

**До проблеми забезпечення значущих результатів педагогічних
вимірювань з використанням комп'ютерних технологій.**

Резюме. У статті розглянуто проблеми отримання значущих результатів педагогічних вимірювань. На емпіричних даних, отриманих при впровадженні тестових технологій у навчальний процес університету обґрунтовано методику оцінювання валідності та надійності.

Постановка проблеми. Формування сучасного спеціаліста – процес проблемний та багатогранний. Складною проблемою забезпечення високої якості навчання є постійна корекція навчальних досягнень. Оперативне виявлення результатів навчання отримують через проведення різнопланових педагогічних вимірювань. Поряд із традиційними формами контролю знань, вищі навчальні заклади впроваджують використання тестових методик визначення рівня досягнутих результатів навчання. Використання тестів є науково обґрунтованим інструментом оцінювання знань. Широко висвітлена у педагогічній літературі і проблема отримання значущих результатів педагогічних вимірювань. Впровадження комп'ютерних технологій і навчально-виховний процес та адаптація їх до проведення педагогічних вимірювань породжує проблему адекватності отриманих результатів.

Призначення статті полягає в тому, щоб опираючись на сучасний досвід педагогічної науки у сфері проведення педагогічних вимірювань із застосуванням комп'ютерних технологій, сформулювати методику підтвердження адекватності отриманих результатів рівню досягнутих навчальних результатів.

Аналіз останніх досліджень і публікацій. Процес навчання нерозривно пов'язаний із проблемою перевірки якості засвоєння вивченого матеріалу. Орієнтація сучасних навчальних методик на використання комп'ютерних технологій створює умови для розвитку нового напрямку педагогічної діяльності – індивідуалізоване навчання. Педагогічне обґрунтування

індивідуалізованого навчання можна виявити в працях Й.Г. Песталоцці, Дж. Керрола, Б. Блумба та інших. Завдання підготовки до індивідуалізованого навчання полягає у створенні психолого-педагогічних умов для повного засвоєння навчального матеріалу студентом. Н.Ф. Тализіна виділяє п'ять рівнів засвоєння навчального матеріалу [6]. Згідно із такою організацією навчального процесу перехід на наступний рівень відбувається за результатами педагогічного вимірювання результатів навчання на попередньому. За основу педагогічних вимірювань прийнято класичні роботи Лорда з статистичної теорії [8]. Ряд вчених Росії та західної Європи приймають за теоретичну основу педагогічних вимірювань математичну теорію вимірювання латентних якостей особистості. Математичний апарат даної теорії застосовують для проведення вимірювань у політології, психології, соціології. Але, вихід на арену широкомасштабних вимірювань із застосуванням комп'ютерних технологій вимагає змін і доповнень у концепції моделі педагогічних вимірювань. За теоретичну основу побудови моделі сучасних педагогічних вимірювань можна прийняти дослідження В.С. Аванесова [1; с. 135].

Подання основного матеріалу. Однією із комп'ютерно-орієнтованих методик педагогічних вимірювань є застосування тестів. Результати комп'ютерного тестування надзвичайно інформативні і дозволяють здійснювати контроль за динамікою навчального процесу, будувати індивідуальні стратегії його продовження, прогнозувати результати. Дані результати можуть бути використані як статистична база для проведення аналітичних досліджень для вивчення сучасних методик формування спеціаліста в умовах вищого навчального закладу.

Під час підготовки і проведення тесту ми виходимо з того, що його результати повинні відображати рівень умінь застосовувати знання, давати об'єктивну оцінку ступеня оволодіння навичками, вміннями та видами діяльності практичного характеру. Проте, «головним аргументом проти того, щоб вважати результати тестів надійним показником якості освіти, є те, що

при акцентуванні уваги на підсумкових результатах не беруться до уваги інші фактори, які впливають на успішність або неуспішність»[4]. Множина неврахованих факторів педагогічного вимірювання носить соціально-економічний, мотиваційний та кадровий характер, що значно нівелюється в рамках одного навчального закладу. Використання комп'ютерних технологій під час проведення тестування виключає суб'єктивну складову оцінки, яка завжди присутня у проведенні оцінювання знань викладачем. Комп'ютерне тестування передбачає проведення об'єктивного кількісного співставлення певної властивості студента із деяким сформованим еталоном. У такій постановці проблеми «педагогічне вимірювання – це процес виявлення відповідності між характеристиками студентів, які оцінюються, і точками емпіричної шкали, в якій відношення між різними оцінками характеристик виражені властивостями числового ряду.» [7; 432] Найбільшу надійність, як показує педагогічна практика мають, мають адаптивні тестові технології. Адаптивні тести, як правило розпочинають із завдань середнього рівня і коригують у відповідності до результатів відповідей на кожному окремому етапі. Використання таких методик можливе тільки при використанні спеціалізованих середовищ із досконало організованим аналітичним блоком оперативного відслідковування результатів. Системний модульний контроль за якістю підготовки спеціалістів у нашому університеті ведеться із використанням програмного комплексу Moodle. Даний комплекс призначений для організації дистанційного навчання. Результати тестування проведеного із використанням даного комплексу можна отримати у вигляді узагальненого документа в форматі Excel. Нажаль цих даних недостатньо для проведення поглиблених педагогічних досліджень запропонованого тесту. Провівши дослідження програмного комплексу Moodle ми прийшли до висновку, що первинні дані, які використовуються для оцінювання результатів, можна самостійно досліджувати, попередньо імпортувавши їх із бази даних у програму Excel. Нескладний запит до бази SQL, на основі якої

функціонує програма, дозволить отримати більш детальні відомості про відповіді кожного студента.

Недоліком сучасної педагогічної тестології є відсутність єдиної класифікації основних категорій вимірювання знань. Низька дослідженість основних характеристик педагогічного тестування, як однієї із категорій вимірювання знань, уповільнює їх впровадження у освітню практику. Серед основних категорій вимірювання знань в педагогічній літературі виділяють:

- метод вимірювання;
- інструментарій вимірювання;
- процедура вимірювання;
- процедура оцінювання.

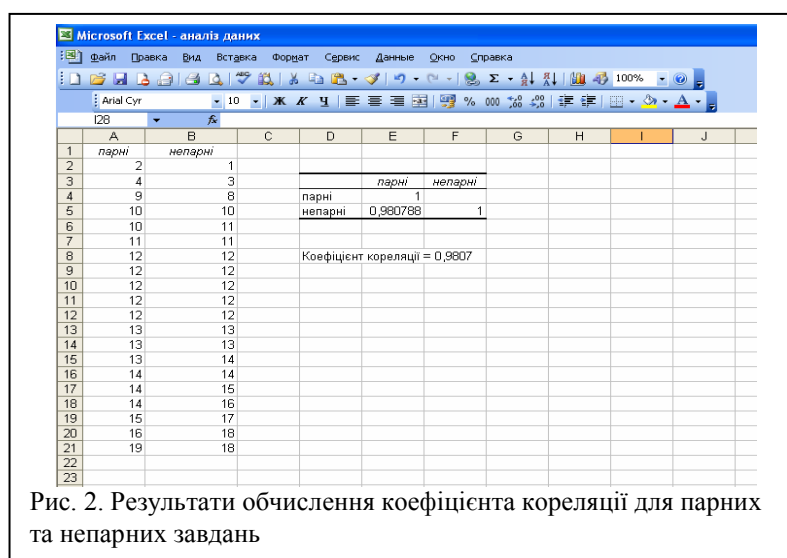
Значущі результати педагогічного вимірювання можна отримати тільки при високій валідності даних категорій. Серед багатьох визначень валідності у педагогічній літературі зупинимось на одному, яке дає А. Анастасі: "поняття, яке вказує, що вимірюють за допомогою тесту і наскільки добре це робиться" [2; 316].

Поняття валідності за своїм змістом складне та багатоаспектне. Окремі автори виділяють валідність – методу, інструментарію, процедури тестування та процедури оцінювання.

Булах І.Є. [3; 221] пропонує проводити аналіз валідності тестових завдань, опираючись на результати статистичного опрацювання даних тестування. Комплексний аналіз результатів опрацювання та інтерпретація дає можливість оцінити якість запропонованих тестів.

Для проведення статистичного аналізу даних, які отримані в результаті використання програми Moodle, формуємо матриці тестових результатів. До матриці включаємо тільки результати правильних відповідей, які позначаємо 1. Допускаємо також, що результати вимірювань які проводились в однакових умовах, збігаються. Для підготовки статистичних даних часто використовують метод повторного тестування, проведення паралельного тестування та ділення результатів тестування пополам. Готуючи статистичну

вибірку, ми використали технологію поділу результатів. Згідно такого підходу до першої вибірки X_n^1 заносимо відповіді студентів на питання з парним номером, до другої X_n^2 – відповіді на непарні питання. Емпіричні дані вимірювань для N студентів, що пройшли тестування, заносимо у двовимірну таблицю програми Excel. Використовуючи команду *Сервіс/Аналіз даних/Кореляція*, обчислюємо кореляційну матрицю для факторів X_n^1 та X_n^2 . За результатами аналізу матриці визначаємо характер зв'язку між незалежними факторами. При значенні коефіцієнта кореляції близькому до 1 результати можна вважати сприйнятливими для подальшого опрацювання.



Графічна інтерпретація емпіричних результатів у вигляді гістограми або зглаженої кривої дає додаткові відомості про нормативно-орієнтований тест. Як показує практика, на основі валідних тестів забезпечується групування до 80% вибірки у центральній частині розподілу. Якщо тест виконали майже всі студенти, або майже ніхто не виконав, то такий тест відбраковується і підлягає переробці, оскільки він визнається недосконалим, або тестування проведено із грубими порушеннями процедури вимірювання.

В педагогічній літературі дається кілька методик для проведення аналізу результатів тестування. За запровадженою в університеті методикою передбачається, що на першому етапі оцінюються міри центральної тенденції сукупності. Для цього визначається мода та середнє арифметичне розподілу

$(\bar{X} = \frac{\sum_{i=1}^N X_i}{N})$. Врахування міри центральної тенденції допомагає оцінити якість тесту та репрезентативність вибірки. При високо валідному тестові забезпечується нормальний розподіл отриманих студентами балів, при якому до 70% емпіричних результатів знаходиться у центральній частині. У випадку, коли мода і середнє арифметичне суттєво відрізняються, тест вважають неправильно сконструйованим і він не може бути визнаним придатним для використання в педагогічних вимірюваннях. Несприйнятливим для використання є тест з бімодальною конфігурацією розподілу результатів. Такий результат свідчить про те, що в тесті використано дві групи питань – легкі та складні, що і приводить до розподілу студентів на дві підгрупи.

Останнім етапом кількісного аналізу є визначення асиметрії та ексцесу кривої розподілу емпіричних даних.

Формула для обчислення асиметрії у розглядуваному випадку має вигляд:

$$A_s = \frac{\sum_{i=1}^N (X_i - \bar{X})^3}{S_x^3 * N},$$

де A_s – асиметрія, X_i – елементи вибірки, \bar{X} – середнє арифметичне вибірки, S_x – стандартне відхилення, що дорівнює кореню квадратному з дисперсії, N – розмір вибірки.

Із використанням вище вказаних позначень ексцес обчислюємо за формулою:

$$E_s = \frac{\sum_{i=1}^N (X_i - \bar{X})^4}{S_x^4 * N}.$$

Обчисливши значення асиметрії та ексцесу, можна зробити висновок про валідність тесту та його недоліки.

Тест вважається валідним при $A_s=0$ та $E_s=0$. Практика експериментальних досліджень великої кількості тестів, що

використовуються в університеті, показує, що такі параметри має невелика кількість з них.

Інші комбінації A_s та E_s оцінюємо за методикою, запропонованою Паращенко Л.І, Леонським В.Д., Леонською Г.І. [5; 217]. Так при $E_s < 0$ тест характеризується високими розподільчими характеристиками, а при $E_s > 0$ – низькими. Валідний тест повинен мати високі розподільчі характеристики, що забезпечує достатню вибірковість під час проведення педагогічних вимірювань.

Показник асиметрії вибірки свідчить про складність тесту. Так при $A_s > 0$ тест вважається складним, при значеннях $A_s < 0$ тест легкий. Наведені на рис результати аналізу тесту дають можливість зробити висновок, що запропонований викладачем тест виявився легким для студентів ($A_s = -0,90$) і має низькі розподільчі характеристики ($E_s = 4,038$).

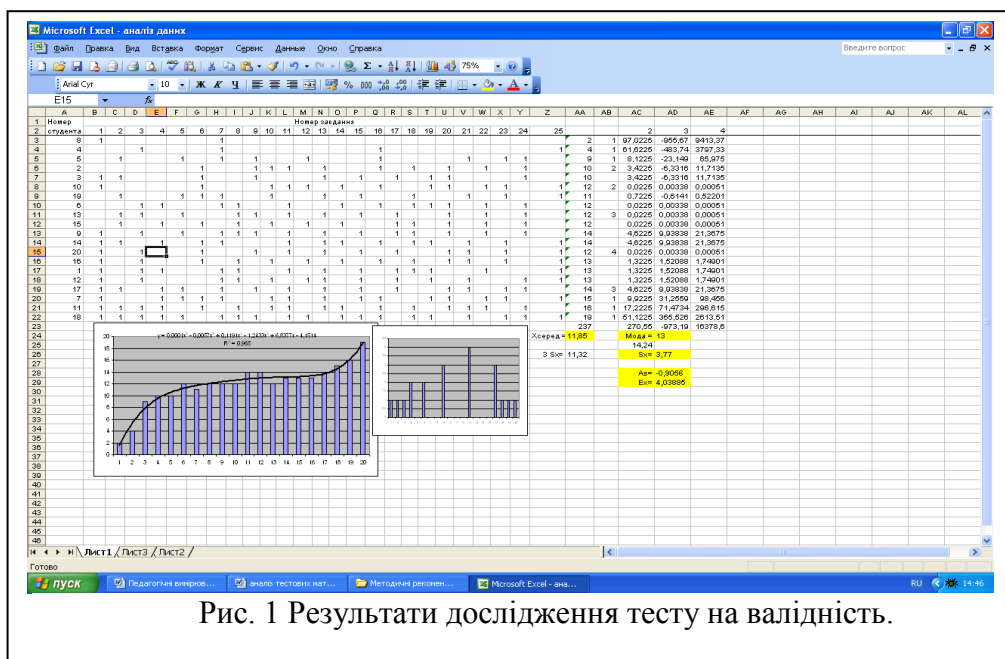


Рис. 1 Результати дослідження тесту на валідність.

Близькість середнього значення і моди дають підставу рекомендувати викладачеві доопрацювати даний тест, зберігши його конфігурацію та спрямованість. Для визначення причин порушення валідності необхідно також проаналізувати педагогічні умови проведення тестів та тривалість тестування.

Завершується аналіз результатів тестування визначенням показника надійності тесту за допомогою прогностичної формули Спірмена-Брауна. Параметром функції для обчислення надійності є коефіцієнт кореляції.

Формула Спірмена-Брауна для даного випадку має вигляд

$$C = \frac{2 * r}{1 + r},$$

де C – коефіцієнт надійності, r – коефіцієнт кореляції, розрахований для даних, отриманих за методикою поділу навпіл. За наведеними даними $r = 0,98$, отже $C = \frac{2 * 0,98}{1 + 0,98} = 0,98$. Даний коефіцієнт характеризує послідовність

вимірювань і дає відповідь на ряд важливих питань. Використаний тест виявився достатньо надійним, це означає що і при іншій серії вимірювань отримали б такі самі результати.

Надійні тести із різними показниками асиметрії можуть використовуватись в моделі адаптивних вимірювань. При отриманні низьких результатів повторне тестування проводиться із використанням легших тестів.

Висновок. Проблема забезпечення значущих результатів педагогічних вимірювань при використанні тестових завдань може бути розв'язана через використання надійних методик статистичного аналізу отриманих результатів на етапі апробації та впровадження. На підставі вірогідних результатів дослідження результатів тестування можна зробити висновок про шляхи удосконалення педагогічних вимірювань в навчальному закладі.

Література

1. Аванесов В.С. Математические модели педагогических измерений. Научное издание. М.: Исс. центр проблем качества подготовки специалистов, 1994. – 26 с.; «Научные проблемы тестового контроля знаний». М.: Исс. Центр, 1994. – 135 с.
2. Анатази А. Психологическое тестирование: Пер. С англ.: В 2-х кн./ Под редакцией К.М. Гуревич, В.И. Лобовского. – М.: Педагогіка, 1982. – Кн. 1. – 316 с.

3. Булах І.Є. Комп'ютерна діагностика навчальної успішності. – К.: ЦМК МОЗ України, УДМУ. – 1995. – 221 с.
4. Драммонд Т. Використання результатів тестування для оцінювання якості освіти: за і проти // Вісник. Тестування і моніторинг в освіті. – 2006. – №6. – С. 31-34.
5. Паращенко Л.І. Тестові технології у навчальному закладі :Метод. Посібник / Паращенко Л.І, Леонський В.Д., Леонська Г.І; Наук. Ред.. О.І. Ляшенко – К.: (ТОВ «Майстерня книги»), 2006. – 217 с.
6. Талызина И.С. Управление процессом усвоения знаний. – М., 1984.
7. Чельшкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. – М.: Логос, 2002. – 432 с.
8. Lord F.M. Novick M.R. Statistical Theories of Mental Test Scores. Reading, MA: Addison-Wesley. 1968.