

5. Петракова Л.Г. Особенности профессионально-ориентированного обучения русскому языку и культуре речи студентов лингвистических специальностей. - Воронежский экономико-правовой институт: Территория науки. - 2014. - № 1., с.20

References

1. Antonova I.S. O sootnoshenii ponáti «iazykovoie soznanie» i «professionalnoe iazykovoie soznanie» // Sosiosfera.- 2010. - №4., s. 52-57. http://psyjournals.ru/sociosfera/2010/n4/35857_full.shtml.)
2. Batraeva O. M. Professionalno-kommúnikativnye kompetensu inostrannyh stýdentov tehnikeskogo výza i vozmojnosti ih formirovaniá v prosese obýchenia iazyký speshialnosti // Naýchno-metodicheski elektronnyi jýrnal «Konsept». – 2014. – T. 26. – S. 256–260. – URL: <http://e-koncept.ru/2014/64352.htm>.
3. Búzen T. «Sýperмышlenie», - Izdatelstvo: Popýrri, 2003, - 304 s.
4. Koreneva A.V. Professionalno orientirovannoe obýchenie rechevoi deiatelnosti stýdentov-nefilologov na osnove mejdisiplinarnoi integrasii v kýrse «Rýsski iazyk i kúltýra rechi». Avtoref. dis. doktora ped. náyk. Orel, 2010. – 48 s.
5. Petrakova L.G. Osobennosti professionalno-orientirovannogo obýchenia rýsskomú iazyký i kúltýre rechi stýdentov nelingvisticheskikh speshialnostei. - Voronejski ekonomiko-pravovoi institút: Teritoria náyk. - 2014. - № 1., s.20

МРНТИ 10.21

10.51889/6328.2022.81.26.004

*Алтыбаева Ш.Б.,¹ Шинетова Л.Е.²

^{1,2}РГКП «Национальный центр тестирования»
Министерства образования и науки Республики Казахстан,
Нур-Султан, Казахстан

ОЦЕНКА НАДЕЖНОСТИ ТЕСТА НА ОПРЕДЕЛЕНИЕ ГОТОВНОСТИ К ОБУЧЕНИЮ ПО РЕЗУЛЬТАТАМ АПРОБАЦИИ

Аннотация

Данная статья посвящена анализу практики проведения комплексного тестирования поступающих на программы послевузовского образования в Республике Казахстан. В статье рассматриваются параметры критерия поступления в магистратуру, а именно, теста на определение готовности к обучению будущих магистрантов, который оценивает способности претендентов к исследовательской деятельности.

Цель исследования - с помощью программных обеспечений, которые позволяют оценить результаты в шкале логитов, переводя значения баллов в значения уровня подготовленности тестируемых, выявить эффективность использования теста на определение готовности к обучению при поступлении в магистратуру. Данные в исследовании получены с участием 2464 студентов выпускных курсов вузов в 2020-2021 учебном году.

Ключевые слова: Апробация, тест, тестовые задания, анализ, RUMM2020

*Алтыбаева Ш. Б.,¹ Шинетова Л. Е.²

^{1,2} "Ұлттық тестілеу орталығы" РМҚК
Қазақстан Республикасы Білім және ғылым министрлігінің,
Нұрсұлтан, Қазақстан

АПРОБАЦИЯ НӘТИЖЕЛЕРІ БОЙЫНША ОҚУҒА ДАЙЫНДЫҒЫН АНЫҚТАУҒА АРНАЛҒАН ТЕСТ СЕНІМДІЛІГІН БАҒАЛАУ

Аңдатпа

Бұл мақала Қазақстан Республикасында жоғары оқу орнынан кейінгі білім беру бағдарламаларына түсушілерге кешенді тестілеу өткізу тәжірибесін талдауға арналған. Мақалада магистратураға түсу критерийінің параметрлері, атап айтқанда үміткерлердің зерттеу қызметіне қабілеттерін бағалайтын болашақ магистранттарды оқытуға дайындығын анықтауға арналған тест қарастырылады.

Зерттеудің мақсаты - логит шкаласындағы нәтижелерді бағалауға, балл мәндерін тестіленушілердің дайындық деңгейінің мәндеріне аударуға, магистратураға түсу кезінде оқуға дайындықты анықтауға арналған тестті қолданудың тиімділігін анықтауға мүмкіндік беретін бағдарламалық жасақтаманың көмегімен. Зерттеу деректері 2020-2021 оқу жылында жоғары оқу орындарының бітіруші курстарының 2464 студентінің қатысуымен алынды.

Түйін сөздер: Аprobация, тест, тест тапсырмалары, талдау, RUMM2020

*Altybaeva Sh.,¹ Shinetova L.²

^{1,2} RSE "National Testing Center"
Ministry of Education and Science of the Republic of Kazakhstan,
Nursultan, Kazakhstan

ASSESSMENT OF THE RELIABILITY OF THE TEST TO DETERMINE READINESS FOR TRAINING BASED ON THE RESULTS OF APPROBATION

Abstract

This article is devoted to the analysis of the practice of conducting comprehensive testing of applicants for postgraduate education programs in the Republic of Kazakhstan. The article discusses the parameters of the admission criterion for a master's degree, namely, a test to determine the readiness for training of future undergraduates, which evaluates the applicants' abilities for research activities.

The purpose of the study is to use software that allows you to evaluate the results in the logit scale, translating the score values into the values of the level of readiness of the test takers, to identify the effectiveness of using the test to determine readiness for training when enrolling in a master's degree. The data in the study were obtained with the participation of 2,464 university graduate students in the 2020-2021 academic year.

Keywords: Approbation, test, test tasks, analysis, RUMM2020

Введение. Одним из определяющих факторов конкурентоспособности государства является уровень образования и науки. Степени магистра и доктора PhD являются наивысшими степенями в науке и образовании Казахстана. Главное предназначение

послевузовского образования – подготовка педагогических кадров для ведения образовательной деятельности в высших учебных заведениях, а также научных кадров. Поскольку комплексное тестирование в магистратуру тест высокой важности, по результатам которого распределяются государственные гранты, необходимо применять лучшие измерители оценки, чтобы подобрать наиболее подготовленных претендентов к исследовательской деятельности.

Одним из параметров критерия поступления в магистратуру и докторантуру PhD является результат теста на определение готовности к обучению [1].

Материалы и методы исследования. Структура теста на определение готовности к обучению (далее – ТГО) повышает эффективность оценивания уровня знаний претендентов и обеспечивает равнозначные по трудности варианты тестов. Тест содержит 30 заданий. Максимальное количество баллов - 30.

Все тестовые задания закрытой формы. Тест состоит из 2-х частей:

- критическое мышление - 15 тестовых заданий.
- аналитическое мышление - 15 тестовых заданий.

Представленная структура теста ТГО позволит выявить претендентов, наиболее подготовленных к исследовательской деятельности.

Секция «Критическое мышление» состоит из математических заданий, соответствующих уровню средней школы. В ней проверяются способности критически рассуждать, решать элементарные математические проблемы, анализировать и интерпретировать информацию, представленную в графическом и табличном виде, консолидировать различные типы информации. Секция «Критическое мышление» раздел «Сравнение значений» состоит из трех заданий, в которых вычисляются два значения, а значение каждого находится и сравнивается на основе информации, предоставленной в задании. Во второй части расположены вопросы на уравнения, неравенства, геометрические задачи, текстовые задачи, состоящая из восьми заданий. Это широко известный формат, в котором необходимо сделать расчет, после выбрать один правильный вариант ответа из представленных вариантов. Третья часть - Таблица, диаграмма, схема, текст - представлена в четырех тестовых заданиях. В ней делается анализ данных и на их основе выводы.

В секцию «Аналитическое мышление» входит проверка способности понимать, анализировать академические и научно-популярные тексты различной сложности, определять главную мысль текста, находить причинно-следственные связи и выявлять скрытые закономерности. Первые три задания в секции «Аналитическое мышление» заключаются в заполнении пропусков в тексте. Задания предназначены для оценки способности осмыслить неполное предложение. Вторая часть - Понимание и анализ текста - состоит из одного текста, к которому представлены девять тестовых задания. Также есть задания, где нужно выявить общую точку зрения из двух текстов, к которому представлены три тестовых задания. Тексты заимствованы из научных дисциплин, которые варьируются от физики и биологии до социальных и гуманитарных наук. Третья часть - Анализ информации на основе различных источников - предлагает три тестовых задания. В этой части оценивается способность восприятия актуальной информации из различных источников.

В целях проверки возможности применения ТГО при проведении комплексного тестирования в магистратуру проводится апробация тестовых заданий (далее - апробация).

Цель достигается посредством анализа статистических характеристик как апробируемых вариантов в целом (надёжность, распределение трудности заданий, корреляционные характеристики, информационные кривые и др.), так и отдельных заданий (трудность, дифференцирующая способность и др.), результатов анализа экспертных характеристик апробируемых заданий (корректность формулировок, значимость для комплексного тестирования в магистратуру). Также результаты теста всегда исследуются на

его валидность – действительно ли тест определяет то, что от него требовалось. На основе результатов анализа проводится доработка заданий, отдельные задания могут отбраковываться. Далее формулируются выводы о возможности применения ТГО в комплексном тестировании в магистратуру.

Результаты и обсуждение. Апробация включает диапазон действий: подготовка и организация исследования, включая время, отбор вузов, которые будут участвовать в апробации, определение технологии (электронный или бумажный формат) процедуры и обработка результатов апробации; предъявление теста и подготовка данных к анализу, подготовка отчётов и принятие решений на основе проведённого исследования.

Надежной мотивацией для участников апробации является возможность ознакомления студентов с процедурой оценивания и форматом тестирования. А для вузов, позволяющих проводить апробацию, это возможность своевременно получать информацию о потенциале обучения студентов в магистратуре. Мотивированность обеспечивает четкие ответы участников, при этом студентам не будут безразличны результаты их работы, а также их участие в апробации.

В апробации приняли участие студенты выпускных курсов высших учебных заведений. Общее количество участников апробации 2464 студентов. Объем выбора позволил сделать выводы о трудности и качестве заданий по ТГО. Согласно структуре теста был разработан вариант, состоящий из 30 тестовых заданий с выбором одного правильного ответа. При этом не менее 400 участников апробации ответили на каждый вариант.

Анализ результатов апробации проводился при помощи программ RUMM2020 (Rasch Unidimensional Measurement Models) и Winstep, которые позволяют оценить результаты в шкале логитов, переведя значения баллов в значения уровня подготовленности тестируемых. А также данные программы позволяют измерить по одной и той же шкале (в логитах) уровень подготовленности студентов и трудность тестовых заданий [2].

Были обработаны матрицы результатов апробации. Для анализа использовались варианты по ТГО с наибольшим числом тестируемых. Для применения программы необходимы объекты с большой выборкой, с целью получения более достоверной картины.

Анализ качества заданий проводился по следующим аспектам:

- определение экстремальных заданий;
- характеристика измерительного инструмента выбранного набора;
- определение соответствия тестового задания с моделью измерения;
- определение соответствия между уровнем подготовленности студентов и трудностью тестовых заданий;
- диапазон варьирования трудности заданий и равномерность распределения заданий по трудности.

Определение экстремальных заданий.

Задания, трудность которых нельзя измерить при использовании модели Раша, или задания на которые все испытуемые отвечали либо правильно, либо неправильно называются экстремальными заданиями. Такие задания в тест не включаются. А также если выявляются испытуемые, способные правильно или не правильно решать все задания, то такие испытуемые также называются экстремальными. Такие испытуемые исключаются из дальнейшего анализа тестовых результатов, как несоответствующие данному уровню трудности теста и уровню подготовленности испытуемых. В качественных педагогических измерениях уровень подготовленности испытуемых должен соответствовать уровню трудности заданий [3].

В исследуемых матрицах экстремальных заданий не оказалось. Средний балл по тесту составил 20,3. Минимальный набранный балл - 2,0 максимальный набранный балл - 29,0.

Характеристика измерительного инструмента выбранного набора.

Рассмотрим статистические характеристики заданий теста, чтобы оценить согласие тестового задания с моделью измерения.

В таблице 2 приведены статистические характеристики тестовых заданий ТГО. Статистика соглашения отображается в последних двух столбцах таблицы.

Простая статистика согласия более чувствительна к экстремально неожиданным ответам, когда сильный испытуемый неожиданно неправильно отвечает на легкое задание или, наоборот, слабый испытуемый неожиданно правильно отвечает на трудное задание. Взвешенная статистика позволяет уменьшить влияние экстремально неожиданных ответов. Допустимым интервалом для статистик согласия является (0,5; 1,5).

В столбцах 2 и 3 таблицы представлены оценки трудности заданий и соответствующие ошибки измерения (в логитах). В следующем столбце приведен коэффициент корреляции между баллами по заданию и уровнем подготовленности испытуемых.

Таблица 2. Статистические показатели тестовых заданий ТГО

Номер задания	Оценка Трудности (Measure)	Ошибка Измерения (SE)	Коэффициент корреляции	Статистики согласия	
				Взвешенная (OUTFIT MNSQ)	Простая (INFIT MNSQ)
Item30	-2,60	0,27	0,37	0,84	0,79
Item25	-2,34	0,25	0,30	0,79	0,97
Item23	-2,02	0,22	0,38	0,77	0,85
Item16	-1,93	0,21	0,37	0,87	0,86
Item24	-1,80	0,2	0,32	1,31	0,89
Item18	-1,45	0,18	0,30	1,13	0,99
Item28	-1,33	0,17	0,30	1,27	1,03
Item26	-1,11	0,16	0,31	1,11	0,97
Item29	-0,95	0,15	0,32	0,99	1
Item22	-0,91	0,15	0,33	1,18	0,98
Item27	-0,88	0,15	0,34	1,1	0,96
Item17	-0,70	0,14	0,24	1,15	1,11
Item19	-0,64	0,14	0,34	1,05	0,99
Item21	-0,46	0,13	0,30	1,02	1,06
Item12	0,19	0,11	0,42	0,98	0,93
Item3	0,24	0,11	0,32	1,33	1,03
Item8	0,32	0,11	0,45	0,9	0,92
Item4	0,48	0,11	0,45	1,05	0,92
Item14	0,52	0,11	0,41	0,94	0,98

Item11	0,60	0,11	0,43	0,92	0,95
Item7	0,84	0,11	0,47	0,93	0,91
Item13	0,87	0,11	0,39	1,06	0,99
Item20	1,23	0,11	0,35	1,06	1,05
Item9	1,42	0,11	0,43	1,04	0,95
Item6	1,49	0,11	0,34	1,24	1,04
Item5	1,61	0,11	0,31	1,25	1,12
Item1	1,68	0,11	0,41	1,1	0,97
Item10	1,88	0,11	0,45	0,93	0,93
Item2	2,34	0,12	0,34	1,87	1
Item15	3,41	0,16	0,18	1,34	1,16

В данной таблице задания расположены в порядке возрастания их реальной трудности.

Согласно таблице 1, в данном тесте Item30 задание является самым легким, его трудность составляет -2,60 логит, а задание Item15 является самым трудным со значением 3,41.

Только два задания имеют коэффициент корреляции ниже значения 0,3 (задания Item15 и Item17, коэффициент соответственно 0,18 и 0,24). Все задания статистики согласия входят в указанный интервал.

Определение соответствия тестового задания с моделью измерения.

Для определения соответствия все испытуемые делятся на три группы по уровню подготовленности. Для каждой группы с низким, средним и высоким уровнем подготовленности вычисляется как среднее (экстремальное) и ожидаемое значение на основе модели Раша. По критерию Хи-квадрат определяется соответствие экстремального значения ожидаемому значению. Если уровень значимости статистики Хи – квадрат для задания меньше чем 0,05, то это задание плохо совместимо с остальными заданиями [4]. Такие задания рекомендуется исключать из теста.

Рассмотрим характеристические кривые заданий теста, чтобы оценить соответствие тестового задания с моделью измерения.

Отдельно рассмотрим задания, находящиеся в согласии с моделью, и задания, не находящиеся в согласии с моделью.

В рисунках с характеристическими кривыми по оси абсцисс отложены значения латентных переменных (в логитах) уровня подготовленности тестируемого и по оси ординат откладывается вероятность ожидаемого ответа тестируемых, которая варьирует от 0 до 1.

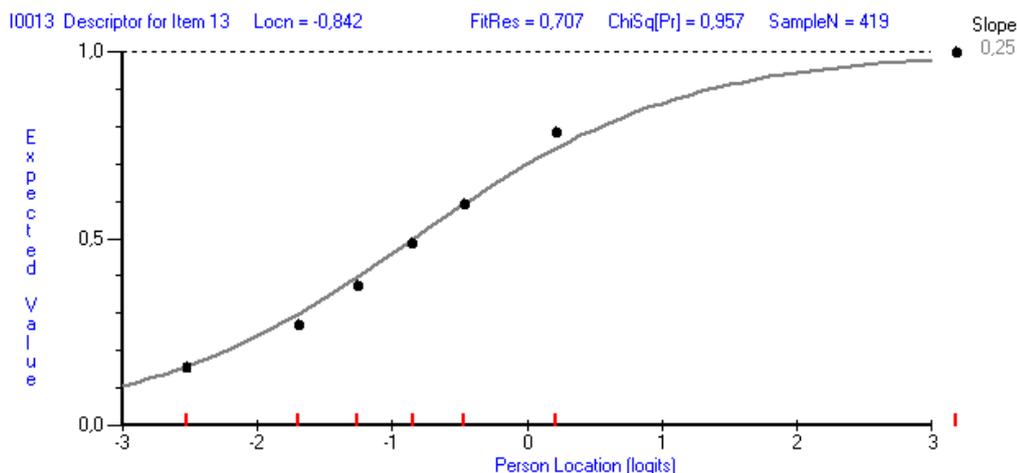


Рис.1 Характеристическая кривая тестового задания Item13

Информация на задания Item13:

- трудность задания (Locn) – -0,842;
- суммарное отклонение ответов тестируемого на данное задание от ожидаемых на основе Раша (Fit Res) – 0,707;
- количество тестируемых – 419.

Из рисунка 1 видно, что экспериментальные точки лежат на характеристической кривой. Это показывает, что характеристическая кривая задания Item13 соответствует модели измерения. Также это видно по эмпирическому уровню значимости статистики Хи-квадрат (ChiSq(Pr)) для задания, которая равна 0,957.

Информация на задания Item27:

- трудность задания (Locn) - 0,777;
- суммарное отклонение ответов тестируемого на данное задание от ожидаемых на основе Раша (Fit Res) - 0,078;
- количество тестируемых - 419.

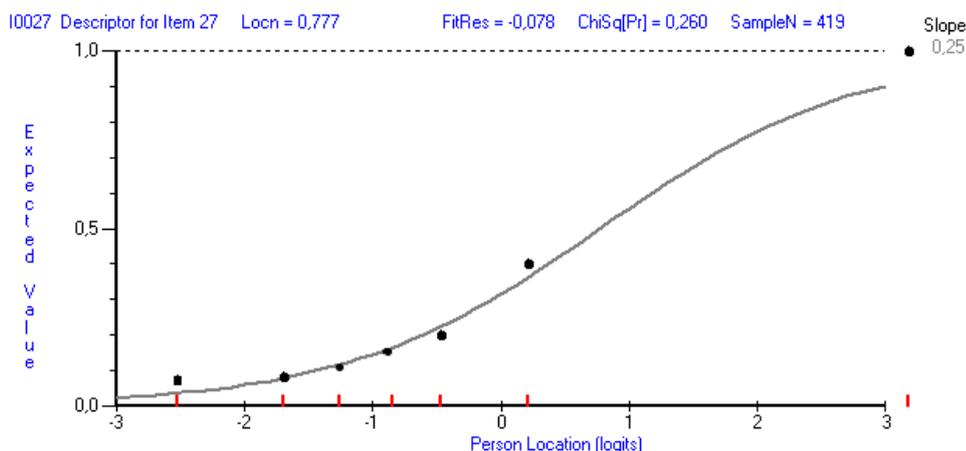


Рис.2 Характеристическая кривая тестового задания Item 27

Из рисунка 2 видно, что экспериментальные точки лежат на характеристической кривой. Эмпирический уровень значимости задания Хи-квадрат 0,260. Характеристическая кривая задания Item27 соответствует модели измерения.

Не соответствуют модели измерения только два задания Item15 и Item17. Характеристические кривые приведены на рисунке 3 и 4 соответственно.

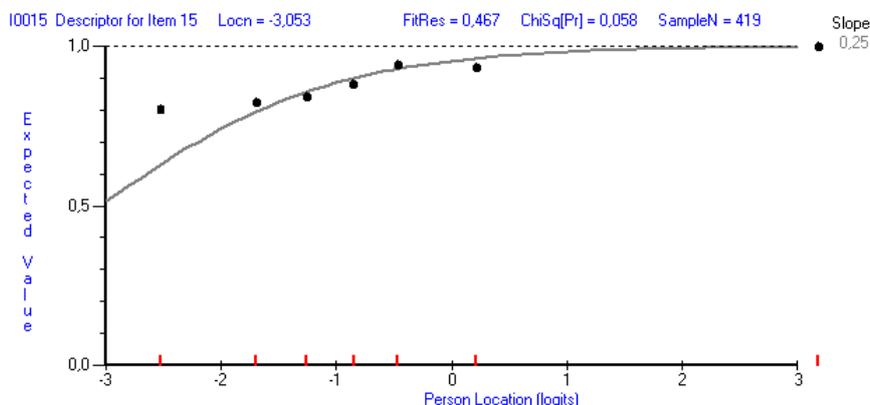


Рис.3 Характеристическая кривая тестового задания Item15

Несоответствие задания Item15 модели измерения проявляется в том, что испытуемые с низким уровнем знаний отвечают на это задание правильно с вероятностью большей, чем предполагается по модели. Это задание совместимо с остальными заданиями на уровне значимости статистики Хи-квадрат больше 0,05, которая равна 0,058.

Задание Item15 проверяет сформированность у тестируемых умение проводить анализ и интерпретацию различных видов информации. Во время решения заданий 25 процентов испытуемых ответили правильно, 45 процентов дали неправильный ответ. Из-за лишней информации задания трудно поддаются логическому анализу, в результате это составило трудность определить правильный ответ. И испытуемые, знакомые с предметом, были вынуждены выбрать неправильный ответ. Тестовое задание требует корректировки.

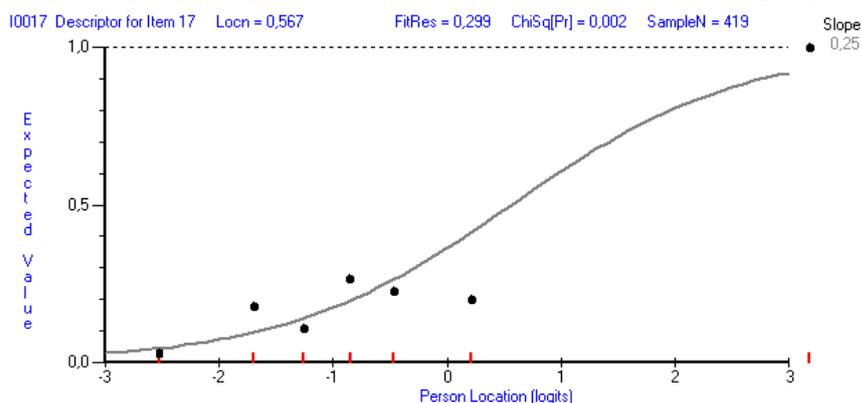


Рис.4 Характеристическая кривая тестового задания Item17

Несоответствие задания Item17 модели измерения проявляется в том, что студенты со средним и высоким уровнем знания отвечают на это задание правильно с вероятностью меньшей, чем предполагается по модели.

Это задание не совместимо с остальными заданиями на уровне значимости статистики Хи-квадрат 0,002. Следовательно, задание должно быть откорректировано.

Соответствие между уровнем подготовленности студентов и трудностью тестовых заданий.

Определим соответствие трудности измерительных материалов уровню подготовленности студентов. На рисунке 6 представлены распределения уровня подготовленности студентов (вверху) и оценок трудности заданий (внизу).

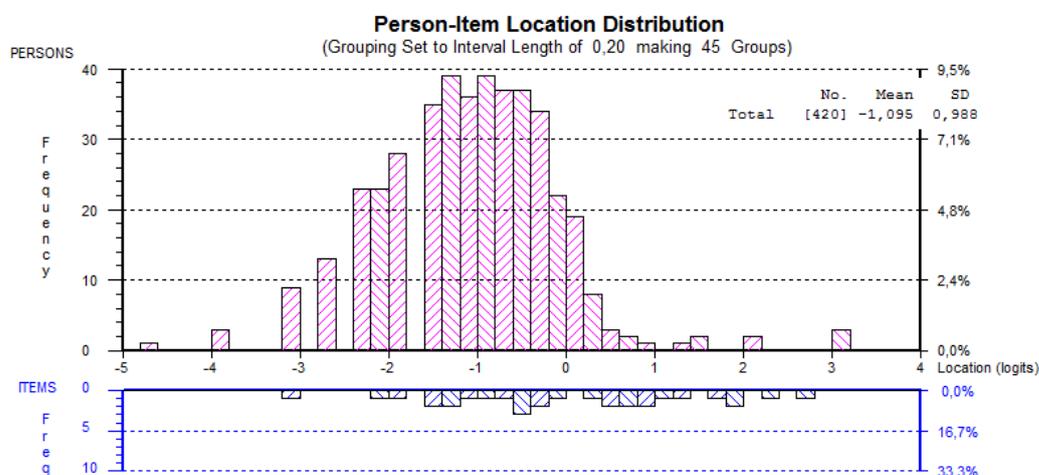


Рис.5. Соответствие между уровнем подготовленности студентов и трудностью тестовых заданий (шаг 0,5 логит)

На рисунке 5 по оси абсцисс откладываются значения латентных переменных (в логитах) уровня подготовленности тестируемых и уровня трудности тестовых заданий, т.е. на одной и той же шкале. По оси ординат с левой стороны указывается число испытуемых (вверху) и число заданий (внизу). С правой стороны – соответствующие величины в процентах.

В нашем случае можно считать, что распределение трудностей заданий близко к равномерному закону. Это обозначает, что представленный набор тестовых заданий с одинаковой точностью позволяет оценить уровень подготовленности студентов на всем диапазоне изменения трудностей тестовых заданий. Тест достаточно хорошо дифференцирует претендентов в магистратуру по уровню их подготовленности. Среднеквадратичное отклонение составило 0,988. Средний уровень знаний тестируемых – -1,095, тест является легким для данного контингента претендентов. Коэффициент надежности (Альфа-Кронбаха) (Alpha) – 0,763 (хорошо).

Диапазон варьирования трудности заданий и равномерность распределения заданий по трудности.

Для рассматриваемой выборки уровень трудности заданий изменяется в диапазоне от -3,2 до 2,8. Длина всего интервала составила 6 логит. Данный тест обеспечивает измерение как низкого, так высокого уровня знаний студентов. В этом случае нет необходимости существенно корректировать выбранный набор. Диапазон изменения уровня подготовленности от -4,8 до 3,2 логит. Таким образом, длина всего интервала составила 8 логит. Этот факт подтверждает дифференцирующую способность теста.

На рисунке 6 выбран мелкий шаг равный 0,20 логит. И можно отметить тот факт, что значения трудностей заполняют почти весь диапазон их изменений. Отсутствуют задания с трудностями в диапазонах от -3 до -2,2 логит, от -2,2 до -2,4 логит и от 0,0 до 0,2 логит, от 1,4 до 1,6 логит, от 2,0 до 2,2 логит и от 2,4 до 2,6 логит. Трудность соседних заданий не должна отличаться более чем на 0,5 логита, в противном случае предполагается, что уровень знаний испытуемых в этом диапазоне не дифференцируются с необходимой точностью. В диапазоне от -3 до -2,2, отличие более чем 0,5 логит. С этой позиции тест требует в этом диапазоне небольшой корректировки.

Заключение. Таким образом, результаты апробации ТГО, подтверждают, что тесты объективные и валидные, так как они строго построены в соответствии со спецификацией. Соблюдены уровни трудности тестовых заданий, поэтому можно судить, что тест правильно определяет функциональную грамотность, что означает овладение инструментарием универсальных навыков и умений тестируемого независимо от области их применения. Были выявлены задания, которые требуют доработки, определены пути их совершенствования. В целом уровень качества тестов достаточно высокий, распределение уровня подготовленности тестируемых соответствует нормальному закону распределения. Тест имеет хорошие показатели надежности.

Тест хорошо дифференцирует тестируемых, как с низким, так и с высоким уровнями знаний. В целом, результаты анализа показали, что разработанные тесты позволяют объективно оценивать уровень знаний претендентов в магистратуру.

Список литературы

1. *Об утверждении Правил проведения комплексного тестирования. Приказ Министра образования и науки Республики Казахстан от 8 мая 2019 года № 190.*
2. Маслак А.А. *Измерение латентных переменных в социально-экономических системах //Монография. - Славянск-на-Кубани: Издательский центр СГПИ, 2006 –22 -28, С.105-106*
3. Аванесов В.С. *Item Response Theory: Основные понятия и положения //Педагогические Измерения, №2, 2007г. с. 3-28.*
4. Маслак А.А., Поздняков С.А. *Анализ качества тестовых заданий с выбором одного правильного ответа //Методические рекомендации. - Славянск-на-Кубани: Издательский центр СГПИ, 2009 – 50 с.*

References

1. *Ob ýtverjdeni Pravil provedenia kompleksnogo testirovaniya. Prikaz Ministra obrazovaniya i nauki Respubliki Kazahstan ot 8 maia 2019 goda № 190.*
2. *Maslak A.A. Izmerenie latentnyh peremennyh v sosialno-ekonomicheskikh sistemah //Monografiya. - Slavánsk-na-Kýbani: Izdatelski sentr SGPI, 2006 –22 -28, S.105-106*
3. *Avanesov V.S. Item Response Theory: Osnovnyye ponátia i polojenia //Pedagogicheskie Izmereniya, №2, 2007g. s. 3-28.*
4. *Maslak A.A., Pozdnákov S.A. Analiz kachestva testovyh zadani s vyborom odnogo pravilnogo otveta //Metodicheskie rekomendasii. - Slavánsk-na-Kýbani: Izdatelski sentr SGPI, 2009 – 50 s.*