

АНАЛИЗ КАЧЕСТВА ЗАДАНИЙ ПЕДАГОГИЧЕСКОГО ТЕСТА ПО УЧЕБНОЙ ДИСЦИПЛИНЕ «МАТЕМАТИКА И ИНФОРМАТИКА»

Галина Смирнова

Славянский-на-Кубани государственный
педагогический институт
smirnova_g_i@mail.ru

В статье рассмотрена методика оценки качества заданий педагогического теста, которая основана на результатах обработки тестовых матриц посредством программы измерения значений латентных переменных RUMM. Предложены варианты усовершенствования тестовых заданий.

Для оценки качества педагогического теста с использованием модели Г. Раша были использованы задания по учебной дисциплине «Математика и информатика»¹, подготовленные в рамках промежуточного тестирования модульно-рейтинговой системы обучения студентов. Протестировано 45 студентов группы 04-Н (специальность – учитель начальных классов и иностранного языка) факультета педагогики и методики начального образования. Студентами изучены первый и второй модуль по дисциплине «Математика и информатика», которые заканчиваются промежуточным тестированием.

1

Автор благодарит Е.П. Бакай за предоставление возможностей анализа первого рабочего варианта исходной матрицы результатов.

Задания рассматривались в следующей последовательности:

1. Выявление и исключение тех заданий, трудность которых нельзя измерить при использовании модели Раша. Если на некоторые задания все испытуемые отвечают правильно, или все отвечают неправильно то определить меру трудности таких заданий невозможно. Их мера трудности выходит за пределы возможностей измерения. В западной литературе такие задания называются экстремальными. В большинстве алгоритмов их обычно исключают из матрицы результатов тестирования.

2. Совместимость тестовых заданий.

Совместимость тестовых заданий является необходимым условием использования модели Раша для измерения уровня подготовленности студентов. Совместимость тестовых заданий проверяется:

- для всей матрицы результатов тестирования;
- для каждого тестового задания (по результатам тестирования для задания);
- для каждого испытуемого (по результатам тестирования для испытуемого);
- для каждого элемента матрицы тестирования.

Во всех четырёх случаях совместимость определяется на основе критерия χ^2 — результа-

ты тестирования сопоставляются с ожидаемыми значениями на основе модели Раша.

Все испытуемые по полученным оценкам своих уровней подготовленности делятся на группы, в данном случае — на три: с низким, средним и высоким уровнем подготовленности. Далее для каждой группы вычисляется среднее значение (экспериментальное) и ожидаемое значение — на основе модели Раша. На основе критерия χ^2 определяется соответствие экспериментальных значений ожидаемым. Если экспериментальное значение статистики χ^2 превышает табличное, то такое тестовое задание должно быть исключено из теста.

Для заданий первого модуля совместимость всей матрицы результатов тестирования на основе критерия χ^2 оказалась равной 0,076239. Полученное значение говорит о том, что степень соответствия данных тестирования модели Раша низкая.

Анализ *совместимости для каждого тестового задания* показал, что задание № 1 и № 10 не соответствуют модели Раша, так как экспериментальное значение статистики χ^2 для каждого из них не соответствует табличному, критическому значению 0,05.

Характеристическая кривая для задания 1.

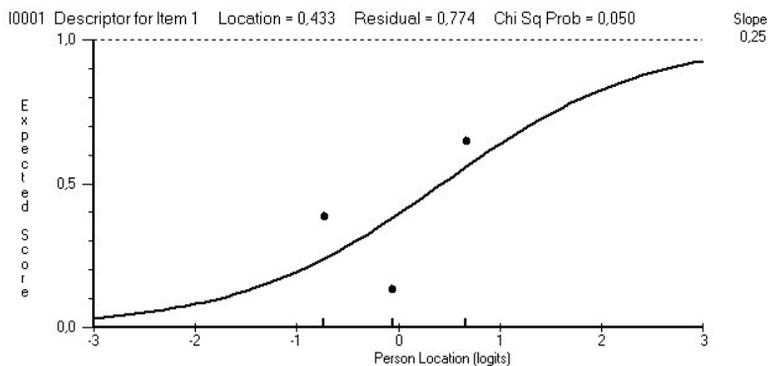


Рис. 1. Характеристическая кривая для тестового задания № 1

По оси абсцисс отложены значения латентной переменной «уровень знаний по математике и информатике». В данном случае латентная переменная варьируется от -3 логита до $+3$ логита. По оси ординат откладывается ожидаемый ответ индивида. Ожидаемый результат (Expected Score) варьируется от 0 до 1. В верхней части рисунка расположена следующая информация:

- код тестового задания (10001);
- название тестового задания, здесь названия задания выбраны по умолчанию. В данном случае это «Descriptor for Item 1»;
- трудность задания (Location = 0,433);
- суммарное отклонение ответов индивидов на данное задание от ожидаемых на основе модели Раша (Residual = 0,774);
- степень соответствия данных модели Раша (Chi Sq Prob = 0,050);
- наклон кривой (Slope = 0,25).

Из рис. 1 видно, что данное тестовое задание можно отнести к заданиям с *«неупорядоченным ответом»*. Студенты с низким и высоким уровнем знаний имеют большую вероятность правильно ответить на это задание, чем студенты со средним уровнем знаний. В целом такое задание плохо «вписывается» в тест, и такой тип заданий также должен быть исключен из набора тестовых заданий.

Посмотрим на кривую другого задания, № 10 (см. рис. 2).

Из рис. 2 видно, что данное тестовое задание можно также отнести к заданиям с *«неупорядоченным ответом»*. Следовательно, задание № 10 также должно быть исключено из набора тестовых заданий.

Для второго модуля *совместимость всей матрицы результатов тестирования* на основе критерия χ^2 составляет 0,029236. Полученное значение говорит о том, что степень

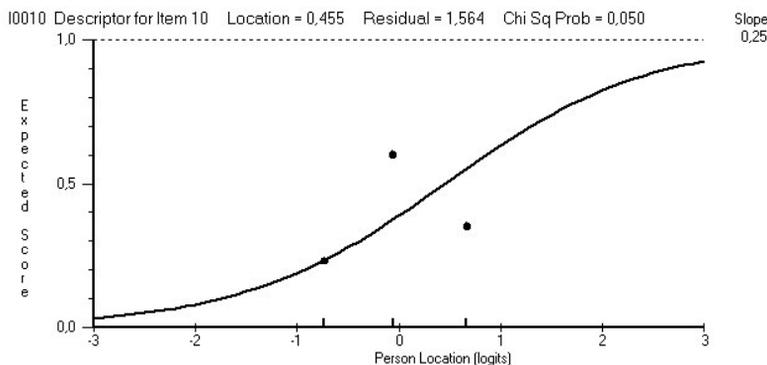


Рис.2. Характеристическая кривая для тестового задания № 10

соответствия данных тестирования модели Раша низкая.

Анализ **совместимости для каждого тестового задания** показал, что задание № 9, № 14 и № 24 не соответствуют модели Раша, так как экспериментальное значение статистики χ^2 для каждого из данных тестовых заданий не соответствует табличному и составляет 0,02822; 0,005291 и 0,030508 соответственно, при критическом значении 0,05.

Задание № 9:

Двоичное число 11001₂

равно десятичному числу (правильный ответ — «а»)

а) 25₁₀

б) 19₁₀

в) 20₁₀

г) 18₁₀

д) 50₁₀

Из рис. 3 видно, что данное тестовое задание можно отнести к заданиям *со сверхвысокой различающей способностью*. Студенты с низким уровнем знаний

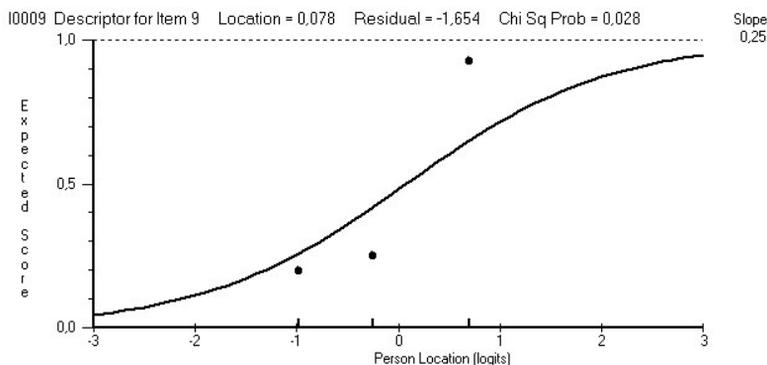


Рис. 3. Характеристическая кривая для тестового задания № 9

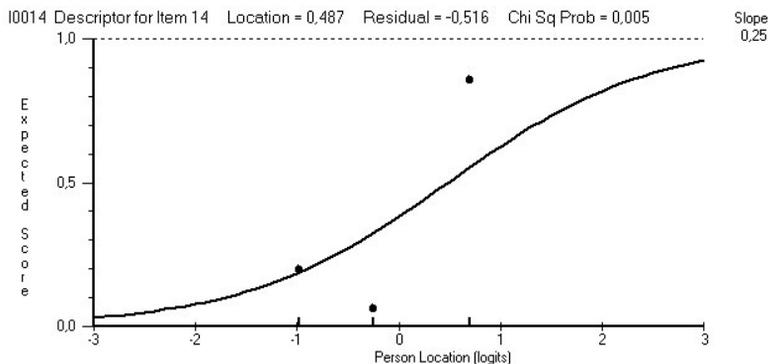


Рис. 4. Характеристическая кривая для тестового задания № 14

отвечают значительно хуже, чем ожидается по модели эксперимента, а студенты с высоким уровнем знаний — значительно лучше. Такой результат можно объяснить тем, что тема очень хорошо усвоена. Однако такой тип заданий также должен быть исключен из теста, что приходится делать из-за отклонения от модели Раша.

Задание № 14:

Восьмеричное число 35_8 равно десятичному (правильный ответ — «а»)
 а) 29_{10}
 б) 25_{10}
 в) 24_{10}
 г) 232_{10}

Из рис. 4 видно, что данное тестовое задание можно отнести к заданиям с «неупорядоченным ответом». Студенты со средним уровнем знаний имеют меньшую вероятность правильно ответить на это задание, чем студенты с высоким

уровнем знаний. В то время как вероятность ответа студентов с низким уровнем знаний соответствует модельной кривой. В целом такое задание плохо «вписывается» в тест, и такой тип заданий также должен быть исключен из набора тестовых заданий.

Задание № 24:

В блок-схеме команда ввода изображается в (правильный ответ — «а»)
 а) параллелограмме
 б) прямоугольнике
 в) ромбе
 г) шестиугольнике

Из рис. 5 видно, что данное тестовое задание можно отнести к заданиям с «неупорядоченным ответом». Студенты со средним уровнем знаний имеют большую вероятность правильно ответить на это задание, чем студенты с высоким уровнем знаний. В то время как вероятность ответа студентов с

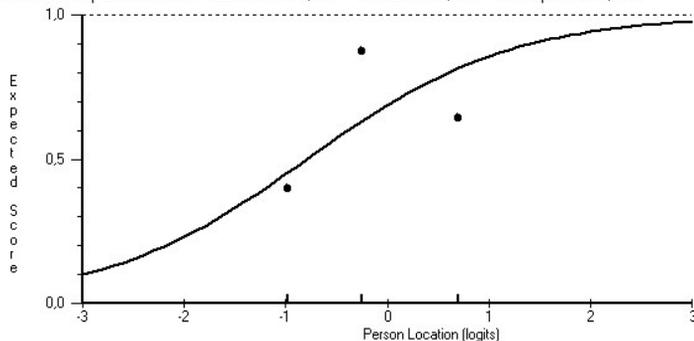


Рис. 5. Характеристическая кривая для тестового задания № 24

низким уровнем знаний соответствует модельной кривой. В целом такое задание плохо «вписывается» в тест, и такой тип заданий также должен быть исключен из набора тестовых заданий.

3. Анализ качества дистракторов. Дистракторы — это неправильные варианты ответа на тестовое задание. Хорошие дистракторы похожи на правильный ответ, и тем самым

позволяют проверить уровень знаний испытуемого. Если ни один из испытуемых не выбирает какой-либо дистрактор, то встаёт вопрос о целесообразности использования такого дистрактора. Насколько «хорошо» работают дистракторы, определяется с помощью критерия χ^2 — неправильные ответы испытуемых должны быть равномерно распределены по всем дистракторам.

Таблица 1

Правильный ответ		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3	
Дистракторы	1	18	22	17	36	43	38	0	34	15	18	34	18	5	32	13	11	15	3
	2	7	2	0	3	0	0	0	2	5	2	0	14	17	0	15	16	15	9
	3	19	16	28	2	2	3	0	8	4	3	9	5	14	0	16	18	14	26
	4	0	2	0	3	0	2	45	1	19	22	1	7	9	7	0	0	0	6

Правильный ответ	2	1	1	3	1	1	3	1	1	2	3	1
Дистракторы	6	19	20	14	18	35	0	24	13	24	14	14
	31	9	12	14	6	6	0	10	15	20	5	5
	0	4	4	5	6	2	36	4	5	0	19	12
	6	11	9	11	14	2	9	6	11	1	7	13

В первом модуле дистракторы имеют следующий вид (табл. 1).

Расчёты по определению значения дистракторов первого модуля в группе 04-Н приведены в Приложении 3. По итогам расчётов можно сказать, что дистракторы не работают для тестовых заданий № 1–3, № 8–11, № 14, № 25 и № 28. Данные задания необходимо пересмотреть. Общее значение χ^2 для всего теста составило 10,03636 при критическом значении = 6. Следовательно, работа дистракторов теста неудовлетворительна.

Во втором модуле дистракторы имеют следующий вид (табл. 2).

Расчёты по определению значения дистракторов аналогичны расчётам, приведённым в приложении 3. По итогам расчётов можно сказать, что дис-

тракторы не работают для тестовых заданий № 3, № 12, № 14, № 16, № 21, № 23, № 28, и № 29. Данные задания необходимо пересмотреть. Общее значение χ^2 для всего теста составило 4,613617 при критическом значении = 6. Следовательно, работа дистракторов теста удовлетворительна.

4. Равномерность распределения заданий по трудности. Трудность соседних заданий не должна отличаться более чем на 0,5 логита. Значение 0,5 выбрано на том основании, что в хорошем тесте ошибка измерений уровня знаний испытуемых находится в пределах 0,25 логитов. Если это условие не выполняется, то тест не является эффективным измерительным инструментом. Это объясняется тем, что испытуемые внутри такого диапазона не дифференцируются.

Таблица 2

Правильный ответ		1	3	1	1	3	1	1	1	1	1	1	1	1	1	1	1	3	
Дистракторы	1	18	13	15	26	6	15	0	18	20	26	22	25	21	16	28	12	22	7
	2	3	6	25	6	10	12	0	5	7	5	7	8	11	4	7	17	7	5
	3	9	10	5	7	24	12	0	10	3	8	8	11	7	9	6	15	12	31
	4	7	11	0	6	5	6	45	7	7	6	8	1	6	16	4	0	4	2

Правильный ответ	2	4	1	4	1	1	3	2	2	1	3	1
Дистракторы	3	6	19	14	16	29	7	3	6	20	11	13
	30	6	7	13	13	7	11	23	28	0	3	11
	7	7	4	5	2	6	15	12	6	11	15	14
	5	26	14	13	14	3	12	7	5	14	16	7

Согласно рис. 6 для *первого модуля* (см. ниже) критерий распределения заданий по трудности не соблюдается, то есть трудность соседних заданий данного набора тестовых заданий превышает порог в 0,5 логитов. Таким образом, анализируемый тест не может являться эффективным измерительным инструментом.

Для *второго модуля* согласно рис. 7 (см. ниже) критерий распределения заданий по трудности соблюдается, то есть трудность соседних заданий данного набора тестовых заданий не превышает порог в 0,5 логитов. Таким образом, анализируемый тест является эффективным измерительным инструментом.

5. Соответствие трудности теста уровню подготовленности испытуемых. Средняя трудность заданий не должна отличаться от средне-

го уровня подготовленности испытуемых более чем на 0,5 логита. Напомним, что при анализе результатов тестирования на основе модели Раша уровень подготовленности испытуемых и трудность тестовых заданий измеряются на одной и той же интервальной шкале. Поэтому если средняя трудность заданий отличается от среднего уровня подготовленности испытуемых более чем на 0,5 логита, то это означает, что уровень подготовленности некоторых испытуемых (в нижней или верхней части шкалы) плохо дифференцируется. Для анализируемого теста (рис. 3) данный критерий не соблюдается.

Для *первого модуля* было получено следующее соответствие между уровнем знаний студентов и трудностью заданий теста для первого модуля (рис. 6).

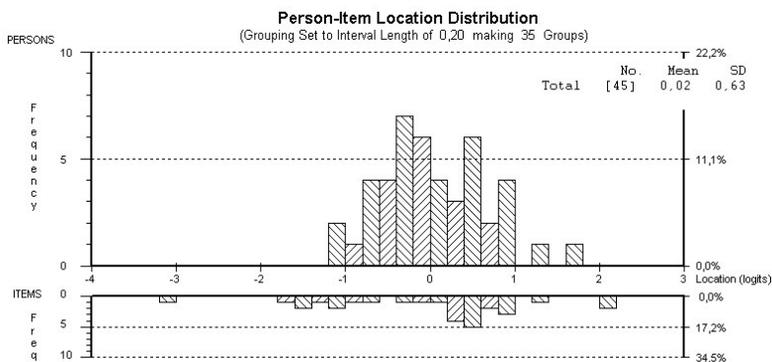


Рис. 6. Соответствие между уровнем знаний студентов и трудностью заданий теста для первого модуля

Данный рисунок является гистограммой. По оси абсцисс откладывается значение латентной переменной (уровень знаний студента и уровень трудности задания), по оси ординат — число индивидов (вверху), и заданий (внизу). Средний уровень знаний индивидов на 0,2 логита выше среднего уровня сложности заданий, это означает, что уровень подготовленности некоторых испытуемых плохо дифференцируется.

В идеальном случае гистограмма распределения ответов испытуемых (верхняя часть рисунка) должна быть близка к нормальному закону распределения: относительно небольшое число студентов с низким и высоким уровнем знаний и относительно много со средним уровнем знаний. Гистограмма распределения трудностей заданий теста (нижняя часть рисунка) должна быть близка к равномерному закону распределения: трудности заданий должны быть равномерно распределены по шкале логитов — это означает, что разработанный набор тестовых заданий позволяет эффективно оценивать уровень знаний.

В данном случае уровни знаний испытуемых распределены практически по нормальному закону распределения. Однако существуют некоторые отклонения. Рассмотрим предыдущую гистограмму в виде карты соответствия между уровнем знаний студентов и трудностью заданий теста. Здесь показаны номера заданий, которые составляют столбцы нижней гистограммы (рис. 7).

Согласно рис. 7 видим, что задание № 5 является лёгким, на него отвечают все испытуемые: и с высоким уровнем подготовленности, и с низким. Задание № 5 следует усложнить.

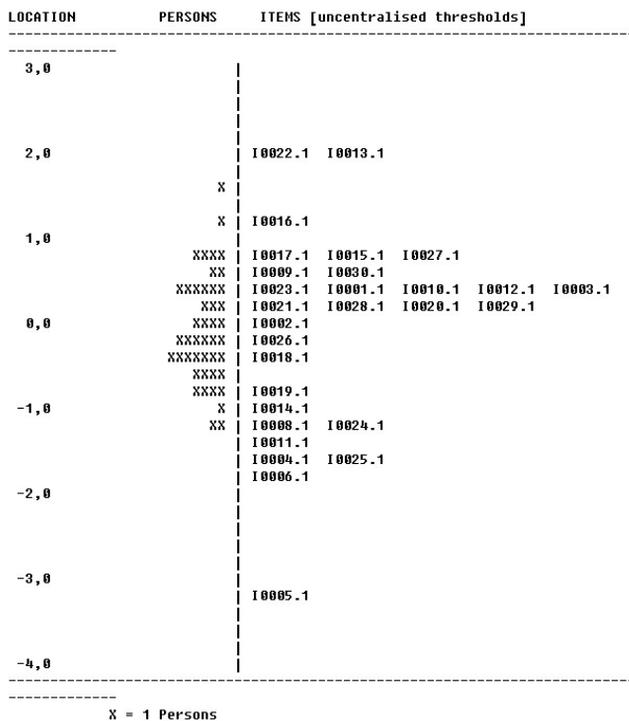


Рис. 7. Карта соответствия между уровнем знаний студентов и трудностью заданий теста для первого модуля

Задание № 5:

Объединение множеств А и В обозначается символом

(правильный ответ — «1»):

- 1) \cup
- 2) \cap
- 3) \in
- 4) \subset

А задания № 10, № 12, № 20, № 29 оказались сложными для испытуемых. Задание № 3 является самым сложным из всего набора тестовых заданий. Это может быть связано с неправильной его формулировкой либо с неверно подобранными дистракторами. Его следует упростить.

Методика

Методика

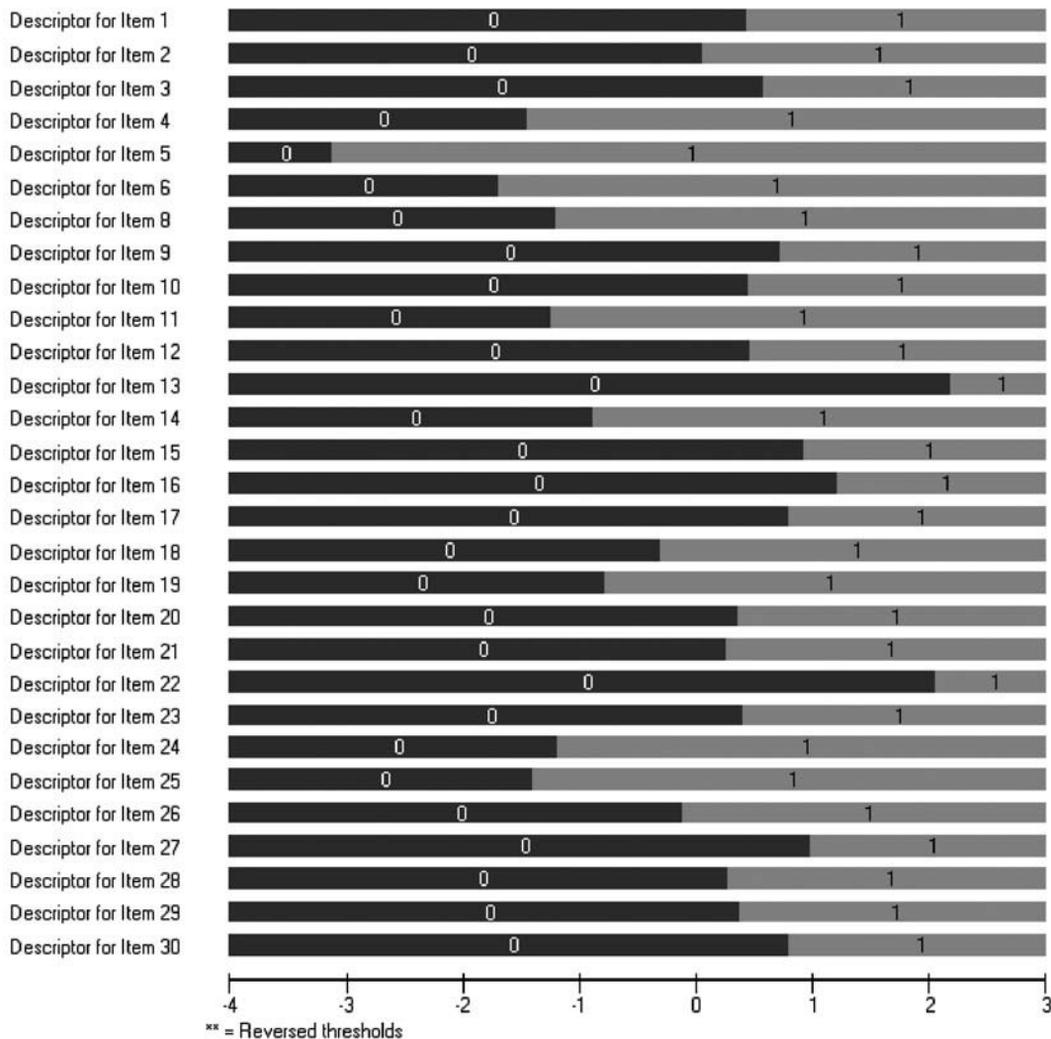


Рис. 8. Работа дистракторов в тестовых заданиях первого модуля

Задание № 3:

Множество, состоящее из общих элементов множества А и множества В, есть (правильный ответ — «1»):

- 1) пересечение множеств А и В
- 2) разность множеств А и В
- 3) объединение множеств А и В
- 4) разность множеств В и А

Так как в анализе сказано, что степень сложности задания зависит от работы дистракторов, рассмотрим, как работают дистракторы тестовых заданий первого модуля (рис. 8).

Из рис. 8 можно сделать вывод, что дистракторы плохо работают только в уже упомянутом тестовом задании № 5. А для заданий № 13, № 16, № 22 и № 27 испытуемые неправильные ответы принимают за верные. Указанные задания нужно переформулировать.

Для второго модуля было получено следующее соответствие между уровнем знаний студентов и трудностью заданий теста (рис. 9).

В данном тесте уровни знаний испытуемых распределены по нормальному закону распределения. Уровни трудности заданий распределены по шкале логитов равномерно.

Задания расположены по мере возрастающей трудности по шкале логитов примерно равномерно, однако гистограммы заметно различаются. Однако существуют некоторые отклонения. Рассмотрим предыдущую гистограмму в виде карты соответствия между уровнем знаний студентов и трудностью заданий теста. Здесь показаны номера заданий, которые составляют столбцы нижней гистограммы (рис. 10).

Из рис. 10 видно, что набор тестовых заданий второго модуля имеет примерно одинаковый уровень трудности. Однако задания № 6, 12 и 29 выделяются как самые сложные. Их следует переформулировать.

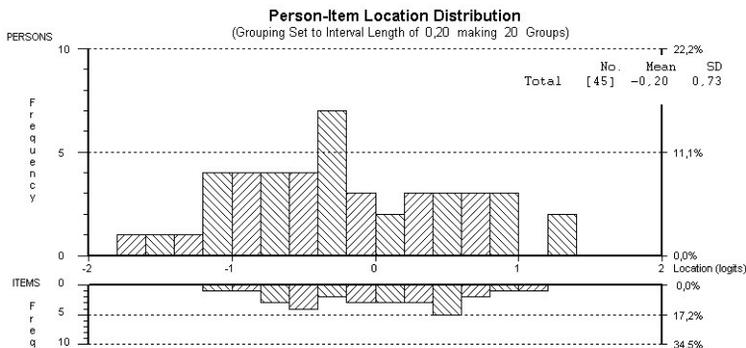


Рис. 9. Соответствие между уровнем знаний студентов и трудностью заданий теста во втором модуле

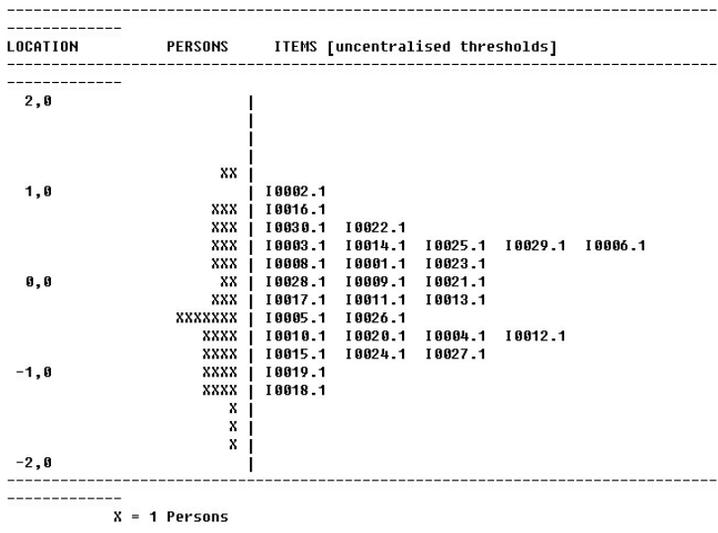


Рис. 10. Карта соответствия между уровнем знаний студентов и трудностью заданий теста для второго модуля

Рассмотрим работу дистракторов данного набора тестовых заданий (рис. 11).

Из рис. 11 можно сделать вывод, что дистракторы плохо работают в тестовых заданиях № 15, № 18, № 19, № 24 и № 27. А для заданий № 2, № 16, № 22 и № 30 испытуемые неправильные ответы принимают за верные. Указанные задания нужно переформулировать.

6. Диапазон варьирования заданий должен иметь значение больше трёх логитов. Если диапазон не достигает данного значения, то необходимо добавлять в тест задания всех уровней сложности. Согласно рис. 3 для первого модуля группы 04-Н диапазон значений данного теста находится на отрезке от $-3,2$ до

$+2,2$ по шкале логитов, следовательно, диапазон варьирования заданий составляет 5,4 логита. Данное значение удовлетворяет выдвинутому критерию, данный тест не следует усложнять.

Для второго модуля диапазон варьирования заданий находится на отрезке от $-1,2$ до $+1,2$ по шкале логитов, следовательно, его значение составляет 2,4 логита, что не удовлетворяет выдвинутому критерию, данный тест следует усложнить.

Приведённые критерии качества тестов являются важным средством повышения эффективности, тестов. Учёт этих критериев позволяет разрабатывать тесты, которые объективно и точно измеряют уровень подготовленности испыту-

емых. Для модуля 1 в группе 04-Н было выявлено, что необходимо переработать следующие тестовые задания: № 1–3, № 8–11, № 14, № 25 и № 28. Для модуля 2 – № 3, № 9, № 12, № 14, № 16, № 19, № 21–24, № 28, и № 29. Так же при сравнении двух тестов модуля 1 и модуля 2 видно, что по крите-

рию 3–5 качество тестовых заданий модуля 2 выше, чем модуля 1.

Выводы

Теоретическая значимость данного исследования заключается в анализе возможностей программы RUMM для оценки качества заданий и теста в целом.

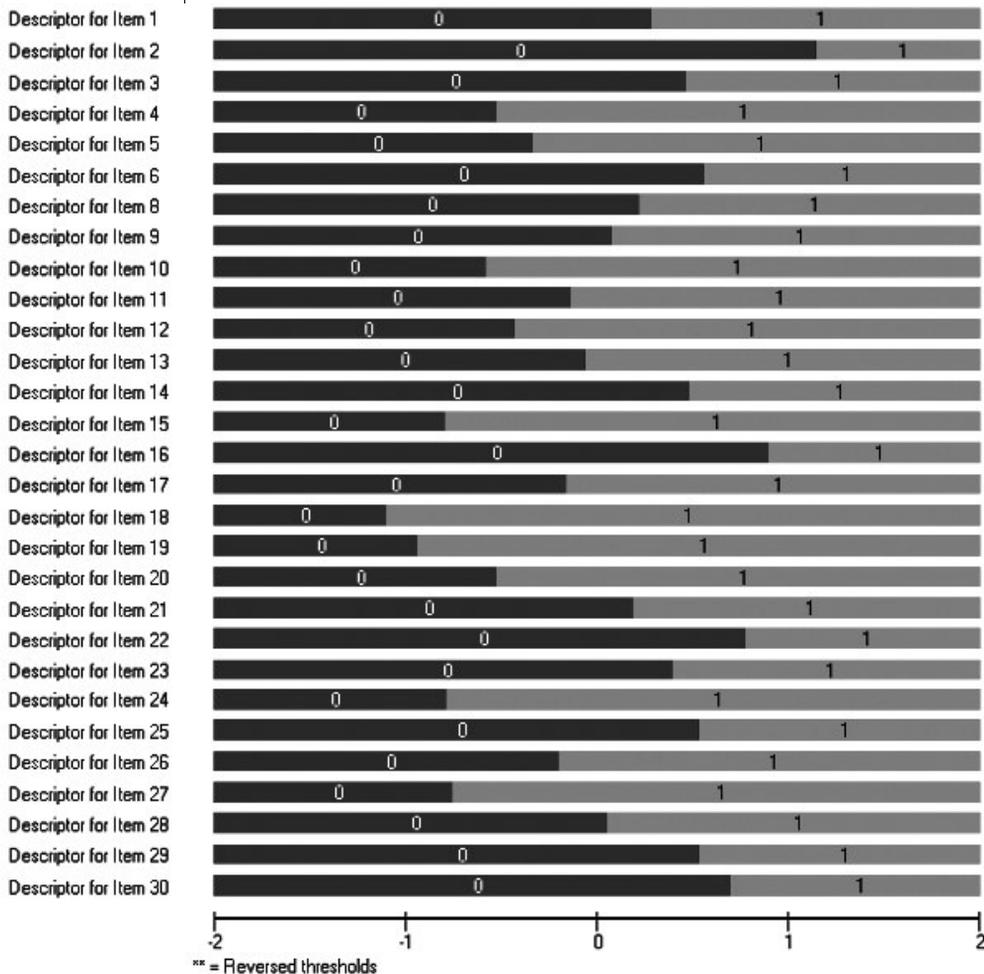


Рис. 11. Работа дистракторов в тестовых заданиях второго модуля

Практическая значимость заключается в анализе набора тестовых заданий на основе модели Раша по дисциплине «Математика и информатика». Предложены варианты усовершенствования заданий. Для модуля 1 необходимо переработать задания № 1–3, № 8–10, № 11, № 14, № 25 и № 28. Для модуля 2 — № 3, № 9, № 12, № 14, № 16, № 19, № 21–24, № 28, и № 29.

Методическая значимость исследования состоит в том, что предложена методика оценки качества теста: с помощью программы — R_{ИММ} (Rash Unidimensional Measurement Models) анализируется матрица ответов испытуемых и измеряется латентная переменная — уровень подготовленности студентов по дисциплине «Математика и информатика». R_{ИММ} позволяет рассматривать полученные результаты с нескольких сторон, например:

1. Выявление и исключение из теста экстремальных заданий, трудность которых нельзя измерить при использовании модели Раша.
2. Совместимость тестовых заданий.
3. Анализ качества дистракторов.
4. Равномерность распределения заданий по трудности.
5. Соответствие трудности теста уровню подготовленности испытуемых.

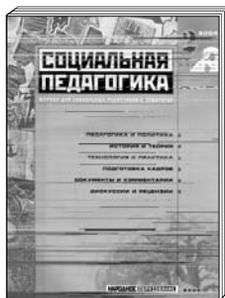
6. Диапазон варьирования заданий.

По данным критериям делается вывод о качестве каждого тестового задания, включённого в тест, качестве всего теста и об уровне выраженности измеряемой латентной переменной. Далее автору теста необходимо исправить указанные недостатки и снова провести тестирование студентов тех же групп, используя обновлённые варианты тестов.

Проведённый анализ является первым этапом исследования. Планируется применить разработанную методику, дополнительно проанализировав результаты тестирования группы испытуемых по третьему модулю и итоговому тесту дисциплины «Математика и информатика». Далее намечено провести повторное тестирование студентов группы по трем модулям и итоговому тесту исследуемой дисциплины и сделать вывод о справедливости выдвинутой гипотезы: процесс проверки уровня подготовленности студентов по дисциплине «Математика и информатика» будет реализовываться более эффективно, если при оценке качества набора тестовых заданий активно и целенаправленно использовать разработанную авторскую систему.

Литература

1. *Анисимова Т.С.* Измерение латентных переменных в образовании. М.: Исследовательский центр проблем качества подготовки специалистов, 2004. С. 148.
2. Закон об образовании. Москва, Кремль. 13 января 1996 г. №12-ФЗ.
3. *Моисеев Н.Н.* **Алгоритмы развития. М.: Наука, 1987.**
4. *Селезнева Н.А.* Качество высшего образования как объект системного исследования. Лекция-доклад. М.: Исследовательский центр проблем качества подготовки специалистов, 2004. С. 95.
5. *Селезнёва Н.А.* Система обеспечения качества и управления качеством высшего образования по специальностям (направлением подготовки) как главные объекты комплексного исследования и модернизации: Доклад в слайдах. М., Уфа: Исследовательский центр проблем качества подготовки специалистов, 2004. С. 45.
6. *Сенашенко В., Жач Г.* **Болонский процесс и качество образования (Российский университет дружбы народов). ALMA MATER. Вестник высшей школы. 2003. № 8.**
7. Толковый словарь В. Даля.
8. *Шишов С.Е., Кальней В.А.* Мониторинг качества образования в школе. М., 1998.



СОЦИАЛЬНАЯ ПЕДАГОГИКА

Профессиональный журнал для социальных педагогов и специалистов по социальной работе.
Четыре номера в год.

Индекс — 81416, 47005