

УДК 658.051.012

М.С. Мазорчук, В.С. Добряк, К.А. Гончарова

Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков

## ОЦЕНКА ПАРАМЕТРОВ ТЕСТА НА ОСНОВЕ МОДЕЛИ IRT

*В работе проанализированы существующие модели оценки качества тестов. На основе результатов проведенного анализа предложена система тестирования, учитывающая индивидуальные способности опрашиваемых. Разработана математическая модель оценки параметров качества тестов с использованием современной теории тестирования IRT. На примере показаны расчеты основных параметров тестовых заданий.*

**Ключевые слова:** классическая теория тестирования, теория моделирования и параметризации тестов, автоматизированная система тестирования, надежность теста, информационная функция теста, характеристическая кривая теста.

### Введение

Актуальность автоматизации образовательного процесса обуславливается необходимостью уменьшения затрат на проведение обучения, увеличением количества студентов, повышением качества обучения за счет применения современных информационных технологий. В настоящее время проблемы создания единой образовательной среды являются приоритетными для многих высших и средних учебных заведений [1].

Одним из наиболее трудоемких этапов образовательного процесса является контроль достигнутого уровня знаний. Системы автоматизированного тестирования решают проблему сокращения временных и трудовых затрат, однако уносят решаемых ими задач, определяемая во многом ограниченностью используемого формально-математического аппарата, отсутствием индивидуального подхода к тестируемому и учета их личностных характеристик, существенно снижают ожидаемый эффект.

Методика адаптивного тестирования позволяет учесть индивидуальные особенности студента, построить гибкую процедуру оценивания знаний, максимально приближенную к устному опросу, лишённую субъективного фактора, решить проблему максимальной информативности результатов тестирования.

Таким образом, проблема разработки и реализации системы, позволяющей проводить эффективную систему оценивания знаний студентов, анализировать полученные результаты и формировать рекомендации по составлению тестовых заданий является актуальной и требует решения.

### 1. Постановка задачи исследования

Целью задачи исследования является создание системы адаптивного тестирования для студентов технических ВУЗов.

Объектом исследования является процесс контроля знаний в техническом ВУЗе.

Предметом исследования являются методы и модели разработки тестовой системы контроля знаний.

Основные задачи исследования следующие:

1. Рассмотреть основные существующие модели компьютерного тестирования, выделить их преимущества и недостатки.
2. Разработать математическую модель системы контроля знаний на основе алгоритма адаптивного тестирования.
3. Разработать алгоритмическую модель компьютерного тестирования.

### 2. Решение поставленной задачи

Для решения поставленной задачи предлагается проанализировать существующие модели оценки параметров теста и тестовых заданий, выделить их преимущества и недостатки.

Согласно классической теории тестов наблюдаемый балл ( $X$ ) включает в себя истинный балл ( $T$ ) испытуемого и некоторую ошибку измерения ( $E$ ). Формула наблюдаемого балла будет выглядеть следующим образом:

$$X = T + E.$$

Истинный балл определяется как показатель испытуемого в гипотетической генеральной совокупности заданий бесконечного теста.

Ведущая идея IRT сводится к обоснованию возможности эффективного прогнозирования результатов тестирования на задания различного уровня трудности. Такой прогноз особенно необходим в системах адаптивного обучения и адаптивного тестового контроля. Прогноз основан на следующих утверждениях:

- 1) существуют латентные параметры личности, недоступные для непосредственного наблюдения. В тестировании это уровень подготовленности испытуемого и уровень трудности задания;

2) существуют индикаторные переменные, связанные с латентными параметрами, доступные для непосредственного наблюдения. По значениям индикаторных переменных можно судить о значениях латентных параметров;

3) оцениваемый латентный параметр должен быть одномерным. Это означает, что, например тест, должен измерять знания только в одной, четко заданной, предметной области.

### 3. Сравнительный анализ классической теории тестирования и теории моделирования и параметризации тестов IRT

Основные различия между классической теорией тестирования и IRT представлены в табл. 1.

Таблица 1

Основные различия между классической теорией тестирования и IRT

| Область  | Классическая теория тестирования  | Теория моделирования и параметризации тестов IRT  |
|--|---|---|
| Модель   | Линейная  | Нелинейная  |
| Рассматриваемый объект   | Тест  | Тестовое задание  |
| Отношение «тестовое задание – уровень подготовки тестируемого» | Не определено   | Характеристическая функция тестового задания  |
| Ошибка   | Стандартная ошибка измерения  | Стандартная ошибка оценки   |
| Статистики тестового задания                                   | $p$ (трудность тестового задания), $g$ (дискриминативность тестового задания) | $a$ (трудность тестового задания), $b$ (дискриминативность тестового задания), $c$ (псевдогауссовский параметр) |
| Размер выборки (необходимый для оценки тестового задания)      | 200 – 500 человек   | Зависит от модели IRT, превышает 500 человек  |

Как видно из таблицы, преимущества IRT над классической теорией тестов заключаются в следующем:

1. IRT превращает измерения, выполненные в дихотомических и порядковых шкалах, в линейные измерения, в результате качественные данные анализируются с помощью количественных методов.

2. Мера измерения параметров модели Раша является линейной, что позволяет использовать ши-

рокий спектр статистических процедур для анализа результатов измерений.

3. Оценка трудности тестовых заданий не зависит от выборки испытуемых, на которых она была получена.

4. Оценка уровня подготовленности испытуемых не зависит от используемого набора тестовых заданий.

5. Неполнота данных (пропуск некоторых комбинаций испытуемый - тестовое задание) не является критичной.

### 4. Математическая модель IRT

Основной задачей IRT является переход от индикаторных переменных к латентным параметрам.

В IRT устанавливается связь между двумя множествами значений латентных параметров. Первое множество составляют значения латентного параметра, определяющего уровень подготовки испытуемых  $\theta_i$ , где  $i$  – номер испытуемого, изменяющийся в интервале от 1 до  $n$  ( $n$  – количество испытуемых). Второе множество составляют значения латентного параметра, характеризующего трудность  $j$ -го задания  $\beta_j$ . Индекс  $j$  меняется в пределах от 1 до  $m$ , где  $m$  – количество заданий в тесте[2].

Георг Раш предположил, что уровень подготовленности испытуемого  $\theta_i$  и уровень трудности задания  $\beta_j$  размещены на одной шкале и измеряются в одних и тех же единицах - логитах. В качестве математической модели, связывающей успех испытуемого с уровнем его подготовки и трудностью задания, выбирается логистическая функция (функция успеха). Аргументом функции успеха испытуемого является разность  $\theta_i - \beta_j$ . Если эта разность положительна и велика, то соответственно высока вероятность достижения успеха  $i$ -го испытуемого в  $j$ -м задании. Если же эта разность отрицательна и велика по модулю, то вероятность достижения успеха  $i$ -го испытуемого в  $j$ -м задании будет низкой. Вид функции успеха зависит от математической модели IRT.

Виды математических моделей IRT:

- а) однопараметрическая модель Г.Раша;
- б) двухпараметрическая модель А.Бирнбаума;
- с) трехпараметрическая модель А.Бирнбаума.

В IRT аналогом стандартной ошибки измерения является стандартная ошибка оценки (COO). Для расчета COO необходима информационная функция задания ( $I_i(\theta)$ ), определяемая для однопараметрической модели как произведение вероятности правильного ответа ( $P_i(\theta)$ ) на данное задание на вероятность неправильного ответа  $Q_i(\theta)$ :

$$I_i(\theta) = P_i(\theta) \cdot Q_i(\theta).$$

График информационной функции тестового задания представлен на рис. 1.

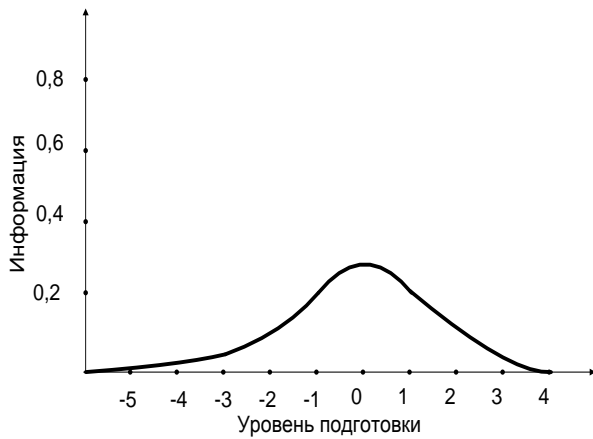


Рис. 1. Информационная функция тестового задания

Тестовое задание, ответ на которое знают все тестируемые, не даст никакой информации, так же, как и задание, ответ на которое не знает никто. Полезную информацию получаем, когда некоторые тестируемые знают ответ на задание, а некоторые – нет.

Информационная функция теста вычисляется как сумма информационных функций тестовых заданий:

$$I(\theta) = D^2 \cdot \sum_{i=1}^n I_i(\theta),$$

где  $D$  – поправочный коэффициент ( $D = 1,7$ ), необходимый для приближения распределения логистической вероятности к нормальному закону распределения.

В компьютерных системах адаптивного тестирования информационная функция теста высчитывается после ответа на каждое тестовое задание и т.о. служит критерием окончания процедуры тестирования.

Одним из центральных понятий теории IRT является характеристическая кривая тестового задания (Item Characteristic Curve – ICC). Графики ICC представляют собой функциональную зависимость вероятности правильного ответа на задания от латентной характеристики  $\theta$ , лежащей в основе выполнения заданий теста (рис. 2).

Как следует из графика, с возрастанием оценки латентной характеристики увеличивается вероятность правильного ответа на задание теста. Преимущество теории IRT состоит в том, что она позволяет нам видеть, как вероятность правильного ответа зависит от латентной характеристики [6].

### 5. Алгоритмическая модель компьютерного тестирования

Для реализации алгоритма адаптированного тестирования использовалась математическая модель IRT.

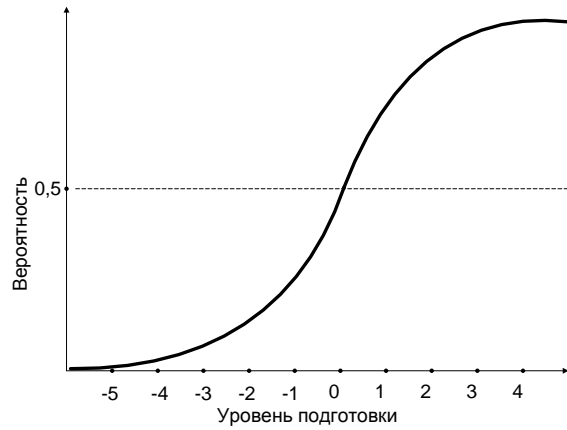


Рис. 2. Характеристическая кривая тестового задания

Основным недостатком при неадаптивном тестировании является то, что для испытуемого каждое последующее задание теста формируется путём случайного выбора заданий из всего множества тестовых заданий. Такой подход может привести к ситуациям, когда слабому студенту попадает сложный тест и как следствие – практическое отсутствие ответов, и на оборот – сильный студент может получить легкий тест. В таком случае испытуемый не сможет реализовать свои способности. Таким образом, для обеспечения максимальной информативности результатов необходимо, чтобы средняя сложность тестового задания для тестируемого соответствовала его гипотетическому уровню подготовки. Реализацию такого подхода это можно осуществить с помощью адаптивного тестирования. В модели адаптированного тестирования очередной шаг совершается только после оценки результатов выполнения предыдущего шага. Так, после выполнения очередного тестового задания каждый раз подбирается такое задание трудность которого рассчитывается исходя из предыдущего ответа (верный или неверный ответ).

Пример процедуры оценки знаний, построенной на алгоритме адаптированного тестирования показан на рис. 3.

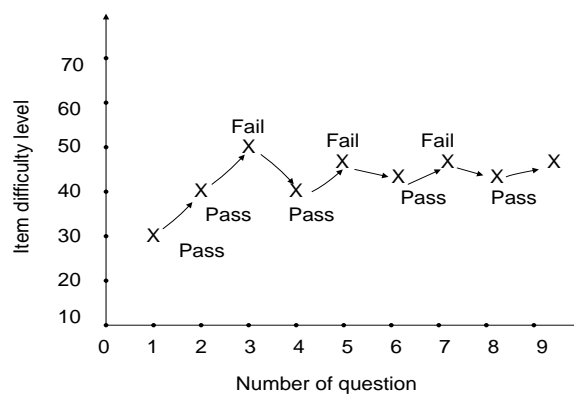


Рис. 3. Пример процедуры оценки знаний, при адаптированном тестировании

Алгоритм отбора заданий строится по принципу обратной связи: при правильном ответе испытуемого очередное задание выбирается более трудным, а при неверном ответе – более легкое задания, чем то, на которое испытуемым был дан неверный ответ [3].

На основе проведенного анализа построена алгоритмическая модель, представленная на рис. 4.

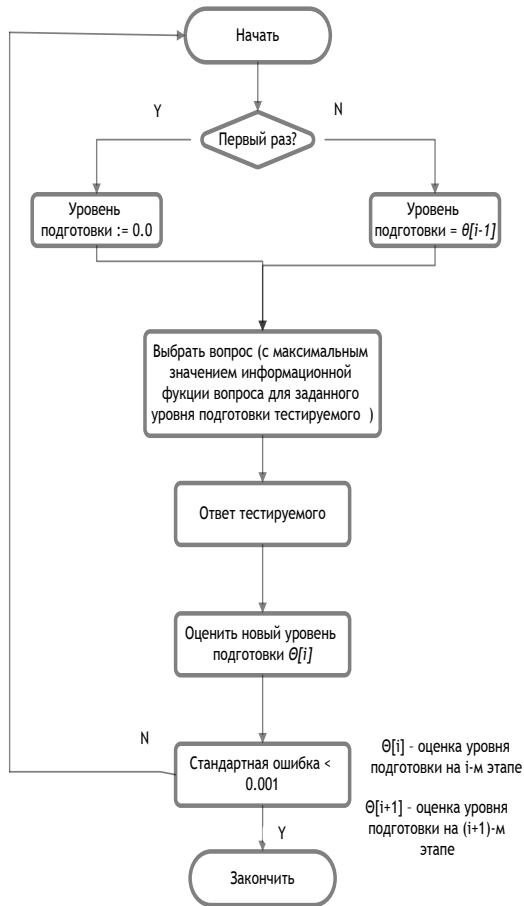


Рис. 4. Блок-схема алгоритма процедуры адаптивного тестирования

Согласно данному алгоритму тестируемый осуществляет переход между тестовыми заданиями с различными уровнями сложности. Критерием остановки тестирования служит значение стандартной ошибки оценки.

### 6. Пример реализации компьютерной системы адаптивного тестирования

На базе разработанной алгоритмической модели была реализована компьютерная система адаптивного тестирования. Примеры работы данной компьютерной системы представлены на рис. 5, 6.

В данном примере осуществлен расчет таких показателей тестовых заданий, как информационные функции, и графическое представление полученных расчетов. Компьютерная система разработана с учётом пользования отдельно как преподавателем, так и студентом. Вход в систему сопровождается вводом пароля и логина.

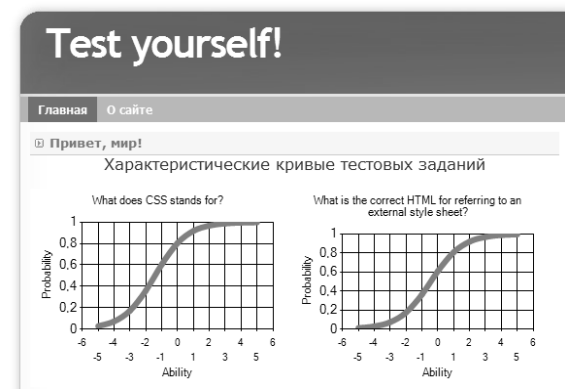


Рис. 5. Пример построение характеристических кривых тестовых заданий



Рис. 5. Пример построения информационных кривых тестовых заданий

Преподаватель имеет возможность просмотреть статистические оценки теста, информационную функцию тестового задания, оценить уровень сложности вопросов теста и т.д. (например рис. 5, 6).

Информация о каждом студенте и тестовые задания хранятся в базе данных системы, если перед прохождением теста студент не был зарегистрирован в БД, система попросит внести необходимые данные об испытуемом. Результаты тестирования также сохраняются в БД.

Таким образом, разработанная компьютерная система обладает такими свойствами как высокая информативностью, простотой в использовании и удобным интерфейсом.

### Заключение

Таким образом, в данной статье рассмотрены существующие теории тестирования, выделены их преимущества и недостатки. На основе проведенного сравнительного анализа разработана алгоритмическая модель оценки уровня подготовки тестируемых, построенная на основе теории моделирования и параметризации тестов IRT. Использование предложенной алгоритмической модели даёт возможность образовательным учреждениям проводить полную и эффективную оценку уровня подготовки испытуемых.

## Список литературы

1. Информатизация образования: направления, средства, технологии: пособие для системы повышения квалификации / Под общ. ред. С.И. Маслова. – М.: Издательство МЭИ, 2004. – 868 с.
2. Нейман Ю.М. Введение в теорию моделирования и параметризации педагогических тестов [Основные концепции классической теории тестирования] / Ю.М. Нейман. – М.: Прометей, 2000. – 300 с.
3. Rash G. Probabilistic model for some intelligence and attainment tests [Basic ideas of IRT ] / G. Rash. – Chicago: Mess Press, 1993.
4. Чельщикова М.Б. Теория и практика конструирования педагогических тестов: учебн. пос. [Адаптированное тестирование] / М.Б. Чельщикова – М.: Логос, 2002. – 432 с.

5. Thissen, D Reliability and measurement precision [Test reliability] / D. Thissen. – Mahwah, NJ: Lawrence Erlbaum associates.

6. Baker F.B. The Basics of Item Response Theory [Rash and Birnbaum models ] / F.B. Baker. – Hieneman, Portsmouth, New Hampshire, 2001. – 185 с.

Поступила в редколлегию 23.09.2010

**Рецензент:** д-р техн. наук, проф. О.Е. Федорович, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков.

## ОЦІНКА ПАРАМЕТРІВ ТЕСТУ НА ОСНОВІ МОДЕЛІ IRT

М.С. Мазорчук, В.С. Добряк, К.А. Гончарова

У роботі проаналізовано існуючі моделі оцінки параметрів тестів. На основі аналізу запропоновано єдину систему тестування та розрахунку головних показників тестових завдань. Розроблена математична модель отримання показників тестів на основі теорії моделювання та параметризації тестів IRT. Наведено приклад розрахунків головних параметрів тестових завдань.

**Ключові слова:** класична теорія тестування, теорія моделювання та параметризації тестів, автоматизована система тестування, надійність тесту, інформаційна функція тесту, характеристична крива тесту.

## TEST PARAMETERS ESTIMATION BASED ON IRT MODEL

M.S. Mazorchuk, V.S. Dobryak, K.A. Goncharova

In this article existing models of test parameters estimation are analyzed. The unified system of testing and calculating of the test basic parameters is offered based on the results of the analysis. A mathematical model for test parameter calculating is based on Item Response Theory. The example shows the calculations of basic parameters of the test tasks.

**Keywords:** Classical Test Theory, Item Response Theory, Automated Test System, test reliability, Test Information Function, Item Characteristic Curve.