

КРИТЕРІЇ ЕФЕКТИВНОСТІ КОМП'ЮТЕРНОГО ТЕСТУВАННЯ У ВИКЛАДАННІ ІНОЗЕМНОЇ МОВИ

Стаття присвячена аналізу та визначенню ключових характеристик ефективності, якості та адекватності комп'ютерних тестових технологій у викладанні іноземної мови. Розкрито зміст понять «валідність, надійність, економність, практичність, автентичність, інтерактивність, квантифікація, визначеність, однозначність, диференціація.

Ключові слова: комп'ютерне тестування, критерії ефективності, тестова валідність та надійність.

Статья посвящена анализу и определению ключевых характеристик эффективности, качества и адекватности компьютерных тестовых технологий в преподавании иностранного языка. Раскрыто содержание понятий «валидность, надежность, экономность, практичность, подлинность, интерактивность, квантификация, определенность, однозначность, дифференциация.

Ключевые слова: компьютерное тестирование, критерии эффективности, тестовая валидность и надежность.

The paper discusses and analyses the key efficiency, quality, adequacy characteristics of the computer assisted language testing (CALT). The terms «validity, reliability, economy, practicality, authenticity, interactivity, quantification, determinacy, mono-semanticity, differentiation» are defined.

Key words: computer assisted language testing, efficiency criteria, test validity and reliability.

Тестування – один з найсучасніших інструментів отримання педагогічної інформації. Більшість дослідників схиляються до думки, що це – найоб'єктивніший і найякісніший спосіб контролю знань, навиків і умінь при вивченні іноземної мови, оскільки він дозволяє створити рівні для всіх умови, запропонувати єдині критерії оцінки та інтерпретації результатів. Використання інформаційних технологій як інструмента для створення нових засобів навчання вивело тестування на якісно новий рівень, розширило сферу його застосування. Як інструменти моніторингу навчального процесу можна використовувати методики, засновані на технології тестування. На відміну від традиційних засобів оцінки якості освіти, комп'ютерне тестування дозволяє значно скоротити часові рамки, необхідні на обробку результатів контрольної перевірки знань, автоматизувати процес перевірки відповідей та звести до мінімуму суб'єктивний вплив викладача на результат виміру.

Проблеми розробки комп'ютерних тестів та їх застосування в практиці навчання іноземних мов досліджували українські та зарубіжні вчені Л. В. Банкевич, В. А. Коккота, С. Ю. Ніколаєва, О. П. Петрашук, Дж. Андерсон, А. Х'ю, Т. Аткинсон, А. Девіс та інші. Проте, багато невдач,

пов'язаних із комп'ютерним тестуванням у викладанні іноземних мов, обумовлено недостатньою увагою та науковим обґрунтуванням основних критеріїв його ефективності.

Відповідно, **мета** нашої статті полягає в аналізі та визначенні ключових характеристик ефективності, якості та адекватності комп'ютерних тестових технологій у викладанні іноземної мови.

Розглядаючи тест як форму контрольнотренувального завдання, призначеного для певного рівня визначення якості навчання, серед основних критеріїв ефективності комп'ютерного тестування визначають, перш за все, надійність, під якою розуміється одноманітність і однозначність отримуваних результатів після проведення чергового тесту в однакових умовах [2]. С. П. Суворов у цій вимозі підкреслює об'єктивність отримуваних результатів: їх аналогічність при проведенні повторного тесту в однакових умовах [7].

Надійністю називається характеристика тесту, що відображає точність тестових вимірів, а також стійкість тестових результатів до дії випадкових чинників. Ця характеристика якості комп'ютерного тестування полягає в забезпеченні стійкості результатів багаторазового тестування одного й того ж студента. Отже, термін «надійність» має

два значення: по-перше, тест вважається надійним, якщо він забезпечує високу точність вимірів, і, по-друге, тест вважається надійним, якщо він дає при повторному виконанні близькі результати за умови, що підготовка тестованого не змінилася за час до повторного виконання [10, 319].

Серед факторів, які впливають на надійність комп'ютерного тестування, доцільно виділити:

- довжина тесту – чим довший тест, тим він надійніший. Оптимальний тест повинен містити 30 завдань зі зростаючою складністю [11];
- рівень складності – чим більше в тесті завдань середньої складності, тим вище його надійність. Проте, сильні чи слабкі студенти погано реагують на такі види завдань. Тому тест повинен включати різні види завдань з домінуючою кількістю завдань середнього рівня;
- мотивація студентів – результати тестування будуть якісні, якщо студенти зацікавлені в результатах тестування.

Крім зазначених вище основних факторів, що впливають на надійність тесту, О. В. Бузулутський виокремлює ступінь однорідності рівня підготовки тестованих (при однорідному контингенті надійність тесту знижується, особливо в тому випадку, якщо тест важчий оптимального (50-60% правильних відповідей)); здатність, що диференціює, завдань тесту (чим вона вища в середньому за тестом, тим надійніший тест); стабільність зовнішніх умов тестування – усних інструкцій, мотивацій, часу, відведеного на тестування; стабільність і стандартизація оцінки суб'єктивних мовних тестів [1, 12-13].

Серед сучасних способів визначення та оцінки надійності комп'ютерного тестування існують три методи.

I. Альтернативне (паралельне) тестування полягає у проведенні спершу якогось одного тестування серед групи студентів, а тоді – паралельного або альтернативного варіанту серед тієї самої групи. Друге тестування може проводитися або того самого дня, або трохи пізніше. Така процедура виявлення надійності найточніша, особливо коли між першим і другим тестуванням минає якийсь час, адже тоді враховано розбіжності між добірками тестових завдань, щоденні зміни та різний темп роботи окремих студентів [8]. Для еквівалентності виміру паралельних форм тесту, необхідно розробити приблизно однакову кореляцію висловлювань (завдань), що виражають одну і ту саму думку. Певна трудність полягає також в тому, що доводиться створювати два тести на одну тему, що подвоює роботу викладача.

II. Тестування-перетестування. Цей підхід передбачає повторне написання того самого тесту тією самою групою студентів через якийсь час або відразу. Такий спосіб часто використовують для тестів, що вимагають коротких відповідей. Переваги цього методу полягають у відносній простоті його використання і в простоті

розрахунків. До недоліків можна віднести невизначеність у виборі тимчасових інтервалів між першим і другим тестуванням.

III. Внутрішня зв'язність (метод ділення та розщеплення тесту на частинки). Цей метод, на відміну від двох попередніх, передбачає одноразове тестування. В основі цього методу лежить допускання паралельності (адекватності) не лише окремих форм, але і частин усередині однієї форми тесту [11].

Вибір способу перевірки надійності тесту безпосередньо залежить від викладача. Не можна сказати, що один спосіб якісніший за інший; до того ж жоден зі способів не забезпечує абсолютної надійності комп'ютерного тестування.

Другим критерієм якості тестування є *валідність*. Під валідністю розуміють здатність тесту вимірювати те, що він повинен вимірювати за задумом. Це поняття пов'язане з «... відповідністю, значущістю та корисністю окремих висновків, що виведені на основі отриманих балів» [12]. К. Інгенкамп виділяє куррікулярну, а також конструктивну, критерійну (статистичну) і зовнішню валідність [4]. С. П. Суворов у валідності виділяє такі аспекти, як адекватність за тестованим матеріалом (content validity), одночасну відповідність (concurrent validity), прогностичну добротність (predicative validity), добротність теоретичної моделі тесту (construct validity) [7].

Змістова валідність (content validity) *залежить від глибини попередньо проведеного авторами тесту, аналізу навчального плану й програми, змісту підручників, від точності й повноти формулювання мети навчання. У мовних тестах змістова валідність визначається використанням в них відповідного мовного матеріалу (лексичних одиниць, що входять до словникового мінімуму, і граматичного матеріалу, передбаченого діючою програмою з іноземної мови).* При цьому, О. В. Бузулутський зазначає, що в мовних тестах навіть внесення деякої кількості слів, що не входять до лексичного мінімуму, але відносяться до потенційного словника, і їхнє значення яких може бути встановлене з контексту, істотно не знижує змістову валідність. Якщо ж йдеться про змістову валідність текстових основ тестів на розуміння прочитаного, то текст може бути зрозумілим тими, що вчать, навіть при 30% незнайомих слів [1, 11].

Порівняльна валідність (concurrent validity) передбачає зіставлення результатів проведеного тесту з результатами іншого тесту, самооцінкою тестованих, оцінками викладачів.

Прогностична валідність (predictive validity / prognostic validity) відображає ступінь упевненості, що отримані за тест оцінки добре прогнозують майбутні досягнення тестованого. Для її виявлення результати одного тесту порівнюють із результатами іншого тесту, який проводять, наприклад, через півроку або рік навчання [5].

Конструктивна валідність (construct validity) полягає у встановленні конструкта – об'єкту

планованого вимірювання (у методиці навчання іноземної мови такими є види мовної діяльності (читання, письмо, мовлення та аудіювання), які, у свою чергу, можуть бути розкладені на більш дрібні конструкти, також доступні вимірюванням) та підборі типу завдань, адекватних цілям вимірювання даного конструкта.

Зовнішня валідність (*face validity*) *свідчить, наскільки привабливими та прийнятними є завдання тесту для викладачів/учителів, учнів/студентів, адміністрації, тобто тих, хто не є експертами-тестологами й можуть мати лише інтуїтивно-суб'єктивне судження про зміст тесту.*

Серед факторів, **що знижують валідність результатів оцінювання, можна назвати:**

- завдання не відповідають цілям тестування;
- дуже мала кількість завдань, яка не охоплює усього змісту навчальної дисципліни або теми, яку перевіряють;
- завдання є неоднозначними за змістом, некоректно сформульовані, у тесті є ключі до відповідей тощо;
- у тесті вміщено завдання без урахування їх складності;
- завдання в тесті неправильно упорядковані (не враховано розташування завдань за складністю, завдання не згруповано за формою);
- наявність нечітких або неповних інструкцій, що перешкоджає тестованим правильно зрозуміти, як треба працювати з тестом;
- суб'єктивність під час оцінювання завдань на надання відповіді або помилки при обчисленні балів у завданнях на вибір відповіді;
- неврахування часу, необхідного для виконання тесту;
- порушення процедур проведення тестування, незадовільні умови контролю [5].

Валідність – це головний критерій виявлення адекватності вимірів результатів комп'ютерного тестування.

Третім показником якості тестування є *економічність*, яка дозволяє за декілька хвилин відповісти на велику кількість питань, а на більш високому рівні протестувати за годину те, що було засвоєне за місяць, за рік або п'ять років навчання.

З іншого боку, економічність тесту полягає не лише в можливості за короткий проміжок часу протестувати досить великий об'єм програми з іноземної мови, але вона також передбачає гранично спрощену перевірку результатів тесту як чисто механічну процедуру [11].

Надійність, валідність та економічність – найбільш важливі, але не єдині вимоги, що пред'являють до тестових завдань. До критеріїв оцінки тестових завдань відносять:

Практичність – характеристика тесту, яка визначає: доступність інструкцій тесту і змісту тестових завдань для розуміння тих, хто виконує тест; простоту організації проведення тестування в різних умовах; простоту перевірки відповідей і визначення результатів та оцінки. Під практич-

ністю тесту також розуміється співвідношення між ресурсами, необхідними для проектування, розробки і використання тестів, і наявними ресурсами. Під ресурсами мають на увазі людські ресурси, матеріальні ресурси та час, затрачений на проведення тестування. Відношення наявних ресурсів до необхідних визначає пороговий рівень практичності тесту. Це значення має бути більшим одиниці для того, щоб тест вважався практичним [9, 32-33].

Диференційна здатність – характеристика тесту, яка вказує на здатність даного тесту виявляти встигаючих і невстигаючих тестованих, тобто з достатнім і недостатнім рівнем володіння іношомовними навичками і вміннями [1, 11].

Стандартизація полягає в уніфікованій процедурі планування, оцінювання й виставлення балів. Мета – забезпечити рівні умови для оцінювання всіх учасників тестування, щоб бали адекватно демонстрували рівень здобутих знань і не підлягали впливу умов, які між собою відрізняються.

Простота тесту означає, що тестові завдання повинні формулюватися якомога чіткіше і простіше. Інколи, щоб ускладнити тестове завдання, викладач удається до туманних формулювань, порушуючи при цьому інші критерії ефективності тестового завдання [11].

Однозначність тесту полягає в тому, що якість виконання тесту повинна однаково оцінюватися різними експертами.

Визначеність. Забезпечення цієї вимоги необхідне для розуміння студентом того, що він повинен зробити, та для виключення правильних відповідей, які не відповідають еталонній у даному тесті [3].

Для отримання точніших результатів мовних тестових вимірів, вимоги до використання мови, що пред'являються тестованим, мають бути такими ж, як і вимоги до використання мови в природній ситуації спілкування. Це відображається у автентичності тесту.

Автентичність – ступінь відповідності характеристик завдань даного мовного тесту характеристикам завдань у мовленнєвій ситуації.

Під *інтерактивністю* тесту розуміється ступінь та характер застосування індивідуальних характеристик тестованого під час виконання тестових завдань. При мовному тестуванні найбільш актуальні такі характеристики тестованого, як його комунікативна і стратегічна компетенції, загальні знання й емоційність.

Ще однією характеристикою тесту є його *дія на суспільство* і систему освіти, а також на індивіда усередині цієї системи. Ця дія відбувається на двох рівнях: на мікрорівні, тобто на рівні індивіда, що виконує цей тест, і на макрорівні, тобто на рівні системи освіти або суспільства в цілому [6].

У зарубіжній педагогіці велика увага приділяється ще *квантифікації* тесту в інтерпретації Н. В. Володіна (в R. Lado це вимога – *scorability*) [2]. Це можливість виразити в цифрових показ-

никах (кількісних оцінках) результати неодноразово проведеного тесту із наявністю ідентичного контингенту студентів у процесі його стандартизації. Ці показники – важливі для практики використання тестів, бо саме обробка даних про їх недотримання дозволяє розкривати найбільш типові помилки і труднощі, аналіз яких дозволяє встановлювати причини їх появи і своєчасно вносити корективи до процесу вивчення іноземної мови.

Висновок. Отже, для максимально ефективної реалізації всіх функцій тесту варто дотримуватись

наступних основних принципів: складання й застосування тільки таких тестів, які є якісними, надійними, валідними згідно цілей і завдань тестування й навчально-виховної роботи. Для визначення якості та ефективності комп'ютерного тесту важливе значення також мають такі характеристики тестів, як економність, практичність, автентичність, інтерактивність, квантифікація, визначеність, однозначність, диференціація.

ЛІТЕРАТУРА:

1. Бузулутський О. В. Тестування як форма педагогічного контролю / О. В. Бузулутський, О. О. Величко // Вісник Національної академії оборони України. – 2009. – № 3 (11). – С. 5–13.
2. Володин Н. В. К методике составления тестов/ Н. В. Володин // ИЯШ – 1972. – № 1.
3. Грубій А. М. Комп'ютерний моніторинг знань учнів і студентів [Електронний ресурс] / А. М. Грубій. – Режим доступу до джерела: www.nbuv.gov.ua/portal/soc_gum/pedp/2010_2/66/nayk/grybij.pdf.
4. Ингенкамп К. Педагогическая диагностика / К. Ингенкамп – М.: Педагогика, 1991.
5. Коваленко Л. Т. Короткий тестологічний словник-довідник/ Л. Т. Коваленко – К.: Грамота, 2008. – 160 с.
6. Овчаренко Виктория Павловна. Компьютерное тестирование как метод оценивания уровня сформированности лингвистической компетенции: дис. ... канд. пед. наук: 13.00.08 / Овчаренко Виктория Павловна. – Таганрог, 2007. – 194 с. – Библиогр.: с. 144-165.
7. Суворов С. П. Педагогические тесты, практика их составления и использование в США / С. П. Суворов // Путь улучшения преподавания иностранных языков. – М.: Мысль, 1970.
8. Тайджман А. Моніторинг стандартів освіти./Альберт Тайджман – Львів: Літопис, 2003. – С. 155-168.
9. Цатурова И. А. Использование метода тестов в современных методических системах обучения иностранному языку [Текст] / И. А. Цатурова // Материалы Всероссийской научно-практической конференции «Преподавание иностранного языка в XXI веке: проблемы и перспективы». – М.: Международная программа образования, 1998. – С. 67–70.
10. Цатурова И. А. Компьютерные технологии в обучении иностранным языкам [Текст] / И. А. Цатурова, А. А. Петухова. – М.: Высшая школа, 2004. – 95 с.
11. Ялаева Наталья Вячеславовна. Компьютерное тестирование как средство интенсификации обучения английскому языку в юридическом вузе: дис. ... канд. пед. наук: 13.00.08 / Ялаева Наталья Вячеславовна. – Екатеринбург, 2003. – 255 с. – Библиогр.: с. 175-190.
12. American Educational Research Association, American Psychological Association and National Council on Measurement in Education. Standards for educational and psychological testing./ American Educational Research Association, American Psychological Association and National Council on Measurement in Education. – Washington DC: American Psychological Association, 1985.