

Определение качества тестовых заданий

Такир Балыкбаев,

Национальная академия образования им. Ы.Алтынсарина

Гульжан Примбетова

Национальная академия образования им. Ы. Алтынсарина

fire-guljan@mail.ru

В статье раскрывается проблема оценки качества заданий, используемых для измерения учебных достижений испытуемых.

Вопросы истории

Теория и практика педагогических измерений развивается на протяжении более ста лет, начиная с конца XIX века до настоящего времени. Обзор истории развития тестов проведен и представлен в работах ученых разных лет. В данной статье мы будем опираться на выводы В.С.Аванесова, которые сделаны на основе анализ научной литературы по соответствующей проблеме. В работе этого автора (1994 г.) отмечены следующие периоды развития возникновения тестов и развития тестового контроля:

- 1) донаучный период (с глубокой древности до конца XIX - начала XX века);
- 2) период научного обоснования теории тестов (с начала XX века до конца семидесятых годов);
- 3) период автоматизации процесса тестового контроля (с конца 70-х годов текущего столетия) ¹.

Процессы информатизации всех сфер жизни человека, в частности, образования позволили развивать оценку качества образования не только в пределах учебного процесса, одной страны, но и проводить на протяжении 20 лет сравнительные международные исследования уровня учебных достижений учащихся более 50 стран мира, обрабатывать огромные массивы результатов тестирования учащихся.

Ситуация в Казахстане

В Казахстане за 20 лет был создан фонд тестовых заданий и разработаны технологии тестирования, благодаря которым проводятся разные формы государственного контроля качества образования, такие как Единое национальное тестирование (ЕНТ) и Промежуточный государственный контроль (ПГК). В 2000 г. была создана Национальная система оценки качества образования (НСОКО) и центры прикладных исследований в области качества образования.

¹ Аванесов В.С. Научные проблемы тестового контроля знаний, Москва, 1994 г., -135 с. с.12-13.

Внедрение ЕНТ, комплексного тестирования абитуриентов и ПГК показало, что существует объективная необходимость оценки качества тестовых заданий.

Критические замечания педагогической, родительской общественности к качеству тестовых материалов на ЕНТ и ПГК дают основание для анализа используемых заданий с целью их совершенствования и обновления в соответствии с принятой новой парадигмой образования и изучения международного опыта в разработке заданий как средств измерения учебных достижений. В частности, на примере заданий, использованных в сравнительных международных исследованиях по программам TIMSS и PISA.

На первый взгляд, исследования по направлениям TIMSS и PISA несколько схожи, но между ними есть существенные отличия. Отличие исследований TIMSS и PISA состоит в следующем:

- TIMSS - сравнительная оценка естественнонаучной и математической подготовки учащихся средней школы в странах с различными системами образования и выявление факторов, влияющих на уровень этой подготовки (оценивается общеобразовательная подготовка по математике и естествознанию учащихся 4 и 8 классов);
- PISA – оценка способности выпускников основной школы применять полученные знания и умения в ситуациях лично и социально значимых, выходящих за пределы чисто учебных (оценивается математическая, естественнонаучная грамотность, грамотность чтения и уровень понимания текстов, умение решать проблемы не предметного содержания).

В 2003 году приоритетным направлением в исследованиях PISA было выявление уровня «математической грамотности». Основной целью этого исследования являлась оценка образовательных достижений учащихся 15-летнего возраста. В исследовании 2003 года проверялась сформированность компетентности на трех уровнях: уровне воспроизведения, уровне установления связей, уровне рассуждений.

Первый уровень (уровень воспроизведения) - это прямое применение известных фактов, стандартных приемов, распознавание эквивалентных представлений, узнавание знакомых математических объектов и свойств, выполнение стандартных процедур, применение известных алгоритмов и технических навыков, работа со стандартными, знакомыми выражениями и формулами, непосредственное выполнение вычислений.

Второй уровень (уровень установления связей) строится на репродуктивной деятельности по решению задач, которые, хотя и не являются стандартными, но все же знакомы учащимся или же выходят за рамки известного лишь в очень малой степени. Обычно в этих задачах больше требований к интерпретации решения, они предполагают установление связей между

разными представлениями ситуации, описанной в задаче, установление связей между данными из условия задачи.

Третий уровень (уровень рассуждений) строится как развитие предыдущего уровня. Для решения задач этого уровня требуются определенная интуиция и творчество в выборе математического инструментария, применение знаний из разных разделов программы, самостоятельная разработка алгоритма действий. Задания, как правило, более комплексные, включают больше данных; от учащихся часто требуется найти закономерность, провести обобщение и объяснить или обосновать полученные результаты².

В 2006 году приоритетным направлением в исследованиях PISA была проверка естественнонаучной грамотности, то есть умения учащихся применять свои естественнонаучные знания в реальной жизни.

Казахстан принял участие в международных сравнительных исследованиях учащихся в 2007 году. Это дает возможность пересмотреть существующий банк тестовых заданий с позиций новых требований к содержанию и формам современных заданий учебных достижений.

В Национальной академии образования им. Ы. Алтынсарина работу по анализу качества тестовых материалов начали проводить с 2001 года. Сравнительный анализ качества тестовых заданий проводился на протяжении трёх лет по следующим параметрам:

- 1) полнота охвата программного материала по дисциплине;
- 2) процентное соотношение представленных тем дисциплины;
- 3) наличие заданий, оценивающих интегрированные знания;
- 4) обеспечение равенства условий для участников тестирования, независимо от языка обучения;
- 5) соответствие требованиям конструирования заданий в тестовой форме;
- 6) обеспечение принципа однозначности формулировки тестовых заданий, корректности формулировки задания;
- 7) соблюдение принципа возрастающей трудности при отборе содержания тестовых заданий (от простого к сложному, в последовательности заданий в тесте);
- 8) наличие правильного ответа среди предложенных дистракторов (вариантов ответа);
- 9) время, требуемое на выполнение задания.

² Баранов В.Ю., Ковалева Г.С. и др. Основные результаты международного исследования образовательных достижений учащихся ПИЗА -2003, Москва, 2004 / Центр оценки качества образования ИСМО РАО. Национальный фонд подготовки кадров, с.11-15

Результаты анализа

В результате анализа количественных и качественных показателей тестовых заданий, предложенных абитуриентам на вступительных экзаменах в высшие учебные заведения в 2001-2003гг. можно сделать следующие выводы:

1) при отборе содержания тестовых заданий не учтены принципы репрезентативности и возрастающей трудности, комплексности и сбалансированности, системности; встречаются многократные повторы заданий в разных вариантах, часто в вариантах тестов отсутствуют значимые элементы содержания программного материала³;

2) при конструировании тестовых заданий не всегда соблюдаются требования к оформлению заданий с выбором одного или нескольких правильных ответов и заданий открытой формы, которые приняты в педагогических измерениях, то есть встречаются задания, где неправильно используется математическая терминология, где текст задания состоит из большого количества слов, что затрудняет восприятие его условия.

С 2006 г. были начаты разработки по научно-методическим основам проектирования системы заданий, используемых для измерения учебных достижений в системе непрерывного образования РК. Учитывая новые требования к содержанию и форме тестовых заданий в условиях реформирования образования и перехода на личностно-ориентированное и компетентностное образование, также изучая опыт международных сравнительных исследований по созданию измерительных материалов, нами были разработаны задания нового содержания.

Эти задания направлены на определение метапредметных или интегрированных знаний, умений и навыков, практикоориентированные и ситуационные задания на примере математики для основной средней школы.

Разработанные тестовые задания были апробированы на экспериментальных площадках в нескольких школах, а затем по полученным результатам был проведен анализ их качества с использованием математических моделей.

Основными показателями качества тестовых заданий являются мера трудности, дифференцирующая способность, а также коэффициента корреляции.

Использование математической модели Г. Раша

Использование компьютера при тестировании позволило проводить экспериментальную проверку качества тестовых заданий путем использования математических моделей.

³ Аванесов В.С. Методологические и теоретические основы тестового педагогического контроля, дисс. д.п.н, С-Пб. Гос ун-ет, 1994.

Нами сконструированы и систематизированы тестовые задания по математике за курс 7-11 классов. Оценка качества сконструированных заданий проводилась посредством математической модели G.Rasch.

Главная гипотеза данной модели заключается в следующем: «вероятность правильного ответа на задание теста зависит от трудности задания и от уровня подготовленности испытуемого».

Математическая модель G.Rasch имеет следующий вид⁴:

$$P_j \{x_{ij} = 1 | \beta_j\} = \exp(\theta - \beta_j) / (1 + \exp(\theta - \beta_j))$$

где $P_j(\theta)$ это вероятность правильного ответа на задание под номером j ;

θ - уровень знаний испытуемого;

β_j - уровень трудности j -го задания теста;

\exp – константа e , равное 2,71828.

Исходный уровень подготовленности испытуемого и исходный уровень трудности задания вычисляются при помощи следующих формул:

$$\theta = \ln \frac{p_i}{q_i}$$

$$\beta_j = \ln \frac{q_j}{p_j}$$

где p_i - доля правильных ответов испытуемого i по всем заданиям теста;

q_i - доля неправильных ответов того же испытуемого i по всем заданиям теста.

Числовые значения доли правильных и неправильных ответов можно получить из матрицы результатов тестирования, которая заполняется после получения ответов испытуемых на задания теста.

Оценка качества тестовых заданий предполагает вычисление коэффициента корреляции, которое по своей сути позволяет определить - выполняет ли задание возложенную на тест миссию?

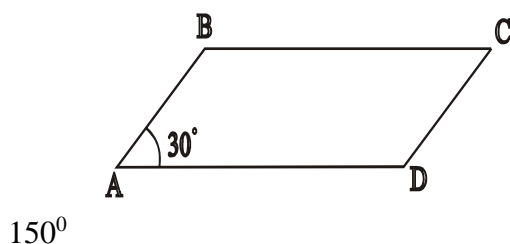
В качестве нижней границы включения заданий в тест обычно рассматриваются значения $r_{xy}=0,300$, и самой нижней, в исключительных случаях⁵, $r_{xy}=0,200$.

Теперь перейдем непосредственно к анализу тестовых заданий из теста по разделу «Четырехугольники» разработанных нами для проверки знаний учащихся 8 класса. Тестовое задание выглядит следующим образом:

⁴ Лекции В.С.Аванесова 2006 год.

⁵ В.С.Аванесов. Основы теории педагогических заданий / Педагогические измерения, №2, 2006 г, с.26-59

1. Найдите углы параллелограмма, если $\angle A = 30^\circ$ (Рис.1)



- А) $50^\circ; 50^\circ; 50^\circ$
 В) $30^\circ; 150^\circ; 150^\circ$
 С) $30^\circ; 30^\circ;$

Рисунок 1

Расчеты, проведенные нами по соответствующим формулам, дали возможность определить коэффициент корреляции $r_{XY} \approx 0,6$.

В соответствии с положениями, принятыми в теории педагогических измерений, полученный коэффициент свидетельствует о наличии у задания системных свойств, присущих для тестового задания. Это дает нам основание к проведению дальнейших вычислений для определения остальных показателей качества задания. Теперь можно построить графический образ задания по эмпирическим данным.

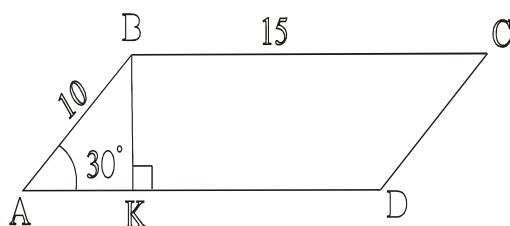


Рисунок 2 – Графический образ задания №1

Данный график можно интерпретировать следующим образом: дифференцирующее свойство задания среднее, а вероятность угадывания правильного ответа в нём достаточно большая, что указывает на определённый дефект.

Аналогичным образом провели расчеты и построили графический образ следующего задания:

5. ПЛОЩАДЬ ПАРАЛЛЕЛОГРАММА (рис.3) РАВНА



- А) $32,5 \text{ м}^2$
 В) 75 м^2
 С) 300 м^2

Рисунок 3

В нашем случае $r_{xy} \approx 0,1$. Это означает, что данное задание не может быть включено в тест и называться тестовым. Для наглядного подтверждения данного утверждения мы воспользовались математической моделью G.Rasch и построили графический образ задания (рис.4). Из которого видно, что по мере роста уровня подготовленности испытуемых вероятность правильного ответа у них странным образом вначале падает, но затем начинает расти.



Рисунок 4 – Графический образ задания №5

Построение графических образов заданий, вычисление коэффициентов корреляции – все эти этапы оценки качества заданий проводятся экспертами, профессиональными разработчиками. И все задания, сконструированные разработчиками должны проходить проверку на качество, прежде чем будут использованы в школах страны.

В нашем исследовании результатом анализа научно-педагогической литературы по проблемам разработки заданий и оценке их качества явилась выработка следующего алгоритма проектирования заданий:

1. Определение целевой функции задания.
2. Выбор содержания задания.
3. Выбор формы задания.
4. Конструирование заданий в соответствии с требованиями, предъявляемыми к данной форме заданий.
5. Систематизация заданий по принципу возрастающей трудности.
6. Апробация разработанных заданий.
7. Оценка качества заданий посредством математических моделей.
8. Коррекция и совершенствование заданий.

Данный алгоритм должен выполняться при разработке всех видов заданий (задачи, вопросы, упражнения, задания в тестовой форме и т.д.). Только таким образом, по нашему мнению, можно достичь положительных результатов в разработке заданий, которые в свою очередь позволят сформировать у школьников в стенах школы функциональные

грамотности (математическая грамотность, естественнонаучная грамотность и т.д.), компетентности.

Для того, чтобы результаты контроля были объективными, необходимо проводить обязательную оценку качества каждого задания. Все задания без исключения должны проходить качественный анализ. Оценкой качества заданий должны заниматься профессионалы, имеющие специальную подготовку. Проектирование системы измерителей должно проводиться в соответствии с алгоритмом, приведенным выше. При таких условиях можно добиться высокого качества измерителей, и вместе с ними и получить объективную информацию об учебных достижениях школьников и студентов.