

ПРОБЛЕМА КАЧЕСТВА КОНТРОЛЬНО-ИЗМЕРИТЕЛЬНЫХ МАТЕРИАЛОВ ПЕДАГОГИЧЕСКОГО ТЕСТИРОВАНИЯ

В статье обсуждаются результаты применения современной теории тестирования (IRT) в рамках однопараметрической модели Раша для моделирования содержания контрольно-измерительных материалов (тестов), нацеленных на выявление уровня подготовки студентов педагогических университетов по дисциплине «прикладная химия»

Качество высшего профессионального образования сегодня возведено в ранг национальной политики государства. Реализация в России Национальной системы оценки качества образования ставит задачу осуществления его мониторинга на основе квалиметрического подхода, в качестве объектов которого выступают учебные планы и программы, учебники и учебные пособия, образовательные технологии, педагогические кадры, результаты образования [1-4]. Известно, что на качество подготовки студентов влияют организация и планомерное управление их учебно-познавательной деятельностью, которая нацелена на качественное усвоение профессиональной системы знаний, умений и навыков (специальных компетенций), овладение опытом экспериментальной работы и творческой деятельности. Для определения результатов образования в последнее время в практику работы вузов широко внедряются методы педагогической диагностики, удовлетворяющие требованиям оперативности, объективности, надежности, возможности соотнесения полученных результатов с известными измерительными шкалами [5, 6, 7, с. 68].

В настоящее время в практике выявления уровня подготовленности школьников и студентов широко используются различные технологии тестирования [8-10]. Так, Единый государственный экзамен стал обязательным условием окончания общеобразовательной школы [11]. А вузы участвуют в Интернет тестировании остаточных знаний студентов по основным дидактиче-

ским единицам учебных дисциплин гуманитарно-социально-экономического и естественнонаучного блоков ГОС ВПО [12], что необходимо при аккредитации их работы. Все эти мероприятия проводятся, с одной стороны, для решения задач мониторинга в системе образования, с другой – продиктованы желанием и необходимостью перехода к более объективным методам оценивания результатов обучения. Для достижения этих целей необходимо тесты стандартизировать. Эта проблема не решена сегодня даже в рамках средней школы. Что касается высшего образования, то банк стандартизованных тестовых заданий по некоторым дисциплинам специальной подготовки, например, учителя химии, вовсе отсутствует. Кроме того, любые результаты тестирования являются «сырыми баллами», абсолютное значение которых зависит как от уровня подготовки учащихся, так и от трудности заданий теста [9, 10]. Таким образом, результат оценивания зависит от качества измерительного материала и от уровня подготовленности испытуемых. Результаты же объективной оценки качества обученности не должны зависеть ни от того, кто и кого тестирует, ни от того, какой используется измерительный инструмент.

Одной из современных теорий тестирования является Item Response Theory (IRT). В общем случае это часть латентно-структурного анализа, предназначенная для оценки латентных свойств испытуемых и параметров заданий теста посредством применения математико-статистических моделей измерения. В отличие от классиче-

ской теории тестирования, для IRT характерно стремление к фундаментальному теоретическому подходу в обосновании целого ряда практических задач педагогического измерения [13].

Попытки измерения таких качеств студентов, как «подготовленность», «знание учебной дисциплины», «способность понимать», «интеллектуальное развитие» и других, на уровне обыденного сознания традиционно оканчиваются словесными или численными оценками, содержащими в себе немалые погрешности [9, 10]. Теория тестирования IRT располагает такими математическими методами измерения латентных качеств личности, которые тесным образом согласуются с эмпирическими результатами, что позволяет на основе реально наблюдаемых данных решить задачу воссоздания исследуемого качества испытуемого по измеряемым параметрам. Наибольшее распространение и применение в практической деятельности получила однопараметрическая модель Г. Раша [13, 14]. Она оперирует с инвариантными латентными параметрами – уровень подготовленности студентов θ_i и уровень трудности заданий теста β_j , между которыми устанавливается связь, то есть взаимодействие двух множеств значений латентных параметров порождает наблюдаемые результаты выполнения теста. Мера соответствия параметров (уровня подготовленности испытуемых и трудности заданий) в модели Г. Раша является линейной: $\theta_i - \beta_j$. Такое их представление дает следующую геометрическую интерпретацию результатов. Величина $|\theta_i - \beta_j|$ характеризует расстояние, на котором находится испытуемый с уровнем подготовки θ_i от задания с трудностью β_j . Если эта разность велика по модулю и отрицательна, то задание бесполезно для измерения уровня знаний i -го ученика, ибо он его наверняка не может выполнить. Большие положительные значения этой разности тоже не представляют интереса ни для обучения i -го ученика, ни для процесса контроля, т.к. задания оказываются очень легкими. Математические процедуры, которые используются в однопараметрической мо-

дели Г. Раша, позволяют получить наряду с начальными оценками параметров θ_i и β_j их стандартизованные оценки, на основании которых удается делать заключение о качестве тестового материала и уровне подготовки испытуемых.

Поскольку оценка трудности тестовых заданий не зависит от выборки испытуемых, на которых она была получена, а оценка уровня знаний испытуемых не зависит от используемого набора тестовых заданий, то это позволяет вести моделирование тестовых заданий, и осуществлять процедуру стандартизации разрабатываемых тестов даже на небольших выборках испытуемых (≥ 10), что является вполне реальным в условиях образовательного процесса.

Существенными преимуществами IRT, по мнению ряда специалистов [13, 14], являются устойчивость и объективность оценок параметров, использование шкалы Раша и, как следствие, возможность одновременного измерения значений обоих параметров. Именно последнее свойство позволяет соотнести уровень знаний любого испытуемого с мерой трудности каждого задания теста.

Согласно ГОС ВПО-2000 в программе подготовки учителя химии есть дисциплина «прикладная химия». При изучении разделов «Химические производства» нами были разработаны предтестовые задания закрытой формы с выбором одного правильного ответа из предложенных четырех. Из этих заданий были сформированы тематические тесты по дисциплине («Производство аммиака», «Производство серной кислоты», «Производство азотной кислоты» «Производство удобрений» и другие). Каждое предтестовое задание имело свой порядковый номер, который мог измениться после оценки трудности задания и выбора определенной стратегии предъявления теста, эталон правильного ответа, оценку за выполнение выставляли по дихотомической шкале. Общее число предтестовых заданий, как правило, не превышало 25. Число испытуемых составляло более 20 в каждом конкретном случае. Процесс тестовых измерений проводили в стандартных усло-

виях (тест включал задания только одной формы, группа тестируемых была выровнена по мотивации, все испытуемые выполняли одни и те же задания, подсказки и списывания были исключены, ни одному испытуемому не давали никаких преимуществ перед другими).

Этап статистической обработки полученных результатов тестирования осуществляли в соответствии с методикой, предложенной в работе [14, с. 259-279]. В каждом конкретном случае была сформирована матрица тестовых результатов, на основании которой рассчитывались статистические параметры теста – среднее значение, дисперсия, корреляционная матрица заданий теста. Кроме того, на основе экспертных оценок были рассчитаны начальные значения параметров подготовленности испытуемых θ_i^0 и трудности заданий теста β_j^0 . Результаты стандартизации векторов уровня подготовки испытуемых и трудности заданий теста (в логитах) по теме «Производство серной кислоты» представлены на рисунке, из которого видно, что стандартизованный вектор уровня подготовки испытуемых θ_i укладывается в диапазон изменения вектора трудности заданий β_j теста. Это указывает на то, что практически все задания теста (исключение 7, 9, 11, 12, 16 и 18) оказались пригодными для измерения уровня подготовки испытуемой выборки студентов. Задание 11 имеет самое отрицательное значение логиты трудности (-2,66), а потому является легким. Тогда как задания 7, 9, 12, 16 и 18 оказались наоборот сложными для большей части студентов и имеют высокие положительные значения логиты трудности (от 1,58 до 2,62). По этой причине на эти вопросы мы обратили особое внимание при анализе правдоподобности дистракторов теста, который представлен в таблице 1.

Первый столбец таблицы указывает номер задания теста. Второй и последующие столбцы содержат число и процент испытуемых, выбравших указанный ответ к заданиям теста. Жирным шрифтом помечен процент, соответствующий правильно-

му ответу на вопрос. Курсивом и звездочкой помечены предтестовые задания, в которых максимальное число ответов не совпадает с правильным. Как видно из таблицы, такими являются 7, 9, 12, 16 и 18 вопросы теста, они имели максимальные значения логиты трудности (см. рис. 1), что может быть обусловлено не только трудностью содержания вопроса, но и большой неправдоподобностью используемых дистракторов.

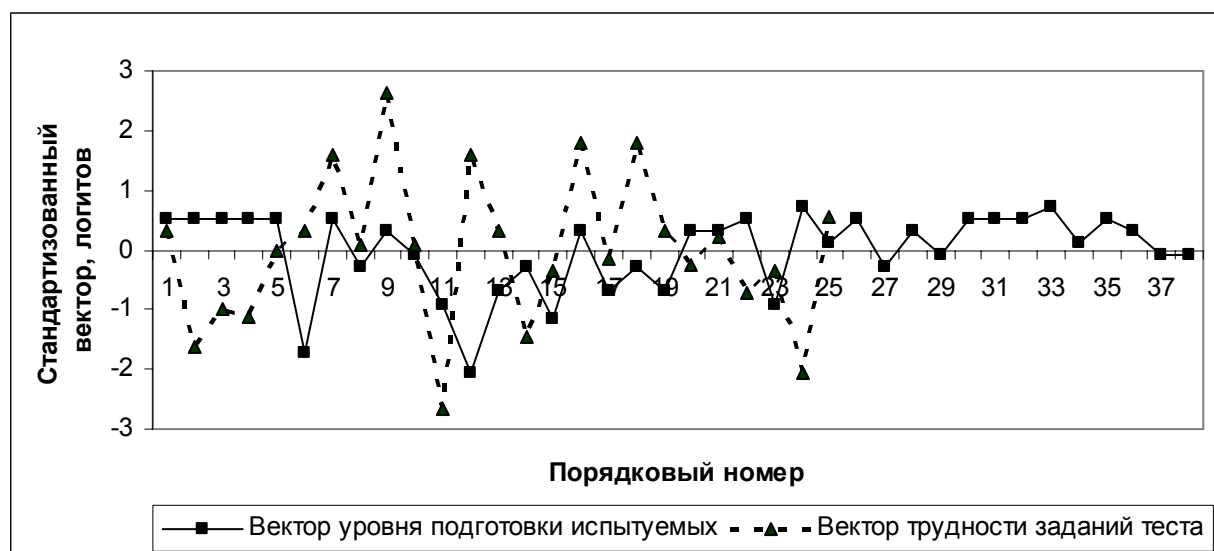
Анализ первой строки таблицы показывает, что правильным является 2-й ответ, и выбрали его 42,1% испытуемых. Согласно классической теории тестирования это задание имеет трудность 42,1 %, то есть отражает долю правильных ответов от общего числа студентов тестируемой группы. Из 38 испытуемых, выполнивших это задание, с ним справились успешно 16, остальные 22 испытуемых распределились между первым (14 испытуемых) и третьим (8 испытуемых) дистракторами. Четвертый дистрактор не выбрал ни один испытуемый. Он оказался непривлекательным, и потому этот ответ как неработающий дистрактор подлежит изменению или удалению из теста. Как следует из таблицы, такие дистракторы выявлены нами в вопросах 7, 11, 13, 14 и 24. Тщательный анализ правдоподобности дистракторов позволяет выделить неработающие, произвести их замену, таким образом моделируя содержание теста. Причем мы заменили и те дистракторы, которые имели небольшой процент ответов (менее 5%). Подобная процедура выполнена для всех разработанных предтестовых заданий по дисциплине.

Современная технология тестирования на основе однопараметрической модели Г. Раша позволяет создавать контрольно-измерительные материалы, которые отвечают требованиям валидности, надежности и стандартизации. Созданные таким образом тесты могут быть использованы для оценки специальных компетенций студентов педагогических университетов, обучающихся по специальности «химия».

Таблица 1

Анализ правдоподобности дистракторов теста по теме «Производство серной кислоты».
Чисто испытуемых – 38 студентов

Номер задания	Распределение ответов тестируемых							
	1-й ответ		2-й ответ		3-й ответ		4-й ответ	
	Число тестируемых	%	Число тестируемых	%	Число тестируемых	%	Число тестируемых	%
1	14	36,8	16	42,1	8	21,1	0	0
2	3	7,9	3	7,9	1	2,6	31	81,6
3	2	5,3	27	71,1	9	23,7	0	0
4	28	73,7	2	5,3	7	18,4	1	2,6
5	11	28,9	5	13,2	3	7,9	19	50
6	13	34,2	5	13,2	16	42,1	4	10,5
7*	7	18,4	20	52,6	11	28,9	0	0
8	1	2,6	2	5,3	18	47,4	17	44,7
9*	17	44,7	7	18,4	11	28,9	3	7,9
10	1	2,6	9	23,7	18	47,4	10	26,3
11	2	5,3	35	92,1	0	0	1	2,6
12*	7	18,4	5	13,2	23	60,5	3	7,9
13	15	39,5	16	42,1	7	18,4	0	0
14	3	7,9	30	78,9	5	13,2	0	0
15	22	57,9	5	13,2	8	21,1	3	7,9
16*	3	7,9	18	47,4	6	15,8	11	28,9
17	2	5,3	8	21,1	20	52,6	8	21,1
18*	6	15,8	17	44,7	10	26,3	5	13,2
19	4	10,5	14	36,8	4	10,5	16	42,1
20	1	2,6	21	55,3	15	39,5	1	2,6
21	3	7,9	12	31,6	6	15,8	17	44,7
22	25	65,8	5	13,2	6	15,8	2	5,3
23	22	57,9	5	13,2	9	23,7	2	5,3
24	4	10,5	33	86,8	1	2,6	0	0
25	6	15,8	14	36,8	4	10,5	14	36,8

Рис. 1. Область взаимодействия латентных параметров θ_i и β_j

Л и т е р а т у р а

1. *Набойченко С., Выварец А., Майбуров И.* Формирование системы мониторинга высшей школы // Высшее образование в России. – 2006. – №6. – С.3-14.
2. *Панасюк В. П., Салова И. Г.* Применение квалиметрических методик и процедур в управлении качеством образования в школе. – СПб., 1997.
3. *Мухартова Н. Е., Лисицкая И. В.* Мониторинг качества образовательной среды // Директор школы. – 2007. – №4. – С.37-40.
4. *Коннычева Г. Г.* Мониторинг образовательного процесса // Директор школы. – 2007. – № 9. – С.21-27.
5. *Мамаева И. А.* Педагогическая диагностика в рамках технологии обучения в вузе // Наука и школа. – 2004. – №4. – С.46-49.
6. *Скопылатов И. А., Ефремов О. Ю.* Система педагогической диагностики в вузе // Педагогика. – 2001. – №7. – С.58-62.
7. *Фарков А. В.* Диагностика обученности и обучаемости математике : моногр. – Архангельск: Изд-во Поморского университета, 2005.
8. *Данилова А. Г.* Оценка качества знаний в системе тематического контроля // Химия в школе. – 2008. – №2. – С.16-20.
9. *Михайлычев Е. А.* Диагностические тесты как средство измерения // Школьные технологии. – 2001. – №3. – С.170-185.
10. *Аванесов В. С.* Педагогические задания в тестовой форме // Школьные технологии. – 2007. – №4. – С.156-167.
11. Единый экзамен станет обязательным. Что приобретет, а что потеряет наше образование? // Директор школы. – 2007. – №4. – С.4-7.
12. *Филипчук О., Фирсова С., Чередниченко О.* Инструмент оценки качества подготовки студентов // Высшее образование в России. – 2008. – №7. – С.155-157.
13. Использование теории тестовых заданий (Item Response Theory) в адаптивном тестировании [Электронный ресурс]. – URL: <http://www.wikiznanie.ru/ru-wz/index> (дата обращения 29.07.2009).
14. *Челышкова М. Б.* Теория и практика конструирования педагогических тестов : учеб. пособие. – М.: Логос, 2002.

