

ЗАСТОСУВАННЯ "ТЕОРІЇ ЛАТЕНТНОЇ ХАРАКТЕРИСТИКИ" ДЛЯ ВДОСКОНАЛЕННЯ АЛГОРИТМУ ПСИХОЛОГІЧНОЇ ДІАГНОСТИКИ ПРОФЕСІЙНОЇ ПРИДАТНОСТІ

В статті йдеться про «теорію латентної характеристики» та її застосування для вдосконалення алгоритму психологічної діагностики професійної придатності. Наведені варіанти тестових моделей, їхні переваги, сфери застосування, розглядається можливість їхнього застосування у практиці психологічного тестування. На основі здійсненого аналізу окреслюються шляхи вдосконалення алгоритму психологічної діагностики професійної придатності та напрямки подальших досліджень.

Ключові слова: «теорія латентної характеристики», психологічна діагностика, професійна придатність, модель Раша, модель Бірнбаума.

В статье идет речь о "теории латентной характеристики" и ее приложения для совершенствования алгоритма психологической диагностики профессиональной пригодности. Приведенные варианты тестовых моделей, их преимущества, области применения, рассматривается возможность их приложения в практике психологического тестирования. На основе осуществленного анализа очерчиваются пути совершенствования алгоритма психологической диагностики профессиональной пригодности и направления дальнейших исследований.

Ключевые слова: "теория латентной характеристики", психологическая диагностика, профессиональная пригодность, модель Раша, модель Бирнбаума.

The article refers to "Item Response Theory" and application of it to improve the algorithm of psychological professional competence diagnostics. Presented variants of test models, their advantages and scope, the possibility of their application in practice of psychological testing. The analysis outlines the ways of improving the algorithm of psychological professional competence diagnostics and directions for further research.

Key words: Item Response Theory, psychological diagnostics, professional competence, Rasch model, Birnbaum model

Постановка проблеми. Розвиток ресурсних можливостей та розробка нового програмного забезпечення сучасних комп'ютерів дозволяє під новим кутом подивитися на здавалось відомі речі. Мова йде про застосування математичних алгоритмів аналізу результатів психологічного тестування або тестування рівня розвитку здібностей. Таке розмежування принципове для нас, адже вимагає різних підходів для їх конструювання. Як правило, в тестах здібностей у запитаннях існує правильна відповідь, на відміну від тестів, що спрямовані на діагностування певної психологічної характеристики (конструкту).

Аналіз останніх досліджень. Розробка та застосування сучасних методів (алгоритмів) аналізу результатів тестування ставала об'єктом дослідження багатьох авторів [3, 5, 7]. Відомо про застосування теорії латентної характеристики з метою розробки та обрахування результатів тестування в інтересах зменшення ефекту спотворення при здійсненні персональної оцінки [8, 10]. Попри це проблему не можна вважати вирішеною остаточно.

Виклад основного матеріалу . «Теорія латентної характеристики», «сучасна або стохастична теорія тестів» або «теорія моделювання та параметризації тестів», в англійському варіанті має аналоги «Item Response Theory» чи «Latent Trait Model». Вона базується на теорії латентно-структурного аналізу (ЛСА), створеної П. Лазарсфельдом та його послідовниками переважно в лабораторіях військових відомств США та університетів. В рамках спеціального психологічного дослідження «American Soldier» застосування ЛСА мало підвищити боєздатність солдатів за рахунок виявлення та усунення тих прихованих факторів, які впливають на поведінку солдат в бойових умовах, а також факторів, що допомагають подолати страх [2].

В англійській літературі цей термін формувався багато десятиріч, шляхом відкидання ряду відмінних характеристик теорії задля скорочення. Нерідко її називають саме сучасною протиставляючи тим самим класичній теорії тестів. В ній оцінка якості тесту традиційно зводиться до визначення міри його надійності й питання валідності отриманих результатів.

Класична теорія тестів - підхід, що домінував у тестології до 60-х років 20 століття - до появи IRT. Її основні положення:

1. Індивідуальний тестовий бал (X_i) і-го досліджуваного виявляється сумою істинного бала випробуваного (T_i) і незалежної помилки виміру

$$X_i = T_i + E_i \quad (1)$$

2. Істинні компоненти (T_i) не корелюють із помилковими (E_i) компонентами виміру. Тобто якщо високим значенням тестових балів відповідають і більш високі значення помилок з певним знаком, то такі помилки не можна вважати випадковими.

Головним інструментом конструювання тестів у межах класичної теорії тестів є забезпечення гомогенності (статистичної узгодженості) тестових завдань, включених в одну тестову шкалу. Первинною інформацією при тестуванні знань є набраний бал досліджуваного (первинний бал), тобто чим більше завдань виконано досліджуваним, тим вище його бал. Достойнством цієї оцінки є її простота й наочність. Проблема полягає в тім, що первинний бал є не абсолютною, а відносною оцінкою, індикатором підготовленості досліджуваного, а не її мірою. Він істотно залежить від важкості завдань тесту й на іншому тесті може виявитися іншим, причому самі важкості тесту у свою чергу визначається всім контингентом досліджуваних. Крім того, через випадковість вибірки не можна задалегідь визначити, які завдання за складності відстануться тому, хто тестується.

Таким чином, до недоліків можна віднести:

- залежність оцінок від важкості завдань тесту;
- залежність оцінок від рівня підготовленості досліджуваного.

У теорії IRT розробник тесту припускає, що відповіді на завдання тесту можуть бути пояснені шляхом аналізу латентних характеристик, яких за кількістю менше, ніж завдань тесту. У Реально ж більшість положень теорії ви-

ходить із того, що єдина латентна характеристика пояснює відповіді на завдання тесту. В "серці" цієї теорії лежить математична модель того, як досліджувані з різними рівнями здібностей за латентними характеристиками повинні відповідати на завдання.

Латентно-структурний аналіз створений для виміру латентних (у тому числі психічних) властивостей особистості. Він є одним з варіантів багатомірного аналізу даних, до яких належать факторний аналіз у його різних модифікаціях, багатомірне шкалювання, кластерний аналіз й ін.

Теорія виміру латентних рис припускає, що:

1. Існує одномірний континуум властивості - латентної змінної (x); на цьому континуумі відбувається імовірнісний розподіл індивідів з певною щільністю $f(x)$.

2. Існує імовірнісна залежність відповіді випробуваного на завдання (пункт тесту) від рівня його психічної властивості, що називається характеристикою кривої пункту. Якщо відповідь має дві градації ("так - ні", "вірно - невірно"), то ця функція є ймовірність відповіді, що залежить від місця, займаного індивідом на континуумі (x).

3. Відповіді досліджуваного не залежать одна від іншої, а пов'язані тільки через латентну рису. Ймовірність того, що, виконуючи тест, досліджуваний дасть певну послідовність відповідей, дорівнює добутку ймовірностей відповідей на окремі завдання.

Первинною моделлю в IRT стала модель латентної дистанції: різниця рівня здібності й важкості тестового завдання ($x_i - b_j$), де x_i - положення i -досліджуваного на шкалі, а b_j - положення j -го завдання на тій же шкалі. Відстань ($x_i - b_j$) характеризує відставання здібності досліджуваного від рівня складності завдання. Якщо різниця велика й від'ємна, то завдання не може бути виконано, тому що для досліджуваного воно занадто складне. Якщо ж різниця велика й позитивна, то завдання також не інформативно, тому що досліджуваний завідомо й правильно його вирішить.

Ймовірність правильного вирішення завдання i -м досліджуваним:

$$P_i(x_{ij}) = f(x_i - \beta_j) \quad (2)$$

В IRT функції x і $f(b)$ називаються функціями вибору пункту. Відповідно перша є характеристичною функцією досліджуваного (властивості), а друга - характеристичною функцією (сили пункту) завдання. Крім "властивості" й "сили пункту" (важкості завдання) в аналітичну модель IRT можуть включатися й інші змінні.

Головна відмінність IRT від класичної теорії тесту в тому, що в ній не ставляться й не вирішуються фундаментальні проблеми емпіричної валідності й надійності тесту: завдання апріорно співвідносяться лише з однією (латентною) властивістю, тобто тест заздалегідь вважається валідним. Вся процедура зводиться до одержання оцінок параметрів завдання й до виміру "здібностей" досліджуваних (створенню "характеристичних кривих" ХК).

Вважається, що латентні змінні x і b нормально розподілені, тому для характеристичної функцій вибирають або логістичну функцію, або інтегральну функцію нормованого нормального розподілу (вони мало відрізняються одна від іншої).

Розробник тесту передбачає, що відповіді на завдання тесту можуть бути пояснені шляхом аналізу латентних характеристик, яких за кількістю менше ніж заданій тесту. Суттю теорії лежить математична модель, того як досліджувані з різним рівнем вираження характеристики мають відповідати на завдання, тобто досліджувані із вищим ступенем вираження риси більш ймовірно дадуть вірну відповідь на завдання. Найбільш поширеними є модель Раша [9] моделі Бірнбаума (двох- та трьох параметрична) [11]. Крім того існує й чотирьох параметрична модель В.С. Аванесова [1].

Одним із центральних понять є характеристична крива завдання, графік якої відображає функціональну залежність вірогідності «правильної» відповіді на тестове завдання від латентної характеристики.

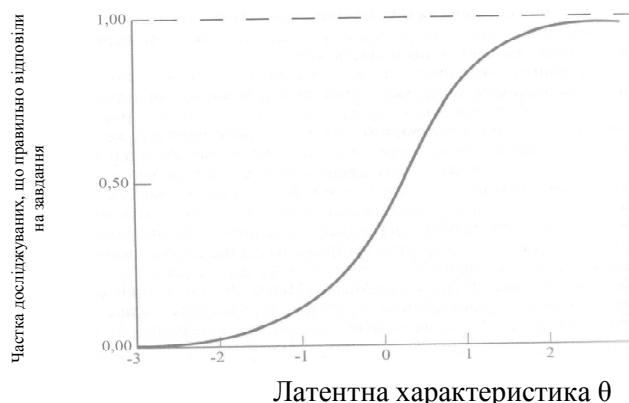


Рис. 1. Графік характеристичної кривої (нормальна огива)

У більшості положень теорії IRT передбачається, що характеристична крива має S-подібну форму. (нормальна огива) Як видно з графіку, зі зростанням оцінки латентної характеристики збільшується ймовірність правильної відповіді на завдання тесту.

1. Якщо рухатися зліва направо, то крива безупинно зростає. Про таку криву говорять, що вона зростає монотонно.

2. Нижня асимптота інтегральної кривої нормального розподілу - це горизонтальна вісь, до якої крива наближається, але ніколи не досягає 0. Верхня асимптота - це горизонтальна пряма, проведена через одиничну точку вертикальної вісі.

3. Нормальна огива безпосередньо пов'язана з нормальним розподілом. Якщо бали обстежуваних виражені у вигляді z-оцінок, ми можемо використати таблицю стандартного нормального розподілу, щоб одержати площу під нормальною кривою ліворуч від будь-якої z-оцінки. Ця площа менше ніж 1, тому її можна інтерпретувати як частку. Інтегральна крива стандартного нормального розподілу зображує ці частки як функцію z-оцінок.

Розглянемо ХК, що показує, що для $\theta=2$ імовірність правильної відповіді дорівнює 0,87. Це можна інтерпретувати в тому розумінні, що випадково обраний із субпопуляції досліджуваних з $\theta = 2$ виконає завдання правильно з імовірністю 0,87. Еквівалентна інтерпретація означає, що для досліджуваних з $\theta = 2$ частка правильних відповідей на завдання дорівнює 0,87.

Хоча S-подібна форма часто використовується, проте вона не є єдиною можливою. Доцільно обговорити й такий варіант ХК як графік ступінчатої функції. Ця ХК означає, що є мінімальна оцінка латентної характеристики, позначена θ' , нижче якої досліджувані не можуть виконувати завдання правильно, але будь-який досліджуваний з рівнем здібності, рівним або більшим чим θ' , обов'язково правильно відповість на завдання тесту. Такі східчасті функції корисні для ілюстрування фундаментальних положень, проте фактично рідко зустрічаються, тому застосовуються рідко.

У межах однопараметричної моделі Г. Раша, завдання характеризуються тільки одним параметром важкості. Імовірність правильної відповіді на j-е завдання з важкістю β_j для i-го досліджуваного з рівнем підготовленості θ_i виражається залежністю:

$$P_{ij} = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \quad (3)$$

При конструюванні тесту необхідно одержати ХК всіх завдань. Характеристичні криві можуть накладатися одна на іншу. У цьому випадку надлишкові завдання видаляються. На певних ділянках осі характеристичні криві завдань можуть бути зовсім відсутніми. Тоді розробник тесту повинен додати завдання відсутньої важкості, щоб рівномірно заповнити ними весь інтервал шкали.

Головним недоліком моделі Раша вважають нехтування "крутизною" ХК: "крутиза" їх вважається однаковою. Завдання з більше "крутими" характеристичними кривими дозволяють краще "розрізнити" досліджуваних (особливо в середньому діапазоні шкали здібностей), ніж завдання з більш "пологими" кривими.

Таким чином модель була вдосконалена - введено a_g - двухпараметрична модель Бірнбаума.

$$P_{ij} = \frac{e^{\alpha_j(\theta_i - \beta_j)}}{1 + e^{\alpha_j(\theta_i - \beta_j)}} \quad (4)$$

Можливий ще варіант: $X = D a_g(\theta - b_g)$ де D константа, що може бути обрана довільно, проте загальноприйнятою нормою є $D=1,7$ адже тоді ординати $P_g(\theta)$ для логістичної кривої та нормальної огиви не будуть відрізнятися більше ніж на 0,01 ні за якого значення θ

$$P_g(\theta) = \frac{e^{D\alpha_g(\theta - b_g)}}{1 + e^{D\alpha_g(\theta - b_g)}} \quad (5)$$

Повернемося до рівняння нормальної огиви, який має вигляд [4]:

$$P_g(\theta) = \int_{-\infty}^w f(z) dz \quad (6)$$

$P_g(\theta)$ це частка досліджуваних з латентною здібністю θ , які відповідають на завдання g правильно. Вираз праворуч - кумулятивна нормальна огива (кумулятивна інтегральна крива нормального розподілу). Це означає, що площа під нормальною огивою між $-\infty$ може бути обчислена. Величина w - число, що визначається рівнянням:

$$w = a(\theta - b)$$

a - параметр диференціюючої здатності завдання; b - параметр важкості того ж завдання. (Параметри a та b мають аналоги, якими є параметри дискримінативності й важкості завдання в класичній теорії тестів.)

Величина w чи еквівалентний їй вираз $a(\theta - b)$ є подібними до z-оцінки в тому розумінні, що $P_g(\theta)$ являє собою площу ліворуч від точки $a(\theta - b)$ в стандартному нормальному розподілі. Таким чином, якщо досліджуваний має оцінку латентної здібності $\theta=2$, а ХК для завдання g має значення параметрів $a=0,5$ $b=1$, то для цього досліджуваного та завдання g маємо:

$$w = a(\theta - b) = 0.5(2 - 1.0) = 0.5$$

Через те що площа ліворуч від точки 0,5 в стандартному нормальному розподілі дорівнює 0,69, то застосування рівняння передбачає що 69% досліджуваних з $\theta=2$ дадуть відповідь на завдання вірно.

Тому, якщо ми маємо наближені оцінки величин a та b для завдання g і хочемо знати частку досліджуваних з певною конкретною латентною здібністю, які мають відповісти на завдання правильно, то потрібно виконати наступні дії: обрахувати значення $a(\theta - b)$; використати це значення як z-оцінку, щоб знайти $P_g(\theta)$ в стандартній z-таблиці нормального розподілу.

Варто пам'ятати, що параметри a та b залежать від завдань, тому ведучи мову про ХК ці значення часто значають як a_g та b_g

Тому рівняння функції нормальної інтегральної кривої має вигляд:

$$P_g(\theta) = \int_{-\infty}^{a_g(\theta - b_g)} f(z) dz, \quad (7)$$

де $w = a_g(\theta - b_g)$.

Для того щоб побачити, як змінюється параметр важкості завдання b_g розглянемо три завдання які мають однакові значення a_g проте різні b_g : $b_1 = 0,5$, $b_2 = 1,0$, $b_3 = 1,5$

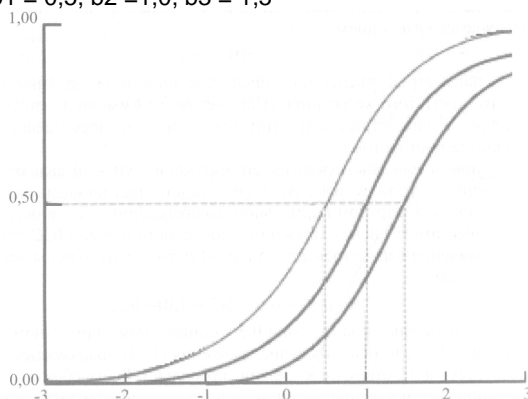


Рис. 2. Характеристичні криві завдання із різними рівнями важкості : $b_1 = 0,5$, $b_2 = 1,0$, $b_3 = 1,5$.

Для кожної ХК вертикальні штрихові лінії вказують на горизонтальній осі оцінку латентної характеристики, для якої $P_g(\theta) = 0,5$. зліва направо 0,5, 1,0 та 1,5 відповідно. Відмітимо, що для першого завдання $P_g(\theta) = 0,5$ при $\theta = b_1$, подібним чином для другого $P_g(\theta) = 0,5$ при $\theta = b_2$, третього $P_g(\theta) = 0,5$ при $\theta = b_3$. Таким чином, значення параметру b_g (важкості завдання) дорівнює оцінці латентної характеристики, за якої половина досліджуваних дасть відповідь на завдання правильно.

Щоб побачити, як змінюється параметр диференціюючої здатності завдання a_g розглянемо три ХК для кожної з яких $b = 1,5$, а значення a буде різним: $a_1 = 0,1$, $a_2 = 1$, $a_3 = 100$.

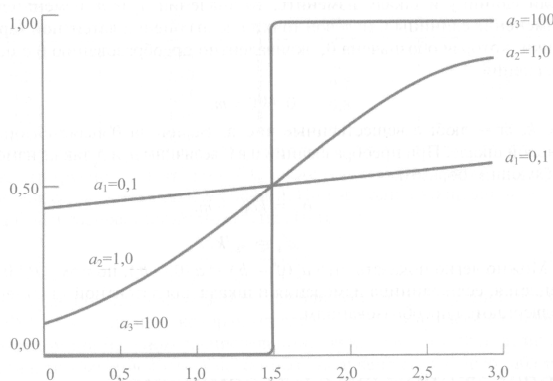


Рис. 3. Характеристичні криві завдання із різним рівнем диференціюючої здатності $a_1 = 0,1$, $a_2 = 1$, $a_3 = 100$

З Рисунка 3 видно, що перше завдання не розділяє досліджуваних достатньо ефективно, тому що частки обстежуваних, що відповідають на завдання правильно майже однакові для всіх оцінок латентної характеристики. В той самий час третє завдання з $a_3 = 100$ дуже ефективно диференціює обстежуваних на тих, у кого оцінки менше 1,45 та тих у кого більше 1,55. Дійсно для тих у кого $\theta \leq 1,45$ частка тих хто відповів вірно менше 0,01, тоді як для тих хто відповів вірно з оцінками $\theta \geq 1,55$, частка більше 0,99.

Нарешті друге завдання з $a_2 = 1$ є проміжним між першим та третім щодо його диференціюючої здатності. Тобто a_g – це параметр, що визначає нахил, чи крутизну ХК. Завдання з пологими кривими мають низьку ефективність для диференціації обстежуваних.

Існує й трьохпараметрична модель Бірнбаума, де вводиться параметр вгадування правильної відповіді:

$$P_j = C_j + (1 - C_j) \frac{e^{\alpha_j(\theta_j - \beta_j)}}{1 + e^{\alpha_j(\theta_j - \beta_j)}} \quad (8)$$

де C_j - вірогідність правильної відповіді на завдання j , в тому випадку якщо досліджуваний вгадував відповідь на запитання. На нашу думку, таку модель варто застосовувати для тестів, що мають єдину правильну відповідь, при здійсненні вибору із варіантів, кількість яких більше за два. Хоча для психологічних тестів також вважається

можливим застосування трьохпараметричної моделі, привласнивши показнику C_j значення соціальної бажаності схвальної відповіді на тестове завдання.

В той же час В. Аванесов додатково ввів параметр «внутрішня валідність завдання» і для нас має радше науковий аніж практичний інтерес [6].

$$P_j = C_j + (1 - C_j) \frac{e^{\alpha_j(\theta_j - \beta_j)}}{1 + e^{\alpha_j(\theta_j - \beta_j)}} \quad (9)$$

Висновки. Завданням професійно-психологічного відбору є виявлення та відсів кандидатів, що не задовольняють параметрам які перевіряються. В той самий час, за умови відсутності таких протипоказань необхідно здійснити добір кандидатів які найбільше відповідають специфічним параметрам. Вбачається доцільним здійснити обрахування параметрів тестових завдань психологічних тестів в інтересах ефективного відбору та добору. Це стосується таких параметрів як диференціююча здатність завдання та параметр важкості. Головним достоїнством такого рішення є визначена (через застосування перевірених «класичних тестів») валідність, що не ставить під сумнів «працездатність» тестового запитання.

Перспективи подальших досліджень вбачаються у визначенні вказаних параметрів тестового завдання із застосуванням відповідного програмного забезпечення (WINSTEPS, R, RUMM 2020) та подальшої розробки інструментарію потенційно, що дозволить досягти скорочення кількості тестових завдань та підвищити ефективність діагностичної процедури. Особливе місце займає діагностування психологічних конструктів (характеристик), в завданнях яких міститься фактор соціальної бажаності (привабливості) тієї чи іншої відповіді. Розробка шляхів його нівелювання проводиться переважно закордонними фахівцями, аналіз досягнень яких представляє великий інтерес. Вдосконалення алгоритму психологічної діагностики професійної придатності дозволить суттєво покращити її процедуру, за умови розробки адекватної технології застосування комп'ютерної техніки та відповідного програмного забезпечення.

1. Аванесов В.С. Педагогическое измерение латентных качеств // Педагогическая диагностика. 2003. № 4. С. 69-78 2. Аванесов В.С., Item Response Theory: основные понятия и положения Назва з екрану. Режим доступу: www.testolog.narod.ru/Theory59.html 3. Карданов Р.С. Разработка и программная реализация методов анализа результатов массового тестирования : автореф. дис. ... канд. техн. наук - Великий Новгород, 2011. - 20 с. 4. Крокер Л Введение в классическую и современную теорию тестов М.: Логос, 2010. - 668 с. 5. Куравский Л.С., Юрьев Г.А. использование марковских моделей при обработке результатов тестирования / Л.С. Куравский, Г.А. Юрьев// Вопросы психологии – 2011. № 2 2011 – с. 112-121. 6. Лаврухина Н.А., Абасова Н.И. Методы оценки качества тестов по результатам тестирования / Н.А. Лаврухина, Н.И. Абасова // Информационные технологии и проблемы математического моделирования сложных систем.– Иркутск: ИИТМ ИрГУПС, 2010, Вып. 8, С. 124 – 134. 7. Arnold Smit, Henk Kelderman, and Henk van der Flier Latent Trait Latent Class Analysis of an Eysenck Personality Questionnaire Methods of Psychological Research Online 2003, Vol. 8, No. 3, pp. 23-50. 8. Chernyshenko, O.S., & Stark, S., Prewett, M. S., Gray, A. A., Stilson, F. R., & Tuttle, M.D. (2009). Normative scoring of multidimensional pairwise preference personality scales using IRT: Empirical comparisons with other formats. Human Performance, 22, 105 – 127. 9. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: The Univ. of Chicago Press, 1980. 10. Stark, S.. A new IRT approach to test construction and scoring designed to reduce the effects of faking in personality assessment [Doctoral Dissertation]. University of Illinois at Urbana-Champaign. 2002. 11. Wright B.D., Masters G.N. Rating scale analysis. Rasch measurements. Chicago: MESA Press, 1982.

Надійшла до редколегії 30.11.12