

## МОДЕЛІ ОЦІНЮВАННЯ ТЕСТОВИХ ЗАВДАНЬ РІЗНИХ ТИПІВ ТА ВИЗНАЧЕННЯ ЇХ РІВНЯ СКЛАДНОСТІ

В. П. Марценюк, Н. М. Гандзюк

*Тернопільський державний медичний університет імені І. Я. Горбачевського*

У статті описується технологія оцінювання тестових завдань, а також розподіл завдань різних типів за рівнями складності. При цьому розподілі пропонується використовувати інтегровану функціональну модель, що дозволяє застосовувати різні способи розподілу тестових завдань, включених в тест, залежно від їх типів.

**Ключові слова:** модель, рівень складності, тестові завдання.

## МОДЕЛИ ОЦЕНИВАНИЯ ТЕСТОВЫХ ЗАДАНИЙ РАЗНЫХ ТИПОВ И ОПРЕДЕЛЕНИЯ ИХ УРОВНЯ СЛОЖНОСТИ

В. П. Марценюк, Н. М. Гандзюк

*Тернопольский государственный медицинский университет имени И. Я. Горбачевского*

В статье описывается технология оценивания тестовых заданий, а также распределение заданий разных типов по уровням сложности. При этом распределении предлагается использовать интегрированную функциональную модель, что позволяет применять разные способы распределения тестовых заданий, включенных в тест, в зависимости от их типов.

**Ключевые слова:** модель, уровень сложности, тестовые задания.

## MODELS OF EVALUATION OF TEST TASKS OF DIFFERENT TYPES AND DETERMINATION OF THEIR LEVEL OF COMPLICATION

V. P. Martsenyuk, N. M. Handzyuk

*Ternopil State Medical University by I. Ya. Horbachevsky*

The paper is devoted to the test items marking technology and distribution of different types items by the levels of complication. At distribution of test items on complication levels it is offered to use the integrated functional model which allows to apply different ways of the test items distribution included in the test.

**Key words:** model, level of complication, test tasks.

**Вступ.** Контроль знань є важливою частиною процесу навчання і дозволяє отримати всесторонню оцінку рівня знань студентів. Однією з форм контролю знань, що добре себе зарекомендувала, є тестування.

Тестування, не дивлячись на велику кількість певних недоліків, є єдиним по-справжньому технологічним засобом для виміру рівня знань і незамінним інструментом, що дозволяє реалізувати ефективний педагогічний контроль і належним чином організувати управління навчальним процесом.

Для визначення об'єктивного стану рівня підготовки студентів вищих медичних (фармацевтичних) навчальних закладів України, для прогнозування розвитку галузі охорони здоров'я ми повинні володіти ефективни-

ми методиками оцінювання теоретичних знань і практичних вмінь студентів [1]. Світовій медичній школі відомий цілий ряд підходів та шляхів для оцінки знань студентів-медиків. Багато з них вже успішно апробовані та використовуються, інші потребують доопрацювання та запровадження в Україні з урахуванням традицій та пріоритетів національної медичної школи.

**Мета дослідження.** Як відомо, тести, що складаються із завдань, не відібраних за ступенем складності відповідно до стандарту розподілу, визначають вимірювану якість (властивість) на ранговій шкалі. Перехід до сильнішої інтервальної шкали вимагає параметризації завдань за ступенем їх складності і відбору відповідно до заданого стандарту розподілу.

Виходячи з класичної теорії тестів, обмеження знімаються у сучасних моделях тестування, наприклад логістичних моделях Раша, Бірнбаума, котрі ми і розглянемо, тобто залежність оцінки складності завдань від підготовленості студентів.

В якості математичної моделі що зв'язує успіх студента з рівнем його підготовленості і складністю завдання вибирається логістична функція. Для моделі Раша вона має вигляд

$$P_j(\theta) = \frac{e^{1,7(\theta-\beta_j)}}{1 + e^{1,7(\theta-\beta_j)}} \quad (1)$$

$$P_i(\beta) = \frac{e^{1,7(\theta_i-\beta)}}{1 + e^{1,7(\theta_i-\beta)}} \quad (2)$$

Масштабний множник 1,7 використовується для сумісності моделі ОЯ^сБ з моделлю А.Рег§^оп, де вірогідність правильної відповіді для завдання виражена інтегралом нормального розподілу (3), що дозволяє використовувати замість логістичних кривих добре вивчену інтегральну функцію нормованого нормального розподілу [2].

$$P_j(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\theta-\beta_j} e^{-\frac{1}{2}x^2} dx \quad (3)$$

Оскільки модель Раша описує вірогідність успіху студента як функцію одного параметра  $(\theta_i - \beta_j)$ , то іноді її називають однопараметричною моделлю IRT.

Взаємодія двох множин  $\theta_i$  і  $\beta_j$  утворює дані, що володіють властивістю «сумісної адитивності» (conjoint additivity). Правильне використання моделі Раша дозволяє відокремити оцінки студентів від оцінок складності завдань, і навпаки. Ця властивість Rasch Measurement носить назву separability parameter estimates - «незалежність оцінок завдань від студентів і оцінок студентів від параметрів завдань».

На рис. 1 показані три характеристичні криві згідно з рівнянням (1) із рівнем складності завдань -2, 0 і +2

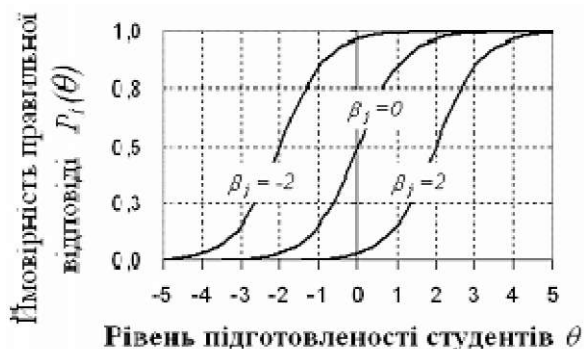


Рис. 1. Характеристичні криві завдань (ICC) в моделі (1PL).

(перше найлегше, друге - середнє, третє найважче). З наведених залежностей ми бачимо, що чим вищий рівень підготовленості  $\theta$  студентів, тим більша вірогідність успіху в тому або іншому завданні. Наприклад, для студента з  $\theta=0$  вірогідність правильно відповісти на перше завдання близька до одиниці, на друге рівна 1/2 і на третє майже дорівнює нулю. Відмітимо, що в точках, де  $\theta = \beta$ , вірогідність правильної відповіді дорівнює 0,5. Тобто, якщо складність завдання дорівнює рівню підготовленості (ability) студента, то він з рівною ймовірністю може справитися або не справитися з цим завданням.

Характеристичні (логічні) криві для завдань тесту в англійській літературі називаються *Item Characteristic Curve (ICC)*.

На рис. 2 показані три характеристичні криві тестування студентів згідно з рівнянням (2) - «Person Characteristic Curve» (PCC). Зображені графіки для трьох студентів з рівнем підготовленості - 2 логіта (найслабший), 0 логітів (середній) і +2 логіта (сильний студент).

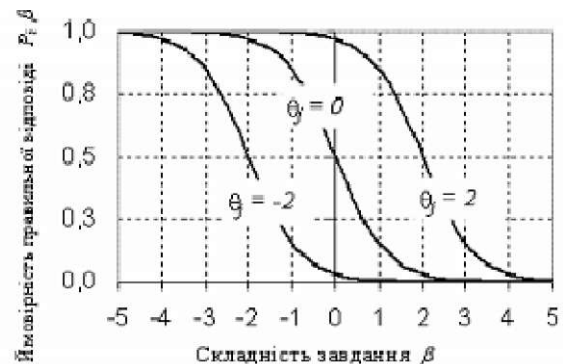


Рис. 2. Характеристичні криві випробувань (PCC) в моделі IPB.

З наведених залежностей можна зробити висновок, що чим вищий рівень підготовки, тим вища вірогідність правильної відповіді на завдання. Наприклад, завдання з складністю  $\beta = 0$  перший студент ( $\theta = -2$ ) практично не зможе виконати, другий  $\theta = 0$  має вірогідність виконання завдання рівну 0,5, третій ( $\theta = +2$ ) легко справиться із завданням, оскільки для нього вірогідність успіху майже дорівнює одиниці.

#### Двопараметрична модель Бірнбаума

З вищенаведених залежностей впливає, що кривизна характеристичних кривих в області  $P_i=0,5$  однакова, тобто диференціююча здатність є константою. Для дихотомічної моделі ця константа дорівнює 0,25.

Якщо тест містить завдання з різною диференціюючою здатністю, то однопараметрична модель IPL не може описати такі емпіричні дані. Для вирішення

цієї проблеми А. Бірнбаум (A. Birnbaum) [3] ввів ще один параметр -  $a$  (item discrimination parameter).

$$P_j(\theta) = \frac{e^{1,7a_j(\theta-\beta_j)}}{1+e^{1,7a_j(\theta-\beta_j)}} \quad (4)$$

$$P_i(\beta) = \frac{e^{1,7a_j(\theta_i-\beta)}}{1+e^{1,7a_j(\theta_i-\beta)}} \quad (5)$$

Параметр  $a$  визначає нахил (кривизну) характеристичної кривої  $j$ -гсі завдання. Приклади характеристичних кривих показані на рисунку 3. З цього рисунка ми можемо зробити висновок, що чим більше  $a$ , тим крутіше йде крива, тим вища диференціююча здатність завдання.

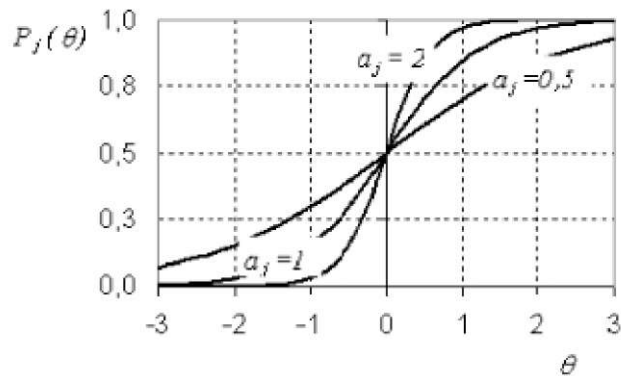


Рис. 3. ICC в двопараметричній моделі 2PL.

Для ще кращої відповідності емпіричним даним А. Бірнбаум ввів третій параметр  $c$  - параметр вгадування.

$$P_j(\theta) = c_j + (1-c_j) \frac{e^{1,7a_j(\theta-\beta_j)}}{1+e^{1,7a_j(\theta-\beta_j)}} \quad (6)$$

$$P_i(\beta) = c_j + (1-c_j) \frac{e^{1,7a_j(\theta_i-\beta)}}{1+e^{1,7a_j(\theta_i-\beta)}} \quad (7)$$

З рівнянь (5) і (6) випливає, що при  $c=0$  і  $a=1$  ці рівняння переходять в однопараметричну модель. З цієї причини інколи говорять, що модель Раша є окремим випадком дво- і трипараметричної моделі Бірнбаума. Формально це так, але по суті невірно. Обговорення цієї проблеми ми будемо проводити пізніше.

На рис. 4 наведені приклади характеристичних кривих для трьох завдань із складністю  $\beta = 1$ , дискримінаційним параметром  $a = 1$  і різними параметрами вгадування  $c_j = 0, c_j = 0,25, c_j = 0,5$ .

З наведених графіків видно, що наявність параметра вгадування приводить до пропорційного зсуву ICC вгору на величину  $c$ .

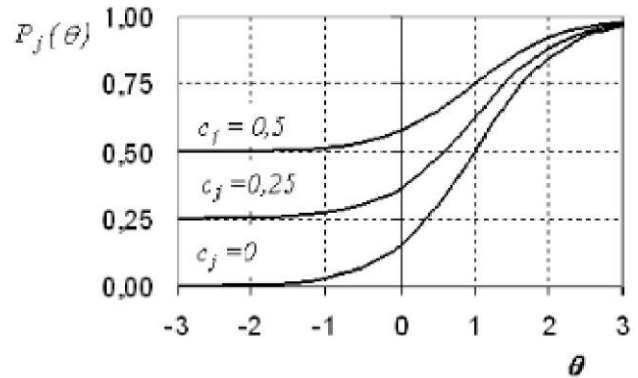


Рис. 4. ICC в трипараметричній моделі 3PL.

В якості теоретичної оцінки  $c$  можна використовувати обернену величину від кількості відповідей в завданнях з вибором. Наприклад, в тесті використовуються завдання з чотирма відповідями, тоді  $c = 1/4 = 0,25$ . Це значення повинне уточнюватися при аналізі емпіричних даних.

#### Модель Rasch Measurement

Обговоримо питання про ступінь придатності моделей IRT для цілей виміру латентних параметрів.

Характерною особливістю моделі Раша є те, що характеристичні криві (ICC) не перетинаються (рис.1). Це означає, що якщо деяке завдання «А» простіше за завдання «Б», то це співвідношення зберігається у всьому інтервалі зміни  $\theta$ .

Абсолютно інша картина спостерігається для дво- і трипараметричної моделей. На рис. 3 це добре видно. Завдання з  $a=0,5$  в області позитивних значень  $\theta$  є найважчим з представлених трьох завдань, тобто вірогідність правильної відповіді на це завдання найнижча. В області ж негативних значень  $\theta$  це ж завдання тепер уже найлегше - вірогідність правильної відповіді на нього найбільша. Отже, для слабких студентів це найлегше завдання, а для сильних студентів - найважче.

Аналогічна картина спостерігається і для трипараметричної моделі. На рисунку 4 показаний окремий випадок непересічних характеристичних кривих, оскільки для них вибрані однакові параметри  $\beta=1$  і  $a=1$ , тобто всі три завдання мають однакову складність і однаковий параметр диференціюючої здатності. На рисунку 5 наведений інший приклад.

Тут у завдання з параметром  $c_j=0$  змінена складність  $\beta_j = -1$ , що і викликало перетин характеристичних кривих. Завдання з  $c_j=0$  в області  $(\theta < -2)$  є найважчим. У області  $-1,5 < \theta < -1$  це завдання легше завдання з  $c_j=0,25$  і важче за завдання з  $c_j=0,5$ . У області  $\theta > -1$  завдання з  $c=0$  є найлегшим.

Подібний перетин ICC практично завжди відбувається для дво- і трипараметричної моделей.

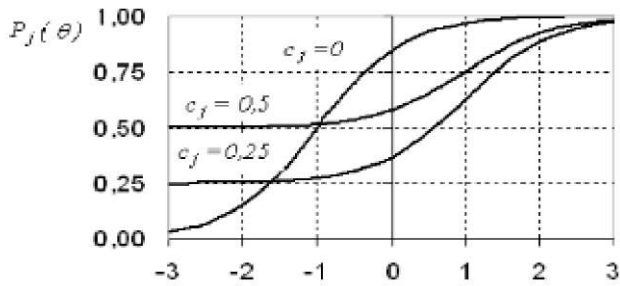


Рис. 5. Пересічні ICC в трипараметричній моделі.

Таким чином, тільки однопараметрична модель Раша відповідає вимогам, що пред'являються до якісного вимірювального інструментарію. Саме модель RASCH MEASUREMENT найбільше придатна для побудови тесту як вимірювального інструменту.

**Висновок.** Педагогічний тест, як засіб виміру навчальних досягнень, може дати достовірний результат тільки в разі його коректного застосування.

#### Література

1. Гулюкіна Н. А. Педагогический тест: этапы и особенности конструирования и использования : пособие для преподавателей / Н. А. Гулюкіна, С. В. Клишина - Новосибирск : - Вид-во НГТУ 2001.
2. Михеев О. В. Математические модели педагогических измерений / О. В. Михеев // Педагогические измерения. -

Коректність застосування тесту - це багатоаспектне поняття, що включає питання конструювання і дизайну тесту, питання розробки і застосування тестів і, зрозуміло, інтерпретації результатів тестування. У даній статті основна увага приділена питанням коректності інтерпретації результатів педагогічного тестування, що проводиться на основі моделей Раша та Бірнбаума. Аналіз результатів зазвичай проводиться на основі класичної теорії тестів або на основі Item Response Theory.

IRT зараз є загально визнаною теорією. В якості латентних параметрів моделі виступають як характеристики тестування, так і самого тесту. Отже можна зробити висновок, що унікальність моделей Раша та Бірнбаума полягає в тому, що вони задають певний механізм перетворення формальних спостережень за результатом подій в об'єктивні виміри на метричній шкалі латентних стимулів цих подій.

2004. - №2 - С75 - 88.

3. Birnbaum A. Some Latent Trait Models and Their Use in Inferring and Examinee's Ability. In Lord F.M., Novick M. Statistical Theories of Mental Test Scores. Addison-Wesley Publ. Co. Reading, Mass, 1968. - P. 397 - 479.