

# Статистический анализ информационной системы измерения знаний на основе IRT-моделей

Г.Л. Гринберг<sup>1</sup>, Л.М. Любчик<sup>1</sup>

**Abstract** – The statistical analysis problem of educational assessment system is considered. Based on IRT approach both experimental test and score scaling characteristics is investigated using real statistical data.

**Keywords** - Educational Measurement, Item Response Theory, Score Scale, Testing.

## I. ВВЕДЕНИЕ

Качество подготовки специалистов в области компьютерных и системных наук в значительной мере определяется уровнем подготовки абитуриентов. В Украине в настоящее время прием в Вузы осуществляется по результатам внешнего независимого оценивания (ВНО), проводимого на основе тестирования, при этом процедура измерения знаний включает в себя собственно тестирование, шкалирование, обработку и интерпретацию результатов. Качество отбора абитуриентов непосредственно связано с точностью измерения уровня их подготовленности, определяемой информационно-статистическими характеристиками системы измерения знаний, анализ которых может быть проведен на основе теории тестирования IRT (Item Response Theory) [1].

## II. СТРУКТУРА СИСТЕМЫ ИЗМЕРЕНИЯ ЗНАНИЙ.

Структура системы измерения знаний, используемой при проведении ВНО, представлена на рис. 1. В блоке тестирования входные латентные переменные "Уровень подготовленности абитуриентов" и "Уровень сложности теста" преобразуются в измеряемые индикаторы – тестовые баллы, которые в свою очередь в блоке шкалирования преобразуются в рейтинговые баллы.



Рис.1. Структура информационной системы измерения знаний

Существенным для дальнейшего анализа является то, что характеристика блока тестирования, описываемая моделью Раша [1], в действительности неизвестна и может быть восстановлена лишь на основе статистических данных результатов тестирования.

Кроме того, характеристика блока шкалирования, определяемая по методике УЦОЯО на основе метода эквипроцентильного преобразования [2], зависит от полученного распределения тестовых баллов, то есть, фактически, от уровня подготовки всего тестируемого контингента. Указанные особенности должны учитываться при проведении анализа точностных характеристик информационно-измерительной системы измерения знаний.

## III. АНАЛИЗ ПРОЦЕССА ТЕСТИРОВАНИЯ

Характеристики процесса тестирования на примере теста по математике были получены на основе двухпараметрической модели Раша с использованием реальных статистических данных из официального отчета о проведении ВНО 2010 г. [3]. Предварительно было выполнено агрегирование данных с разбиением множества тестовых заданий на группы близких уровней сложности, для которых вычислялся средний процент тестируемых, верно решивших задания соответствующей группы, и множества абитуриентов на группы с близким уровнем подготовленности, для каждой из которых вычислялся средний тестовый балл. Параметры модели теста оценивались по методу максимального правдоподобия с использованием реальных статистических данных.

На рис. 2 приведены полученные характеристические кривые теста (а)

$$P_R(\bar{\beta}^*, \theta) = \sum_{i=1}^N w_i \cdot \frac{e^{\alpha(\beta_i^* - \theta)}}{1 + e^{\alpha(\beta_i^* - \theta)}} \quad (1)$$

и информационные функции теста (б)

$$I_R(\theta, \bar{\beta}^*) = \sum_{i=1}^N w_i \cdot \frac{e^{\alpha(\theta - \beta_i^*)}}{[1 + e^{\alpha(\theta - \beta_i^*)}]^2} \quad (2)$$

для сбалансированного В и реального R теста по математике ВНО 2010 г. Здесь  $w_i$  - относительные веса групп тестовых заданий,  $\beta_i^*$  - оценки уровней сложности соответствующих групп,  $\theta$  - уровень подготовленности,  $P_R$  - средняя доля правильных ответов.

Анализ полученных результатов позволяет сделать вывод о несбалансированности теста, а именно, реальный средний уровень подготовленности выпускников оказался, видимо, ниже ожидаемого, что проявилось в том, что тестовые задания оказались слишком сложны для значительной доли проходивших тестирование.

<sup>1</sup> Национальный технический университет «Харьковский политехнический институт», ул. Фрунзе, 21, Харьков, 61002, УКРАИНА, E-mail: [glngrinberg@gmail.com](mailto:glngrinberg@gmail.com), [lyubchik.leonid@gmail.com](mailto:lyubchik.leonid@gmail.com)

Этот вывод, к сожалению, подтверждается и существенной левой асимметрией приведенной в [3] гистограммы распределения выпускников по количеству набранных тестовых баллов.

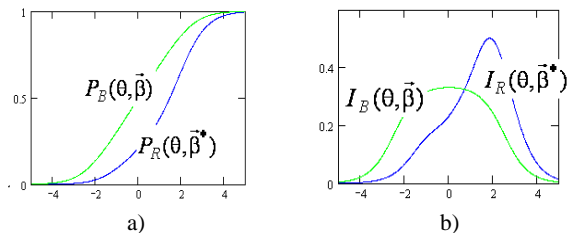


Рис.2. Характеристические кривые и информационные функции теста по математике

#### IV. АНАЛИЗ ПРОЦЕССА ШКАЛИРОВАНИЯ

На рис. 3 представлены построенные по статистическим данным функции шкалирования (а)  $P_{100-200}(\theta, \vec{\beta}^*) = R(P_{53}(\theta, \vec{\beta}^*))$ , выполняющие конвертацию тестовых баллов по шкале (0-53) в рейтинговые баллы по шкале (100-200), и нормированные характеристические кривые (b) сбалансированного и реального теста по шкале тестовых  $P_{53}(\theta, \vec{\beta}^*)$  и рейтинговых  $P_{100-200}(\theta, \vec{\beta}^*)$  баллов.

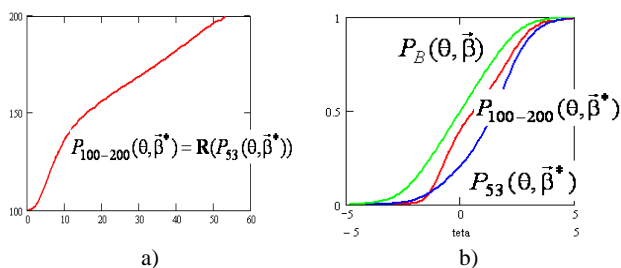


Рис.3. Функция шкалирования и характеристические кривые теста для шкал тестовых и рейтинговых баллов

Из полученных результатов следует, что реализованная функция шкалирования является монотонной, и, следовательно, не нарушает порядок ранжирования абитуриентов в зависимости от их уровня подготовленности. В то же время указанная функция является существенно нелинейной и неравномерной, что приводит к искусственному завышению рейтинговых баллов для абитуриентов с относительно низким уровнем подготовленности и должно учитываться при анализе и интерпретации окончательных результатов тестирования.

Полученные зависимости позволяют также исследовать влияние различных факторов на качество процесса измерения знаний. В частности, большой интерес представляет анализ влияния фактора угадывания [4].

На рис. 4 представлены полученные нормированные характеристические кривые сбалансированного В (а) и реального R (b) теста по математике 2010 г. для различных методов шкалирования в зависимости от

вероятности угадывания  $\rho$ , построенные для трехпараметрической модели Бирнбаума [1]:

$$P_B^p(\theta, \vec{\beta}) = \rho + \frac{(1-\rho)}{N} \cdot \sum_{i=1}^N \frac{e^{\alpha(\theta-\beta_i)}}{1 - e^{\alpha(\theta-\beta_i)}}, \quad (3)$$

$$P_R^p(\theta, \vec{\beta}^*) = \rho + (1-\rho) \cdot \sum_{i=1}^N w_i \cdot \frac{e^{\alpha(\theta-\beta_i^*)}}{1 - e^{\alpha(\theta-\beta_i^*)}}. \quad (4)$$

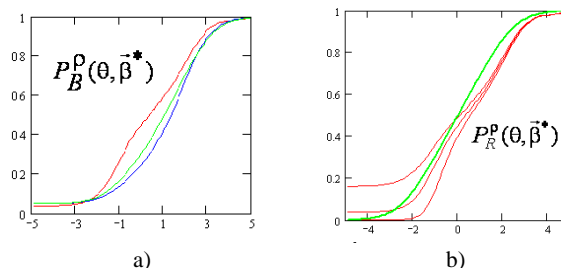


Рис.4. Характеристические кривые тестов с учетом вероятности угадывания

Результаты расчетов показывают, что применение эквипроцентильной рейтинговой шкалы (100-200) делает характеристику несбалансированного теста крайне чувствительной к фактору угадывания и, как следствие, снижает ее разрешающую способность в области низких уровней подготовленности.

#### V. ЗАКЛЮЧЕНИЕ

Полученные результаты позволяют сделать вывод о том, что используемая в системе ВНО методика измерения знаний требует определенных коррекций и уточнений. В первую очередь это касается структуры тестов и их информационных характеристик, а также методов шкалирования результатов тестирования. При этом использование моделей IRT совместно со статистическими методами анализа данных тестирования позволяет выявить существующие проблемы и предложить пути их решения.

#### СПИСОК ССЫЛОК

- [1] Ким В.С. Тестирование учебных достижений. - Усурийск: Изд-во УГПИ, 2007. - 214 с.
- [2] О. І. Ляшенко, С. А. Раков. Тестові технології і моніторинг в системі освіти України: стан і перспективи розвитку // Вісник ТІМО. - 2008. - № 11-12. - С. 67-70.
- [3] Офіційний звіт про проведення зовнішнього незалежного оцінювання знань випусників загальноосвітніх навчальних закладів України в 2010 році // [Електронний ресурс]. - Режим доступу: <http://www.testportal.gov.ua/index.php/text/vidp/>
- [4] Чельшкова М. Б. Теория и практика конструирования педагогических тестов. М.: Логос, 2002. - 432 с.