

КЛАССИФИКАЦИЯ И СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ОЦЕНИВАНИЯ КАЧЕСТВА ПЕДАГОГИЧЕСКИХ ТЕСТОВ

К.А. Литвинов

В настоящее время тестирование как средство проверки знаний все глубже интегрируется в образовательный процесс. Как и любой другой инструмент, тест должен соответствовать определенным критериям качества. В противном случае его применение станет неэффективным или даже вредным.

Актуальность исследования обуславливается отсутствием четкой классификации методов оценивания качества теста. Определение взаимосвязи между методами позволит упростить их понимание, а также повысить эффективность и результативность практического использования.

Целью настоящей статьи является описание, классификация, систематизация и сравнительный анализ методов оценивания качества тестов. Использование подобной классификации позволит тестологу облегчить понимание задачи оценивания качества теста, упростить выбор наилучшего и наиболее подходящего метода и облегчить его реализацию.

Качество теста определяется оценками его основных параметров. На рис. 1 приведена схема классификации методов оценивания качества теста.

По нашему мнению, методы оценивания параметров можно разделить по этапу применения на априорные и апостериорные.

Сразу после написания заданий их анализ проводится экспертными методами. После проведения пробного тестирования применяются математическая обработка данных. Соответственно, существующие методы также можно декомпозировать на экспертные и статистические.

Экспертные методы

Суть экспертного оценивания качества теста – субъективный анализ тестовых заданий с последующим формированием качественной или

количественной оценки отдельных заданий и/или теста в целом. Наиболее активно экспертные методы применяются на этапе разработки теста.



Рис. 1. Признаки классификации методов оценивания качества

Статистические методы

Статистические методы [2-11] сегодня наиболее популярны благодаря хорошо проработанному математическому аппарату и повсеместному распространению мощной вычислительной техники. Существует две основополагающих теории тестирования: классическая и современная.

Классическая теория тестирования

Создателем классической теории тестов (Classical Theory of mental tests) является Ч. Спирмен (Charles Edward Spearman) [11]. Большой вклад в развитие теории внес Л. Гуттман (Louis Guttman) [9]. Однако впервые всесторонне и полно классическая теория тестов изложена труде Г. Гулликсена (Harold Gulliksen) [8]. Из отечественных исследователей впервые описание этой теории дал В.Аванесов [1]. В работе М. Чельшковой [6] рассматривается методика статистического обоснования качества теста.

В классической теории измеряемый балл испытуемого складывается из суммы истинного балла и случайной ошибки.

Современная теория тестирования

Принято считать, что классическая теория имеет ряд недостатков, главный из которых – зависимость результатов измерения от инструмента измерения (конкретного теста). Неудовлетворенность такой ситуацией и привела к созданию Item Response Theory (IRT).

Основным предположением IRT является наличие взаимосвязи между наблюдаемыми результатами учебных достижений, латентными свойствами самих тестируемых и характеристиками используемых тестовых заданий.

Уровень сложности задания и уровень подготовленности испытуемого в IRT рассматриваются как переменные, которые влияют на результат выполнения задания испытуемым. В рамках IRT применяются несколько моделей оценивания качества теста:

Однопараметрическая модель Раша

Разработана Г. Рашем (Georg Rasch) [10]. С точки зрения оценивания качества теста помощью этой модели можно определить только сложность задания и сложность теста в целом.

Двухпараметрическая модель Бирнбаума

Разработана американским статистиком А. Бирнбаумом (Allan Birnbaum) [7]. Имеет дополнительный параметр тестового задания, характеризующий дифференцирующую способность задания.

Трехпараметрическая модель Бирнбаума

Имеет еще один дополнительный параметр, характеризующий правильность ответа, если ответ угадан.

Четырехпараметрическая модель Аванесова

По мнению В.С. Аванесова, даже отлично знающие студенты могут иногда терпеть неудачу по причинам, связанным не со знанием предмета, а с какими-то другими факторами, которые иногда трудно определить. На подобные случаи рассчитана четырехпараметрическая модель измерения латентных качеств личности.

Существующие математические разработки применимы только для бинарной модели определения истинности ответов.

Следующий признак классификации – модель определения истинности ответа. Рассмотрим существующие на сегодняшний день модели подробнее:

- бинарная;
- алгебраическая;
- расчетная;
- нечеткая.

Бинарная модель – наиболее распространенная, простая и популярная модель определения истинности ответов на тестовое задание. Согласно бинарной модели каждый вариант ответа однозначно определен в категориях «правильно – неправильно».

Алгебраическая модель предоставляет автору тестового задания возможность определить степень истинности каждого варианта ответа. Каждый из предлагаемых ответов характеризуется числом от 0 до 1, при этом 0 соответствует абсолютно неправильному ответу, а 1 – исчерпывающе правильному.

Расчетная модель оценивания истинности ответов предложена И.Д. Рудинским. Суть модели заключается в использовании для расчета истинности вариантов ответов информации о структурных свойствах различных типов заданий.

Нечеткая модель также предложена И.Д. Рудинским [5]. Эта модель позволяет использовать более широкую шкалу оценивания. Степень истинности каждого ответа определяется функцией принадлежности, заданной на лингвистической переменной, значения которой используются в качестве оценочных категорий.

Для каждой из упомянутых выше моделей разработана своя методология статистической обработки результатов тестирования. В классической теории тестирования его создателями стали Спирман, Гутман, Гулликсен [8, 9, 11], а в IRT заметных результатов добились Раш и Бирнбаум [7, 12]. Для

алгебраической и нечеткой моделей такой механизм был предложен авторами настоящей статьи в 2009 году [3, 4]. Тогда же была разработана универсальная модель оценивания качества теста, инвариантная к применяемым моделям оценивания истинности ответов [3].

На практике возможна ситуация, когда в рамках одной модели можно оценить один и тот же параметр разными методами. Рассмотрим показатели качества теста и методы их оценивания. Обратимся к схеме классификации на рис. 2.



Рис. 2. Показатели качества теста

Результаты классификации методов оценивания и параметров качества теста сведены в таблицу 1.

Классификация методов оценивания качества теста

Признаки	По виду		По применению в теориях тестирования		По модели оценивания истинности ответа				По этапу применения	
	Экспертные	Статистические	Классическая	Современная	Бинарная	Алгебраическая	Нечеткая	Универсальная	Априорные	Апостериорные
Параметры										
Дисперсия		●	●		●	●	●	●		●
Сложность	●	●	Вычисление соотношения правильных и неправильных ответов	Метод получения коэффициента из логистической функции	●	●	●	●	●	●
Корреляция		●	Метод Пирсона		●	●	●	●		●
Объективность	●		●	●					●	
Репрезентативность		Метод сравнения графиков распределения	●		●	●	●	●		●
Дискриминативность		●	Дисперсионный метод; метод разницы достижений; точечно-бисериальная корреляция	Метод определения крутизны характеристической кривой в двухпараметрической модели	●	●	●	●		●
Надежность		●	Метод Спирмена-Брауна		●	●	●	●		●
Валидность	Содержательная экспертиза	Корреляция по критерию	●		●	●	●	●	●	●

В ячейках указаны названия методов, используемых при оценивании соответствующего параметра.

Пустая ячейка означает отсутствие метода оценивания указанного параметра по соответствующему признаку.

Символ ● означает наличие формулы либо метода для оценивания параметра.

Предлагаемая классификация упрощает задачу выбора параметров и методов, используемых для оценивания качества теста, а также повышает наглядность связей и проясняет генезис описанных методов и моделей. Подавляющее большинство методов и параметров оценивания качества теста предлагались и развивались разными авторами. Однако непрерывное возрастание востребованности тестовых технологий и повышение интереса к ним со стороны педагогического сообщества наряду с все более очевидным требованием стандартизации применяемых тестовых материалов обуславливает необходимость интеграции этих методов в универсальных системах автоматизированного контроля учебных достижений, основанных на различных моделях оценивания истинности ответов.

Литература

1. Аванесов В.С. Основы научной организации педагогического контроля в высшей школе. М.: 1989. 167 с.
2. Ким В.С. Тестирование учебных достижений: монография. Уссурийск: Издательство УГПИ, 2007. 214 с.
3. Литвинов К.А., Рудинский И.Д. Универсальная модель оценивания качества тестового контроля знаний // Сборник трудов Международной научно-практической конференции «Информационные технологии в образовании, науке и производстве». Серпухов: 2009. С. 292-297.
4. Литвинов К.А., Рудинский И.Д. Оценивание качества тестового контроля знаний при использовании нечеткой модели оценивания истинности ответов // Известия БГА РФ. Психолого-педагогические науки. 2009. № 4(8). С. 67-79.
5. Рудинский И.Д. Структурные основы тестологии. Калининград: Изд-во ФГОУ ВПО «КГТУ», 2010. 249 с.
6. Чельшкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. М.: Логос, 2002. 432 с.

7. Birnbaum A. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F.M. Lord and M.R. Novick. Statistical Theories of Mental Test Scores. Reading Mass.: Addison-Wesley, 1968.

8. Gulliksen H. Theory of mental tests. N-Y. Wiley, 1950. 486 pp.

9. Guttman, L. A special review of Harold Gulliksen, Theory of mental tests. Psychometrika. 1953. 18. P. 123-130.

10. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, 1960, Danish Institute of Educational Research. (Expanded edition, Chicago, 1980, The University of Chicago Press).

11. Spearman C. Correlation calculated from faulty data // British Journal of Psychology. 1910. Vol. 3. № 2. P. 271-295.