Oqu-agartu ministrinin 2025 zhylғy 23 қantaradagy № 12 buyrygyna 2-kosymsha. Kazakhstan Respublikasynyn Adilet ministrliginde 2025 zhylғy 24 қantarda № 35670 bolyp tirkeldi. (2025). [Электронды ресурс] URL: https://adilet.zan.kz/kaz/docs/V2500035670#z211 (Accessed: 05.09.2025).

7. Tokayev, K.-Zh. (2025, March 14). Memleket basshysy Kasym-Zhomart Tokayevtin Ulttық kuryltaydyn tortinshi otyrysynda soylegen sozi. Akorda. [Электронды ресурс] URL: https://www.akorda.kz/kz/memleket-basshysy-kasym-zhomart-tokaevtyn-ulttyk-kuryltaydyn-tortinshi-otyrysynda-soylegen-sozi-14242 (Accessed: 15.09.2025).

8. Ekosistema [Ekosuye] Vikipediya ashyq enciklopediyadan alyndy. [Электронды ресурс] URL: https://surl.li/zjcvfo (Accessed: 15.03.2024).

9. Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. American Psychologist, 32 (7), 513-531.

10. Arthur, J., & Kristjánsson, K. (2022). The Jubilee Centre framework for character education in schools. University of Birmingham.

11. Van den Akker, J. (2013). Curriculum Perspectives and Design in Educational Ecosystems. DOI: 10.1007/978-94-6209-143-6_1.

12. Care, E., Kim, H., Vista, A., & Anderson, K. (2018, November). Education system alignment for 21st century skills: Focus on assessment. Optimizing Assessment for All (OAA). Center for Universal Education at the Brookings Institution.

13. Mukesh, H.V., & Rajasekharan Pillai, K. (2020). «Role of Institutional Ecosystem in Entrepreneurship Education: An Empirical Reiteration». Journal of Entrepreneurship and Innovation in Emerging Economies, 29(1), 176-205. DOI: 10.1177/2393957519890526.

14. Kovaliuk, T., & Kobets, N. (2021). The Concept of an Educational Ecosystem for the Digital Transformation of the Ukrainian Economy.

15. Rokhim, A., Suryadi, S., & Supadi, S. (2021). Evaluation of The Implementation of The Inclusion Program. International Journal of Elementary Education, 5, 675. DOI: 10.23887/ijee.v5i4.37217.

16. Kicherova, M. N., & Trifonova, I. S. (2023). Educación no formal: Una revisión de los estudios modernos. Obrazovanie i nauka = Educación y Ciencia, 25 (2): 46–67. DOI: 10.17853/1994-5639-2023-2-46-67.

17. Lickona, T. (1996). Character Education in the Educational Ecosystem: A Value-based Approach. DOI: 10.1080/0305724920190104.

18. Senge, P., et al. (2000). Schools as Learning Organizations: A Systems Thinking Approach. DOI: 10.1016/j.sbspro.2011.11.256.

19. Groff, J. (2013). Technology and the Learning Ecosystem: An Interconnected Approach. DOI: 10.2139/ssrn.2386422.

Авторлар туралы мәлімет:

Сагиндиков Исатай Умирзакович - педагогика ғылымдарының кандидаты, Ы. Алтынсарин атындағы Ұлттық білім академиясының жетекші ғылыми қызметкері, Астана қ., Қазақстан Республикасы, Isatay60@mail.ru

Танирбергенова Анар Шынболатовна - педагогика ғылымдарының кандидаты, Ы.Алтынсарин атындағы Ұлттық білім академиясының вице-президенті, Астана қ. Қазақстан Республикасы, anar-tanirbergenova@mail.ru

Мағауия Айжан – магистр, Ы.Алтынсарин атындағы Ұлттық білім академиясының аға ғылыми қызметкер, Астана қ. Қазақстан Республикасы, mag-aizhan@mail.ru

Сведения об авторах:

Сагиндиков Исатай Умирзакович - кандидат педагогических наук, главный научный Ведущий научный сотрудник Национальной академии образования им. И. Алтынсарина, г. Астана, Республика Казакстан, Isatay60@mail.ru

Танирбергенова Анар Шынболатовна - кандидат педагогических наук, вице-президент Национальной академии образования им. И.Алтынсарина, г. Астана, Республика Казакстан, anar-tanirbergenova@mail.ru

Магауия Айжан - магистр, старший научный сотрудник Национальной академии образования им.И. Алтынсарина,г. Астана, Республика Казахстан, mag-aizhan@mail.ru

Information about authors:

Sagindikov Issatay - Candidate of Pedagogical Sciences, Leading Researcher at the National Academy of Education named after I.Altynsarin, Astana, Republic of Kazakhstan, Astana, Republic of Kazakhstan, Isatay60@mail.ru

Tanirbergenova Anar - Candidate of Pedagogical Sciences, Vice President of the National Academy of Education named after I. Altynsarin, Astana, Republic of Kazakhstan, anar-tanirbergenova@mail.ru

Magauiya Aizhan - Master's degree, Senior Researcher at the I. Altynsarin National Academy of Education, Astana, Republic of Kazakhstan, mag-aizhan@mail.ru

МРНТИ 50.03.05 https://doi.org/10.63597/UTO3105-4161.2025.3.3.005

Ш.Б. Алтыбаева[1*], А.К. Каркенова[2], М.Д. Каргалдаева[3]
[1,2,3]Национальный центр тестирования, Астана, Казахстан
*e-mail: shugyla.altybayeva@gmail.com

[1]ORCID 0000-0003-0306-861X, [2]ORCID 0009-0003-0678-1987, [3]ORCID 0009-0006-1138-4160

ПСИХОМЕТРИЧЕСКАЯ ВЕРИФИКАЦИЯ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ УПРАВЛЕНИЯ БАЗОЙ ТЕСТОВЫХ ЗАДАНИЙ

В данной статье рассматриваются результаты комплексной модернизации автоматизированной системы управления базой тестовых заданий, которая используется в национальных экзаменационных процедурах, включая Единое национальное тестирование и комплексное тестирование в магистратуру. Основное внимание уделено влиянию внедрения технических и организационных решений на психометрические характеристики тестов и повышение качества образовательных измерений. В рамках исследования проанализированы такие меры, как внедрение механизмов кэширования, обновление вычислительной инфраструктуры, использование технологии шардирования баз данных, интеграция DLP-системы для обеспечения информационной безопасности, а также применение инструмента автоматизированной проверки качества заданий CheckerWord. Для оценки эффективности проведенных изменений выполнен психометрический анализ с использованием модели Раша и программного обеспечения Winsteps (версия 5.8.3.0). Результаты показали, что модернизация положительно сказалась на надежности тестов, уровне дифференциации участников и соответствии заданий модели. Отдельное внимание уделено вопросам защиты персональных данных и устойчивости функционирования базы. Полученные результаты подтверждают, что комплексный подход, включающий как технические, так и психометрические меры, способствует обеспечению валидности, надежности и воспроизводимости образовательных измерений в национальной системе оценки качества образования.

Ключевые слова: автоматизированная система управления тестовыми заданиями, психометрическая верификация, модель Раша, надежность теста, валидность, DLP-система, CheckerWord.

Введение

В условиях модернизации национальной системы образования особую роль занимают экзамены республиканского масштаба, каждый из которых выполняет свою уникальную функцию и обладает высокой значимостью как для образовательной системы в целом, так и для отдельных обучающихся. К числу таких процедур относятся Единое национальное тестирование (далее - ЕНТ), комплексное тестирование в магистратуру (далее - КТ). Эти экзамены обеспечивают справедливый отбор, открывают доступ к последующим уровням образования и выступают основным инструментом распределения государственных образовательных грантов [1]. Их результативность напрямую связана с качеством используемых тестовых заданий, что подчеркивает необходимость системного подхода к формированию, хранению и обновлению базы тестовых материалов.

Функционирование экзаменационных процедур невозможно без надежной базы тестовых заданий, которая является фундаментом всей системы оценивания. Современные вызовы требуют перехода от фрагментарного формирования тестовых заданий к комплексному управлению базой тестовых заданий, где каждая операция – от создания и экспертизы до отбора и компоновки вариантов тестов – подлежит автоматизации [2]. Внедрение специализированного программного обеспечения позволяет обеспечить не только эффективное администрирование

базы тестовых заданий, но и прозрачность, надежность и воспроизводимость всех этапов тестологического процесса.

Особое значение в системе национального оценивания приобретает проблема валидации тестовых заданий, под которой понимается подтверждение их качества, соответствия целям экзамена и способности объективно оценивать уровень знаний обучающихся. Автоматизированные системы управления базой тестовых заданий (далее - АСУ ТЗ) играют ключевую роль в обеспечении данных процедур, так как позволяют проводить комплексный анализ надежности, валидности и дискриминативности каждого задания. При этом такие системы должны не только гарантировать бесперебойное функционирование и масштабируемость, но и обеспечивать высокое психометрическое качество тестового материала.

Предшествующий аудит функционирования АСУ ТЗ, применяемой в рамках национальных экзаменационных процедур ЕНТ и КТ и проведенный в рамках исследовательского проекта «Улучшение качества экзаменационных материалов с помощью цифровой платформы» Национального центра тестирования, выявил ряд критических уязвимостей, требующих немедленной реакции [3]. В частности, результаты экспертной оценки, указали на области, характеризующиеся низкими средними значениями и высоким разбросом экспертных суждений, что свидетельствовало о наличии системных проблем. К числу наиболее уязвимых элементов были отнесены скорость отклика системы, обеспечение правовой и технологической защиты конфиденциальных данных, и ограниченные возможности работы с большими массивами данных.

Полученные результаты стали отправной точкой для разработки и внедрения комплекса корректирующих мер, направленных не только на устранение выявленных технических проблем, но и на повышение качества конечного продукта – тестовых заданий, генерируемых системой.

Цель исследования заключается в описании и анализе результатов внедрения комплекса мер по модернизации АСУ ТЗ, основанных на рекомендациях предыдущего этапа работы, а также в эмпирической верификации их влияния на психометрическое качество тестов, формируемых системой.

Материалы и методы исследования

В качестве основного метода использован кейс-стади, позволяющий рассмотреть не всю систему АСУ ТЗ в целом, а конкретный фрагмент ее функционирования – рекомендации по модернизации [3], выработанные по итогам предшествующего аудита. Дополнительно применялся метод системного анализа для оценки взаимосвязей между техническими, организационными и психометрическими аспектами.

В рамках исследования были проанализированы и апробированы следующие мероприятия:

– модернизация инфраструктуры (серверное кэширование, аудит программного обеспечения, обновление аппаратных ресурсов);

– внедрение DLP-системы для защиты информации и обеспечения чистоты данных для психометрического анализа;

– оптимизация работы с большими данными через шардирование базы тестов по годам и блокам;

– разработка программы автоматизированной проверки качества «CheckerWord», обеспечивающей предварительную верификацию тестовых заданий и повышение показателей надежности и валидности.

Для оценки качества и валидности тестов применялся психометрический анализ (в том числе на основе модели Раша), что позволило количественно подтвердить надежность и устойчивость автоматизированной системы. Использование данного метода обеспечило объективность выводов и выявление факторов, влияющих на достоверность результатов тестирования.

Таким образом, сочетание кейс-стади, системного анализа и психометрической экспертизы придало исследованию междисциплинарный характер и позволило не только выявить технические и организационные аспекты функционирования АСУ ТЗ, но и определить конкретные пути ее совершенствования в целях повышения устойчивости, валидности и доверия к национальным экзаменационным процедурам.

Несмотря на успешные результаты, настоящее исследование имеет определенные ограничения. Анализ был проведен на основе данных, полученных в результате одного среза после внедрения изменений. Для подтверждения устойчивости и долгосрочной эффективности внедренных мер рекомендуется проведение исследования в динамике, что позволит отслеживать изменения психометрических показателей.

Результаты и обсуждение

Комплексная апробация предложенных мер модернизации АСУ ТЗ позволила зафиксировать ряд значимых результатов, демонстрирующих как технологические, так и методологические улучшения в системе тестирования.

Для решения проблемы низкой скорости отклика системы, которая ранее имела критически низкую среднюю оценку, был реализован комплексный подход, включающий три взаимосвязанных направления.

Во-первых, была внедрена технология локального кэширования в оперативной памяти, которая обеспечивает хранение наиболее часто запрашиваемых данных без необходимости многократного обращения к диску. Это решение позволило существенно повысить скорость выборки тестовых заданий и снизить нагрузку на серверное оборудование [4]. Эффективность механизма кэширования проверялась путем моделирования повторных запросов: после первичной выборки данные сохранялись в памяти, что обеспечивало их мгновенную загрузку при последующих обращениях.

Во-вторых, в рамках решения проблемы лицензионных рисков, имевшей одну из самых низких оценок, был проведен аудит программного обеспечения. Этот систематический процесс позволил выявить фактическое использование лицензионного и нелицензионного ПО, обеспечив соответствие условиям лицензионных соглашений и минимизировав юридические и финансовые риски. Результаты аудита способствовали высвобождению ресурсов, которые были направлены на дальнейшее развитие системы.

В-третьих, для повышения вычислительной мощности была осуществлена установка современных высокопроизводительных вычислительных узлов, при этом планируется поэтапное обновление всей вычислительной инфраструктуры. В контексте АСУ ТЗ эти машины выполняют функции серверов баз данных и рабочих станций, обеспечивая быструю обработку и хранение информации, что является фундаментальным условием для эффективного функционирования системы. Повышение скорости отклика было достигнуто не за счет одного, а трех взаимодополняющих и усиливающих друг друга решений, охватывающих оптимизацию работы с данными, увеличение вычислительных ресурсов и совершенствование правовой базы.

Одним из ключевых направлений модернизации стало обеспечение правовой и технологической защиты конфиденциальных данных. Особое внимание было уделено системе автоматической генерации вариантов тестов, которая была признана потенциально уязвимой с

точки зрения воспроизводимости. Для решения этой проблемы была внедрена система *DLP (Data Loss Prevention)* [5]. DLP-система предназначена для контроля и предотвращения утечек конфиденциальной информации. Она анализирует и при необходимости блокирует передачу данных из внутренней сети во внешнюю среду [6]. Ее функционал включает мониторинг и контроль каналов передачи данных (электронная почта, мессенджеры, внешние носители), классификацию данных как конфиденциальных и обнаружение файлов, которых не должно быть на рабочих станциях. В контексте образовательного тестирования, внедрение DLP-системы имеет решающее значение для обеспечения валидности психометрического анализа. Модель Раша, как и другие подходы в психометрии, исходит из предположения, что результаты испытуемых отражают их истинные способности, а не предварительное знание ответов. Утечка тестовых заданий могла бы нарушить это допущение, что привело бы к некорректным результатам психометрического анализа. Таким образом, DLP-система выполняет не только защитную функцию, но и обеспечивает «чистоту» данных, что является основополагающим условием для получения достоверных выводов.

Для устранения проблем, связанных с работой с большими данными, было принято решение о применении шардирования базы данных [7]. Шардирование – это техника горизонтального масштабирования, которая предполагает разделение большой логической базы данных на несколько независимых частей (шардов) [8]. Каждый шард функционирует как отдельная база данных, что позволяет эффективно распределять нагрузку и значительно повышать производительность при обработке больших объемов информации.

В рамках данного проекта было реализовано диапазонное шардирование через разделение каталогов тестовых заданий по годам и блокам. Такой подход обеспечивает не только ускорение выполнения запросов, но и является стратегическим шагом для проведения эффективного психометрического мониторинга. Анализ надежности и валидности тестов требует работы с большими историческими массивами данных о результатах тестирований. Разделение по годам и блокам позволяет быстро извлекать данные для конкретного периода или предметной области, что делает регулярный и оперативный анализ возможным.

Для повышения качества тестовых заданий на этапе их создания был разработан и внедрен инструмент «CheckerWord». Эта программа автоматизирует процесс контроля качества, снижая нагрузку на экспертов. Процесс работы инструмента следующий: пользователь загружает тесты в формате .doc, после чего система проверяет наличие обязательных элементов (дистракторы, дескрипторы), соответствие формата правилам базы тестовых заданий, выявляет повторяющиеся вопросы и идентичные дистракторы, а также оценивает корректность структуры. При обнаружении ошибок формируется сообщение, а после их исправления возможна повторная проверка. После успешной верификации задания конвертируются в рабочий формат для интеграции в систему.

Внедрение «CheckerWord» является превентивной мерой, напрямую влияющей на психометрическое качество тестов. Модель Раша, которая использовалась в дальнейшем анализе, основана на допущении одномерности, т.е. тест должен измерять одну и ту же скрытую переменную. Такие проблемы, как повторяющиеся вопросы или некорректные дистракторы, которые выявляет «CheckerWord», нарушают это допущение. Их устранение на входе обеспечивает высокое качество данных, что, в свою очередь, приводит к улучшению показателей соответствия модели в последующем анализе.

Внедрение комплекса решений, направленных на оптимизацию инфраструктуры и управление данными, позволило достичь статистически значимых улучшений по ряду ключевых показателей производительности системы. Проведенный сравнительный анализ (таблица 1) выявил снижение среднего времени отклика и времени обработки запросов базы

данных более чем в два раза, что свидетельствует о повышении эффективности обработки информации и рациональном использовании вычислительных ресурсов.

Таблица 1 - Сводные показатели производительности до и после внедрения оптимизаций

Показатель	До внедрения (M, SD)	После внедрения (M, SD)
Экспертная оценка скорости отклика (баллы)	3,91; 1,29	4,55; 0,35
Экспертная оценка времени обработки запроса БД (баллы)	3,93; 1,15	4,80; 0,42
Примечание - до внедрения (M, SD) - взяты из предыдущей работы: «Эффективность управления базой тестовых заданий в цифровой среде», журнал «Педагогические измерения» (Национальный центр тестирования, МНВО РК).		

Полученные результаты показывают, что системные улучшения привели не только к сокращению времени отклика, но и к снижению дисперсии значений. В частности, средние значения (M) по показателям существенно возросли: для скорости отклика – с 3,91 до 4,55, а для времени обработки запросов БД – с 3,93 до 4,80. Учитывая, что максимальное значение шкалы равно 5 (отлично), данные результаты свидетельствуют о приближении системы к оптимальному уровню функционирования.

Данный факт имеет принципиальное значение: уменьшение стандартного отклонения (SD) подтверждает рост стабильности и предсказуемости работы системы. Если на предварительном этапе наблюдалась выраженная вариативность производительности (скачкообразные задержки, зависящие от нагрузки), то после внедрения оптимизаций система демонстрирует устойчивое время отклика и предсказуемость поведения даже при пиковых нагрузках.

Таким образом, достигнутый эффект отражает не только количественные улучшения (ускорение обработки и рост средних оценок до уровня «хорошо – отлично»), но и качественные – устранение нестабильного пользовательского опыта, что напрямую влияет на надежность и доверие к системе тестирования.

Результаты психометрического анализа. По завершении всех системных улучшений был разработан вариант теста по профильной дисциплине «Математика» для группы образовательной программы М010 «Подготовка педагогов математики». Данный тест был использован при проведении КТ. В апробации приняли участие 977 испытуемых. По результатам тестирования был проведен психометрический анализ с использованием программы Winsteps (таблица 2).

Таблица 2 представляет сводные психометрические характеристики теста. Согласно полученным данным, все ключевые показатели имеют высокие значения, что подтверждает надежность и качество измерительного инструмента. Средняя стандартная ошибка измерения трудности заданий, составляющая 0,08 логит, свидетельствует о высокой точности оценки трудности заданий на уровне всей выборки.

Таблица 2 - Сводные психометрические характеристики теста

Сводка по 30 измеренным (неэкстремальным) элементам

	Сумма-рный балл	Коли-чество изме-рений	Изме-рение	Стандартная ошибка	Показатель соответствия INFIT		Показатель внешнего соответствия OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
Среднее значение	481.5	977.0	0.00	0.08	0.99	-0.13	1.00	0.12
Стандартная ошибка среднего	40.3	0.0	0.24		0.02	0.68	0.05	0.71
Стандартное отклонение популяции	216.9	0.0	1.27	0.01	0.13	3.67	0.26	3.84
Стандартное отклонение выборки	220.6	0.0	1.29	0.01	0.13	3.73	0.26	3.90
Макс.значение	882.0	977.0	1.79	0.11	1.28	9.44	1.45	8.60
Мин.значение	181.0	977.0	-2.63	0.07	0.82	-7.01	0.53	-6.35
Реальная среднеквадратичес-кая ошибка		0.08	Истин-ное стандарт ное отклоне-ние	1.27	Коэффи-циент разделения	15.37	Коэффи-циент надеж-ности теста	0.89
Модельная среднеквадратичес-кая ошибка измерения		0.08	Истин-ное стандарт ное отклоне-ние	1.27	Коэффи-циент разделения	15.69	Надеж-ность испытуе-мых	0.91
Стандартная ошибка среднего значения элемента				0.24				

Надежность является важнейшим показателем качества теста, определяющим точность и стабильность измерений, свободных от случайных ошибок. По сути, надежность отражает устойчивость и согласованность результатов, получаемых при повторном тестировании одних и тех же лиц в разное время, при использовании альтернативных вариантов теста или изменении условий проведения [9]. В данном исследовании показатели надежности подтверждают высокую качество инструмента: надежность участников (0,91) в сочетании с высоким коэффициентом разделения (15,69) свидетельствует о том, что тест эффективно и устойчиво различает участников по уровню их знаний. Надежность теста, оцененная с помощью коэффициента альфа Кронбаха (0,89), демонстрирует высокую воспроизводимость результатов и точность установления иерархии сложности заданий [10].

Значения показателей соответствия модели (Infit и Outfit MNSQ) для всех заданий находятся в допустимых пределах (0.5–1.5), что отвечает критерию «продуктивности для измерения» и подтверждает согласованность эмпирических данных с моделью Раша [11]. Это указывает на высокое психометрическое качество используемых заданий.

Диапазон трудности заданий варьируется от -2.63 до 1.79 логита, что составляет интервал в 4.42 логита. Такой разброс уровней сложности свидетельствует о способности теста эффективно дифференцировать респондентов с разным уровнем подготовленности.

Таким образом, совокупность представленных статистических характеристик подтверждает, что тест обладает высокой надежностью и соответствует требованиям современной психометрии.

В таблице 3 представлены результаты анализа отдельных тестовых заданий, полученные с использованием модели Раша. Психометрические характеристики каждого из 30 тестовых заданий включают уровень трудности, стандартную ошибку, показатели соответствия модели Раша и коэффициент корреляции.

Таблица 3 - Оценка параметров заданий

№	Балл	Кол-во	Изме-рение	Станд. ошибка	INFIT		OUTFIT		Точечно-бисер. корреляция		% правильных ответов	
					MNSQ	ZSTD	MNSQ	ZSTD	Набл.	Ожид.	Набл.	Ожид.
I0001	286	977	1,08	0,08	1,21	1,5	1,45	1,55	0,29	0,39	69,2	74,4
I0012	415	977	0,38	0,07	1,28	1,44	1,4	1,6	0,26	0,42	58,9	68,7
I0020	302	977	0,99	0,08	1,19	1,29	1,38	1,88	0,37	0,39	68,3	73,4
I0027	266	977	1,2	0,08	1,15	1,81	1,35	1,83	0,33	0,39	74	75,7
I0017	201	977	1,64	0,09	1,12	1,37	1,34	1,64	0,34	0,36	79,7	80,7
I0023	273	977	1,16	0,08	1,13	1,51	1,29	1,21	0,35	0,39	72,1	75,2
I0025	199	977	1,65	0,09	1	-0,09	1,19	2,1	0,33	0,36	82	80,9
I0029	288	977	1,07	0,08	1,01	0,16	1,16	2,59	0,22	0,4	75,9	74,3
I0005	181	977	1,79	0,09	0,96	-0,86	1,15	1,55	0,36	0,35	83,6	82,4
I0016	465	977	0,13	0,07	1,12	1,29	1,13	1,17	0,32	0,42	62,8	68
I0008	307	977	0,96	0,08	1,1	1,84	1,12	2,01	0,31	0,4	70,1	73,1
I0024	232	977	1,42	0,08	0,99	-0,15	1,1	1,38	0,37	0,37	78,7	78,2
I0021	477	977	0,07	0,07	1,07	1,79	1,09	2,29	0,36	0,42	64,7	68
I0028	461	977	0,15	0,07	1,04	1,65	1,06	1,39	0,38	0,42	65,9	68,1
I0011	291	977	1,05	0,08	0,97	-0,77	1,05	0,76	0,4	0,39	76,6	74,1
I0030	578	977	-0,44	0,07	1	0,15	1,01	0,17	0,41	0,42	68,9	69,5
I0007	438	977	0,26	0,07	0,99	-0,53	1	-0,08	0,43	0,42	68,9	68,2
I0014	447	977	0,22	0,07	0,98	-0,62	1	0,03	0,43	0,42	68,3	68,1
I0009	673	977	-0,96	0,08	0,93	-1,95	0,83	-1,01	0,47	0,4	72,8	73,5
I0004	563	977	-0,37	0,07	0,89	-1,13	0,85	-1,82	0,52	0,42	73,3	69,1
I0019	790	977	-1,73	0,09	0,89	-1,33	0,69	-1,72	0,48	0,36	82,4	81,9
I0013	401	977	0,45	0,07	0,88	-1,4	0,87	-1,05	0,51	0,42	74	69
I0026	882	977	-2,63	0,11	0,88	-1,45	0,58	-1,33	0,41	0,29	90,9	90,5
I0006	868	977	-2,45	0,11	0,86	-1,98	0,62	-1,19	0,44	0,3	89	89,1
I0018	528	977	-0,19	0,07	0,86	-1,67	0,82	-1,75	0,55	0,42	74,6	68,4
I0022	636	977	-0,75	0,07	0,86	-1,52	0,79	-1,31	0,53	0,41	77	71,7
I0002	833	977	-2,09	0,1	0,84	-1,77	0,7	-1,91	0,47	0,33	87,4	85,7
I0010	850	977	-2,26	0,1	0,84	-1,56	0,53	-1,59	0,49	0,32	87,8	87,4
I0015	748	977	-1,42	0,08	0,84	-1,07	0,69	-1,5	0,53	0,38	81	78,5
I0003	565	977	-0,38	0,07	0,82	-1,01	0,75	-1,35	0,58	0,42	75,5	69,2

Согласно таблице 3, большинство заданий соответствуют модели Раша и эффективно различают уровень подготовки. Все стандартные ошибки находятся в пределах 0,07-0,11, что указывает на достаточную точность измерения уровня сложности задания.

Все задания демонстрируют показатели INFIT и OUTFIT в пределах допустимых значений, что указывает на их удовлетворительное соответствие модели Раша и подтверждает адекватность использования этих заданий для оценки участников.

Задания I0012 и I0029 демонстрируют показатели 0,22 и 0,26 соответственно, что немного ниже ожидаемых значений для экзаменов с высокой ставкой (для мониторинга такое значение считается допустимым). Это указывает на их слабую связь с измеряемой конструкцией. Рекомендовано провести содержательный анализ этих заданий, и задания были пересмотрены в результате. В целом, большинство заданий демонстрируют положительную связь с общей способностью участников, что подтверждает их валидность.

Прогнозирование результатов показывает, что наблюдаемые и ожидаемые проценты правильных ответов находятся в близком соответствии, что свидетельствует о том, что модель Раша адекватно описывает ответы участников.

Соотношение между уровнем подготовки участников и сложностью тестовых заданий иллюстрируется на карте переменных «Испытуемые – Задание» (Рисунок 1). Анализ карты демонстрирует распределение подготовленности респондентов (слева) и уровней сложности заданий (справа) в диапазоне примерно от -4 до +4 логит.

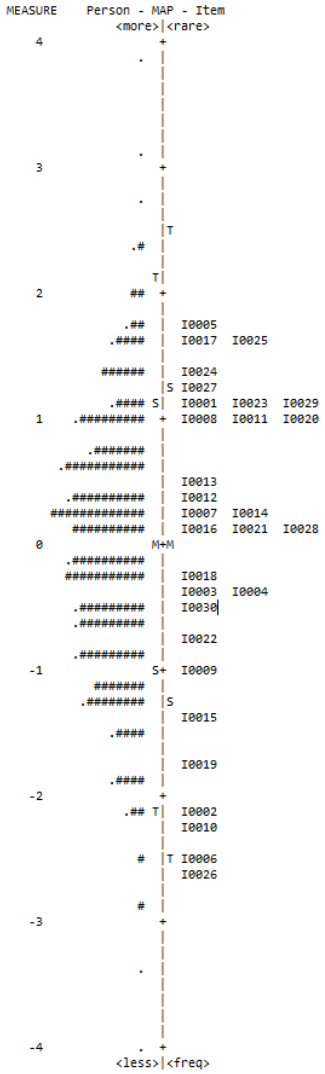


Рисунок 1 - Карта переменных. Испытуемые – Задание

Примечание – составлено автором на основе источника [результаты данных программы Winsteps]

Анализ карты показал, что уровни сложности тестовых заданий варьируются относительно среднего уровня подготовки участников. Задания охватывают диапазон сложности от легких (I0002, I0010, I0006, I0026) до сложных (I0005, I0017, I0025), с большинством заданий среднего уровня (I0001, I0008, I0011, I0020, I0023, I0024, I0029), что обеспечивает дифференциацию участников с различными способностями. Большинство участников сосредоточено вокруг среднего уровня (Measure 0–2), при этом встречаются единичные участники с очень низкими или очень высокими способностями. Карта показывает, что распределение заданий соответствует уровню участников, обеспечивая адекватное измерение и дифференциацию.

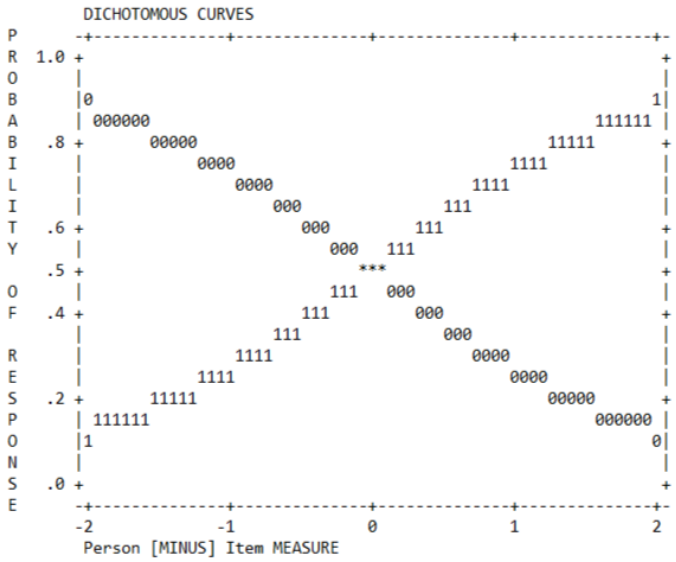


Рисунок 2 - Дихотомическая кривая характеристик задания (Item Characteristic Curve, ICC)

Примечание – составлено автором на основе источника [результаты данных программы Winsteps]

Характеристическая кривая задания (Item Characteristic Curve, ICC) демонстрирует зависимость вероятности правильного ответа от уровня способности участника (Рисунок 2). По оси X откладывается мера способности в логитах, отражающая латентную переменную «способность», а по оси Y — вероятность верного ответа, варьирующаяся от 0 до 1. На представленной кривой S-образная форма отображает закономерное увеличение вероятности правильного ответа с ростом способности: участники с низкой способностью редко отвечают верно, участники со средней способностью имеют около 50% вероятности, а участники с высокой способностью практически всегда отвечают правильно. Такая визуализация позволяет оценить сложность задания, его соответствие модели Раша и эффективность дифференциации участников с различным уровнем знаний.

Таким образом, результаты исследования свидетельствуют о высоком качестве тестовых заданий по математике, включая адекватное распределение их сложности. В дальнейшем целесообразно проведение лонгитюдного исследования для оценки стабильности и долгосрочной эффективности внедренных мер, что позволит отслеживать динамику психометрических показателей.

Техническая оптимизация процессов, связанных с формированием и управлением базой тестовых заданий, оказала существенное влияние на качество итоговых тестов. Стабильность функционирования базы и высокая скорость обработки информации, обеспеченные применением кэширования, нового оборудования и шардирования, создают условия для

бесперебойного доступа к заданиям и корректной генерации вариантов тестов. Это позволяет исключить технические задержки и сбои при формировании контрольно-измерительных материалов, что, в свою очередь, минимизирует риск внешних факторов, способных снизить качество педагогического измерения.

Ключевое значение имеет также обеспечение целостности и защищенности базы заданий. Использование шардирования и DLP-систем предотвращает дублирование и утечку материалов, гарантируя, что в процессе генерации применяются только оригинальные, аутентичные задания. Это является необходимым условием для поддержания валидности и надежности получаемых результатов.

Не менее важным элементом стала автоматизация контроля качества. Инструмент CheckerWord выполняет предварительную проверку заданий на этапе их разработки и хранения в базе. В ходе исследования было выявлено, что внедрение CheckerWord позволило устранить до 15% повторяющихся вопросов и некорректных дистракторов на этапе создания, что положительно сказалось на дискриминационной способности тестов и соответствии их психометрическим моделям (модель Раша). Таким образом, внедрение "CheckerWord" является превентивной мерой, напрямую влияющей на психометрическое качество тестов.

Эффективность инструмента особенно наглядно проявляется при последующем анализе заданий I0012 и I0029, которые показали низкие значения точечно-бисериальной корреляции. Хотя "CheckerWord" формально не выявил эти задания как критически проблемные (они соответствовали шаблону проверки), их содержательный разбор выявил скрытые дефекты: в I0012 был обнаружен "ловушковый" дистрактор с двойным смыслом, а в I0029 — неоднозначность в условии задачи.

Этот опыт демонстрирует важность комплексного подхода: автоматизированная проверка эффективно устраняет очевидные ошибки, а углубленный психометрический анализ выявляет более тонкие содержательные недочеты, что в совокупности значительно повышает качество банка заданий.Таким образом, именно комплексная оптимизация базы тестовых заданий и процедур генерации обеспечила измеримые улучшения качества тестов. Эмпирические данные подтверждают эффективность интеграции технических и педагогических решений, направленных на обеспечение надежности, валидности и воспроизводимости итоговых измерений.

Заключение

В настоящей работе был представлен анализ результатов комплексной модернизации АСУ ТЗ, проведенной на основе рекомендаций предыдущего исследования. Успешное внедрение технических и организационных мер, таких как кэширование, обновление оборудования, шардирование базы данных и внедрение DLP-системы, позволило значительно повысить производительность и безопасность системы. Разработка и использование инструмента "CheckerWord" обеспечили высокий уровень качества тестовых заданий на входе в систему.

Психометрический анализ, проведенный с помощью программного обеспечения Winsteps (версия 5.8.3.0) [12], [13], эмпирически подтвердил, что эти системные улучшения привели к достижению «отличного» психометрического качества тестов. Высокие значения надежности, разделения и соответствия модели свидетельствуют о том, что тестовые задания обладают высокой достоверностью и могут служить надежным инструментом для образовательных измерений. Данное исследование является примером успешной реализации синергетического подхода, где технические и психометрические меры взаимно усиливают друг друга для обеспечения надежности и валидности в образовательном тестировании.

Литература

1. Об утверждении Типовых правил приема на обучение в организации образования, реализующие образовательные программы высшего и послевузовского образования. Приказ Министра образования и науки Республики Казахстан от 31 октября 2018 года №600 [Электронный ресурс]. - URL: https://adilet.zan.kz/rus/docs/V1800017650 (Дата обращения: 31.08.2025).

2. Балашова И. Ю., Волынская К. И., Макарычев П. П. Методы и средства генерации тестовых заданий из текстов на естественном языке // Педагогические системы. – 2018. – № 3. – С. 45–52. [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/metody-i-sredstva-generatsii-testovyh-zadaniy-iz-tekstov-na-estestvennom-yazyke (Дата обращения: 31.08.2025).

3. Алтыбаева Ш.Б., Каргалдаева М.Д., Алибеков Р.Д. Эффективность управления базой тестовых заданий в цифровой среде. [Электронный ресурс]. URL: https://testcenter-zhurnal.kz/index.php/sap/article/view/20 (Дата обращения: 31.08.2025).

4. What are In-Memory Caches? [Электронный ресурс]. URL: https://www.geeksforgeeks.org/system-design/what-are-in-memory-caches/?utm_source (Дата обращения: 31.08.2025).

5. How DLP Helps Educational Institutions Protect Their Data. [Электронный ресурс]. URL: https://www.endpointprotector.com/blog/how-dlp-helps-educational-institutions-protect-their-data/?utm_source (Дата обращения: 31.08.2025).

6. Cisco Systems, Inc. Data Loss Prevention. [Электронный ресурс]. URL: https://www.cisco.com/site/us/en/learn/topics/security/what-is-data-loss-prevention-dlp.html (Дата обращения: 31.08.2025).

7. Шардирование. [Электронный ресурс]. URL: https://yandex.cloud/ru/docs/glossary/sharding (Дата обращения: 31.08.2025).

8. Масштабирование баз данных. URL: https://platformv.sbertech.ru/blog/masshtabirovanie-baz-dannyh (Дата обращения: 31.08.2025).

9. Анастази А. Психологическое тестирование: Книга 1 (К.М. Гуревич, В.И. Лубовский, Ред.; Пер. с англ.). – Педагогика -1982.

10. Альфа-коэффициент Кронбаха. [Электронный ресурс]. URL: https://www.sciencedirect.com/topics/nursing-and-health-professions/cronbach-alpha-coefficient (Дата обращения: 31.08.2025).

11. Bond, T.G., & Fox, C.M. (2015). *Applying the Rasch model fundamental measurement in the human sciences (*3rd ed.). DOI: 10.1111/j.1745-3984.2003.tb01103. (Дата обращения: 31.08.2025).

12. Linacre, M. Winsteps: Tutorial 1. Software operation and basic concepts, instructor - June 2012, https://www.winsteps.com/tutorials.htm.

13. Fit diagnosis: infit outfit mean-square standardized. [Электронный ресурс]. URL: https://www.winsteps.com/winman/misfitdiagnosis.htm?utm_source (Дата обращения: 31.08.2025).

Ш.Б. Алтыбаева, А.К. Каркенова, М.Д. Каргалдаева

ТЕСТ ТАПСЫРМАЛАРЫНЫҢ ДЕРЕККОРЫН АВТОМАТТАНДЫРЫЛҒАН БАСҚАРУ ЖҮЙЕСІНІҢ ПСИХОМЕТРИКАЛЫҚ ВЕРИФИКАЦИЯСЫ

Бұл мақалада ұлттық емтихандық рәсімдерде, соның ішінде Ұлттық бірыңғай тестілеу мен магистратураға кешенді тестілеуде пайдаланылатын тест тапсырмалары базасын басқарудың автоматтандырылған жүйесін кешенді жаңғыртудың нәтижелері қарастырылады. Негізгі назар техникалық және ұйымдастырушылық шешімдерді енгізудің тесттердің психометриялық сипаттамаларына және білімді бағалау сапасын арттыруға әсеріне аударылған. Зерттеу аясында кэштеу механизмдерін енгізу, есептеу инфрақұрылымын жаңарту, деректер базасын шардтау технологиясын қолдану, ақпараттық қауіпсіздікті қамтамасыз ету үшін DLP-жүйесін интеграциялау, сондай-ақ тапсырмалардың сапасын автоматтандырылған тексеру құралы CheckerWord-ты пайдалану сияқты шаралар талданды. Жүргізілген өзгерістердің тиімділігін бағалау үшін Раш моделін және Winsteps

бағдарламалық жасақтамасын (нұсқасы 5.8.3.0) қолдана отырып психометриялық талдау жасалды. Нәтижелер жаңғыртудың тесттердің сенімділігіне, қатысушыларды саралау деңгейіне және тапсырмалардың модельге сәйкестігіне оң әсер еткенін көрсетті. Жеке деректерді қорғау және базаның тұрақты жұмыс істеуі мәселелеріне ерекше көңіл бөлінген. Алынған нәтижелер техникалық және психометриялық шараларды біріктіретін кешенді тәсілдің білім сапасын ұлттық деңгейде бағалаудың валидтілігін, сенімділігін және қайта жаңғыртылуын қамтамасыз етудегі маңыздылығын дәлелдейді.

Түйінді сөздер: тест тапсырмаларын автоматтандырылған басқару жүйесі, психометриялық верификация, Раш моделі, тест сенімділігі, жарамдылық, DLP-жүйесі, CheckerWord.

Sh.B. Altybayeva, A.K. Karkenova, M.D. Kargaldayeva

PSYCHOMETRIC VERIFICATION OF THE AUTOMATED TEST ITEM MANAGEMENT SYSTEM

This article examines the results of a comprehensive modernization of the automated test item management system used in national examination procedures, including the Unified National Test and the Comprehensive Test for master's degree admission. The main focus is on the impact of implementing technical and organizational solutions on the psychometric characteristics of tests and the overall improvement of educational measurement quality. The study analyzed measures such as the introduction of caching mechanisms, upgrading computing infrastructure, applying database sharding technology, integrating a DLP system to ensure information security, as well as using the automated item quality control tool CheckerWord. To evaluate the effectiveness of the implemented changes, a psychometric analysis was conducted using the Rasch model and Winsteps software (version 5.8.3.0). The results showed that modernization had a positive effect on test reliability, participant differentiation, and item fit to the model. Particular attention was given to the protection of personal data and the stable functioning of the database. The findings confirm that a comprehensive approach combining both technical and psychometric measures contributes to ensuring the validity, reliability, and reproducibility of educational assessments within the national quality assurance system.

Keywords: automated test item management system, psychometric verification, Rasch model, test reliability, validity, DLP system, CheckerWord.

References

1. Ob utverzhdenii Tipovykh pravil priema na obuchenie v organizatsii obrazovaniia, realizuiushchie obrazovatelnye programmy vysshego i poslevuzovskogo obrazovaniia. Prikaz Ministra obrazovaniia i nauki Respubliki Kazakhstan ot 31 oktiabria 2018 goda №600 [Elektronnyi resurs]. - URL: https://adilet.zan.kz/rus/docs/V1800017650 (Data obrashcheniia: 31.08.2025).

2. Balashova I. Iu., Volynskaia K. I., Makarychev P. P. Metody i sredstva generatsii testovykh zadanii iz tekstov na estestvennom iazyke // Pedagogicheskie sistemy. – 2018. – № 3. – S. 45–52. [Elektronnyi resurs]. URL: https://cyberleninka.ru/article/n/metody-i-sredstva-generatsii-testovyh-zadaniy-iz-tekstov-na-estestvennom-yazyke (Data obrashcheniia: 31.08.2025).

3. Altybaeva Sh.B., Kargaldaeva M.D., Alibekov R.D. Effektivnost upravleniia bazoi testovykh zadanii v tsifrovoi srede. [Elektronnyi resurs]. URL: https://testcenter-zhurnal.kz/index.php/sap/article/view/20 (Data obrashcheniia: 31.08.2025).

4. What are In-Memory Caches? [Elektronnyi resurs]. URL: https://www.geeksforgeeks.org/system-design/what-are-in-memory-caches/?utm_source (Data obrashcheniia: 31.08.2025).

5. How DLP Helps Educational Institutions Protect Their Data. [Elektronnyi resurs]. URL: https://www.endpointprotector.com/blog/how-dlp-helps-educational-institutions-protect-their-data/?utm_source (Data obrashcheniia: 31.08.2025).

6. Cisco Systems, Inc.. Data Loss Prevention. [Elektronnyi resurs]. URL: https://www.cisco.com/site/us/en/learn/topics/security/what-is-data-loss-prevention-dlp.html (Data obrashcheniia: 31.08.2025).

7. Shardirovanie. [Elektronnyi resurs]. URL: https://yandex.cloud/ru/docs/glossary/sharding (Data obrashcheniia: 31.08.2025).

8. Masshtabirovanie baz dannykh. URL: https://platformv.sbertech.ru/blog/masshtabirovanie-baz-dannyh (Data obrashcheniia: 31.08.2025).

9. Anastazi A. Psikhologicheskoe testirovanie: Kniga 1 (K.M. Gurevich, V.I. Lubovskii, Red.; Per. s angl.). – Pedagogika -1982.

10. Alfa-koeffitsient Kronbakha. [Elektronnyi resurs]. URL: https://www.sciencedirect.com/topics/nursing-and-health-professions/cronbach-alpha-coefficient (Data obrashcheniia: 31.08.2025).

11. Bond, T.G., & Fox, C.M. (2015). Applying the Rasch model fundamental measurement in the human sciences (3rd ed.). DOI: 10.1111/j.1745-3984.2003.tb01103. (Data obrashcheniia: 31.08.2025).

12. Linacre, M. Winsteps: Tutorial 1. Software operation and basic concepts, instructor - June 2012, https://www.winsteps.com/tutorials.htm.

13. Fit diagnosis: infit outfit mean-square standardized. [Elektronnyi resurs]. URL: https://www.winsteps.com/winman/misfitdiagnosis.htm?utm_source (Data obrashcheniia: 31.08.2025).

Сведения об авторах***:***

Алтыбаева Шугыла Болатовна - магистр экономических наук, руководитель управления РГП на ПХВ «Национальный центр тестирования» МНВО РК, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: shugyla.altybayeva@gmail.com.

Каркенова Айна Казиевна - магистр педагогических наук, главный эксперт РГП на ПХВ «Национальный центр тестирования» МНВО РК, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: a.karkenova@mail.ru.

Каргалдаева Молдир Дуйсенбаевна - PhD, руководитель отдела РГП на ПХВ «Национальный центр тестирования» МНВО РК, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: moldirk2018@gmail.com.

Авторлар туралы мәлімет:

Алтыбаева Шугыла Болатовна - экономика ғылымдарының магистрі, басқарма басшысы, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Родниковая көш., 1/1, 010008, Астана, Қазақстан, e-mail: shugyla.altybayeva@gmail.com.

Каркенова Айна Казиевна - педагогика ғылымдарының магистрі, бас сарапшы, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Родниковая көш., 1/1, 010008, Астана, Қазақстан, e-mail: a.karkenova@mail.ru.

Каргалдаева Молдир Дуйсенбаевна - PhD, бөлімінің басшысы, ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Родниковая көш., 1/1, 010008, Астана, Қазақстан, e-mail: moldirk2018@gmail.com.

Information about authors:

Altybaeva Shugyla - master of economics, Head of Department RSE at PCV National Testing Center of the Ministry of Internal Affairs of the Republic of Kazakhstan, Rodnikovaya St., 1/1, 010008, Astana, Kazakhstan, e-mail: shugyla.altybayeva@gmail.com.

Karkenova Aina - Master of Education, Chief Expert RSE at PCV National Testing Center of the Ministry of Internal Affairs of the Republic of Kazakhstan, Rodnikovaya St., 1/1, 010008, Astana, Kazakhstan, e-mail: a.karkenova@mail.ru.

Kargaldayeva Moldir - PhD, Head of Department RSE at PCV National Testing Center of the Ministry of Internal Affairs of the Republic of Kazakhstan, Rodnikovaya St., 1/1, 010008, Astana, Kazakhstan, e-mail: moldirk2018@gmail.com.

IRSTI 04.51.53 https://doi.org/10.63597/UTO3105-4161.2025.3.3.006

D. Chuteneva
Nazarbayev University, Astana, Republic of Kazakhstan
e-mail: Dana.chuteneva9@gmail.com

ORCID 0000-0002-9142-5834

GENDER, MULTILINGUALISM, AND CAREER CHOICE IN KAZAKHSTAN: AN INTERSECTIONAL VIEW OF UNDERGRADUATE ASPIRATIONS

This article synthesizes qualitative findings on how gender and multilingualism intersect to shape the career aspirations of first-year undergraduates in Kazakhstan. Drawing on Feminist Theory [1], Sociocultural Theory [2, 3], and Intersectionality [4], the analysis situates student narratives within Kazakhstan's trilingual policy context—Kazakh, Russian, and English [5]—and wider sociolinguistic hierarchies. Interviews with ten students (five women, five men) reveal persistent gendered steering (e.g., "STEM for men, humanities for women"), the uneven valuing of language skills across fields and genders, and the differential translation of multilingual competence into leadership capital. While multilingualism is widely perceived as an asset, women report its recognition mainly in "supportive" roles (e.g., teaching, translation), whereas men describe language as leverage for mobility in high-status domains (e.g., IT, engineering, international business). Some women strategically select women-dominant fields for psychological safety, even at an income trade-off. The article concludes with implications for policy and practice: gender-responsive career guidance; equitable recognition of multilingual competencies; targeted mentorship for women aiming at multilingual leadership; and curricular/organizational reforms that address the gendered distribution of linguistic capital.

Keywords: Gender studies, Multilingualism, Career choice, Career aspirations, Intersectionality, Kazakhstan.

Introduction

Global discourses often frame multilingualism as a pathway to cognitive flexibility, intercultural competence, and expanded career options [6, 7]. Yet language resources never operate in a social vacuum: which languages "count," how they are recognized, and who benefits from them are patterned by power, ideology, and identity. Gender, in particular, shapes access to opportunity and the valuation of skills across labor markets [8]. In multilingual, post-Soviet Kazakhstan—where a state-led trilingual project promotes Kazakh, Russian, and English [5]—these dynamics converge in distinctive ways.

The thesis on which this article is based explores how gender and multilingual repertoires jointly inform undergraduates' early career choices. It focuses on first-year students at a large urban university—an ideal vantage point because students have chosen majors but are still consolidating identities and aspirations. The article's contribution is threefold. First, it traces how gendered expectations continue to steer choices despite officially inclusive policies [9]. Second, it shows how multilingual competence is stratified by language ideology—Kazakh for identity and civic belonging, Russian for interethnic and bureaucratic practicality, English for elite global mobility [10]. Third, it demonstrates that the returns to multilingualism are gendered: similar linguistic assets accrue different symbolic and economic value to women and men [11].

Guided by three questions—(1) the role of gender, (2) the role of multilingual repertoire, and (3) their intersection—this article integrates relevant scholarship and the study's qualitative findings to illuminate how students interpret, embrace, or resist the overlapping pressures of language hierarchies and gender norms.

Materials and Research Methods

The research employed an instrumental qualitative case study [12] at a multilingual urban university. Using purposeful maximum-variation sampling [13], ten first-year students (five women, five men) from varied majors participated in 60–90-minute semi-structured interviews [14] conducted in Kazakh, Russian, or English. Interviews were transcribed verbatim and thematically analyzed [15]. Ethical approval followed institutional protocols; pseudonyms protected identities.

Results and discussion

Conceptual and Policy Background

Multilingualism as Resource and Hierarchy

Contemporary sociolinguistics defines multilingualism not simply as a count of languages but as a dynamic repertoire of situated resources [16, 17, 18]. Individuals deploy languages unevenly across domains; competence is fluid and context-bound. In Kazakhstan, this fluidity is institutionalized through the trilingual policy: Kazakh (state), Russian (interethnic communication), and English (global integration) [5]. These languages carry stratified symbolic value—Kazakh as national identity, Russian as practical capital, English as global prestige [10]. Access to high-level proficiency remains uneven across urban–rural, class, and schooling lines, with implications for competitive careers [19].

Gender as Social Structure and Performance

Gender is socially constructed and continually enacted through norms and institutional practices [1, 20]. In post-Soviet contexts, women remain overrepresented in care-oriented fields (education, social work) while men dominate high-status technical sectors (engineering, IT, economics) [21, 9]. Self-assessment biases reinforce segregation: even with equal performance, men are more likely to rate themselves competent in STEM, while women underrate their abilities [22, 23].

Intersectionality and Linguistic Capital

Intersectionality [4] highlights how gender, language, class, and ethnicity interlock to produce compounded advantage or constraint. Language skills function as cultural capital [24], but their conversion into status depends on who wields them, where, and for what purpose [25]. In Kazakhstan, women's strong language skills may be recognized in "support" roles; the same skills in men are read as leadership readiness, intensifying unequal career trajectories [26].

Findings

Gendered Career Perceptions and Practices

"STEM for men, humanities for women." Across accounts, STEM was widely identified with masculine traits—logic, toughness, competitiveness—mirroring the gender performativity that normalizes segregation [1]. Several male participants described STEM as a default choice because it promises prestige and good salary. Several women recounted tacit discouragement from STEM by peers, teachers, or family, and described fatigue at the prospect of having constantly to "prove" belonging in male-dominated spaces. Even when women acknowledged increased salary in engineering or IT, some prioritized psychological safety, opting for professions with supportive peer cultures (e.g., education, linguistics). These patterns echo evidence of stereotype-based self-efficacy gaps and "chilly climate" dynamics that sap persistence in STEM [22, 23, 8, 21].

The parental script. Students repeatedly cited parental expectations—often framed as "secure, respectable" gender-appropriate careers. Male students described explicit nudges toward engineering or economics; women described advice to pursue linguistics, pedagogy, or biology, aligned with caregiving and communication. Such domestic socialization, backed by curricular images of gendered roles (Temirzakhkyzy et al., 2023), perpetuates occupational channeling [27].

Anticipatory discrimination. Women anticipated stricter performance scrutiny and lower recognition in male-dominated fields, steering choices even before labor-market entry. This anticipatory calculus illustrates how gender regimes operate through expectation and emotion as much as through formal exclusion [28].

Multilingual Repertoires as Career Capital

Language as strategic asset. Students framed multilingualism as essential for diplomacy, education, international business, and selective public-sector roles. English was consistently perceived as high-leverage capital for mobility and status; Russian as indispensable for interethnic pragmatics; Kazakh as a marker of belonging and civic duty, increasingly important in public institutions. These valuations reflect Kazakhstan's language ideology and the sociopolitical stratification of repertoires [5. 10].

Gendered orientations to language learning. Many women described intrinsic motivation and relational affordances in language study (communication, empathy, intercultural sensitivity), while several men articulated instrumental rationales (salary, leadership prospects, international placement). In classrooms, gendered discourse norms and teacher expectations can subtly reinforce these orientations [29, 30].

Identity work. Students used languages to perform career-aligned identities—projecting cosmopolitan professionalism via English, institutional credibility via Kazakh, and network fluency via Russian. Such "identity positioning" aligns with sociocultural views of language as a mediational means that shapes cognition and affiliation [2, 3, 31].

The Intersection: Who Gets Credit for the Same Skill?

Women's multilingualism - visible but undervalued. Women frequently reported that strong language skills were acknowledged yet channeled toward support functions: teaching, translation, client service—fields with lower pay and limited pathways to leadership. Their multilingualism was treated as expected rather than exceptional, consistent with research on the gendered devaluation of feminized skills [25].

Men's multilingualism - read as leadership. In contrast, men described language competence as a distinctive advantage signaling readiness for international teamwork and managerial tracks in engineering, finance, or tech. This differential interpretation transforms similar competence into uneven symbolic capital—a classic case of intersectional stratification [4, 24, 26].

Strategic compliance and resistance. Some women deliberately selected women-dominant programs to avoid hostile climates. Others resisted channeling and pursued law, diplomacy, or international business, leveraging multilingualism to claim visibility in male-dominated arenas. These resistant narratives echo poststructural feminist accounts of identity negotiation and agency [25].

Discussion: Mapping Findings to Theory

Feminist Theory: Performing "Appropriate" Futures

The normalization of "STEM for men, humanities for women" illustrates gender performativity [1]: repeated social cues define what futures feel natural. Women's anticipatory self-exclusion from male-dominated fields to avoid "proving oneself" reflects the affective labor of navigating gendered gaze and doubt [8]. Feminist analysis thus exposes how institutional climates, home discourses, and curricular images [9] align to narrow the horizon of legitimate ambition.

Sociocultural Theory: Language as Mediation of Identity and Opportunity

Language is not merely a skill but a mediational tool that structures thought, affiliation, and learning [2, 3]. Students' identity positioning via English, Kazakh, and Russian demonstrates how repertoires enable membership in particular communities of practice and opportunity structures [30]. Where high-quality trilingual instruction is uneven, repertoires—and thus career affordances—are unequally distributed [10].

Intersectionality: Converting Linguistic Capital into Status—For Whom?

Similar language skills were read differently across genders. Women's multilingualism accrued value in feminized roles; men's, in leadership tracks—evidence that capital conversion is filtered by gender [24, 25]. Intersectionality [4] helps explain compounded effects for women from rural or less resourced schools who face both limited access to elite English pathways and strong gendered steering.

Implications for Policy and Practice

The findings of this study carry important implications for educational policy, institutional practice, and broader social attitudes in Kazakhstan and similar multilingual, post-Soviet contexts. At the level of higher education, the research suggests that career guidance must explicitly address the ways in which gender stereotypes shape students' perceptions of professional suitability. When female students avoid STEM fields due to anticipated discrimination or discomfort in male-dominated environments, they are not only making individual trade-offs but also reinforcing structural inequalities in the labor market. Accordingly, secondary and tertiary institutions should provide gender-responsive career counseling that encourages students to critically reflect on social norms and recognize a wider range of professional opportunities as accessible to all genders.

Equally significant are the implications for language education. Multilingualism emerged as a highly valued resource, yet its recognition was uneven across gender lines. Whereas male students' linguistic abilities were often interpreted as markers of leadership potential, female students' comparable skills were perceived as "expected" or relegated to supportive positions. This discrepancy underscores the need for universities and employers to adopt more equitable frameworks for evaluating multilingual competencies. Language should be recognized as a form of cultural and professional capital across all career domains, not solely in feminized professions such as teaching or translation. Policies that ensure multilingual proficiency is valued in leadership and technical fields could help mitigate gendered disparities in the labor market.

At the societal level, the study also points to the importance of engaging families and communities in the transformation of gendered expectations. Many participants described parental influence as a decisive factor in shaping their career paths. Without efforts to shift such attitudes, institutional reforms alone may have limited effect. This suggests that awareness campaigns, community dialogues, and parental engagement programs should accompany educational initiatives to dismantle entrenched assumptions about gender-appropriate professions.

Finally, the study highlights the need for a more holistic policy approach that integrates gender equity and multilingual education. Current frameworks often treat these priorities in isolation, focusing either on promoting trilingualism or on addressing gender imbalance. Yet this research demonstrates that the two dimensions are deeply intertwined. Effective policy must therefore consider how gendered norms influence the acquisition and valuation of language skills, and how multilingual repertoires can be leveraged to expand—not restrict—career opportunities for all students.

Limitations and Directions for Future Research

While the study offers valuable insights into the intersection of gender and multilingualism in shaping career choices, it is also subject to several limitations that must be acknowledged. The first limitation concerns the scope and size of the sample. With ten participants drawn from a single multilingual, urban university, the findings cannot be generalized to all Kazakhstani undergraduates. Students in rural areas, monolingual institutions, or vocational pathways may experience language access and gendered expectations in markedly different ways. Future research would benefit from including a larger and more diverse cohort across multiple universities and regions to provide a more comprehensive picture.

A second limitation relates to the focus on first-year students. While this population was chosen deliberately to capture the early stages of career identity formation, it also means that the study captures aspirations rather than long-term outcomes. Students' choices may shift significantly as they

gain professional experience, encounter workplace cultures, or reassess their ambitions in light of structural barriers and opportunities. Longitudinal studies are therefore necessary to trace how the dynamics identified here evolve over time and influence actual career trajectories.

Another limitation lies in the reliance on self-reported data through interviews. Although this method is well suited to exploring perceptions and meanings, it is inevitably shaped by participants' willingness to share openly and by the researcher's interpretive lens. Some participants may have downplayed or omitted experiences of gendered or linguistic discrimination due to social desirability or discomfort, while others may have emphasized certain narratives to align with perceived expectations. Although strategies such as member checking and reflexivity were employed to enhance trustworthiness, the findings must still be understood as constructed accounts rather than objective representations of reality.

Finally, the study focuses primarily on the interplay of gender and multilingualism while acknowledging but not systematically analyzing other intersecting identities such as ethnicity, class, or rural–urban background. Intersectionality theory suggests that these dimensions are crucial in shaping compounded advantage or disadvantage. For example, a rural female student with limited access to English instruction may experience constraints that differ markedly from those of an urban male counterpart. Future research should therefore expand the intersectional scope to capture how these additional layers of identity interact with gender and multilingualism in structuring career opportunities.

Conclusion

Kazakhstan's trilingual vision positions language as a lever of modernization, yet the benefits of multilingualism are unevenly realized. This study's narratives show that gender remains a powerful organizer of aspiration, confidence, and recognition. Women's multilingualism is widely visible but too often confined to roles with limited power and pay; men's similar skills are framed as markers of leadership, especially in technical and high-status domains. These differences do not arise from ability but from the social reading of competence at the intersection of gender and language.

The way forward is not merely to expand language teaching or to proclaim gender equality in the abstract. It is to redesign systems—curricula, counseling, hiring, promotion, and everyday pedagogical practice—so that linguistic capital translates equitably into status, pay, and leadership for *all* students. Feminist, sociocultural, and intersectional perspectives together illuminate how this can be done: by transforming climates, leveling access to high-value repertoires, and revaluing multilingualism as strategic capital across every rung of the career ladder.

References

1. Butler, J. (2011). Gender trouble. In *Routledge eBooks*. [https://doi.org/10.4324/9780203824979]
2. Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. [https://doi.org/10.2307/j.ctvjf9vz4]
3. De Costa, P. I. (2007). J. P. Lantolf and S. L. Thorne: Sociocultural Theory and the Genesis of Second Language Development. Oxford University Press, 2006. *Applied Linguistics*, *28*(3), 477–480. [https://doi.org/10.1093/applin/amm027]
4. Crenshaw, K. (1989). Demarginalizing the intersection of race and sex. *University of Chicago Legal Forum, 1989*(1), 139–167. [https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8/]
5. Mehisto, P., Kambatyrova, A., & Nurseitova, K. (2014). Three in one? Trilingualism in policy and educational practice. In D. Bridges (Ed.), *Educational reform and internationalisation: The case of school reform in Kazakhstan* (pp. 152–177). Cambridge University Press.

[https://www.researchgate.net/publication/357048538_Three_in_One_Trilingualism_in_Policy_and_Educational_Practice]

6. Bialystok, E. (2010). *Bilingualism: Consequences for mind and brain*. *Trends in Cognitive Sciences, 16*(4), 240–250. [10.1016/j.tics.2012.03.001]

7. Dewaele, J.-M., & Li Wei. (2012). Multilingualism, empathy and multicompetence. *International Journal of Multilingualism, 9*(4), 352–366. [https://doi.org/10.1080/14790718.2012.714380]

8. Bloodhart, B., Balgopal, M. M., Casper, A. M. A., Sample McMeeking, L. B., & Fischer, E. V. (2020). Outperforming yet undervalued: Undergraduate women in STEM. *PLOS ONE, 15*(6), e0234685. [https://doi.org/10.1371/journal.pone.0234685]

9. Temirzakhkyzy, K., Ryskulova, M., & Abenov, Y. (2023). *Study of gender discrimination in the universities of Kazakhstan* (Policy analysis exercise, Nazarbayev University School of Public Policy). Nazarbayev University. [https://nur.nu.edu.kz/bitstreams/77d32781-f90e-4493-84a9-b2454d41dff5/download]

10. Smagulova, J. (2008). Language policies of Kazakhization and their influence on language attitudes and use. *International Journal of Bilingual Education and Bilingualism, 11*(3–4), 440–475. [https://doi.org/10.1080/13670050802148798]

11. Pavlenko, A., & Piller, I. (2008). Language education and gender. In *Springer eBooks* (pp. 57–69). [https://doi.org/10.1007/978-0-387-30424-3_5]

12. Stake, R. E. (1995). *The art of case study research*. Sage Publications.

13. Locke, K. (2002). Book Review: Qualitative Research and Evaluation Methods, 3rd ed., by Michael Quinn Patton (2001). Thousand Oaks, CA: Sage, 688 pages. *Organizational Research Methods*, *5*(3), 299–301. [https://doi.org/10.1177/10928102005003006]

14. Ginn, G. M., & Munn, S. L. (2019b). Interviews: Learning the Craft of Qualitative Research Interviewing, Third Edition, by Svend Brinkmann and Steinar Kvale. Thousand Oaks, CA: Sage, 2015. 405 pages, $60.00 (paperback). *New Horizons in Adult Education and Human Resource Development*, *31*(2), 67–69. [https://doi.org/10.1002/nha3.20251]

15. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. [https://doi.org/10.1191/1478088706qp063oa]

16. Wei, L. (2000). *The bilingualism reader*. Routledge. [https://doi.org/10.4324/9780203461341]

17. Blommaert, J. M. E., & Backus, A. (2013). *Repertoires revisited: 'Knowing language' in superdiversity*. (Working Papers in Urban Language & Literacies; Vol. 67). King's College. [http://www.kcl.ac.uk/sspp/departments/education/research/ldc/publications/workingpapers/67.pdf]

18. Cenoz, J., & Gorter, D. (2011). Focus on multilingualism: A study of trilingual writing. *The Modern Language Journal, 95*(3), 356–369. [https://doi.org/10.1111/j.1540-4781.2011.01206.x]

19. Naizabayeva, G., Garifolla, Y., & Mukhanbet, A. (2022). Changes in the traditional occupations and discourse of women: New element of human capital in redesigned Kazakh society. *Journal of Ethnic and Cultural Studies*, *9*(4), 55–69. [https://doi.org/10.29333/ejecs/1295]

20. West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender & Society*, *1*(2), 125–151. [https://doi.org/10.1177/0891243287001002002]

21. Lazarides, R., & Lauermann, F. (2019). Gendered paths into STEM-Related and Language-Related careers: girls' and boys' motivational beliefs and career plans in math and language arts. *Frontiers in Psychology, 10*. [https://doi.org/10.3389/fpsyg.2019.01243]

22. Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology, 106*(6), 1691–1730. [https://doi.org/10.1086/321299]

23. Correll, S. J. (2004). Constraints into preferences: Gender, status, and emerging career aspirations. *American Sociological Review, 69*(1), 93–113. [https://doi.org/10.1177/000312240406900106]

24. Barsky, R. F. (1991). Language and symbolic power. *Modern Language Quarterly, 52*(4), 466–468. [https://doi.org/10.1215/00267929-52-4-466]

25. Piller, I., & Pavlenko, A. (2007). Globalization, gender, and multilingualism. In *Encyclopedia of Language and Education* (pp. 57–69). Springer. [https://doi.org/10.1007/978-0-387-30424-3_5]

26. Radke, H., & Versloot, A. (2023). Referential multimodality, multilingualism and gender. In *Routledge eBooks* (pp. 170–202). [https://doi.org/10.4324/9781003166634-10]

27. Kalilayeva, S. (2020). *Teachers' perceptions of gender and academic achievement: Gender Stereotypes among teachers in single-gender schools in Kazakhstan* [Master's Thesis, Nazarbayev University]. Nazarbayev University Repository. [https://nur.nu.edu.kz/items/e38d3716-ff29-4d10-af0f-09078246c1e0]

28. Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology, 21*, 461–490. [https://doi.org/10.1146/annurev.an.21.100192.002333]

29. Sunderland, J. (2000). Issues of language and gender in second and foreign language education. *Language Teaching, 33*(4), 203–223. [https://doi.org/10.1017/S0261444800015688]

30. Duff, P. A. (2007). Second language socialization as sociocultural theory. *Language Teaching, 40*(4), 309–319. [https://doi.org/10.4324/9780203827147]

31. Burck, C. (2004). Living in several languages: implications for therapy. *Journal of Family Therapy, 26*(4), 314–339. [https://doi.org/10.1111/j.1467-6427.2004.00287.x]

Д. Чутенева

ҚАЗАҚСТАНДАҒЫ ГЕНДЕР, КӨПТІЛДІЛІК ЖӘНЕ МАНСАП ТАҢДАУЫ: СТУДЕНТТЕРДІҢ ҰМТЫЛЫСТАРЫНА ИНТЕРСЕКЦИЯЛЫҚ КӨЗҚАРАС

Бұл мақала Қазақстандағы бірінші курс студенттерінің кәсіби армандарына жыныс пен көптілділіктің тоғысуы қалай ықпал ететінін сапалық деректер негізінде талдайды. Гендерлік теорияға (Батлер, 1990), социомәдени теорияға (Выготский, 1978; Де Коста, 2007) және интерсекционалдылық тұжырымдамасына (Креншоу, 1989) сүйене отырып, талдау студенттердің әңгімелерін Қазақстанның үштілді саясат (қазақ, орыс және ағылшын тілдері) (Мехисто ж.т.б., 2014) пен кеңірек социолингвистикалық иерархиялар аясында қарастырады. Он студентпен (бес әйел, бес ер) жүргізілген сұхбаттар гендерлік бағыттаудың сақталып отырғанын көрсетті (мысалы, «STEM — ерлерге, гуманитарлық ғылымдар — әйелдерге» тұжырымдамасы), тілдік дағдылардың әр сала мен жынысқа қарай әртүрлі бағаланатынын, сондай-ақ көптілді құзыреттіліктің лидерлік капиталға айналуының теңсіздігін айқындады. Көптілділік жалпы артықшылық ретінде қабылданғанымен, әйелдер оның мойындалуын көбіне «қолдаушы» рөлдерде көреді (мысалы, оқыту, аударма), ал ерлер тілді жоғары мәртебелі салаларда (мысалы, IT, инженерия, халықаралық бизнес) ілгерілеудің құралы ретінде сипаттайды. Бұл табыс азаюымен байланысты болса да, кейбір әйелдер өздерінің психологиялық қауіпсіздігі үшін әйелдер басым салаларды саналы түрде таңдайды. Мақала саясат пен практикаға қатысты ұсыныстармен түйінделеді: гендерлік ерекшелікті ескеретін кәсіби бағдар беру; көптілді құзыреттерді әділ мойындау; көптілді лидерлікке ұмтылған әйелдерге бағытталған тәлімгерлік; сондай-ақ тілдік капиталдың гендерлік бөлінісін ескеретін оқу бағдарламасы мен ұйымдастырушылық реформалар.

Түйін сөздер: Гендерлік зерттеулер, Көптілділік, Мансап таңдауы, Кәсіби ұмтылыстар, Интерсекцияналдылық, Қазақстан.

Д. Чутенева

ГЕНДЕР, МУЛЬТИЛИНГВИЗМ И ВЫБОР ПРОФЕССИИ В КАЗАХСТАНЕ: ИНТЕРСЕКЦИОНАЛЬНЫЙ ВЗГЛЯД НА УСТРЕМЛЕНИЯ СТУДЕНТОВ

Эта статья обобщает качественные результаты исследования о том, как пересечение гендера и многоязычия формирует карьерные устремления студентов первого курса в Казахстане. Опираясь на феминистскую теорию (Батлер, 1990), социокультурную теорию (Выготский, 1978; Де Коста, 2007) и концепцию интерсекциональности (Креншоу, 1989), анализ рассматривает студенческие нарративы в контексте триязычной политики Казахстана — казахский, русский и английский языки (Мехисто ж.т.б., 2014) — и более широких социолингвистических иерархий. Интервью с десятью студентами (пять женщин и пять мужчин) выявили устойчивое гендерное направляющее влияние (например, «STEM для

мужчин, гуманитарные науки для женщин»), неравную ценность языковых навыков в разных сферах и для разных полов, а также различия в превращении многоязычной компетенции в лидерский капитал. Несмотря на то, что многоязычие в целом воспринимается как преимущество, женщины отмечают его признание в основном в «поддерживающих» ролях (например, преподавание, перевод), тогда как мужчины описывают владение языками как ресурс для продвижения в высокостатусных сферах (например, IT, инженерия, международный бизнес). Некоторые женщины сознательно выбирают женские доминирующие сферы ради психологической безопасности, даже если это связано с потерей дохода. Статья завершает обсуждение выводами для политики и практики: гендерно-чувствительное профориентирование; справедливое признание многоязычных компетенций; целевое наставничество для женщин, стремящихся к многоязычному лидерству; а также реформы в учебных планах и организациях, учитывающие гендерное распределение лингвистического капитала.

Ключевые слова: Гендерные исследования, Мультилингвизм, Выбор профессии, Карьерные устремления, Интерсекциональность, Казахстан.

Авторлар туралы мәлімет:

Чутенева Дана Олеговна – Назарбаев университет, Астана, Республика Казахстан, e-mail: Dana.chuteneva9@gmail.com.

Сведения об авторах:

Чутенева Дана Олеговна – Назарбаев университет, Астана, Қазақстан Республикасы, e-mail: Dana.chuteneva9@gmail.com.

Information about authors:

Chuteneva Dana – Nazarbayev University, Astana, Republic of Kazakhstan, e-mail: Dana.chuteneva9@gmail.com.

МРНТИ 14.91 https://doi.org/10.63597/UTO3105-4161.2025.3.3.007

Джафарзода С.Ф.[1], Шодиниё А.Ш.[2], Мирзоаминзода Х.М.[3]
[1,2,3]Национальный центр тестирования при Президенте Республики Таджикистан, г. Душанбе, Республика Таджикистан
email: [1]s-jafarov@yandex.ru, [2]abduvosid@list.ru, [3]mxm.ntc@mail.ru

РАЗВИТИЕ НАЦИОНАЛЬНОЙ СИСТЕМЫ ТЕСТИРОВАНИЯ В РЕСПУБЛИКЕ ТАДЖИКИСТАН

[1]*ORCID* 0009-0004-6894-1277, [2]*ORCID* 0009-0001-9042-9251, [3]*ORCID* 0009-0001-0249-9840

В статье рассматривается развитие национальной системы тестирования в Республике Таджикистан в контексте модернизации системы оценивания, образования и интеграции в международное образовательное пространство. Описана институциональная структура системы оценивания Республики Таджикистан, в которой ключевая роль отводится Национальному центру тестирования (НЦТ). Отмечены ключевые вызовы, связанные с развитием этой системы. Особое внимание уделено роли Национального центра тестирования как центрального звенья в сфере оценивания уровня знаний и способностей школьников и абитуриентов. Проанализировано участие Таджикистана в международных сравнительных исследованиях PISA и PIRLS. Показано, что данные, получаемые в результате национальных и международных оценок, имеют критическое значение. Они используются для пересмотра государственных образовательных стандартов, учебных программ, учебников и программ подготовки учителей, а также для построения национальной системы мониторинга качества образования. Описаны технологические и институциональные направления дальнейшего развития системы оценивания, включая переход к компьютерному тестированию, создание единой национальной базы данных результатов, применение методов интеллектуального анализа данных и подготовку национальных экспертов в области тестологии и психометрии.

Ключевые слова: качество образования, централизованные вступительные экзамены, международные сравнительные исследования, мониторинг образовательных результатов, цифровизация оценивания, образовательная политика Республики Таджикистан.

Введение

Национальная система образования Республики Таджикистан находится на новом и важном этапе своего развития. Её главной целью является повышение качества и эффективности обучения, приведение содержания образования в соответствие с требованиями современного мира, внедрение инновационных подходов, а также участия в международных сравнительных исследованиях и оценочных процедурах. Эти цели и задачи четко определены и отражены в стратегических документах, в числе которых Национальная стратегия развития образования Республики Таджикистан на период до 2030 г. [1], а также Национальная концепция и Национальная стратегия оценки качества образования на период до 2044 г. [2, 3]. Указанные документы задают долгосрочные ориентиры, принципы и механизмы формирования целостной национальной системы оценивания, обеспечивающей получение объективных и сопоставимых данных о результатах обучения и их использование при разработке и корректировке образовательной политики. При этом ключевой акцент делается на применении современных подходов к оценке учебных достижений, внедрении интеллектуальных технологий и аналитики данных, что делает систему более прозрачной, доступной и управляемой, а также ориентированной на повышение результатов каждого ученика и каждой школы.

Нормативно-правовая база системы оценивания в Республике Таджикистан опирается, прежде всего, на Закон Республики Таджикистан «Об образовании» [4], а также на специальные подзаконные акты, регламентирующие порядок проведения централизованных вступительных экзаменов в образовательные учреждения среднего и высшего профессионального образования Республики Таджикистан, проведения творческих экзаменов и экзаменов по специальным навыкам в образовательные учреждения среднего и высшего профессионального образования, порядок предоставления льгот абитуриентам на централизованных вступительных экзаменах в образовательные учреждения среднего и высшего профессионального образования, а также мониторинга качества образования [5, 6, 7, 8]. Тем самым создаётся единое правовое поле, в рамках которого реализуются различные формы внешнего оценивания.

Материалы и методы исследования

Институциональную структуру системы оценки качества образования в Республике Таджикистан формируют несколько государственных органов. Министерство образования и науки Республики Таджикистан уполномочено на формирование и реализацию государственной политики в сфере науки и образования, а также на нормативно-правовое регулирование деятельности образовательных учреждений. Комитет по начальному и среднему профессиональному образованию при Правительстве Республики Таджикистан обеспечивает проведение государственной политики в соответствующем сегменте образовательной системы. Агентство по надзору в сфере образования и науки при Президенте Республики Таджикистан осуществляет государственный надзор и контроль качества образования, в том числе на основе результатов внешних оценочных процедур. Особое место занимает Национальный центр тестирования при Президенте Республики Таджикистан, на который возложены функции организации и проведения централизованных вступительных экзаменов в образовательные учреждения среднего и высшего профессионального образования Республики Таджикистан (ЦВЭ), распределения абитуриентов по образовательным учреждениям высшего и среднего профессионального образования (ОУ СПО и ВПО) на основе ЦВЭ, организация мониторинга качества образования и участие в международных сравнительных исследованиях.

Следует отметить, что при этих госорганах созданы специализированные подразделения, занимающиеся оценкой качества образования и аналитикой результатов. Такой распределенный институциональный подход позволяет охватить разные уровни системы образования, однако механизм координации между этими структурами пока недостаточно отлажен. На практике отсутствует единый центр, обеспечивающий согласование действий всех участников системы оценивания. В результате наблюдается дублирование функций в одних областях и пробелы в других, что снижает эффективность мониторинга качества образования. Преодоление данной фрагментации и создание целостной, интегрированной системы оценки является одной из первоочередных задач дальнейшего развития.

Результаты и обсуждение

Национальный центр тестирования (НЦТ) является ключевым звеном национальной системы оценки качества образования в Республике Таджикистан. НЦТ наделён особыми полномочиями, обеспечивающими прозрачность и равный доступ к централизованным вступительным экзаменам (ЦВЭ), а также объективность оценки знаний школьников и абитуриентов. Основные цели НЦТ состоят в обеспечении справедливой, объективной и надёжной оценки знаний абитуриентов и равноправия при поступлении в ОУ СПО и ВПО РТ, а также проведении исследований и оценки уровня знаний учащихся образовательных

учреждений страны. При этом особое внимание уделяется как предметному содержанию тестов, так и технологии проведения ЦВЭ (стандартизированным процедурам), что направлено на достижение максимально возможной надёжности и валидности оценочных результатов. Также, деятельность НЦТ охватывает организацию и проведение масштабных тестирований по всей республике, разработку тестовых материалов, а также аналитическую обработку и хранение полученных данных.

Нормативную основу для ЦВЭ составляют специальные положения и порядки, утверждённые Правительством и Министерством образования и науки Республики Таджикистан [5, 6, 8, 9].

Содержание ЦВЭ формируется на основе государственных образовательных стандартов. Структура тестовых материалов включает задания двух основных типов: закрытые тестовые задания (две формы – задания с выбором одного правильного ответа из четырёх предложенных вариантов и задания на установление соответствия) и открытые задания (с кратким ответом – требующие самостоятельного решения задания). Такое сочетание позволяет оценивать у тестируемых одновременно знания фактов и способность применять их в новых ситуациях, анализировать и интерпретировать информацию. Для обеспечения прозрачности и объективности процедуры все задания проверяются в автоматизированном режиме, с использованием сканирования бланков и специализированного программного обеспечения. Это минимизирует влияние субъективного фактора и обеспечивает оперативность обработки: результаты экзамена публикуются в сжатые сроки, что повышает доверие участников и общества в целом к процедурам оценивания.

Помимо организации ЦВЭ, НЦТ осуществляет и мониторинговые оценки. В 2021–2022 гг. было проведено национальное исследование образовательных достижений учащихся 5-го и 11-го классов, охватившее базовые уровни школьного образования. Полученные данные заложили основу для последующих мониторингов качества обучения и позволили сопоставить результаты различных ступеней образования.

Значимым этапом развития национальной системы тестирования стало участие Таджикистана в международных сравнительных исследованиях качества образования, что способствует модернизации национальной системы образования в целом. В 2025 г. Республика Таджикистан впервые приняла участие в исследовании PISA, ориентированном на оценку функциональной грамотности 15-летних учащихся. В исследовании приняли участие 5778 учащихся из 160 образовательных учреждений. На момент подготовки статьи проводится анализ полученных данных, который позволит оценить уровень сформированности функциональной грамотности таджикских школьников в сравнении с другими странами, а также выявить сильные и слабые стороны национальной системы образования в глобальном контексте, требующие приоритетного внимания.

В области начального образования Таджикистан присоединился к программе PIRLS, оценивающей читательскую грамотность учащихся начальных классов. Основное исследование PIRLS запланировано на 2026 г., а в марте 2025 г. проведена апробация инструментария на выборке из 1180 учащихся 42 школ. Пилотное исследование позволило протестировать организационные процедуры, проверить корректность адаптации тестовых материалов и подготовить инфраструктуру для основного исследования. Анализ пилотных данных служит основой для дальнейшей доработки инструментов оценивания читательской грамотности младших школьников.

Результаты участия Республики Таджикистан в международных исследованиях используются также как инструмент внутренней модернизации. На их основе проводится пересмотр государственных образовательных стандартов и учебных программ, разрабатываются новые учебники, корректируются подходы к подготовке и повышению квалификации педагогов.

Данные международных исследований позволяют формировать национальные индикаторы качества образования, сопоставимые с международными, и выстраивать систему мониторинга, ориентированную на достижение измеримых результатов. Таким образом, внешние оценочные процедуры становятся частью внутреннего цикла управления качеством образования.

Параллельно с содержательными и организационными изменениями планируется масштабная технологическая модернизация системы оценивания в Республике Таджикистан. В перспективе предполагается поэтапный переход от бумажных форматов тестирования к компьютерному тестированию на централизованных экзаменах и в рамках мониторинговых исследований. Это создаёт условия для повышения эффективности управления процедурами (включая регистрацию участников, формирование вариантов тестов, обработку и распространение результатов) и для усиления безопасности данных. Введение компьютерного тестирования открывает возможности использования адаптивных моделей, повышающих точность измерения уровня подготовки учащихся.

Ключевым элементом новой инфраструктуры должна стать единая национальная база данных образовательных результатов учащихся образовательных учреждений Республики Таджикистан. В ней предполагается аккумулировать результаты ЦВЭ, мониторинговых исследований и иных форм внешней оценки, что позволит отслеживать динамику переходов учащихся между ступенями образования, анализировать различия между регионами и типами образовательных учреждений, выявлять группы риска и факторы, влияющие на успеваемость.

Для анализа больших массивов данных планируется использовать методы интеллектуального анализа данных и элементы искусственного интеллекта. Это создаёт предпосылки для более точной диагностики проблем, прогнозирования результатов и формирования адресных рекомендаций для образовательных организаций.

Институциональное развитие системы оценивания предполагает не только изменение процедур и создание новых информационных ресурсов, но и формирование кадрового потенциала. Важной задачей является подготовка национальных экспертов в области тестологии и психометрии, а также специалистов по анализу образовательных данных. Реализация программ подготовки и повышения квалификации для сотрудников профильных подразделений, методистов и преподавателей позволит обеспечить устойчивое функционирование системы оценивания на научно обоснованной основе. Одновременно планируется развитие онлайн-платформ для диагностики и самооценки, что повысит доступность оценочных инструментов для учащихся и педагогов и сделает оценивание частью повседневной учебной практики.

Развитие национальной системы тестирования в Республике Таджикистан в значительной степени опирается на адаптацию передового международного опыта. Важную роль играют методические подходы и рекомендации таких организаций, как OECD, IEA, UNICEF, которые предоставляют концептуальные рамки и технические решения для построения систем оценки качества образования. При этом принципиальное значение имеет учёт национального контекста: языковых и культурно-исторических особенностей, ресурсных возможностей и приоритетов образовательной политики. Адаптация международных методик в сочетании с формированием собственных подходов позволяет выстраивать систему оценивания, одновременно соответствующую мировым стандартам и ориентированную на реалии Республики Таджикистан.

Заключение

В целом, национальная система оценки качества образования в Республике Таджикистан находится в стадии активной трансформации. Обновление нормативно-правовой базы,

формирование институциональной архитектуры, развитие Национального центра тестирования, участие в международных исследованиях и планируемая технологическая модернизация создают базу для построения прозрачной, объективной и управляемой системы оценивания. Эффективное использование накопленных и вновь получаемых данных, а также дальнейшее укрепление кадрового и ресурсного потенциала системы оценивания будут способствовать повышению качества образования и конкурентоспособности человеческого капитала страны.

Литература

1. Национальная стратегия развития образования Республики Таджикистан на период до 2030 года: постановление Правительства Республики Таджикистан от 29.09.2020 г. №526.
2. Национальная концепция оценивания качества образования в Республике Таджикистан на период до 2044 года: постановление Правительства Республики Таджикистан от 29.04.2024 г. №240.
3. Национальная стратегия оценки качества образования в Республике Таджикистан на период до 2044 года: постановление Правительства Республики Таджикистана от 28.02.2025 г. №141.
4. Республика Таджикистан. Закон «Об образовании» от 22.07.2013 г. №1004.
5. Порядок проведения централизованных вступительных экзаменов в образовательные учреждения среднего и высшего профессионального образования Республики Таджикистан: утв. постановлением Правительства Республики Таджикистан от 29.04.2023 г. № 188.
6. Порядок проведения мониторинга качества образования в образовательных учреждениях: постановление Правительства Республики Таджикистан от 24.07.2025 г. №405.
7. Порядок проведения творческих экзаменов…: утв. распоряжением Министра образования и науки Республики Таджикистан от 12.10.2016 г. №3836; зарегистрирован в Минюсте РТ 23.11.2016 г. №830.
8. Справочник абитуриента: информационно-методическое пособие. – Душанбе, 2025. – 231с.
9. Порядок предоставления льгот абитуриентам на централизованных вступительных экзаменах…: утв. распоряжением Министра образования и науки Республики Таджикистан от 12.10.2016 г. № 3839; зарегистрирован в Минюсте РТ 23.11.2016 г. №832.

Джафарзода С. Ф., Шодиниё А. Ш., Мирзоаминзода Х. М.

ТӘЖІКСТАН РЕСПУБЛИКАСЫНДА ҰЛТТЫҚ ТЕСТІЛЕУ ЖҮЙЕСІН ДАМЫТУ

Мақалада Тәжікстан Республикасындағы ұлттық тестілеу жүйесінің дамуы бағалау жүйесін, білім беруді жаңғырту және халықаралық білім беру кеңістігіне кіру аясында қарастырылады. Тәжікстан Республикасының бағалау жүйесінің институционалдық құрылымы сипатталған, онда Ұлттық тестілеу орталығына (ҰТО) негізгі рөл берілген. Осы жүйенің дамуына байланысты негізгі қиындықтар атап өтілген. Оқушылар мен талапкерлердің білім деңгейі мен қабілеттерін бағалау саласындағы орталық буын ретіндегі Ұлттық тестілеу орталығының рөліне ерекше көңіл бөлінген. Тәжікстанның PISA және PIRLS халықаралық салыстырмалы зерттеулеріне қатысуы талданған. Ұлттық және халықаралық бағалау нәтижесінде алынған деректердің өте маңызды екендігі көрсетілген. Олар мемлекеттік білім беру стандарттарын, оқу бағдарламаларын, оқулықтарды және мұғалімдерді даярлау бағдарламаларын қайта қарау, сондай-ақ білім беру сапасын ұлттық мониторингілеу жүйесін құру үшін қолданылады. Бағалау жүйесін одан әрі дамытудың технологиялық және институционалдық бағыттары сипатталған, оның ішінде компьютерлік тестілеуге көшу, нәтижелердің бірыңғай ұлттық деректер базасын құру, интеллектуалды деректерді талдау әдістерін қолдану және тестология мен психометрия саласындағы ұлттық сарапшыларды даярлау қамтылған.

Түйінді сөздер: білім сапасы, орталықтандырылған қабылдау емтихандары, халықаралық салыстырмалы зерттеулер, білім беру нәтижелерінің мониторингі, бағалауды цифрландыру, Тәжікстан Республикасының білім беру саясаты.

Jafarov S.F., Shodiniyo A.Sh., Mirzoaminzoda Kh.M.

DEVELOPMENT OF THE NATIONAL TESTING SYSTEM IN THE REPUBLIC OF TAJIKISTAN

The article examines the development of the national testing system in the Republic of Tajikistan in the context of modernizing the assessment system, education, and integration into the international educational space. The institutional structure of the Republic of Tajikistan's assessment system is described, with a key role assigned to the National Testing Center (NTC). Key challenges related to the development of this system are noted. Special attention is paid to the role of the National Testing Center as the central link in assessing the level of knowledge and abilities of schoolchildren and applicants. Tajikistan's participation in the international comparative studies PISA and PIRLS is analyzed. It is shown that the data obtained from national and international assessments are of critical importance. They are used to revise state educational standards, curricula, textbooks, and teacher training programs, as well as to build a national system for monitoring the quality of education. The technological and institutional directions for the further development of the assessment system are described, including the transition to computer-based testing, the creation of a unified national database of results, the application of intelligent data analysis methods, and the training of national experts in the field of testology and psychometrics.

Key words: quality of education, centralized entrance examinations, international comparative studies, monitoring of educational outcomes, digitalization of assessment, educational policy of the Republic of Tajikistan.

References

1. Natsional’naya strategiya razvitiya obrazovaniya Respubliki Tadzhikistan na period do 2030 goda: postanovleniye Pravitel’stva Respubliki Tadzhikistan ot 29.09.2020 g. №526.
2. Natsional’naya kontseptsiya otsenivaniya kachestva obrazovaniya v Respublike Tadzhikistan na period do 2044 goda: postanovleniye Pravitel’stva Respubliki Tadzhikistan ot 29.04.2024 g. №240.
3. Natsional’naya strategiya otsenki kachestva obrazovaniya v Respublike Tadzhikistan na period do 2044 goda: postanovleniye Pravitel’stva Respubliki Tadzhikistana ot 28.02.2025 g. №141.
4. Respublika Tadzhikistan. Zakon «Ob obrazovanii» ot 22.07.2013 g. №1004.
5. Poryadok provedeniya tsentralizovannykh vstupitel’nykh ekzamenov v obrazovatel’nyye uchrezhdeniya srednego i vysshego professional’nogo obrazovaniya Respubliki Tadzhikistan: utv. postanovleniyem Pravitel’stva Respubliki Tadzhikistan ot 29.04.2023 g. № 188.
6. Poryadok provedeniya monitoringa kachestva obrazovaniya v obrazovatel’nykh uchrezhdeniyakh: postanovleniye Pravitel’stva Respubliki Tadzhikistan ot 24.07.2025 g. №405.
7. Poryadok provedeniya tvorcheskikh ekzamenov...: utv. rasporyazheniyem Ministra obrazovaniya i nauki Respubliki Tadzhikistan ot 12.10.2016 g. №3836; zaregistrirovan v Minyuste RT 23.11.2016 g. №830.
8. Spravochnik abituriyenta: informatsionno-metodicheskoye posobiye. – Dushanbe, 2025. – 231s.
9. Poryadok predostavleniya l’got abituriyentam na tsentralizovannykh vstupitel’nykh ekzamenakh...: utv. rasporyazheniyem Ministra obrazovaniya i nauki Respubliki Tadzhikistan ot 12.10.2016 g. № 3839; zaregistrirovan v Minyuste RT 23.11.2016 g. №832.

Сведения об авторах:

Джафарзода С.Ф. – доцент, кандидат технических наук, директор Национального центра тестирования при Президенте Республики Таджикистан, г. Душанбе, Республика Таджикистан, e-mail: s-jafarov@yandex.ru

Шодиниё А.Ш. – первый заместитель директора Национального центра тестирования при Президенте Республики Таджикистан, г. Душанбе, Республика Таджикистан, e-mail: abduvosid@list.ru

Мирзоаминзода Х.М. – кандидат физико-математических наук, начальник управления по планированию, мониторингу, исследованию и международным отношениям Национального центра тестирования при Президенте Республики Таджикистан, e-mail: mxm.ntc@mail.ru

Авторлар туралы мәлімет:

Джафарзода С.Ф. – техника ғылымдарының кандидаты, доцент, Тәжікстан Республикасы Президенті жанындағы Ұлттық тестілеу орталығының директоры, Душанбе қ., Тәжікстан Республикасы, e-mail: s-jafarov@yandex.ru

Шодиниё А.Ш. – Тәжікстан Республикасы Президенті жанындағы Ұлттық тестілеу орталығы директорының бірінші орынбасары, Душанбе қ., Тәжікстан Республикасы, e-mail: abduvosid@list.ru

Мирзоаминзода Х.М. – физика-математика ғылымдарының кандидаты, Тәжікстан Республикасы Президенті жанындағы Ұлттық тестілеу орталығының жоспарлау, мониторинг, зерттеу және халықаралық қатынастар басқармасының бастығы, e-mail: mxm.ntc@mail.ru

Information about the Authors:

Jafarov S.F. – Associate Professor, Candidate of Technical Sciences, Director of the National Testing Center under the President of the Republic of Tajikistan, Dushanbe, Republic of Tajikistan, e-mail: s-jafarov@yandex.ru

Shodiniyo A.Sh. – First Deputy Director of the National Testing Center under the President of the Republic of Tajikistan, Dushanbe, Republic of Tajikistan, e-mail: abduvosid@list.ru

Mirzoaminzoda Kh.M. – Candidate of Physical and Mathematical Sciences, Head of the Department for Planning, Monitoring, Research, and International Relations of the National Testing Center under the President of the Republic of Tajikistan, e-mail: mxm.ntc@mail.ru

ЖАСАНДЫ ИНТЕЛЛЕКТ ЖӘНЕ БІЛІМ БЕРУДІ БАҒАЛАУДАҒЫ ИННОВАЦИЯ

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ИННОВАЦИИ В ОЦЕНИВАНИИ ОБРАЗОВАНИЯ

ARTIFICIAL INTELLIGENCE AND INNOVATIONS IN EDUCATIONAL ASSESSMENT

IRSTI 14.15.01 https://doi.org/10.63597/UTO3105-4161.2025.3.3.008

N. Baizhanov[1*], A. Sabyrov[2], A. Makhmutova[3,4], S. Kadyrov[4]
[1,2]National Testing Centre, Ministry of Science and Higher Education, Astana, Kazakhstan
[3]SDU University, Kaskelen, Kazakhstan
[4]New Uzbekistan University, Tashkent, Uzbekistan
*e-mail: nurbaizhanov@gmail.com

[1]ORCID 0009-0008-5302-9858, [2]ORCID 0009-0006-1470-5433, [3]ORCID 0000-0002-8597-7667, [4]ORCID 0000-0002-8352-2597

LARGE LANGUAGE MODELS IN EDUCATIONAL MEASUREMENT OF KAZAKH LANGUAGE PROFICIENCY

This study evaluates the performance of large language models (LLMs) in assessing Kazakh language proficiency within the context of the Unified National Test (UNT) in Kazakhstan. The primary objective is to examine the accuracy, error patterns, and psychometric characteristics of five state-of-the-art LLMs—Gemini 2.5 Pro Preview, Claude 3.7 Sonnet, Deepseek R1, Qwen, and Llama 3.1-405B-Instruct—on 138 multiple-choice questions (MCQs) from the 2024 UNT Kazakh language test. The methodology involved a zero-shot evaluation with standardized prompts, ensuring no external data access, and employed statistical analyses, including Cochran's Q test, McNemar's tests, and Generalized Estimating Equations (GEE) logistic regression, to assess model performance across difficulty levels and linguistic topics. Results indicate that Gemini achieved the highest accuracy (90.6%), significantly outperforming other models, while Llama showed the lowest (37.7%). Performance varied by difficulty and topic, with Gemini excelling across all categories and others showing strengths in specific areas like complex linguistic reasoning. The study highlights the potential of LLMs for educational assessment in low-resource languages like Kazakh, while identifying gaps in model optimization, fairness, and reliability, necessitating targeted fine-tuning and culturally relevant data curation.

Keywords: Large Language Models, Educational Assessment, Kazakh Language, Unified National Test, Pedagogical Measurement, Artificial Intelligence.

Introduction

Education in Kazakhstan places strong emphasis on assessing student proficiency through the Unified National Test (UNT), which plays a central role in university admission. In recent years, increasing numbers of test takers have chosen to complete the UNT in Kazakh, reflecting both policy priorities and the broader use of the national language in education [1]. At the same time, advances in artificial intelligence, particularly large language models (LLMs), have created new opportunities for educational measurement. LLMs such as GPT-3.5, GPT-4, and other contemporary models are increasingly applied to tasks including automated grading, item generation, and diagnostic assessment [2]. These technologies promise efficiency and scalability, but their performance varies across domains, item types, and especially languages.

In language proficiency assessment, particularly for less-resourced languages, key questions remain about the validity, fairness, and reliability of LLM outputs [3]. For the Kazakh language category of the UNT, no published study has systematically evaluated LLM performance on multiple-choice questions. This gap is significant, since Kazakh proficiency is not only an academic requirement but also a cultural and policy priority. Therefore, this study aims to evaluate the performance of several state-of-the-art LLMs on the Kazakh language multiple-choice portion of the UNT. By examining their accuracy, error patterns, and psychometric characteristics, the study provides evidence on the potential and limitations of LLMs in supporting educational assessment in Kazakhstan.

Literature review

The Role of AI and LLMs in Educational Assessment and Testing. Over the past decade, advancements in artificial intelligence (AI) have dramatically transformed the educational assessment landscape. LLMs—characterized by their billions of parameters and training on vast, heterogeneous datasets—are now capable of generating human-like text and offering real-time, adaptive feedback on a variety of tasks. These systems facilitate automated test item generation, adaptive scoring, and personalized formative feedback, thereby reducing the burden on educators and allowing for scalable assessment solutions [4, 5]. For example, state-of-the-art models such as GPT-4 have demonstrated strong performance on numerous high-stakes examinations by replicating expert reasoning and delivering coherent explanations, a capability that signifies a major shift from traditional manual assessment methods [6, 7].

LLMs have been deployed to automatically generate draft test items based on textbook content or curricular themes. By extracting and synthesizing key concepts from large corpora, these models produce multiple-choice questions, true/false items, and even open-ended prompts that align with predetermined learning objectives [8]. Such automated generation not only accelerates the development cycle of test items but also enables educators to focus on qualitative review and refinement rather than on initial content creation [8]. In computerized adaptive testing environments, LLMs contribute further by assisting in real-time calibration of question difficulty based on examinee responses, thus tailoring tests dynamically to individual ability levels and ensuring that assessments are both engaging and diagnostically informative [9].

Additionally, the integration of LLMs into automated scoring systems has been transformative. Natural language processing (NLP) techniques within these models are used to evaluate essays and open-ended responses by considering not only surface features such as vocabulary usage and syntactic complexity but also deeper semantic coherence [9, 10]. The immediate feedback provided by automated systems helps students identify their learning gaps and enables teachers to design targeted interventions, thus democratizing access to high-quality, personalized assessments across diverse educational settings [9, 11]. LLMs also play a supportive role in ensuring test security through anomaly detection during online exams, flagging irregular response patterns that may indicate academic dishonesty [5, 11]. Together, these developments illustrate how AI-driven systems are gradually shifting educational assessment from traditional, labor-intensive processes to dynamic, technology-enabled paradigms that are scalable, consistent, and highly adaptable [9, 12].

LLM Performance on Multiple-Choice Question Answering in High-Stakes Examinations. A significant stream of research has focused on the performance of LLMs in multiple-choice question (MCQ) settings, especially on high-stakes examinations used in professional licensure and academic selection. Early iterations of generative models such as GPT-3.5 exhibited inconsistent performance on domain-specific MCQs, largely due to limitations in their reasoning abilities and the necessity for extensive prompt engineering [6]. However, subsequent models, notably GPT-4, have achieved remarkable improvements, attaining scores that place them within the top decile on examinations such as the Uniform Bar Exam and the United States Medical Licensing Examination (USMLE) [7, 13].

Researchers have investigated not only raw accuracy but also the calibration of confidence estimates in LLM outputs when answering medical and legal MCQs. Calibration analyses for GPT-4, for instance, have indicated that the model's probability assignments closely track actual correctness, a feature considered critical in high-stakes testing environments where score margins have significant implications [14]. In parallel, chain-of-thought prompting techniques—which encourage the model to articulate intermediate reasoning steps—have been shown to boost accuracy by making the decision process more transparent and interpretable [11, 15].

Another focus of research has been on the sensitivity of LLMs to the ordering of answer options. Positional bias, where models may disproportionately favor options presented in certain positions, has been documented, and techniques such as majority voting over multiple reordered presentations have been developed to mitigate these biases [16, 17]. These ensemble methods not only help in refining the final answer but also improve reliability by confirming that the chosen answer is robust across different prompt formulations [18, 19]. They further underscore the necessity of continuous refinements in prompting and calibration to ensure that LLMs' performance is both optimal and ethically reliable for high-stakes decision-making [20, 21].

Research on LLMs in Low-Resource Languages. While breakthroughs in LLM performance have predominantly emerged from high-resource language contexts, there is a growing recognition that these models must be extended to accommodate low-resource languages. Kazakh, a language characterized by agglutinative morphology, vowel harmony, and complex syntactic structures, epitomizes the challenges encountered in this area [22]. Existing multilingual LLMs, although designed to handle several languages simultaneously, often exhibit diminished performance when processing Kazakh due to limited training data and tokenization schemes optimized for Indo-European languages [23].

Recent efforts have sought to address these challenges by developing culturally and linguistically tailored benchmarks such as the KazMMLU dataset. This benchmark comprises thousands of multiple-choice questions designed specifically to assess competencies in Kazakh and Russian, thereby reflecting the bilingual nature of the region's educational system [23]. Evaluations using KazMMLU have revealed that while proprietary models like GPT-4o and DeepSeek V3 perform robustly on Russian test items—often achieving accuracies well above 75%—their performance on Kazakh items lags significantly, highlighting a critical imbalance in model efficacy across languages [22, 23].

The performance gap is not solely a function of data scarcity but also arises from the inherent linguistic complexities present in Kazakh. Standard subword segmentation techniques, which work effectively for languages with relatively simple morphology, struggle to capture the nuances of Kazakh grammar and syntax, resulting in poorer understanding and lower quality outputs [5, 24]. Moreover, research has indicated that prompts written in English often lead to higher accuracy than those written entirely in Kazakh, suggesting that current models are biased toward high-resource language representations that were more abundantly available during training [21, 22]. To mitigate these limitations, research in transfer learning from closely related languages such as Russian and specialized prompt engineering techniques is being explored, though these measures have only partially closed the performance gap [11].

Furthermore, the evaluation benchmarks developed so far for low-resource languages have primarily focused on narrow NLP tasks such as sentiment analysis or named entity recognition, with limited attention to the comprehensive reasoning abilities required in educational assessments [22, 24]. This gap underscores the necessity for developing comprehensive datasets and evaluation protocols that not only assess factual recall but also complex problem-solving and critical thinking in the context of educational measurement for languages like Kazakh [23].

Challenges Related to Fairness, Validity, and Reliability in AI-Driven Educational Assessment. A significant area of concern in the deployment of LLMs in educational assessment relates directly to the issues of fairness, validity, and reliability. These concerns are particularly acute when the training data for LLMs is derived from sources that may not represent the full diversity of student populations, potentially leading to biased outcomes that disadvantage underrepresented groups [25]. The opaqueness that characterizes many deep learning models makes it challenging to diagnose the specific sources of bias, which can range from demographic inequities to cultural misinterpretations, consequently affecting the fairness and reliability of automated scoring processes [26, 27].

Fairness concerns are not limited to demographic biases but extend to the methodological biases that occur in automated scoring systems. Frequently, these systems rely on proxy measures such as essay length, vocabulary complexity, or punctuation frequency, which might correlate with writing proficiency in a biased manner [8, 10]. Such methods often fail to capture more abstract elements of quality, including critical thinking, creativity, and coherence, thereby potentially reducing construct validity in high-stakes exams [9, 11]. These challenges are compounded when high-stakes decisions such as licensure or university admissions are based on scores generated by AI systems, where even small biases can have disproportionate impacts on individual educational trajectories [8].

Reliability also emerges as a critical challenge due to the “black-box” nature of many LLM architectures. The uncertainty inherent in deep neural network algorithms raises questions about the consistency of scores produced by AI systems over time. Various studies have attempted to address this by using ensemble methods, such as majority voting or calibration techniques, to ensure that LLM outputs remain stable across multiple runs and different prompt formulations [11, 16]. Although such techniques can improve the overall reproducibility of the results, concerns persist regarding the interpretability of the decision-making processes used by these models [12, 21].

Moreover, academic integrity issues further complicate the adoption of LLM-based assessments. The ease with which students can leverage LLMs to generate entire responses or essays—sometimes without proper attribution—has sparked debate regarding whether such practices constitute plagiarism or represent a legitimate aid in learning [23, 25]. These issues necessitate a balanced approach wherein the benefits of rapid, automated feedback are weighed against the risk of undermining authentic learning and critical thinking skills [8]. Finally, ongoing interdisciplinary collaborations are required to develop robust ethical frameworks and standardized guidelines that ensure transparent, fair, and reliable use of LLMs in educational assessment [26, 28].

Gaps in the Literature and Justification for Investigating LLM Performance in Kazakh Language Testing (UNT). Despite significant progress in applying LLMs to educational assessments in high-resource language settings, several critical gaps justify a closer investigation of model performance for low-resource languages such as Kazakh, especially in the context of the Unified National Testing (UNT) system. First, while numerous studies have detailed the impressive achievements of models like GPT-4 in handling MCQ tasks and adaptive testing in English, there is a marked paucity of research addressing the performance of these systems in languages that exhibit complex morphological features and limited digital resources [23].

Benchmark evaluations such as those based on the KazMMLU dataset have provided early insights into this gap. These evaluations demonstrate that while proprietary models (e.g., GPT-4o, DeepSeek V3) achieve commendable performance in Russian—a language for which relatively abundant training data exist—their accuracies drop significantly when processing Kazakh test items [23]. Furthermore, studies have indicated that English-language prompts lead to superior performance relative to Kazakh prompts, suggesting that the underlying training paradigms and tokenization methods are not sufficiently optimized for languages with complex agglutinative morphology [22]. Such disparities highlight the urgent need for dedicated research on prompt engineering, data curation, and model adaptation that better capture the linguistic nuances of Kazakh [5, 24].

Another notable gap is the lack of research on the longitudinal impacts of integrating LLM-driven assessments in low-resource settings. Although many studies have focused on short-term performance metrics, the long-term educational outcomes—such as changes in student learning trajectories, retention of critical thinking skills, and overall academic growth—remain poorly understood, particularly in the context of culturally-specific and linguistically diverse environments like Kazakhstan [5, 11]. There is also minimal insight into how LLM-driven scoring systems interact with locally defined educational standards and curricula, an issue that becomes critical when assessments are used for high-stakes decisions such as university admissions and licensure [23].

Moreover, existing benchmarks for low-resource languages are often limited to basic NLP tasks such as sentiment analysis or named entity recognition, without addressing the comprehensive reasoning and domain-specific knowledge required for educational assessments [22, 24]. The current literature thus falls short of providing a complete picture of LLM performance on intricate tasks such as MCQ answering, essay grading, and cognitive diagnostics in a low-resource language context [21, 22]. This situation is exacerbated by the lack of culturally relevant training data and evaluation protocols that reflect the authentic educational experiences of Kazakh-speaking students [23].

The need for comprehensive, bilingual benchmarks that cover a wide range of subjects—from STEM to social sciences—is therefore imperative. Such benchmarks would not only enable a detailed assessment of LLM performance but also inform the development of tailored model architectures and fine-tuning strategies that bridge the performance gap between high- and low-resource languages [23]. Importantly, these efforts should also address issues of fairness and reliability by incorporating mechanisms for differential item functioning analysis and cross-cultural validation [26, 28]. Ultimately, investigating LLM performance on Kazakh language testing is not merely a technical challenge but also an ethical imperative to ensure that the benefits of AI-driven assessments are distributed equitably across diverse linguistic communities [23].

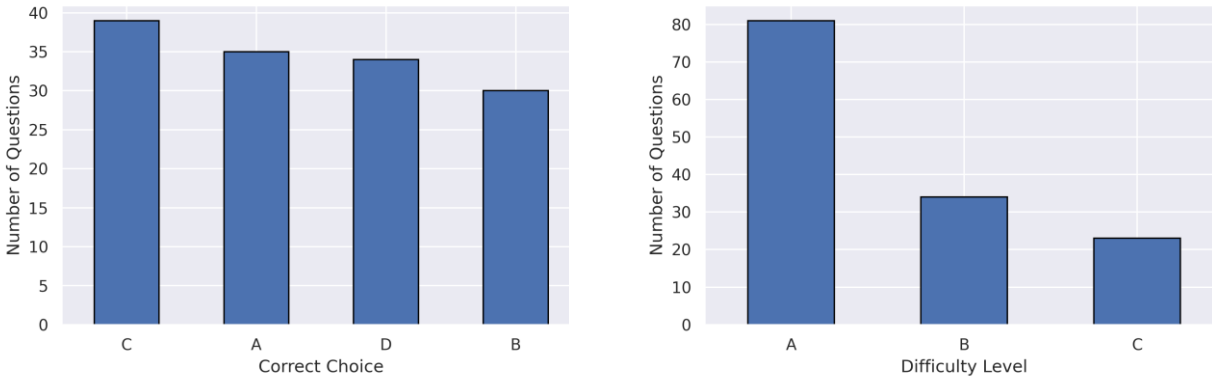
In summary, the integration of AI and large language models into educational assessment systems marks a transformative shift in how learning is measured and evaluated. LLMs have demonstrated exceptional prowess in generating content, calibrating adaptive tests, and scoring high-stakes examinations with performance metrics that are increasingly competitive with those of human experts [4, 7]. However, while high-resource language contexts have benefited immensely from these advancements, significant disparities persist when these systems are applied to low-resource languages, particularly Kazakh.

The literature reviewed here underscores not only the remarkable progress that has been achieved with MCQ answering and adaptive testing in high-stakes examinations but also identifies critical shortcomings in model performance, especially in handling complex morphological nuances and culturally specific content [22, 23]. Moreover, concerns regarding fairness, validity, and reliability in automated assessment systems remain paramount, necessitating continued research into ethical frameworks, prompt engineering, and culturally tailored data curation [26, 28].

Materials and methods

In UNT, the Kazakh Language component consists of 40 MCQs designed to assess reading comprehension, grammar, vocabulary, and linguistic reasoning across multiple subdomains. The methodology encompasses data preparation, model selection, experimental design, and evaluation procedures to address the research questions on LLMs' ability to handle low-resource language assessments, with a focus on improving preparation for the UNT in Kazakhstan's educational context. All LLMs relied on internal reasoning via prompt engineering, with no access to online searches or external corpora, ensuring that their performance reflected inherent model capabilities rather than external retrieval.

To construct the evaluation dataset, we collected 138 UNT Kazakh Language MCQs from the 2024 test cycle, originally presented in Kazakh. These questions were used in their original language to ensure fidelity to the source material and to assess the multilingual and low-resource language capabilities of the selected LLMs. This approach avoided potential distortions from translation and allowed for a more authentic evaluation of model performance in a real-world national testing setting. Due to confidentiality constraints, sample questions are not included here, but representative UNT Kazakh language test items are available from the National Test Center (2024) [29].



a) Distribution of Correct Answer Choices

b) Distribution of Questions by Difficulty Level

Figure 1 - Descriptive Statistics of the UNT Kazakh Language Dataset

Questions were categorized into three difficulty levels—A (easy), B (moderate), and C (hard)—based on historical UNT performance data, expert judgment, and expected percent-correct thresholds (Figure 1a). The dataset contained 81 level A questions, 34 level B questions, and 23 level C questions. This distribution ensured a balanced representation across difficulty levels and aligned with national testing standards. The distribution of correct answers was relatively balanced across alternatives, indicating no systematic bias in item construction (Figure 1b).

The dataset further covered a wide range of linguistic domains, distributed across 22 topics. The most represented topics included punctuation (n = 22), main and secondary sentence parts (n = 18), synonyms, antonyms, and homonyms (n = 10), orthography of compound and hyphenated words (n = 9), and assimilation rules (n = 9). Less frequent but still relevant categories included verb forms (n = 6), sentence phrases (n = 5), adjectives, interjections, and nouns (n = 4 each), with specialized categories such as archaic words, neologisms, and professional terms represented by one item each. This broad coverage ensured that the dataset reflected the curricular diversity of the Kazakh language test.

Five state-of-the-art LLMs were selected for evaluation based on their multilingual performance, reasoning capabilities, and accessibility through the OpenRouter platform. The models represent a diverse set of providers and architectures. Gemini 2.5 Pro Preview (Google) offers a million-token context window and is particularly strong in mathematical reasoning. Qwen (Alibaba) uses a Mixture of Experts architecture with 235 billion parameters in total, of which 22 billion are active during inference. Deepseek R1 (DeepSeek) is specialized in technical reasoning tasks, while Claude 3.7 Sonnet (Anthropic) emphasizes high performance in natural language proficiency. Finally, Llama 3.1-405B-Instruct (Meta AI) represents an open-source, large-scale model suitable for broad research applications. For performance evaluation, each LLM was provided with a standardized prompt directing it to select the single best answer to a UNT multiple-choice question (e.g., “Answer: C”). The evaluation was conducted in a zero-shot setting, ensuring that the models had no prior exposure to the test items. Each prompt contained the full question text and four answer choices, with explicit instructions for the model to reply using only the corresponding letter (A, B, C, or D). Model outputs were processed using regular expressions to extract the selected letter, with non-standard responses resolved by interpreting the first character or marking them as invalid.

To analyze model performance, we applied a combination of nonparametric and regression-based statistical techniques suitable for binary outcomes. Cochran’s Q test was used to evaluate whether accuracy differences across the five LLMs were statistically significant when answering the same set of questions. For pairwise comparisons, McNemar’s tests with Bonferroni correction were conducted to

identify which models differed significantly from each other. In addition, a Generalized Estimating Equations (GEE) logistic regression model with an exchangeable correlation structure was employed to account for within-question clustering and to estimate the relative performance of each model compared to a baseline. Together, these methods provided both omnibus and pairwise insights into LLM performance, while controlling for dependence across repeated measures on the same test items.

Results and discussion

Figure 2 presents the overall accuracy of the five evaluated LLMs on the UNT Kazakh Language dataset. Gemini achieved the highest score, correctly answering 90.6% of the questions, which indicates near-human level performance in this low-resource language setting. Claude followed with 61.6%, demonstrating moderate proficiency but with noticeable gaps compared to Gemini. Deepseek and Qwen performed similarly, with accuracies of 53.6% and 49.3%, respectively, reflecting only partial capability in handling the linguistic and reasoning demands of the assessment. Llama obtained the lowest score at 37.7%, suggesting that even large-scale open-source architectures may struggle without targeted optimization for low-resource languages. These differences highlight both the potential and the limitations of current LLMs in supporting high-stakes national testing contexts.

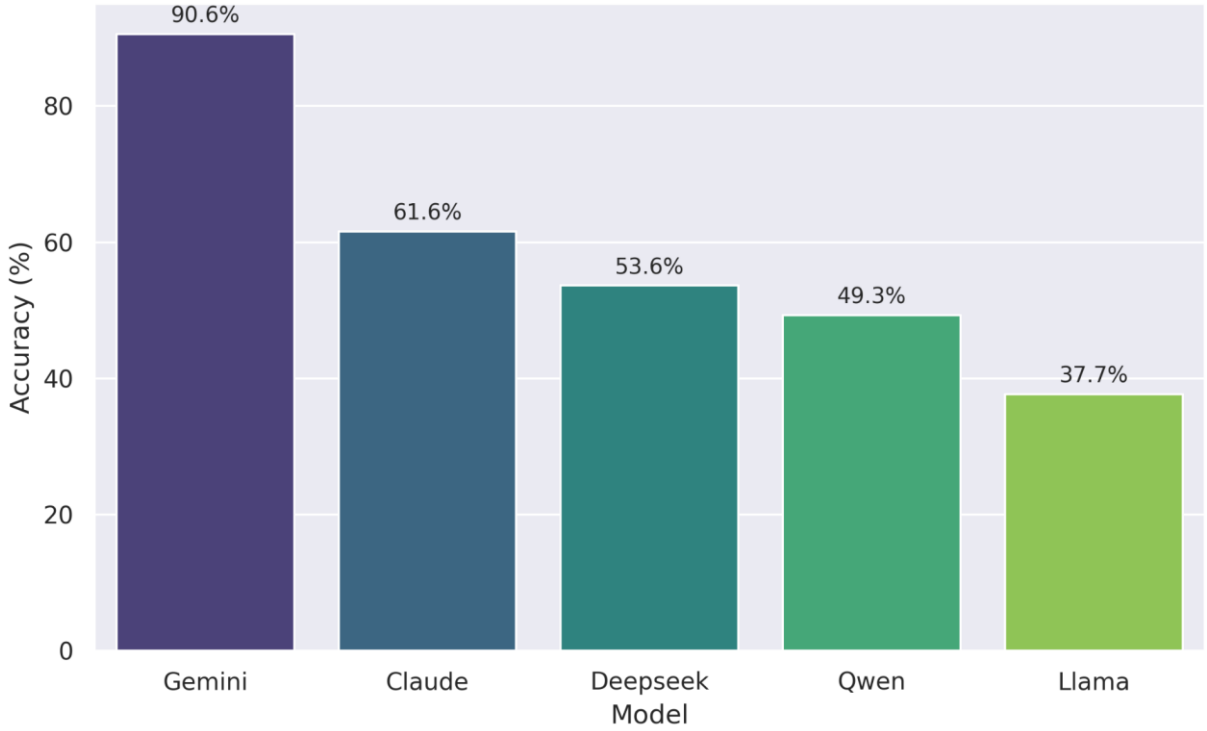


Figure 2 - Performance of LLM's on Kazakh language test in UNT

Figure 3 illustrates model accuracy across different difficulty levels. Gemini consistently achieved the highest accuracy, maintaining strong performance on easy (91.4%) and moderate (91.2%) items, with only a slight decline on hard items (87.0%). Claude displayed a different trend, performing moderately on easy questions (58.0%) but improving with increasing difficulty, reaching 69.6% on the hardest items. Deepseek followed a similar pattern, starting lower on easy items (51.9%) but showing its best performance on difficult questions (73.9%). Qwen demonstrated relatively stable results, hovering around 43–50% on easy and moderate items before a marked increase to 69.6% on hard

items. Llama, in contrast, showed weaker performance overall, with accuracy declining from 42.0% on easy items to 29.4% on moderate items, before partially recovering to 34.8% on hard items. These results suggest that while Gemini excels across all levels, other models—particularly Claude, Deepseek, and Qwen—may be more resilient on challenging items than on simpler ones, highlighting distinct strengths in handling complex linguistic reasoning.

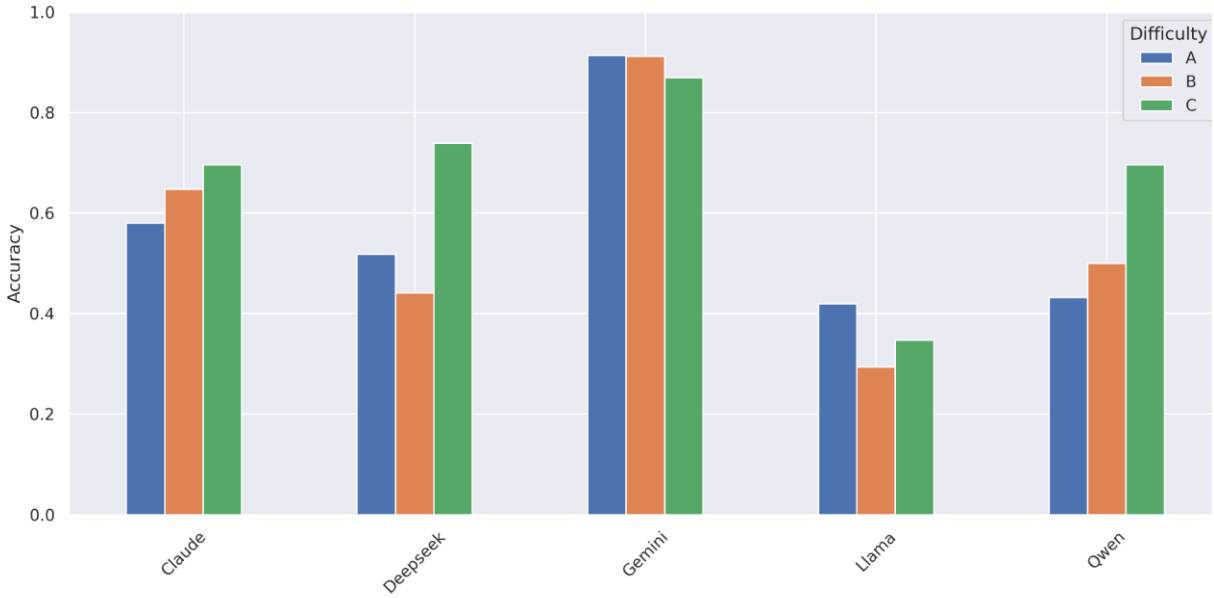


Figure 3 - Accuracy of LLMs on the Kazakh language test by difficulty level

The statistical analysis revealed substantial variation in LLM performance across the 138 UNT Kazakh language questions. Cochran's Q test confirmed that the models' accuracies were not equal ($Q = 108.30$, $df = 4$, $p < .001$), highlighting meaningful differences in their ability to handle low-resource language items. Post-hoc McNemar tests with Bonferroni correction showed that Gemini consistently outperformed Claude, Deepseek, Llama, and Qwen in pairwise comparisons, while some contrasts such as Claude vs Deepseek and Llama vs Qwen were not significant. The GEE logistic regression reinforced these findings, with Gemini achieving significantly higher accuracy than the baseline ($\beta = 1.79$, $p < .001$), corresponding to an estimated improvement of around 15–20 percentage points in correct responses. By contrast, Llama ($\beta = –0.98$, $p < .001$) and Qwen ($\beta = –0.50$, $p = .012$) performed significantly worse, suggesting they struggled with the nuances of Kazakh grammar and vocabulary. Deepseek showed a marginal trend toward lower performance ($p = .099$), positioning it between Claude and the weaker-performing models.

Table 1 summarizes accuracy across major linguistic topics. Gemini consistently achieved perfect accuracy across all categories, underscoring its strong coverage of Kazakh grammar, vocabulary, and syntax. Claude also performed robustly, particularly on professional terms, numerals, pronouns, and sentence-level constructions, though its accuracy was somewhat lower on polysemous words and adjective-related items. Deepseek demonstrated uneven performance, excelling on numerals and phrases but failing on pronouns and struggling with core grammatical categories such as nouns and adjectives. Qwen showed mixed results, performing well on compound sentences and pronouns but less consistently on other areas. Llama displayed the lowest overall accuracy, with performance dropping sharply on complex and sentence-level tasks. These findings highlight that while Gemini shows broad mastery, other models reveal topic-specific strengths and weaknesses, which could guide targeted applications in exam preparation and language learning.

Table 1 - Accuracy of LLMs on Kazakh language test by topic area (%)

Topic (Simplified English)	Claude	Deepseek	Gemini	Llama	Qwen
Professional / taboo terms	100.0	100.0	100.0	100.0	100.0
Numerals and their functions	100.0	100.0	100.0	33.3	66.7
Pronouns and their functions	100.0	0.0	100.0	100.0	100.0
Polysemous words (multiple meanings)	85.7	57.1	100.0	57.1	42.9
Complex sentences (subordinate clauses)	83.3	66.7	100.0	16.7	66.7
Phrases (noun and verb phrases)	80.0	80.0	100.0	80.0	40.0
Compound sentences (coordination)	75.0	50.0	100.0	37.5	75.0
Interjections and mimetic words	75.0	50.0	100.0	25.0	25.0
Nouns and their functions	75.0	25.0	100.0	50.0	50.0
Adjectives and degrees of comparison	75.0	0.0	100.0	25.0	50.0

From a practical perspective, Gemini's relative advantage translates into a substantially higher likelihood of selecting the correct option on UNT-style items, which could make it a more reliable tool for educational support in Kazakhstan's context. Meanwhile, the weaker performance of Llama and Qwen emphasizes the limitations of even large-scale models when applied to low-resource languages without fine-tuning. These findings illustrate both the promise and the current gaps of LLMs for national language assessments: while state-of-the-art models like Gemini can approach usable levels of performance, significant disparities remain, reinforcing the need for targeted adaptation and training on Kazakh data to ensure fairness and effectiveness in real-world applications.

Conclusion

The evaluation of five state-of-the-art LLMs on the UNT Kazakh language MCQs reveals significant disparities in their performance, with Gemini demonstrating near-human accuracy (90.6%) and robustness across difficulty levels and linguistic topics, while models like Llama and Qwen struggled, particularly without targeted optimization for Kazakh's complex morphology. These findings underscore the promise of LLMs in supporting scalable, efficient educational assessments for low-resource languages but highlight critical limitations in fairness, validity, and reliability. To ensure equitable application in high-stakes contexts like the UNT, future work should focus on fine-tuning models with Kazakh-specific datasets, developing culturally tailored benchmarks, and addressing biases through advanced prompt engineering and ethical frameworks. Such efforts are essential to bridge performance gaps and enhance the role of AI in Kazakhstan's educational landscape.

Acknowledgments

This research was funded by the National Testing Centre under the Ministry of Science and Higher Education of the Republic of Kazakhstan.

References

1. Nurseitova, K., Kaliyeva, A., Denst-Garcia, E., & Kussainova, R. (2017) Impact Of Language Of Instruction On Progress In Kazakhstan. *European Proceedings of Social and Behavioural Sciences*.

2. Teckwani, S. H., Wong, A. H. P., Luke, N. V., & Low, I. C. C. (2024). Accuracy and reliability of large language models in assessing learning outcomes achievement across cognitive domains. *Advances in Physiology Education*, *48*(4), 904-914.

3. Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., ... & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), 90-112.

4. Abd-Alrazaq, A., AlSaad, R., Alhuwail, D., Ahmed, A., Healy, P. M., Latifi, S., ... & Sheikh, J. (2023). Large language models in medical education: opportunities, challenges, and future directions. *JMIR medical education*, *9*(1), e48291.

5. Alekseev, A., & Turatali, T. (2024, October). KyrgyzNLP: challenges, progress, and future. In *International Conference on Analysis of Images, Social Networks and Texts* (pp. 3-39). Cham: Springer Nature Switzerland.

6. Bommarito II, M., & Katz, D. M. (2022). GPT takes the bar exam. *arXiv preprint arXiv:2212.14402*.

7. Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2024). Gpt-4 passes the bar exam. *Philosophical Transactions of the Royal Society A*, *382*(2270), 20230254.

8. Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *Eurasia journal of mathematics, science and technology education*, *19*(8), em2307.

9. Caines, A., Benedetto, L., Taslimipoor, S., Davis, C., Gao, Y., Andersen, O., ... & Buttery, P. (2023). On the application of large language models for language teaching and assessment technology. *arXiv preprint arXiv:2307.08393*.

10. Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education sciences*, *13*(7), 692.

11. Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274.

12. Johnson, M. S., & McCaffrey, D. F. (2023). Evaluating fairness of automated scoring in educational measurement. *Advancing natural language processing in educational assessment*, *142*.

13. Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., ... & Tseng, V. (2023). Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS digital health*, *2*(2), e0000198.

14. Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of gpt-4 on medical challenge problems. *arXiv preprint arXiv:2303.13375*.

15. Law, A. K., So, J., Lui, C. T., Choi, Y. F., Cheung, K. H., Kei-ching Hung, K., & Graham, C. A. (2025). AI versus human-generated multiple-choice questions for medical education: a cohort study in a high-stakes examination. *BMC Medical Education*, *25*(1), 208.

16. Pezeshkpour, P., & Hruschka, E. (2024, June). Large Language Models Sensitivity to The Order of Options in Multiple-Choice Questions. In *Findings of the Association for Computational Linguistics: NAACL 2024* (pp. 2006-2017).

17. Rouzegar, H., & Makrehchi, M. (2024). Generative AI for enhancing active learning in education: A comparative study of GPT-3.5 and GPT-4 in crafting customized test questions. *arXiv preprint arXiv:2406.13903*.

18. Sari, A. N. (2024). Exploring the Potential of Using AI Language Models in Democratising Global Language Test Preparation. *International Journal of TESOL & Education*, *4*(4), 111-126.

19. Stribling, D., Xia, Y., Amer, M. K., Graim, K. S., Mulligan, C. J., & Renne, R. (2024). The model student: GPT-4 performance on graduate biomedical science exams. *Scientific Reports*, *14*(1), 5670.

20. Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., ... & Xie, X. (2024). A survey on evaluation of large language models. *ACM transactions on intelligent systems and technology*, *15*(3), 1-45.

21. Maeda, H. (2025). Field-testing multiple-choice questions with AI examinees: English grammar items. *Educational and Psychological Measurement*, *85*(2), 221-244.

22. Maxutov, A., Myrzakhmet, A., & Braslavski, P. (2024, August). Do LLMs speak Kazakh? A pilot evaluation of seven models. In *Proceedings of the First Workshop on Natural Language Processing for Turkic Languages (SIGTURK 2024)* (pp. 81-91).

23. Togmanov, M., Mukhituly, N., Turmakhan, D., Mansurov, J., Goloburda, M., Sakip, A., ... & Koto, F. (2025). KazMMLU: Evaluating Language Models on Kazakh, Russian, and Regional Knowledge of Kazakhstan. *arXiv preprint arXiv:2502.12829*.

24. Karibayeva, A., Karyukin, V., Abduali, B., & Amirova, D. (2025). Speech Recognition and Synthesis Models and Platforms for the Kazakh.

25. Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, *20*(2), 1-24.

26. Bulut, O., Beiting-Parrish, M., Casabianca, J. M., Slater, S. C., Jiao, H., Song, D., ... & Morilova, P. (2024). The Rise of Artificial Intelligence in Educational Measurement: Opportunities and Ethical Challenges. *Chinese/English Journal of Educational Measurement and Evaluation*, *5*(3), 3.

27. Walsh, A. (2024). Cusco Quechua and the world of AI: a case study on low resource languages and large language models.

28. Li, S. (2025, April). Trustworthy AI Meets Educational Assessment: Challenges and Opportunities. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 27, pp. 28637-28642).

29. National Test Center. (2024). Sample UNT Kazakh language test questions. Available online at: https://testcenter.kz/upload/iblock/ba5/_-_.pdf (Accessed September 17, 2025).

Baizhanov N., Sabyrov A., Makhmutova A., Kadyrov S.

ІРІ ТІЛ МОДЕЛЬДЕРІНІҢ ҚАЗАҚ ТІЛІНІҢ БІЛІМ ДЕҢГЕЙІН БАҒАЛАУДАҒЫ РӨЛІ

Бұл зерттеу Қазақстандағы Ұлттық бірыңғай тестілеу (ҰБТ) контекстінде қазақ тілін меңгеруді бағалаудағы ірі тіл модельдерінің (LLM) өнімділігін бағалайды. Негізгі мақсат — бес заманауи LLM-нің: Gemini 2.5 Pro Preview, Claude 3.7 Sonnet, Deepseek R1, Qwen және Llama 3.1-405B-Instruct — 2024 жылғы БҰТ қазақ тілі тестінен алынған 138 таңдаулы сұрағындағы дәлдігін, қателіктер сипатын және психометриялық сипаттамаларын зерттеу. Әдістемеге сыртқы деректерге қол жеткізусіз, стандартталған нұсқаулар негізінде нөлдік оқыту режиміндегі бағалауды, сондай-ақ Кохран Q тесті, МакНемар тесттері және модельдердің күрделілік деңгейлері мен лингвистикалық тақырыптар бойынша өнімділігін талдау үшін жалпыланған бағалау теңдеулерімен (GEE) логистикалық регрессияны қоса алғандағы статистикалық талдау кірді. Нәтижелер Gemini-дің ең жоғары дәлдікке (90,6%) қол жеткізгенін және басқа модельдерден айтарлықтай жоғары көрсетеді, ал Llama ең төменгі нәтиже көрсетті (37,7%). Өнімділік күрделілік пен тақырыпқа байланысты өзгеріп отырды, бұл ретте Gemini барлық санаттарда басымдылық танытты, ал басқа модельдер күрделі лингвистикалық ойлау сияқты нақты белгілі бір салаларда өздерінің мықты жақтарын көрсетті. Зерттеу қазақ тілі сияқты ресурстары шектеулі тілдерде білімді бағалау үшін LLM -нің әлеуетін атап көрсетеді, алайда модельдерді оңтайландырудағы, әділдік пен сенімділіктегі олқылықтарды анықтайды, бұл мәдени бейімделген деректерді мақсатты түрде нақтылауды және пайдалануды талап етеді.

Түйін сөздер: Ірі тіл модельдері, Білім беру бағалауы, Қазақ тілі, Бірыңғай ұлттық тестілеу, Педагогикалық өлшеу, Жасанды интеллект.

Baizhanov N., Sabyrov A., Makhmutova A., Kadyrov S.

РОЛЬ КРУПНЫХ ЯЗЫКОВЫХ МОДЕЛЕЙ В ОЦЕНКЕ ВЛАДЕНИЯ КАЗАХСКИМ ЯЗЫКОМ

Данное исследование оценивает производительность крупных языковых моделей (LLM) в оценке владения казахским языком в контексте Единого национального тестирования (ЕНТ) в Казахстане. Основная цель – изучить точность, характер ошибок и психометрические характеристики пяти современных LLM: Gemini 2.5 Pro Preview, Claude 3.7 Sonnet, Deepseek R1, Qwen и Llama 3.1-405B-Instruct – на 138 вопросах с множественным выбором (MCQ) из теста по казахскому языку ЕНТ 2024 года. Методология включала оценку в режиме нулевого обучения с использованием стандартизированных подсказок без доступа к внешним данным, а также статистический анализ, включая тест Кохрана Q, тесты МакНемара и логистическую регрессию с обобщенными оценочными уравнениями (GEE) для анализа производительности моделей по уровням сложности и лингвистическим темам. Результаты показывают, что Gemini достигла наивысшей точности (90,6%), значительно превосходя другие модели, в то время как Llama показала наименьший результат (37,7%). Производительность варьировалась в зависимости от сложности и тематики, при этом Gemini демонстрировала превосходство во всех категориях, а другие модели показывали сильные стороны в специфических областях, таких как сложное лингвистическое мышление. Исследование подчеркивает потенциал LLM для образовательной оценки в языках с ограниченными ресурсами, таких как казахский, но выявляет пробелы в оптимизации моделей, справедливости и надежности, что требует целенаправленной доработки и использования культурно адаптированных данных.

Ключевые слова: Крупные языковые модели, Образовательная оценка, Казахский язык, Единое национальное тестирование, Педагогическое измерение, Искусственный интеллект.

Сведения об авторах:

Байжанов Нурсейт Абсаттарович - РГП на ПХВ «Национальный центр тестирования» МНВО РК, первый заместитель директора, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: nurbaizhanov@gmail.com.

Сабыров Аслан Асхатович - РГП на ПХВ «Национальный центр тестирования» МНВО РК, старший эксперт-аналитик отдела аналитики и искусственного интеллекта, ул. Родниковая, 1/1, 010008, Астана, Казахстан, e-mail: aslansabyrov@gmail.com.

Махмутова Альфира Абдукалыковна - Университет SDU, старший преподаватель, факультет образования и гуманитарных наук, ул. Абылайхана, 1/1, Каскелен, 040900, Казахстан. e-mail: alfira.makhmutova@sdu.edu.kz.

Кадыров Ширали Маратжанович - Новый Узбекский университет, доцент, кафедра общих дисциплин, Мовараннахр, 1, 100000, Ташкент, Узбекистан. e-mail: sh.kadyrov@newuu.uz.

Авторлар туралы мәлімет:

Байжанов Нұрсейт Абсаттарұлы - ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, директордың бірінші орынбасары, Родниковая көш., 1/1, 010008, Астана, Қазақстан, e-mail: nurbaizhanov@gmail.com.

Сабыров Аслан Асхатұлы - ҚР ҒЖБМ «Ұлттық тестілеу орталығы» ШЖҚ РМК, Талдау және жасанды интеллект бөлімінің аға сарапшы-талдаушысы, Родниковая к-сі, 1/1, Астана, 010008, Қазақстан, e-mail: aslansabyrov@gmail.com.

Махмутова Альфира Абдукалыковна - SDU университеті, аға оқытушы, Білім беру және гуманитарлық ғылымдар факультеті, Абылайхан көшесі, 1/1, Қаскелең, 040900, Қазақстан. e-mail: alfira.makhmutova@sdu.edu.kz.

Кадыров Ширали Маратжанович - Жаңа Өзбекстан университеті, доцент, Жалпы білім беру кафедрасы, Мовараннахр, 1, 100000, Ташкент, Өзбекстан. e-mail: sh.kadyrov@newuu.uz.

Information about authors:

Baizhanov Nurseit Absattarovich - 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, First Deputy Director, 1/1 Rodnikovaya St., 010008, Astana, Kazakhstan, e-mail: nurbaizhanov@gmail.com.

Sabyrov Aslan - 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, Senior Data Analyst, 1/1 Rodnikovaya St., Astana, 010008, Kazakhstan, e-mail: aslansabyrov@gmail.com.

Makhmutova Alfira - SDU University, Senior Lecturer, Faculty of Education and Humanities, 1/1 Abylaykhan st, Kaskelen, 040900, Kazakhstan, e-mail: alfira.makhmutova@sdu.edu.kz.

Kadyrov Shirali - New Uzbekistan University, Associate professor, Department of General Education, Movarounnahr, 1, 100000, Tashkent, Uzbekistan, e-mail: sh.kadyrov@newuu.uz.