

УДК: 378.091.26:004

**В.О. ЖУЛАЄВА**

## **ЗАБЕЗПЕЧЕННЯ ВАЛІДНОСТІ Й НАДІЙНОСТІ ТЕСТІВ ПРОТЯГОМ ЖИТТЄВОГО ЦИКЛУ**

***Резюме.** Стаття присвячена проблемам валідності й надійності тестових питань і відповідей, з якими стикається автор при створенні тесту для контролю знань за інформаційно-комунікативних технологій. Життєвий цикл тестових завдань розглядається від початку й до кінця на прикладі втіленої інформаційної системи «Logit».*

***Ключові слова:** тест, валідність, надійність, життєвий цикл, питання, відповіді.*

**Постановка проблеми.** Інформаційне суспільство на сучасному етапі перебудови освітнього простору в Україні, вимагає нових підходів до засобів діагностики отриманих знань в процесі навчання. Західноєвропейські країни, а на самперед США давно й достатньо успішно використовують практику тестового контролю в закладах освіти. Наша країна в цьому сенсі має не багатого досвіду, тому в освітньому просторі дана тема є актуальною.

Якщо розглянути нормативне підґрунтя цього питання, то можливо зробити висновок, що законодавча база досить ретельно описала в нормативних документах засобів діагностики якості освіти. З 2008 року МОН України запровадило по закінченню шкіл зовнішнє незалежне оцінювання, що дає можливість говорити про очевидні переваги тестового контролю, ніж звичний нам іспит.

З квітня 1993 року Україна приєдналася до міжнародного проекту, Темпус, що спрямований на вдосконалення управління закладами освіти, оновлення чи розробку нових навчальних програм і підвищення кваліфікації викладачів, особливо у таких дисциплінах, як економіка, іноземні мови, суспільні науки, європейські студії й право.[1]

Програма Темпус – це програма зовнішньої допомоги Європейського Союзу. Її метою є сприяння в модернізації системи вищої освіти у країнах-партнерах (країнах-сусідах) ЄС через активізацію співпраці між вищими навчальними закладами держав-членів ЄС та країн-партнерів. Програма Темпус надає фінансування з метою заохочення взаємодії та збалансованого співробітництва між вищими навчальними закладами у країнах-партнерах та в країнах-членах Європейського Союзу. [1]

Координаторами проекту “Освітні вимірювання адаптовані до стандартів ЄС” в Україні є три праведні вузи – Національний педагогічний університет ім. М.П. Драгоманова, Кіровоградський педагогічний університет ім. В. Винниченка., Ніжинський державний університет ім. М. Гоголя. [1]

Результатом їх праці стала низка розробок в галузі освітніх вимірювань й тестового контролю. Розвиток інформаційно-комунікаційних, зокрема WEB – орієнтованих технологій потребують розгляду проблеми забезпечення валідності й надійності тестів у цьому контексті.

Багато фахівців і навчальних закладів розробляють, за допомогою сучасних інформаційних технологій, свою систему для перевірки знань. Наповнення інтернет ресурсів тестовими завданнями, загалом непогане явище бо надає можливість вибрати різні тести з різних дисциплін, однак, оцінити якісні характеристики таких тестів досить важко, тому що про валідність чи надійність тестових завдань дуже рідко хто згадує.

**Метою дослідження** є аналіз можливостей для забезпечення валідності й надійності тестів у WEB сервісі протягом життєвого циклу тесту.

**Аналіз останніх досліджень.** Питаннями валідності й надійності тестів займалися багато дослідників. Можна вважати, що ці питання виникли разом з першими кроками

становлення тестології як науки. Першими дослідниками були англійський психолог Френсіс Гальтон, американський психолог Дж. Кеттелл, американський психолог и педагог Едвард Торндайк та інші видатні науковці. В сучасному світі питання валідності й надійності також дуже актуальні ним займаються: Д. Аллахвердієва, А. Анастасі, С. Урбіна, М. Єрмакова, Л. Андреева, В. Куклін, В. Мешалкін, В. Наводнов, Б. Савельєв, А. Майоров, Є. Міхайличев, І. Підласий, А. Саміловський та інші.

**Виклад основного матеріалу.** Термін “валідність” походить від англійського слова *valid* — перекладається, як обґрунтований, дійсний, придатний, цей термін використовують для загальної характеристики тесту в аспекті відповідності меті та вимогам оцінювання.

*Валідність* – комплексна характеристика, яку визначають як параметрами інструменту й процедури вимірювання, так і властивостями ознаки, яку діагностують. Критерій валідності встановлює сферу дійсності, для якої метод дає статистично достовірні результати. [2]

1. Змістовна валідність характеризується на скільки тестове завдання відповідає змісту програми, яку вивчають протягом звітного періоду, чи відповідають данні, які запропоновані змісту підручника, або набору ситуацій та інше. Сучасні дослідники вважають, що змістовна валідність поділяється на концептуальну й програмну. Концептуальна валідність перевіряє тестове завдання на змістовність, структуру й термінологію матеріалу, який вивчається. Вона може бути реалізована завдяки повної систематизації навчальних програм за підтримки держави. Концептуальна й програмна валідність відображають теоретичну основу тестових завдань. Важливим критерієм тестових вимірювань який відображає доцільність, тобто значну роль тесту є очевидна валідність. Очевидна валідність визначається завдяки назві тестового завдання, або інструкції до цього тесту, яке дозволяє зняти негативне відношення тестувальника до завдання.

2. Конструктивна валідність відображає теоретичні аспекти курсу, тобто вимірюється ступінь відображення у тестовому завданні ключові елементи засвоєного матеріалу. Вона може бути виявлена за допомогою аналізу змістовності тестів, об'єкти тестування та характер запропонованих завдань.

3. Критеріальна валідність дає загальну оцінку якості запропонованих тестів, для дисциплін таких, як математика, фізика, хімія, креслення. За думкою дослідників, критеріальна валідність контролює рівень досягнень отриманих шляхом незалежного оцінювання.

4. Поверхнева валідність визначає тестове завдання шляхом поверхневого ознайомлення зі змістом тесту.

5. Порівняльна валідність використовується лише у випадках коли новий інструмент для діагностики, якісно відрізняється від того, яким користуються зараз. Тобто його властивості дають більш широкі й повні можливості визначення результату ніж поточний.

6. Поточна валідність визначає ступінь відставання студентів від прогнозованого рівня засвоєння навчальної дисципліни. Тому можливо говорити про те, що дана валідність відображає рубіжний контроль засвоєних знань з навчальної дисципліни. Прогнозованість засвоєння матеріалу, виходить з попереднього контроль знань, який тримали на початку вивчення запропонованого матеріалу.

7. Узагальнююча валідність дає змогу оцінити збалансованість тестового завдання, тобто змістовної частини й кількісної частини, яка відображає всі змістовні елементи. Таким чином, можна сказати, що узагальнююча валідність дає нам змогу оцінити ступінь збігу між різними показниками, які виявляються при проходженні двох різних тестових завдань. [2, 3]

Розглянувши основні критерії валідності тестових завдань, можливо зробити висновок, що валідність вимірюється лише експериментальним шляхом за допомогою дослідної групи, яка разом з експертами визначає дидактичні методи запропонованих тестових завдань. Проте, важливо не тільки вимірювати не тільки вимірювати валідність, але головне, на етапі

розробки забезпечити її. Проблема валідності прямо визначає ефективність тесту. Виявлення факту не валідності у кінці розробки фактично перекреслює всю виконану роботу. Тому на етапі розробки слід передбачити можливі заходи забезпечення валідності.

Надійним тест вважається, коли багаторазове проведення тих самих тестових завдань з урахуванням умовно однакових умов дають результат без значних відхилень. Якщо тест проводить багато разів, то середнє значення буде характеристикою «істинної» величини параметра. Найбільш часто проводиться оцінка *ретестової* надійності. Чим тісніше корелюють результати початкового і повторного (зазвичай відстроченого на кілька місяців) проведення тесту, тим він надійніший.

Передбачається, що існує необмежена кількість завдань, які можуть «працювати» при вимірюванні. Тест є лише вибіркою завдань з їх генеральної сукупності. В ідеалі можна створити скільки завгодно еквівалентних форм тесту, тому визначення надійності тесту можна провести шляхом кореляції паралельних форм або еквівалентних рівних частин, отриманих шляхом розщеплення тестового завдання на дві частини. Оскільки в реальному тесті число завдань обмежене (не більше 100), то оцінка надійності тесту завжди приблизна. Надійність тестових завдань зазвичай перевіряється за допомогою коефіцієнта кореляції К. Пірсона, тобто за допомогою статистичних показників. Якщо коефіцієнт кореляції результатів становить не менше 0,75, то тест вважається статистично надійним. [5]

Розмаїття характеристик і показників надійності тесту таке ж значне, як і розмаїття умов, що можуть впливати на результат тесту. Однак найбільш відомими на практиці застосування набули тільки характеристики надійності: об'єктивність, надійність застосованих методик, відносна стабільність запропонованих завдань.

Розглянемо, як поділяються можливості та вимоги до забезпечення й визначення валідності й надійності тесту протягом його життєвого циклу.

Термін “життєвий цикл” використовується в різних наукових областях. Цей термін запозичений із таких наук, як – біологія, психологія, інформатика, маркетинг та на виробництві. Життєвий цикл має лише часовий вимір, тобто життя тесту розпочинається з моменту прийняття рішення про створення самого завдання й закінчується коли завдання повністю вилучається з експлуатації, мається на увазі не зміни в тексті або доповнення нових питань. [4]

З нашого погляду життєвий цикл сучасного тесту складається з наступних етапів:

1. Ідея тесту; з грецької слово “ідея” перекладається як початок. Перший етап створення тесту починається з анотації, визначення тематики завдання, яка мета й ціль тестування, на яку аудиторію розраховується завдання. Всі ці питання вирішуються автором тесту на цьому рівні. З точки зору забезпечення валідності на цьому етапі необхідно чітко визначити категорію респондентів (учні сьомого класу, студенти першого курсу гуманітарного напрямку та інші.). Визначається предмет чи дисципліна загалом, або обмежена галузь діяльності, ставиться мета контролю, тобто атестація, поточний контроль знань, контроль самостійної роботи та інші види контролю. Застосування тесту без відповідності заданим вище параметрам призведе до невалідності. Однак причина буде не у самому тесті, а в його невірному, не цільовому застосуванні.

2. Розробка структури тесту; другий етап створення тестового завдання передбачає вирішення таких питань, як визначення розділів, що містить завдання, написання анотацій до розділів. Цей етап вводиться також автором тесту. Виконання даного проміжку на етапі створення тесту, можливий лише на специфікації попереднього етапу.

3. Рецензування тесту; третій етап створення тестового завдання вже переходить до рецензента. Рецензент може бути один або декілька, на цьому етапі визначається поверхнева валідність тестового завдання. Ключовий етап для забезпечення валідності. Оцінка й бачення автора можуть бути суб'єктивними й обмеженими власним баченням і досвідом. Сторонній погляд рецензента має зняти ці обмеження.

4. Підготовка розділів тестового завдання; автором тесту проходить наповнення визначених розділів питаннями й відповідями. Для більш повної перевірки знань пропонується від 30 до 50 питань у одному розділі.

5. Рецензування розділів; на останньому робочому етапі рецензентом або групою рецензентів визначається змістовна валідність тестових питань й запропонованих відповідей. Питання повинні цілком покривати матеріал, який вивчається не виходячи за його межі. Водночас питання повинні мати різну складність. Ця частина дуже трудомістка для рецензента, тому що оцінювач повинен розглядати тест очима фахівця й респондента. Головна увага при цьому приділяється однозначності запитання й відповідей. Респондент має швидко й однозначно зрозуміти текст та запропоновані варіанти відповідей.

6. Випробування розділів; на цьому етапі життєвий цикл тесту переходить до наступної стадії “draft”. Запропоновані тестові завдання проходять апробацію на експериментальній групі.

7. Оцінювання питань; на цьому етапі автор проводить розрахунок профілів кожного питання, де визначається складність й диференційована здатність питань. За цими даними можливо чисельно оцінити валідність й надійність тесту. Проте це вже швидше за все підсумок ніж спонування до дії. Зазвичай встановлення факту невалідностей чи ненадійності тесту свідчить про неналежну увагу до цих параметрів на попередніх етапах. За нашими оцінками трудомісткість виправлень ситуації можливо скласти від 30% до 100% початкових затрат. В подальшому можливо уточнення отриманих оцінок на більших вибірках.

8. Тестування; наступний рівень тестового завдання “recommended”. На цьому етапі учитель з учнями проводять тестування попередньо сформованих тестів. Проводяться індивідуальні й групові тестування з реєстрацією результатів.

9. Уточнення оцінок питань; автор тесту розраховує профілі з урахуванням робочих результатів отриманого тестування.

10. Визначення рейтингу тесту; менеджером тесту проходить підрахунок активності використання тесту й проходить оцінка надійності.

Запропоновані етапи життєвого циклу тестового завдання переважно реалізовані при створенні інформаційній системі підтримки тестового контролю “Logit”, адреса сайту: <http://logit.kdpu.edu.ua/>. При цьому питанням валідності й надійності з боку автора приділено особливу увагу. Також забезпечено належний сервіс рецензування й контролю за цим процесом.

Висновки. В практичному аспекті розробки дидактичних тестів особливу увагу слід приділяти забезпеченню валідності й надійності саме на попередніх етапах створення тесту. Забезпечення валідності й надійності тестових завдань слід розглядати в процесах життєвого циклу. Надважливим інструментом забезпечення цих параметрів є процес рецензування питань й відповідей в процесі створення тестових завдань.

Виявлення невалідності або ненадійності тесту на останніх етапах створення призводить до значних витрат, які можуть перевищувати вартість початкової розробки. Всі зазначені ідеї, зокрема рецензування й контролю було втілено при розробці інформаційної системи “Logit”.

#### ЛІТЕРАТУРА

1. Національний Темпус-офіс в Україні. Офіційна сторінка [Електронний ресурс]. – / Режим доступу: <http://www.tempus.org.ua/uk/tempus.html>. – Назва з екрану.
2. Луценко О.Л. Лекція № 6 з дисципліни «психодіагностика». Валідність тесту. Факторизація тесту / О.Л. Луценко. – [Електронний ресурс]. – Режим доступу: <http://dspace.univer.kharkov.ua/bitstream/123456789/1291/2/%D0%9B%D0%B5%D0%BA%D1%86%D0%B8%D1%8F%20%D0%9F%D0%94%20%E2%84%966%20%D1%83%D0%BA%D1%80.pdf>. – Назва з екрану.

3. Портал знань [Електронний ресурс]. – Режим доступу: <http://www.znannya.org/?view=concept:463>. – Назва з екрану.
4. Поняття життєвого циклу інформаційної системи [Електронний ресурс]. – Режим доступу: <http://www.virtual.ks.ua/essays-term-papers-and-diplomas/347-the-concept-of-life-cycle-information-system.html>. – Назва з екрану.
5. Коновалова М. Д. Експериментальна психологія: конспект лекцій / М. Д. Коновалова. – Електронний ресурс. – Режим доступу: <http://co.zapitai.com.ua/text/15221/index-1.html?page=20>. – Назва з екрану.

#### **В.А. ЖУЛАЕВА. ОБЕСПЕЧЕНИЕ ВАЛИДНОСТИ И НАДЕЖНОСТИ ТЕСТОВ В ТЕЧЕНИИ ЖИЗНЕННОГО ЦИКЛА**

*Резюме.* Стаття посвящена проблемам валидности и надежности тестовых вопросов и ответов, с которыми сталкивается автор при создании теста для контроля знаний с помощью информационно-коммуникативных технологий. Жизненный цикл тестовых заданий рассматривается от начала и до конца на примере воплощенной информационной системы «Logit».

*Ключевые слова:* тест, валидность, надежность, жизненный цикл, вопросы, ответы.

#### **V.O. ZHULAYEVA. GUARANTEES OF TESTING LIFE CYCLE'S VALIDITY AND RELIABILITY**

*The summary.* This article deals with the problems of test's questions and answers validity and reliability witch faced by when the author creates the test for controlling knowledge in the information and communication technologies. The tests' life cycle considered from the beginning to the end on the example of embodied information system «Logit».

*Key words:* test's validity, reliability, life cycle, questions, answers.

Одержано редакцією 27.12.2013 р.