

УДК 811.512.122'1

На правах рукописи



КУЗЕМБЕКОВА ЖАННА ЖАЛАЙЫРОВНА

**Тестирование чтения на казахском языке:
теория – разработка – апробация
(по стандартам *The Interagency Language Roundtable*)**

Автореферат
диссертации на соискание ученой степени доктора философии (PhD)
по специальности *Лингвистика, общая психолингвистика, билингвистика*

Республика Казахстан
Астана, 2011



Работа выполнена на кафедре иностранной филологии и общего языкознания Казахского национального университета имени аль-Фараби

Отечественный
научный руководитель:

доктор филологических наук
профессор Э.Д. Сулейменова

Зарубежный
научный консультант:

доктор филологических наук
профессор М.Д. Лекич

Официальные рецензенты:

доктор филологических наук
профессор З.С. Кузеева

доктор филологических наук
доцент Г.И. Байгунисова

Защита состоится «18» ноября 2011 года в ____ часов на заседании Диссертационного совета гуманитарных наук – 6D020000 при Евразийском национальном университете им. Л.Н. Гумилева по адресу: 010008, г. Астана, ул. Мунайтпасова, ауд. 302. www.enu.kz

С диссертацией можно ознакомиться в библиотеке «Отырар» Евразийского национального университета имени Л.Н. Гумилева.

Ученый секретарь
доктор филологических наук
профессор

М.Б. Амалбекова

ВВЕДЕНИЕ

Общая характеристика работы. Диссертационное исследование посвящено описанию особенностей тестовых заданий по чтению на казахском языке, выполненных в соответствии со стандартами The Interagency Language Roundtable (ILR) для on-line тестирования языковой компетенции во втором языке.

Основанием для теоретической и практической разработки проблемы тестирования чтения послужили базовые исследования отечественных (Е. Абдрасилова, З. Кузековой, Г.С. Пазыловой и др.) и зарубежных (В.С. Аванесова, А. Анастаси, Г.Ю. Айзенка, Т.М. Балыхиной, В.П. Беспалько, Ю.З. Гильбуха, Н.Ю. Гутаревой, А.Е. Гуревича, Л.П. Добраева, В.А. Коккоты, А.В. Коньшевой, А.Н. Леонтьева, А.Н. Майорова, Г.Ю. Костюка, И. Рапопорта, Р. Сельги, И. Соттера, Л.П. Смирновой, М.Б. Чельшиковой, Д.Б. Эльконина, J.D. Brown, L. Bachman, A. Davies, D.E. Davidson, M.D. Lekic, T. McNamara, J.L.D. Clark, N.E. Gronlund, R. Lado, D.P. Harris и др.) ученых¹. Разработка тестовых заданий по чтению и статистическая проверка их валидности были осуществлены в Отделе тестирования ACTR/ACCELS (The American Council of Teachers of Russian / American Councils for International Education) на основе разработанных им стандартов ILR.

Процесс тестирования состоит из трех обязательных этапов: на начальном – создается задание в тестовой форме, на промежуточном – само тестовое задание, на последнем (после апробации) – тестовое задание включается в тест. Понятия *тест* и *тестовое задание* концептуально и процедурно связаны между собой: тестовые задания различных уровней сложности только после апробации формируют тесты, дифференцированные по степени овладения языковыми навыками.

В настоящее время значительно усилилась роль тестирования как при обучении языку, так и при контроле языковой компетенции, сложилась острая необходимость детального и подробного изучения всего процесса создания и апробации формы и содержания тестовых заданий, а также мониторинга эффективности самого тестирования. Однако, если теоретическим и практическим проблемам тестирования казахской, русской, английской или немецкой языковой компетенции внимание еще уделяется, то этого совершенно нельзя сказать о тестировании знаний по казахскому языку как иностранному.

Актуальность диссертационного исследования определяется несколькими важными моментами: а) до настоящего времени не осуществлен всесторонний научный анализ используемых в стране тестовых систем по

¹ В качестве рабочих используются следующие определения основных понятий: «Тест – система заданий стандартной формы, выполнение которых проходит в равных для всех тестируемых условиях, поддается количественному учету (оценке), позволяет установить уровень сформированности знаний, навыков, умений тестируемого» [1, 83-84]; *тестовое задание* – минимальная единица сложного теста, которая дает ясную и значимую информацию для объективного / субъективного оценивания [2]. Иначе говоря, тестовое задание – это отдельный вопрос (задача, задание), на который тестируемому предлагается дать конкретный ответ и к числу важнейших параметров которого относятся информативность (диагностичность), трудность, надежность, валидность, дискриминативность, достоверность.

определению уровней знаний казахского языка по чтению; б) широко применяемые тестовые задания по казахскому языку, построенные, в основном, по требованиям TOEFL, не учитывают возможностей иных систем (в частности, ILR), а также интернет-технологий; в) существует острая необходимость в разнообразных валидных тестовых системах, востребованных при тестировании казахской языковой компетенции разных категорий граждан, в том числе иностранцев, изучающих казахский язык.

Цель диссертационного исследования состоит в разработке в соответствии со стандартами The Interagency Language Roundtable валидных тестовых заданий (в рамках субтеста *чтение*), ориентированных на выявление уровня языковой компетенции иностранных граждан, изучающих казахский язык.

Для достижения объявленной цели были сформулированы следующие исследовательские задачи:

– систематизировать и обобщить мировой и казахстанский опыт тестирования по определению казахской, английской и русской языковой компетенции по субтесту *чтение*;

– разработать по стандартам The Interagency Language Roundtable пакет тестовых заданий по чтению для разных уровней владения казахским языком как иностранным;

– выполнить анализ правдоподобности дистракторов тестовых заданий, составленных по стандартам ILR для разных уровней;

– используя инструкцию по администрированию тестирования по субтесту *чтение*, разработанную ACTR/ACCELS, осуществить базовое on-line тестирование иностранцев, изучающих казахский язык в Казахстане, предварив его анализом характеристик тестируемых;

– для определения валидности тестовых заданий в сотрудничестве с ACTR/ACCELS обработать результаты тестирования в программах ACTFL и SPSS, а также осуществить статистический анализ полученных данных по тестированию чтения на казахском языке.

Объектом исследования стал пакет тестовых заданий по субтесту *чтение*, разработанный автором в отделе тестирования ACTR/ACCELS в соответствии с американскими стандартами ILR.

Предметом научного осмысления явились содержание, форма, способы разработки и администрирования, валидность и надежность тестовых заданий, ориентированных на определение уровня владения казахским языком (субтест *чтение*).

Материал исследования представлен тремя блоками:

Первый блок – это широко распространенные тестовые системы (TOEFL, TestDaF, IELTS, ТРКИ, КАЗТЕСТ, а также ILR), применяемые для выявления уровня владения вторым языком. Данный материал был использован для сопоставительного анализа тестовых систем.

Второй блок – это тестовые задания в сопровождении с описанием особенностей текстов, использованных для их составления; типы и способы формирования дистракторов к этим заданиям. Пакет тестовых заданий формировался на основе экспертизы 60 единиц, из которых сотрудниками

ACTR/ACCELS отбраковывались или корректировались 'неработающие' дистракторы. Окончательный пакет тестовых заданий составил 40 единиц, использованных в ходе пилотного и базового тестирования для определения разных уровней казахской языковой компетенции (субтест *чтение*).

Третий блок – это статистические данные, полученные по результатам on-line тестирования 25 студентов-иностранцев Казахского национального университета имени аль-Фараби, изучающих казахский язык. 1000 ответов были введены в программу SPSS, 258 правильных ответов составили 25,8% от общего числа.

Основными источниками при составлении тестовых заданий по чтению на казахском языке послужили тексты, размещенные на казахстанских официальных сайтах периодических изданий «Ана тілі», «Қазақ әдебиеті», «Түркістан», «Егемен Қазақстан», «Алаш», а также в других казахскоязычных интернет-ресурсах. Отбор текстов для создания тестовых заданий по субтесту *чтение* осуществлялся исходя из строгих требований ILR к соответствию сложности текста определенному уровню языковой компетенции.

Поставленные цель и задачи исследования обусловили необходимость использования следующих общенаучных и собственно лингвистических методов и приемов исследования:

- в качестве основного был использован метод научного описания, включающий анализ и синтез, наблюдение, систематизацию, селекцию, приемы интроспекции, абстрагирования и идеализации, а также теоретическое осмысление и объяснение;
- сопоставительный анализ применялся для изучения особенностей тестовых систем TOEFL, TestDaF, IELTS, ТРКИ, КАЗТЕСТ, а также ILR;
- контекстуальный анализ был использован для отбора корректного текстового материала для тестовых заданий;
- статистический и корреляционный анализ осуществлялся с использованием программ ACTFL и SPSS;
- ретестовый метод использовался для определения надежности разработанных тестовых заданий.

Новизна диссертационного исследования определяется тем, что впервые:

- для составления тестовых заданий и определения уровня казахской языковой компетенции иностранцев была использована тестовая система ILR;
- разработаны тестовые задания по чтению на казахском языке для иностранцев, осуществлено подробное описание процесса создания, апробации, на основе статистического анализа определена валидность предложенных тестовых заданий по чтению;
- успешно внедрен способ on-line тестирования с использованием американской тестовой системы ILR и интернет-ресурсов;
- внедрены способы проверки надежности и валидности тестовых заданий по чтению на казахском языке.

Основные положения диссертации, выносимые на защиту:

1. Стандарты The Interagency Language Roundtable позволили систематизировать и унифицировать все этапы процесса создания (в соответствии с едиными критериями, примененными как к выбору текстов, так и к формированию тестовых заданий), апробации и статистического анализа тестовых заданий по тестированию чтения на казахском языке.

2. Разработанные тестовые задания (40 единиц) были апробированы при проведении on-line тестирования с использованием американской тестовой системы ILR и интернет-ресурсов. Комплексный анализ тестовых заданий позволил установить их практичность, доступность и обоснованность (в корреляции уровня владения языком с числом правильных ответов), а также надежность и валидность.

3. Статистическая интерпретация пакета тестовых заданий обнаружила соответствие требованиям валидности 45% единиц, т.е. только 18 из апробированных 40 тестовых заданий преодолели необходимый порог корреляционного коэффициента $r_{xy} - 0,400$ и получили соответствующую квалификацию. Представлена дифференцированная оценка валидности тестовых заданий в виде возрастающей шкалы от нижнего корреляционного коэффициента $-0,01125$ до высшего корреляционного коэффициента $0,633691$.

4. Правдоподобность 120 дистракторов определялась статистически в соответствии с общим правилом: сильным считается дистрактор, попавший в группу выше 5%, слабым – оказавшийся в группе, составляющей всего 5% общего числа дистракторов. Выявлено, что правдоподобными являются 26% (32 единицы) дистракторов.

Теоретическая значимость исследования состоит во введении в научный оборот новых материалов, которые могут быть использованы при разработке и конкретизации надежности внедрения компетентностного подхода к оценке уровня владения языком, при определении критериев отбора равноуровневого текстового материала и создании тестовых заданий по субтесту *чтение*, а также в обосновании положений, рассматривающих текст как объект чтения в рамках коммуникативного, прагматического, когнитивного и лингвистического аспектов.

Практическая значимость работы заключается в возможности использования полученных материалов и результатов в подготовке тестовых заданий и администрировании on-line тестирования языковой компетенции, а также в процессе преподавания вузовских базовых и элективных курсов (по лингвистике текста, контрастивной лингвистике, методике преподавания второго языка, теории усвоения второго языка, современному казахскому языку). Возможна экстраполяция полученных эвристических результатов и выводов по созданию, апробации и статистическому анализу тестирования чтения на казахском языке на тестирование чтения на казахском языке носителей иных (в первую очередь, русского) языков, а также на тестирование иных видов речевой деятельности.

Апробация работы. Экспериментально-аналитическая часть диссертации, как и создание тестовых заданий и пилотное тестирование, выполнялись во

время стажировок в Отделе тестирования ACTR/ACCELS (июль-октябрь 2009, июнь-июль 2010 – Вашингтон, США) при непосредственном научном консультировании профессора М.Д. Лекич.

Результаты и основные положения диссертационного исследования докладывались на научных семинарах докторантов PhD, заседании кафедры общего языкознания КазНУ им. аль-Фараби, а также на Международном симпозиуме «Слово в межкультурном и межличностном общении» (Вашингтон, 2009), Международной конференции «Русский язык и литература в международном образовательном пространстве: современное состояние и перспективы» (Гранада, 2010), Международной конференции «Русский язык в современном мире: традиции и инновации в преподавании русского языка как иностранного и в переводе» (Салоники, 2011), XI Международной конференции «Полилингвизм: Язык – Сознание – Культура» (Алматы, 2008), Международном симпозиуме «Тюркский мир: интеграция языка и культуры на рубеже тысячелетий» (Алматы, 2009), Международной конференции «Инновационные технологии в теории и практике преподавания языка и литературы: проблемы и пути решения» (Астана, 2009), XIII Международной конференции «Язык – Общество – Время» (Алматы, 2010), VII Международной конференции «Образование и наука XXI века» (Болгария, 2011).

Результаты исследования нашли отражение в 12 публикациях автора.

Диссертационная работа состоит из введения, двух разделов, заключения, списка использованной литературы и приложений (А, Б, В, Г, Д). Диссертация иллюстрирована 27 таблицами и 19 рисунками.

ОСНОВНАЯ ЧАСТЬ

Во введении обосновываются актуальность темы, цели и задачи, объект и предмет исследования, раскрываются научная новизна, теоретическая и практическая значимость работы, указываются фактический материал, источники и методы исследования, обобщаются теоретические положения, выносимые на защиту.

В первом разделе «Теория и практика тестирования чтения на казахском языке» рассматриваются терминологическая база и теоретические объекты теории тестирования, анализируются мировые и отечественные тестовые системы, уточняются критерии селекции текстов к тестовым заданиям.

В подразделе «Виды тестовых заданий и тестирование чтения» анализируются различные толкования термина *тест* (С.И. Воскерчян, С.Г. Геллерштейн и С.И. Федоров, И.А. Рапопорт, Р. Сельг, И. Соттер, Т.А. Ильин, Э.А. Штульман, И.А. Цатуров, Т.М. Балыхина, В.С. Аванесов и др.), а также систематизируются разнообразные виды тестовых заданий.

Тест как контрольный документ представляет собой систему, которая может делиться на субтесты для тестирования отдельных видов речевой деятельности – аудирования, говорения, чтения, письма.

исторических артефактах, требующих глубокого знания культуры, истории, этнографии, национальной психологии и менталитета.

На общее понимание прочитанного, как правило, влияет такие общие характеристики текста, предписывающие существенные ограничения к их использованию при составлении тестовых заданий: связность и когерентность текста (нельзя допускать отсутствия смысловой связи внутри текста, смысловой незавершенности, излишней компрессии текста, обилия личных местоимений и служебных слов и др.); содержательная сложность текста (нельзя допускать обилия в тексте цифр, исторических дат, статистических данных, имен и фамилий, названий населенных пунктов и т.п.); информационная перенасыщенность фактами, специальной терминологией, абстрактными понятиями и абстрактной лексикой; скучность текстов.

При отборе текста для тестирования необходимо также учитывать уровень общих фоновых знаний тестируемых, и содержание текста не должно быть хорошо знакомо тестируемому.

2. Формально-структурные критерии селекции текста.

К формально-структурным критериям отбора текстов относятся внешнее оформление текста: длина, удобочитаемость (шрифт, длина строки, расстояние между строками и т.д.) и видимость (качество печати, бумаги, яркость шрифта).

К этой же группе можно отнести объем текста. Вместе с тем текст не должен быть слишком простым и легким, тестируемый должен продемонстрировать способность к преодолению трудностей различного характера как показатель владения определённым комплексом элементарных умений, развитие которых осуществляется в работе с текстами.

Обобщая, можно дополнить сказанное следующим списком формально-структурных критериев селекции текстов: обязательное указание источника текста; качество печати (размер, четкость / размытость шрифта); размер или длина текста (количество абзацев, предложений, слов); наличие/отсутствие заголовка; отсутствие выделений в тексте (подчеркиваний, полужирного шрифта, курсива); отсутствие подстрочных примечаний; отсутствие рисунков, таблиц, схем, диаграмм, чертежей и пр.; отсутствие стилистически перегруженных конструкций; соблюдение орфографических и стилистических ортологических норм.

3. Функционально-языковые критерии селекции текста.

К функционально-языковым критериям текста относятся стиль, жанры и типы речи текста, предписывающие ряд ограничений: стиль текста (нельзя использовать специальные научные тексты, например, отрывки из монографий, статей, научных докладов, официально-политические тексты, например, конституций, коммюнике, международных соглашений, официально-деловые тексты, например, указы, постановления, приказы); могут использоваться такие жанры, как реклама, объявление, короткая газетная заметка, научно-популярные тексты, автобиографии, мемуары, новости (местные); ограничений к типам речи нет (это может быть описание, повествование, рассуждение, объяснение, включающее диалоги и монологи и др.).

В подразделе «Роль и особенности дистракторов в тестовых заданиях» обосновывается роль дистракторов в составлении тестовых заданий. В результате комплексного анализа обширного теоретического материала [6, 7, 8], а также собственных выявленных ошибок при составлении тестовых заданий требования к дистракторам были обобщены следующим образом: а) все дистракторы должны быть правдоподобными, убедительными, привлекательными и внушающими доверие, не должны содержать частичную правду; б) внешний вид и грамматическая структура дистракторов в пределах тестового задания должны быть подобными (аналогичными, похожими); в) тестовые дистракторы должны быть краткими и компактными; визуально все дистракторы должны быть, насколько это возможно, одинаковыми по длине, чтобы с первого взгляда, не вникая в смысл ответов, было неясно, какой из них является правильным; г) при формулировке дистракторов желательно избегать использования ключевых слов, так они могут служить подсказкой для тестируемого; д) нельзя использовать такие варианты ответов, как «ни один из перечисленных», «все перечисленные» и другие, провоцирующие угадывание правильного ответа; е) все повторяющиеся слова включать в тестовое задание, но не использовать их в дистракторах; ж) дистракторы должны содержать исключительно тот материал, который представлен в тексте, и не апеллировать к фоновым знаниям тестируемого, или содержать незнакомые слова или термины; з) все дистракторы грамматически и семантически следует согласовывать с основной частью тестового задания; и) формулировка основной части задания, должна быть лаконичной и законченной; к) в ответы целесообразно включать не более двух-трех важных ключевых слов, при этом правильный ответ не должен отличаться какими-либо формальными признаками от неправильных дистракторов; л) из текста задания необходимо исключить все словесные ассоциации, способствующие выбору правильного ответа с помощью догадки и др.

Во втором разделе «On-line тестирование чтения на казахском языке и статистическая обработка результатов» описываются результаты оценки языковой компетенции казахского языка как иностранного студентов-иностранцев, полученные в результате on-line тестирования, а также статистического анализа степени валидности тестовых заданий.

Таблица 2 – Гендерная представленность тестируемых

	Frequency	Percent	Valld Percent
Valid male	20	80,0	80,0
female	5	20,0	20,0
Total	25	100,0	100,0

Из таблицы 2 видно, что выборка тестируемых не была пропорциональной по гендерному признаку: женщины составили лишь 20%, мужчины – 80%. Для того, чтобы узнать, какой гендерной группе удалось удачней избежать ловушек дистракторов, необходимо было определить среднее число правильных

ответов. Оказалось, что на одну женщину приходится – 7,6%, а на одного мужчину – 10,1% правильных ответов.

В таблице 3 продемонстрирована возрастная дифференциация тестируемых, объединенных в три группы:

Первая группа – тестируемые, родившиеся в период с 1962 по 1967 гг.

Вторая группа – тестируемые, родившиеся в период с 1986 по 1989 гг.

Третья группа – тестируемые, родившиеся в период с 1990 по 1992 гг.

Таблица 3 – Возраст тестируемых

	Frequency	Percent	Valid Percent
Valid missing	3	12,0	12,0
1990	6	24,0	24,0
1991	5	20,0	20,0
1962	1	4,0	4,0
1989	4	16,0	16,0
1986	2	8,0	8,0
1988	2	8,0	8,0
1967	1	4,0	4,0
1992	1	4,0	4,0
Total	25	100,0	100,0

Для чистоты эксперимента были отобраны студенты с различной языковой биографией, разной этнической и гражданской принадлежностью: граждане Афганистана – 14, Китая – 3, Кореи – 6, США – 2 человека. Данные о месте рождения представлены в таблице 4.

Таблица 4 – Место рождения

	Frequency	Percent	Valid Percent
Valid China	3	12,0	12,0
Afganistan	14	56,0	56,0
USA	2	8,0	8,0
Korea	6	24,0	24,0
Total	25	100,0	100,0

Для шестерых тестируемых родным языком заявлен дари, для пятерых – корейский, для четверых – фарси, для троих – пушту, для двух – китайский, для одного – английский. Согласно полученным данным, более высокий уровень владения навыками чтения на казахском языке оказался свойствен гражданам Афганистана.

Далее дается обоснование практичности тестовых заданий, под которой в современной тестологии понимают такие качества, как доступность инструкции теста и содержания заданий для понимания; сравнительная простота организации тестирования; возможность проведения теста в обычных для вуза

условиях; сравнительная простота проверки опросных листов, обработки, подсчета результатов теста и подведения итогов.

При оценке надежности и валидности тестовых заданий по субтесту *чтение* определены основные критерии, в соответствии с которыми строятся тестовые задания. О высокой надежности тестовых заданий свидетельствует то, насколько точно тест измеряет конкретное свойство или уровень владения материалом, для измерения которых он и был создан.

Для применения статистических методов была создана матрица³, в которой фиксировались ответы 25 тестируемых на 40 заданий теста по чтению (Приложение В).

Формирование матрицы начинается с выбора определенного правила для оценки ответов тестируемых на задания теста. Обычно результаты ответов оцениваются дихотомически: за каждый правильный ответ тестируемый получает один балл, за неправильный ответ или за пропуск задания – ноль баллов. Если символом x_{ij} обозначить результат выполнения i -м тестируемым j -го задания теста, то в сокращенной форме приведенное выше правило можно записать в виде:

$$x_{ij} = \begin{cases} 1, & \text{если ответ } i\text{-го испытуемого на } j\text{-е задание верный} \\ 0, & \text{если ответ } i\text{-го испытуемого на } j\text{-е задание неверный} \end{cases}$$

После выбора оценочного правила эмпирические данные сводятся в матрицу. Строки матрицы, состоящие из единиц, двоек, троек и четверок, соответствуют ответам тестируемых на различные задания теста. По столбцам располагаются профили ответов тестируемых на каждое задание теста.

Индивидуальный балл тестируемого получается суммированием всех единиц, полученных им за правильные выполненные задания теста. Например, 4-ый тестируемый (ID 8976998) выполнил 15 заданий правильно, поэтому его индивидуальный балл равен $X_4=15$. В строке 25-го тестируемого (ID 8610721) стоят всего две единицы – его индивидуальный балл $X_{25}=2$. Для удобства полученные индивидуальные баллы X_i ($i=1,2, \dots, 25$) приводятся в последнем столбце матрицы результатов. Число правильных ответов на задания R_j также получается суммированием единиц, но уже расположенных по столбцам. Например, в 1-м столбце стоят 6 единиц – число тестируемых, правильно ответивших на 1-е задание. На последнее 40-е задание ответили правильно только двое тестируемых, поэтому $R_{40} = 2$. Число каждого задания также помещается в матрицу результатов, обычно оно располагается в последней строке под номером соответствующего задания теста. Эти же результаты для наглядности продемонстрированы в диаграмме (см. рисунок 3).

Среди тестируемых наиболее слабые результаты были выявлены при выполнении тестовых заданий, в которых содержались вопросы высокого уровня.

³ Матрицей называется компактная форма записи множества элементов, связанных некоторой общностью содержания [7].

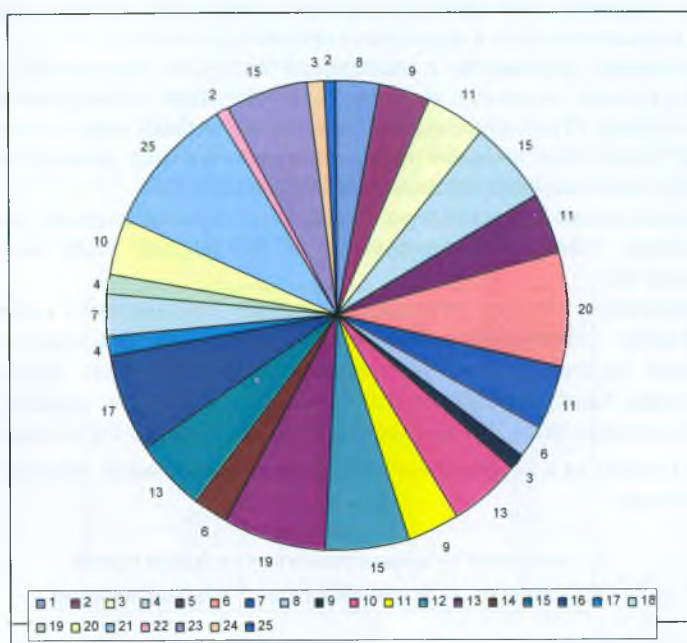


Рисунок 3 – Результаты с индивидуальными баллами тестируемых и количество правильных ответов на задания теста

Результаты тестирования можно увидеть в графическом представлении полигона частот, для построения которого по горизонтальной оси откладываются тестовые баллы, а по вертикальной – частота появления у тестируемых каждого балла.

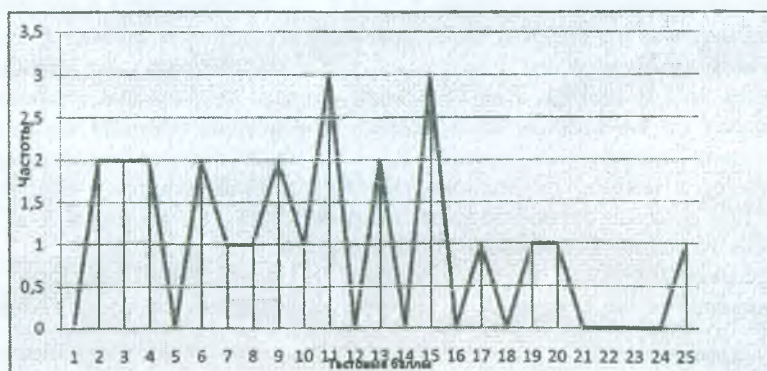


Рисунок 4 – Полигон частот тестовых результатов

В подразделе «Характеристика тестовых заданий» дается рассмотрение расчетных характеристик тестового задания, направленных на выявление и проверку тестовых свойств заданий. По В.С. Аванесову, тестовое задание – это составная единица теста, отвечающая требованиям технологичности, формы, содержания, а также следующим статистическим требованиям: известной трудности; достаточной вариации тестовых баллов; положительной корреляции баллов задания с баллами по всему тесту [7, 163].

Технологичность является одним из важных критериев, предъявляемых к тестовым заданиям. Технологичность предусматривает использование при тестировании технических средств. Кроме того, как отмечает В.С. Аванесов, задания становятся технологичными, если их содержание четко и быстро понимается тестируемыми, а форма заданий способствует процессу компьютеризации тестирования [7, 165].

Трудность тестовых заданий. Наличие определенной меры трудности является обязательным требованием при составлении тестовых заданий. Оценка трудности тестовых заданий проводится по-разному в зависимости от выбранной теории методов обработки эмпирических результатов. Представление о трудности заданий теста дает анализ гистограммы, представленной на рисунке 5.

В классической теории оценка трудности j -го задания исчисляется по формуле

$$p_j = R_j / N \quad (1)$$

где p_j - доля правильных ответов на j -е задание; R_j – количество тестируемых, выполнивших j -е задание правильно; N – число тестируемых в тестируемой группе; j – номер задания теста.

В тесте из N заданий j принимает значения $j=1, 2, \dots, N$. Трудность задания нередко выражают в процентах, тогда формула для оценки трудности имеет вид

$$P_j = R_j / N \cdot 100\% \quad (2)$$

По сложившейся традиции в рамках классической теории тестов за трудность задания принимается именно доля p_j .

Для примера сошлемся на матрицу в Приложении Д, где доля правильных ответов на первое задание

$$p_1 = 6/25 = 0,24,$$

а доля неправильных ответов

$$q_1 = 19/25 = 0,76$$



Рисунок 5 – Гистограмма трудности задания теста

Гистограмма получена после обработки эмпирических данных. На первое задание дали ответ 24% тестируемых, а на второе и третье задания – 36% тестируемых. Четвертое задание оказалось слишком легким: с ним справились 56% тестируемых. Недостаток данного теста заключается в том, что задания более легкого уровня должны предшествовать заданиям повышенной сложности, иначе говоря, трудность тестовых заданий в процессе тестирования должна возрастать постепенно.

Вариация тестовых баллов. Вариация баллов является обязательным требованием к тестовым заданиям. Если на какое-то задание правильно отвечают все тестируемые, то такое задание является неоправданно легким, оно теряет существенное свойство быть тестовым заданием. Поскольку цель разработки теста – измерение уровня подготовленности испытуемых, ее можно достичь только с помощью системы заданий возрастающей трудности. Если же все, без исключения, тестируемые отвечают на задание одинаково правильно, то это означает, что задание не дифференцирует знающих тестируемых от незнающих. Соответственно отсутствуют и вариации в значениях ответов. По данному заданию в матрице будут стоять одни единички.

Нет вариации и в очень трудном задании, на которое нет ни одного правильного ответа: в матрице стоят, соответственно, одни нули. Вариация по нему тоже равна нулю, что означает необходимость и его удаления из проектируемого теста. Такое задание не является тестовым.

Корреляция задания с критерием (r_{xy}) является еще одним важным статистическим требованием, предъявляемым к тестовым заданиям. По мнению В.С. Аванесова, такое требование достигается при корреляции оценок, полученных тестируемыми в задании, с оценками, полученными ими же по кому-либо иному критерию, например, по сумме баллов. Данная корреляция определяет меру дифференцирующей способности задания. Нельзя назвать

тестовое задание таковым, если оно не коррелирует с отмеченной суммой [7, 188].

Для расчета r_{xy} создаются два числовых вектор-столбца, один из них показывает баллы по заданию (X_j), другой – по сумме индивидуальных баллов (Y). Между значениями этих двух векторов и устанавливается мера связи. Формулы для расчета коэффициентов корреляции даются ниже. Проверим, например, меру связи ответов тестируемых по заданию (1) с суммой баллов тестируемых по всему тесту. При расчете последовательно используются четыре формулы:

1) Находится сумма квадратов отклонений баллов тестируемых от среднего арифметического балла в интересующем задании (SS_1).

$$SS_1 = \sum X_1^2 - \frac{(\sum X_1)^2}{N} = 6 - \frac{6^2}{25} = 4.56 \quad (3)$$

2) Следом по приведенной формуле находится сумма квадратов отклонений тестовых баллов тестируемых от среднего арифметического балла по всему тесту (SS_y):

$$SS_y = \sum Y^2 - \frac{(\sum Y)^2}{N} = 3556 - \frac{66564}{25} = 893.4 \quad (4)$$

3) По нижеследующей формуле находится так называемая скорректированная на середине значения сумма попарных произведений X и Y :

$$sp_{xy} = \sum XY - \frac{\sum X \sum Y}{N} = 87 - \frac{6 \cdot 258}{25} = 25.08 \quad (5)$$

В последней формуле $\sum XY$ представляет собой сумму произведений баллов каждого тестируемого по первому заданию и по Y , тестовому баллу тестируемых. Вторая часть формулы представляет собой коррекцию на средние значения по X и по Y .

4) Рассчитывается коэффициент корреляции по формуле:

$$r_{xy} = \frac{sp_{xy}}{\sqrt{SS_x \cdot SS_y}} \quad (6)$$

Подставляя в эту формулы результаты проведенных расчетов, получаем

$$r_{xy} = \frac{sp_{xy}}{\sqrt{SS_x \cdot SS_y}} = \frac{25.08}{\sqrt{4.56 \cdot 893.4}} = 0.397 \quad (7)$$

Чем выше значения r_{xy} , тем больше вероятность превращения задания в тестовой форме в тестовое задание. Корреляционный коэффициент каждого задания получаем с помощью программы Excel.

Анализ полученных результатов выяснил, что только 18 заданий из предложенных 40 соответствуют данному требованию (см. рисунок 6).

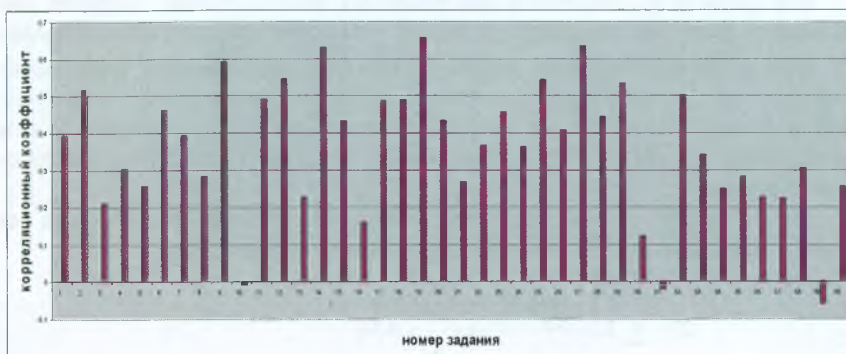


Рисунок 6 – Корреляционный коэффициент тестовых заданий

Статистическая интерпретация пакета тестовых заданий обнаружила соответствие требованиям валидности 45% единиц: только 18 из апробированных 40 тестовых заданий преодолели необходимый порог корреляционного коэффициента r_{xy} – 0,400 и получили необходимую квалификацию.

В подразделе «Анализ правдоподобности дистракторов» представлен статистический анализ ее степени для 1, 1+, 2, 2+, 3 уровней владения языком.

Приведем фрагмент анализа правдоподобности дистракторов для 1-го уровня (см. таблицу 5), к которому было разработано 8 тестовых заданий, на которые дали ответ 25 тестируемых: в первом столбце указаны номера заданий теста, во втором – число тестируемых, выполнивших каждое из заданий первого столбца, все последующие столбцы содержат число тестируемых, выбравших каждый из дистракторов, сопровождающих тестовые задания.

Таблица 5 – Правдоподобность дистракторов (для 1-го уровня)

Номер задания	Всего	Распределение ответов тестируемых			
		1-й ответ	2-й ответ	3-й ответ	4-й ответ
		тест	тест	тест	тест
1	25	6	13	3	3
2	24	9	3	6	6
3	23	9	4	8	2
4	25	14	3	2	6
5	25	5	7	4	9
6	25	9	1	12	3
7	25	9	1	6	9
8	25	12	4	2	7

Прокомментируем отдельные цифры таблицы: 1) Анализ первой строки позволяет получить необходимую информацию о качестве задания первого теста (правильным является 1-й ответ). Из 25 тестируемых шестеро справились с ним успешно, а остальные 19 распределились следующим образом: тринадцать тестируемых выбрали 2-й, трое – 3-й, трое – 4-й дистрактор. Очень похож на правильный 2-й дистрактор: об этом свидетельствует выбор тринадцати (из 19) тестируемых. 2) Ответы на третье задание 23 тестируемых позволяют увидеть, что четвертый дистрактор сформулирован неверно, поскольку в качестве правильного ответа его выбрали всего 2 тестируемых, это является очень низким показателем степени правдоподобности дистрактора. 3) Пятое задание сформулировано корректно, поскольку число выбравших правильные и неправильные ответы оказалось равнозначным. 4) Несомненно, 2-й и 3-й дистракторы в заданиях 6 и 7 нуждаются в переработке: их выбрал только один тестируемый, следовательно, оба дистрактора не работают, а просто создают иллюзию наличия четырех ответов.

Приведем фрагменты анализа дистракторов на 1+ уровне тестовых заданий (см. таблицу 6).

Таблица 6 – Правдоподобность дистракторов (для 1+ уровня)

Номер задания	Всего	Распределение ответов тестируемых			
		1-й ответ	2-й ответ	3-й ответ	4-й ответ
		тест	тест	тест	тест
9	25	12	8	3	2
10	25	12	3	7	3
11	24	5	6	4	9
12	24	11	5	5	3
13	24	3	3	2	16
14	24	9	4	1	10
15	24	12	8	4	0
16	24	4	6	2	12

Анализ полученных данных показывает (комментируются только некоторые данные реферируемого раздела диссертации): 1) На девятое задание правильный ответ был выбран 12 тестируемыми, среди неправильных ответов восемь тестируемых выбрали 2-й, трое – 3-й и двое – 4-й дистрактор. 2) На десятое задание правильно ответили 12 тестируемых, остальные распределились следующим образом: трое выбрали 2-й, семеро – 3-й и трое – 4-й дистрактор. 3) На тринадцатое задание правильно ответили только трое, 2-й дистрактор был выбран тремя, 3-й дистрактор – двумя, 4-й дистрактор шестнадцатью тестируемыми, что позволяет характеризовать 4-й дистрактор тринадцатого задания как очень сильный. 4) На пятнадцатое задание ответили 24 тестируемых, из них правильный ответ выбрали двенадцать, 2-й дистрактор отмечен восемью, 3-й дистрактор – четырьмя, 4-й дистрактор остался не выбранным, следовательно, он нуждается или в переработке, или в удалении.

Аналогичный анализ был осуществлен и для других уровней владения казахским языком. Данные правдоподобности дистракторов для 2-го уровня представлены в таблице 7.

Таблица 7 – Правдоподобность дистракторов (для 2-го уровня)

Номер задания	Всего	Распределение ответов тестируемых			
		1-й ответ	2-й ответ	3-й ответ	4-й ответ
		тест	тест	тест	тест
17	24	6	3	5	10
18	24	4	6	1	13
19	24	9	3	7	5
20	24	4	12	5	3
21	24	6	8	7	3
22	24	8	4	2	10
23	24	5	5	12	1
24	23	6	4	8	5

Анализ правдоподобности дистракторов для 2-го уровня также показал, что при составлении тестовых заданий следует избегать интуитивных оценок, а опираться на статистические данные: 1) На семнадцатое задание ответили 24 тестируемых из 25, при этом только шестеро ответили правильно, трое выбрали 2-й, пятеро – 3-й, десятеро – 4-й дистрактор (т.е. десять из восемнадцати тестируемых неправильно выбрали 4-й дистрактор). 2) На восемнадцатое задание правильный ответ выбрали 4 тестируемых, остальные распределились следующим образом: шестеро выбрали 2-й, один – 3-й и тринадцать – 4-й дистрактор (видим, что 3-й дистрактор очень слабый: его выбрал только один тестируемый). 3) На девятнадцатое задание ответили также 24 тестируемых: правильный ответ выбрали девятеро, трое остановили свой выбор на 2-м, семеро – на 3-м, пятеро – на 4-м дистракторе. 4) На двадцатое и на двадцать первое задание ответы были выбраны тестируемыми корректно: все дистракторы этих заданий были сформулированы точно. 5) Нуждаются в серьезной переработке 3-й дистрактор двадцать второго и 4-й дистрактор двадцать третьего задания.

В реферируемой диссертации анализу правдоподобности дистракторов для тестовых заданий 1, 1+, 2, 2+, 3 уровней владения языком отведено существенное место, поскольку именно дистракторы, с одной стороны, могут вводить в заблуждение тестируемых, содержать в себе подсказку, провоцировать на разного типа ошибки и др., с другой, корректные дистракторы позволяют правильно оценить понимание прочитанного и соответственно уровень владения языком.

Анализ убедил, что правдоподобность дистракторов должна определяться статистически, а не интуитивно, исходя из языковой компетенции и чутья составителя тестовых заданий. Так, несмотря на тщательность подготовки тестовых заданий и их проверку на всех этапах работы (составление, пилотное тестирование, консультации в Отделе тестирования ACTR/ACCELS),

статистический анализ обнаружил, что по-настоящему правдоподобными являются лишь 26% (32 единицы) дистракторов из 120.

Приведенные данные и их интерпретация показывают, насколько серьезной и тщательной должна быть подготовка дистракторов с обязательным сопровождением статистического анализа их правдоподобности.

ЗАКЛЮЧЕНИЕ

Повышение роли тестирования при обучении казахскому языку и при контроле и оценке казахской языковой компетенции послужило предпосылкой комплексного и детального изучения создания, апробации и прогнозирования формы и содержания тестовых заданий и процесса тестирования.

В диссертационном исследовании с учетом содержательных, формально-структурных и функционально-языковых критериев селекции текстов разработан пакет тестовых заданий по казахскому языку для проверки навыков чтения иностранных граждан.

Исследование основано на хорошо зарекомендовавших себя стандартах ILR, которые являются основой создания языковых тестов в США и многих европейских странах. При отборе текстов, создании и апробации тестовых заданий по субтесту *чтение* были использованы соответствующие критерии именно названных стандартов.

1. Осуществлена систематизация и унификация всех этапов создания, апробации и статистического анализа тестовых заданий по тестированию чтения иностранных граждан, изучающих казахский язык в Казахстане.

2. Было отобрано большое число разнообразных аутентичных текстов для первого, второго и третьего уровней владения казахским языком. Соблюдение стандартов ILR сказалось как на тематике всего пакета текстов, так и на ориентированности созданных на их базе тестовых заданий на все элементы каждого уровня владения языком.

3. Американская тестовая система ILR использовалась при создании тестовых заданий, первоначальный пакет которых превышал 100 единиц. Первичный отбор тестовых заданий, а также их последующая селекция и отбраковка на основе пилотного тестирования в Отделе тестирования ACTR/ACCELS позволили минимизировать данный пакет до 40 единиц, которые на следующих этапах были подвергнуты апробации и комплексному анализу.

4. Тестирование испытуемых проходило в on-line режиме при поддержке Отдела тестирования ACTR/ACCELS. On-line тестирование позволило оптимизировать его как по времени, так и по форме прохождения. Кроме того, on-line тестирование позволило оперативно ввести все данные о результатах тестирования в статистические программы для дальнейшего корреляционного статистического анализа.

5. Комплексный анализ тестовых заданий установил их практичность, доступность и обоснованность (в корреляции уровня владения языком с числом правильных ответов), а также надежность и валидность.

6. Статистическая интерпретация пакета тестовых заданий обнаружила соответствие требованиям валидности 45% единиц, т.е. только 18 из апробированных 40 тестовых заданий преодолели необходимый порог корреляционного коэффициента $r_{xy} - 0,400$ и получили соответствующую квалификацию.

7. Была получена дифференцированная оценка валидности тестовых заданий в виде возрастающей шкалы от нижнего корреляционного коэффициента $-0,01125$ до высшего корреляционного коэффициента $0,633691$.

8. Правдоподобность 120 дистракторов была определена статистически в соответствии с общим правилом: сильным считается дистрактор, попавший в группу выше 5%, слабым – оказавшийся в группе, составляющей всего 5% общего числа дистракторов. Выявлено, что правдоподобными являются 26% (32 единицы) дистракторов. Дистракторы в процессе тестирования навыков чтения, как известно, играют ключевую роль. При их создании любой разработчик тестовых заданий должен соблюдать ряд требований, принятых в мировой тестологии. На примере анализа выявленных ошибок в созданных и апробированных тестовых заданиях внесены некоторые уточнения к рекомендациям при составлении тестовых заданий, которые обсуждались в Отделе тестирования ACTR/ACCELS и составили список часто встречающихся нарушений требований и определенных критериев к составлению тестовых заданий. Выполненная работа позволила,

во-первых, увидеть достоинства и недостатки использующихся в стране систем тестирования;

во-вторых, апробировать новую для казахстанской тестологии американскую тестовую систему ILR в on-line режиме;

в-третьих, создать и оценить пакет тестовых заданий с точки зрения строгих статистических требований современной тестологии по таким важнейшим параметрам, как валидность и надежность тестовых заданий, правдоподобность дистракторов;

в-четвертых, увеличить возможности изучения казахского языка как средства международного общения;

в-пятых, приблизиться к решению задач стандартизации контроля усвоения казахским языком (в части государственных требований по определению уровней владения и степени сформированности казахской языковой компетенции) в условиях вариативности и многообразия способов его изучения;

в-шестых, обнаружить перспективы экстраполяции результатов и выводов исследования на создание тестовых заданий по иным видам речевой деятельности разных категорий тестируемых;

наконец, в-седьмых, учитывая национальное своеобразие казахского языка и опираясь на опыт ведущих мировых тестовых систем, приблизиться к более полному описанию особенностей тестовых заданий по субтесту *чтение*.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Балыхина Т.М. Словарь терминов и понятий тестологии. – М., 2000. – 160 с.
2. Brown J.D. Testing in language programs: a comprehensive guide to English language assessment. – Hill, 2005. – 324 p.
3. Якушева Л.С. Создание тестовых заданий для контроля умений и навыков учащихся при обучении иностранному языку в школе // <http://www.google.kz/search?scIient=psvab&hl=ru&source=hp&q=yamkino45.na.rod.ru%2F2010->
4. Кожевникова Л.А. Определение конструкта при составлении тестов по чтению: Дис... канд. филол. наук – М., 2009. – 262 с.
5. Носонович Е.В. Методическая аутентичность учебного текста: Дис... канд. пед. наук. – Тамбов, 1999. – 197 с.
6. Чельщикова М.Б. Теория и практика конструирования педагогических тестов – М., 2002. – 423 с.
7. Аванесов В.С. Композиция тестовых заданий. – М., 2002. – 240 с.
8. Майоров А.Н. Теория и практика создания тестов для системы образования. – М., 2000. – 352 с.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ДИССЕРТАЦИИ

1. Некоторые особенности казахских тестовых заданий по аудированию для англоговорящих // Международный симпозиум «Слово в межкультурном и межличностном общении». – Вашингтон, США, 2009. – С. 62-63.
2. Педагогический тест как форма проверки знания // Международный симпозиум «Тюркский мир: интеграция языка и культуры на рубеже тысячелетий». – Алматы, 2009. – С. 254-258.
3. Тестовые системы в определении уровней владения вторым языком: к проблеме соответствия // «Тілтаным». А. Байтұрсынұлы атындағы Тіл білімі институты. – Алматы, 2010. – № 4. – С. 208-213.
4. О применении стандартов ACTFL для тестирования казахской языковой компетенции (Reading Comprehension) // Международная научная конференция «Инновационные технологии в теории и практике преподавания языка и литературы: проблемы и пути их решения». – Астана, 2009. – С. 296-300.
5. Оқылым тест тапсырмаларын құрастырудағы негізгі қағидалар // XII Международная конференция «Язык – Общество – Время». – Алматы, 2010. – С. 63-67.
6. Разработка тестовых заданий по субтесту *чтения* (по дескрипторам ILR) // Вестник КазНУ. Серия филологическая. – Алматы, 2011. – №2 (132). – С. 313-316.
7. Критерии отбора текстов для тестирования чтения // Русский язык и литература в международном образовательном пространстве: состояние и перспективы. – Испания, Гранада, 2010. – С. 495-500.

8. Тест тапсырмаларының тұтастығы // Вестник КазНУ. Серия филологическая. – Алматы, 2011. – № 2 (132). – С. 330-333.
9. Роль дистракторов в процессе создания тестовых заданий // Международная научно-практическая конференция «Русский язык в современном мире: традиции и инновации в преподавании русского языка как иностранного и в переводе. – Греция, Салоники, 2011. – С. 385-389.
10. Результаты тестирования и анализ правдоподобности дистракторов (по субтесту *чтения*) // Вестник КазНУ. Серия филологическая. – Алматы, 2011. – №3 (133). – С. 265-270.
11. On the Use of ILR Standards for Testing of Kazakh Language Competence in Reading Comprehension // Международная научная конференция «Образование и наука XXI века». – Болгария, София, 2011. – С. 8-13.
12. Интерпретация тестовых результатов on-line тестирования // Международная научная конференция «Образование и наука XXI века». – Болгария, София, 2011. – С. 16-21.

Күзембекова Жанна Жалайырқызы

**ҚАЗАҚ ТІЛІНДЕГІ ОҚЫЛЫМДЫ ТЕСТІЛЕУ:
ТЕОРИЯ – ЖАСАЛЫМ – АПРОБАЦИЯ
(THE INTERAGENCY LANGUAGE ROUNDTABLE СТАНДАРТЫ НЕГІЗІНДЕ)**

Лингвистика, жалпы психолингвистика, билингвистика мамандығы бойынша философия докторы (PhD) ғылыми дәрежесін алу үшін дайындалған диссертацияның авторефератына

ТҮЙІН

Зерттеу тақырыбының өзектілігі. Тілді оқытудағы ең өзекті мәселенің бірі игерілген білімді тексеру және ғылыми нәтиженің шынайылығына көз жеткізу. Қазіргі таңда тестілеу арқылы өзге тілді игеру нәтижелеріне объективті, нақты, дәл қол жеткізуге болады. Бұл қазақ тілін меңгеруші әр түрлі ұлт өкілдері үшін тестілеудің жаңа мазмұнын, сонымен қатар, тестілеу үдерісінің барлық деңгейлерін анықтауды көздейді.

Зерттеу нысаны – оқылым бойынша ACTR/ACCELS тест бөлімінде ILR стандарты негізінде жасалған тест тапсырмалары.

Зерттеу пәні – қазақ тілін меңгеру деңгейлерін анықтауға бағытталған оқылым бойынша тест тапсырмаларының мазмұны, формасы, жасалу жолдары, тест тапсырмаларының сенімділігі және талаптарға жарамдылығы.

Зерттеудің мақсаты – ILR стандарты негізінде өзге ұлт өкілдерінің қазақ тілін меңгеру деңгейін бақылауға бағытталған оқылым негізіндегі тест тапсырмаларын жасап, олардың сенімділігі мен талаптарға жарамдылығын анықтау. Осы мақсатқа сай төмендегі міндеттерді шешу көзделді:

- а) оқылым бойынша әлемдік және қазақстандық тест жүйелерін сараптау;
- б) ILR стандарты бойынша тест тапсырмаларын қазақ тілін өзге тіл ретінде әр түрлі деңгейлер бойынша жасау;
- в) әр түрлі деңгейлерге арналған тест тапсырмаларына арналған дистракторлардың шынайылығына талдау жасау;
- г) ACTR/ACCELS орталығында жасалған тест тапсырмаларын қазақ тілін меңгерген американдықтарға алдын ала өткізу арқылы базалық on-line тестілеу жүйесін өзге ұлт өкілдеріне өткізу;
- д) ACTR/ACCELS орталығының тест бөлімімен бірлесіп, тестілеу нәтижелеріне қол жеткізу, сонымен қатар жасалған тест тапсырмаларына SPSS бағдарламалары негізінде талдау жүргізу.

Диссертациялық зерттеудің ғылыми жаңалығы. Ең алғаш рет өзге ұлт өкілдерінің қазақ тілін меңгеру деңгейін анықтауға бағытталған оқылым негізіндегі тест тапсырмаларын жасауда ILR стандартын қолдану; оқылым бойынша тест тапсырмаларын жасап, тестілеу үдерісінің жүйелі сипатталуы, сондай-ақ алғаш рет интернет жүйесін пайдаланып, тест тапсырмаларының сенімділігі мен талаптарға жарамдылығын анықтаудың жолдарын енгізу.

Қорғауға ұсынылатын негізгі тұжырымдар.

1. The Interagency Language Roundtable стандарты қазақ тілінде оқылым бойынша жасалған тест тапсырмаларының нәтижелерін, барлық кезеңдерін (мәтін іріктеуде, тест тапсырмаларын жасауды бірізді қағидалар негізінде) жүйелендірді.

2. Америкалық PR стандарты бойынша жасалған тест тапсырмалары (40 бірлік) on-line тестілеу барысында апробациядан өтті. Тест тапсырмаларының кешенді талдануы олардың тиімділігін, қол жетімділігін, негізділігін, сонымен қатар сенімділігі мен талаптарға жарамдылығын анықтауға мүмкіндік берді.

3. Тест тапсырмаларының статистикалық талдануы оқылым незіндегі тест тапсырмаларының 45%-ның, яғни 18 (бірлік) тест тапсырмасының қажетті корреляциялық коэффициентіне $r_{xy} = 0,400$ ие болып, талаптарға жарамды екені анықталды.

4. Дистракторлардың шынайылығы, яғни талаптарға жарамдылығына жасалған талдау негізінде 120 дистрактордың 26%-ы (32 бірлік) іріктеліп алынды.

Зерттеу жұмысының теориялық маңызы алынған нәтижелер мен ұсынылатын тұжырымдар негізінде тіл меңгеру деңгейін бағалау барысындағы қадамдар түсінігін нақтылады. Сонымен қатар, мәтінді оқылым негізінде коммуникативтік, когнитивтік, прагматикалық және лингвистикалық аспектілер аясында қарастырды.

Зерттеу жұмысының практикалық маңызы алынған нәтижелерді келешекте өзге ұлт өкілдеріне оқылым бойынша жасалатын қазақ тіліндегі тест тапсырмаларын құрастыруда және on-line тестілеу үдерісінде жүйелендіруде, сонымен қатар контрастивтік лингвистика, екінші тілді меңгеру, қазіргі қазақ тілін меңгеру дәрістерін оқыту барысында қосымша материал ретінде пайдалануға болады.

Жұмыстың аналитикалық бөлімі, сондай-ақ тест тапсырмаларын жасау АСТР/ACCELS бөлімінде ғылыми тағлымдамадан өту барысында орындалды (шілде-қазан 2009, маусым-шілде 2010 – Вашингтон, АҚШ).

Жұмыстың негізгі нәтижелері 12 ғылыми мақалада жарық көрді.

Диссертациялық жұмыстың құрылымы қойылған мақсат пен міндеттермен анықталған. Зерттеу жұмысы кіріспеден, екі бөлімнен, қорытындыдан және пайдаланылған әдебиеттер тізімі және қосымшалардан тұрады.

SUMMARY

of the thesis submitted for the Scientific Degree of Doctor of Philosophy (PhD)
Specialization: *Linguistics, general psycholinguistics, bilingualism*

Kuzembekova Zhanna Zhalaiyrovna

TESTING READING IN THE KAZAKH LANGUAGE: THEORY - DEVELOPMENT – APPROBATION (BY THE STANDARDS OF *THE INTERAGENCY LANGUAGE ROUNDTABLE*)

The topicality of this research. One of the central problems of modern foreign language teaching is to improve the monitoring and verification of knowledge. Technology of an estimation of foreign language acquisition, based on testing, meets modern requirements to objectivity, impartiality and accuracy. It leads to searching new forms and contents of all stages and procedures of testing of the Kazakh language acquisition by different groups of the population.

The object of the research was a package of test items in reading subtest, designed by the author in the testing department ACTR/ACCELS in accordance with American standards ILR.

The subject of the research was the content, form, methods of administration and development, validity and reliability of tests, aimed at determining the level of the Kazakh language acquisition (reading subtest).

The purpose of the research is to develop valid test items (in the reading subtest) in accordance with the standards of The Interagency Language Roundtable, focused on validation of knowledge and identification of the level of linguistic competence of foreign citizens studying the Kazakh language. To achieve the declared goals the following research objectives were formulated: a) to systematize and generalize the world and Kazakhstani experience of testing to determine the Kazakh, Russian and English languages competence by reading subtest; b) to develop a package of test items in reading according to the standard of The Interagency Language Roundtable for different levels acquisition of the Kazakh language as a foreign language; c) to analyze the plausibility distracter test items drawn on standard ILR for different levels; d) to conduct basic on-line testing with foreigners studying the Kazakh language in Kazakhstan, applying the instructions for administering tests in reading subtest developed by ACTR/ACCELS and supplying it with the analysis of the test participants' characteristic features; e) to process the results of testing by programs SPSS and to implement statistical analysis of the data on testing reading in the Kazakh language in cooperation with the ACTR/ACCELS to determine the validity of the test tasks.

The novelty of the research is determined by the fact that for the first time: a) testing system ILR was applied to produce test items and determine the foreigners' level of the Kazakh language acquisition; b) test items in reading in the Kazakh language for foreigners were developed, a detailed description of the process of

creating, testing were carried out, on the basis of statistical analysis the validity of test items in reading were determined; c) on-line testing system of the American ILR and the Internet were implemented successfully; d) ways of checking the reliability and validity of test items in reading in the Kazakh language were implemented.

Conclusions to the defense:

1. Standards The Interagency Language Roundtable allowed to systematize and unify all stages of process of creation, approbation and the statistical analysis of test tasks on reading testing in the Kazakh language.

2. The developed test tasks (40 units) were approved at carrying out on-line testing with using of American test system ILR and Internet resources. Complex analysis of test items allowed to establish their practicality, accessibility and validity (correlation of the level of proficiency with the number of correct answers), as well as reliability and validity.

3. Statistical interpretation of a package of test tasks has found out conformity to validity requirements only 45 % of units, i.e. only 18 from the approved 40 test tasks have overcome a necessary threshold of correlation factor $r_{xy} - 0.400$ and have received qualification of test tasks.

4. Plausibility 120 distracters was defined statistically according to the general rule: a strong distracter is considered that one which got to group distracters above 5 %, a weak one appeared in the group making only 5 % of total number distracters. It is revealed that 26 % (32 units) distracters are plausible.

The theoretical significance of the research consists of introducing new materials to the scientific use that can be used in designing and specification of reliability of implementing competence-based approach to assessing level of language acquisition, in determining the criteria for selection of varied leveled texts and in creating test tasks in reading subtest, as well as in justifying rules considering the text as an object of reading within a communicative, pragmatic, cognitive and linguistic aspects.

The practical significance of the research lies in the possibility of using the materials and results in the preparing test items and administrating on-line test of linguistic competence, as well as in the teaching basic and elective courses at high school. The heuristic results and conclusions can be extrapolated from creating, testing and statistical analysis of test reading in the Kazakh language to testing reading in the Kazakh language for speakers of other languages.

Approbation of the research. The analytical part of the thesis, as well as the creation of test items during the internship in the Department of Testing ACTR / ACCELS (July-October 2009, June-July 2010 - Washington, USA).

Results of the study were reflected in 12 publications by the author.

Structure of the research consists of an introduction, two chapters, conclusions, bibliography.

КУЗЕМБЕКОВА ЖАННА ЖАЛАЙЫРОВНА

**Тестирование чтения на казахском языке:
теория – разработка – апробация
(по стандартам *The Interagency Language Roundtable*)**

Автореферат
диссертации на соискание ученой степени доктора философии (Ph.D.)
по специальности *Лингвистика, общая психолингвистика, билингвистика*

Подписано в печать 11.11.2011 г.
Формат 60 x 85 1/16
Бумага «Multilaser». Печать – RISO
Гарнитура «Таймс». Объем 1, 5 п.л. Тираж 50 экз.
Заказ № 237. обложка – Colorit.

Отпечатано в типографии «Эрекет-Принт»
050036, г.Алматы, 12 мкр., д.16, кв.69
Тел. 221-84-55