

КОНСТРУЮВАННЯ Й АПРОБАЦІЯ ПЕДАГОГІЧНОГО ТЕСТУ ДЛЯ ВИМІРЮВАННЯ РІВНЯ ІНЖЕНЕРНО-ГРАФІЧНОЇ ПІДГОТОВКИ МАЙБУТНІХ УЧИТЕЛІВ ТЕХНОЛОГІЙ

У статті розкрито основні етапи конструювання педагогічного тесту для вимірювання рівня інженерно-графічної підготовки майбутніх учителів технологій. Особливу увагу приділено апробації тестових завдань, що передбачає: визначення трудності і дискримінативності завдань; виявлення завдань, що потребують доопрацювання; визначення науково-обґрунтованих критеріїв якості тесту (надійності, валідності); встановлення часових норм, необхідних для виконання як окремих завдань, так і тесту в цілому; визначення вагових коефіцієнтів тестових завдань; інтерпретацію і шкалювання одержаних результатів тестування.

Ключові слова: інженерно-графічна підготовка, вчитель технологій, педагогічне вимірювання, педагогічний тест.

Постановка проблеми. Нині тестові технології набувають широкого розповсюдження у системі освіти. Педагогічний тест, на думку В. Аванесова, – це система паралельних завдань зростаючої складності і специфічної форми, яка дає змогу якісно й ефективно виміряти рівень та структуру підготовленості досліджуваних [1].

У контексті наукового пошуку, спрямованого на перевірку ефективності методичної системи навчання інженерно-графічних дисциплін майбутніх учителів технологій, тестування виступає основним методом науково-педагогічного дослідження. Завдяки стандартизованій процедурі проведення та заздалегідь спроектованій технології обробки й аналізу результатів, тестування уможливує одержання достовірних відомостей про перебіг дослідно-експериментальної роботи, встановлення об'єктивного рівня інженерно-графічної підготовки студентів на різних етапах навчання.

Аналіз останніх досліджень і публікацій. Фундаментальні основи психолого-педагогічного діагностування відображені у численних працях зарубіжних та вітчизняних учених. Впровадженню діагностичних методик у навчально-пізнавальний процес присвячені наукові роботи Г. Айзенка, В. Аванесова, В. Беспалька, А. Піскунова, Л. Паращенко, М. Челишкової, Д. Чернилевського та ін. Проблема використання тестових технологій для вимірювання якості навчальних досягнень учнівської та студентської молоді з графічних дисциплін постає предметом наукового пошуку Н. Вересоцької, І. Головачука, А. Корнєвої, О. Маркової, М. Репетенка, Н. Титової та ін.

Мета статті – дослідити основні етапи конструювання й апробації педагогічного тесту для вимірювання рівня інженерно-графічної підготовки майбутніх учителів технологій.

Виклад основного матеріалу. Встановлення кількісної характеристики рівня інженерно-графічної підготовки майбутніх учителів технологій на усіх етапах дослідно-експериментальної роботи здійснювалося за допомогою педагогічного тестування, що передбачало використання різномірних тестових завдань, представлених у відповідній формі.

Процес розробки педагогічних тестів здійснювався відповідно до основних наукових положень тестології – зумовлював конструювання, апробацію (експертизу) та впровадження тестової методики, визначення процедури вимірювання, а також обробку, аналіз й оцінювання одержаних результатів (виявлення рівня сформованості інженерно-графічних знань й умінь студентів).

У процесі формування змісту тестових завдань враховувалися основні вимоги, окреслені відомими вченими-тестологами [1; 2; 3; 4; 5]: правильність розташування елементів завдання (умови, варіантів відповідей, додаткових відомостей); логічна форма викладу умови завдання, її лаконічність й однозначність; правильність форми представлення завдання; оптимальна кількість варіантів відповідей; раціональність способів оцінювання.

Процес створення педагогічних тестів для виявлення початкового й вихідного (кінцевого) рівнів інженерно-графічної підготовки майбутніх учителів технологій передбачав попередню процедуру апробації, необхідну для коригування окремих завдань (заміни, доповнення, видозміни та ін.) та перевірки науково-обґрунтованих критеріїв якості (надійності, валідності) тестів з метою забезпечення оптимальної ефективності педагогічного вимірювання.

Етап апробації тестових завдань передбачав: 1) визначення трудності і дискримінативності завдань – для оцінювання їх придатності цілям тестування; 2) виявлення завдань, що потребують доопрацювання (містять помилки, неточності, спірне трактування та ін.); 3) визначення науково-

обґрунтованих критеріїв якості тесту (надійності, валідності); 4) встановлення часових норм, необхідних для виконання як окремих завдань, так і тесту в цілому; 5) визначення вагових коефіцієнтів тестових завдань; 6) інтерпретацію і шкалювання одержаних результатів.

Трудність тестового завдання визначається кількістю і характером розумових операцій, необхідних для успішного його виконання. Проте, такий спосіб не дає достовірної інформації про трудність завдання, оскільки неможливо точно встановити перебіг мисленнєвого процесу особистості. Тому, на практиці трудність тестового завдання знаходять емпірично – підрахунком частки правильних і неправильних відповідей [1]. Відповідно до цього, у процесі апробації тестових завдань формується бінарна матриця результатів, яка відображає успішність (неуспішність) знаходження правильної відповіді суб'єктами тестування й слугує базою для арифметичного визначення трудності кожного завдання.

Для позначення трудності тестових завдань використовують спеціальний показник – індекс трудності (I_T), який вказує на частку досліджуваних (студентів), що не змогли правильно виконати завдання. Індекс трудності обчислюють за формулою [3]:

$$I_T = 100 \cdot \left(1 - \frac{n}{N}\right),$$

де I_T – індекс трудності завдання у відсотках;

n – кількість студентів, які змогли успішно виконати завдання;

N – загальна кількість студентів, залучених до апробації тесту.

Трудність тестового завдання визначає його придатність (непридатність) для педагогічного вимірювання. На думку В. Аванесова [1], завдання з коефіцієнтом трудності більшим за 80 % і меншим за 20 % необхідно переглядати (коригувати, замінювати), оскільки вони мало інформативні й погано диференціюють студентів за рівнем навчальних досягнень (рівнем інженерно-графічної підготовки).

Необхідно враховувати, що трудність тестового завдання зумовлюється не лише характером мисленнєвої діяльності особистості, спрямованої на його розв'язання, але й формою та способом представлення завдання [3]. Відповідно, коригування трудності окремих завдань потрібно здійснювати як за змістом, так і за формою (відкрита, закрита) чи способом відображення (словесна, графічна, символічна).

Одержане значення показника трудності тестового завдання, а також характер навчально-пізнавальної діяльності студентів у процесі його розв'язання, слугують базисом для встановлення необхідної ваги тестового завдання (оптимальної кількості балів для оцінювання).

У процесі апробації тесту, окрім трудності завдань визначають їх дискримінативність (диференційну здатність), що полягає у відсіюванні неякісних завдань, які погано диференціюють студентів за рівнем успішності (рівнем інженерно-графічної підготовки).

Дискримінативність завдання встановлюється за величиною індексу диференційної здатності, який вказує на точність виявлення найбільш (або найменш) підготовлених студентів, тобто характеризує ступінь розмежування студентів з різною продуктивністю навчальної діяльності [3].

Індекс диференційної здатності тестових завдань обчислюється методом крайніх груп: враховуються результати студентів, які найбільш і найменш успішно пройшли тестування. Відповідно до цього, індекс диференційної здатності тестових завдань як різницю частки досліджуваних (студентів) із сильної та слабкої груп можна визначити за формулою [3]:

$$I_D = \frac{H_1}{N_1} - \frac{H_2}{N_2},$$

де: H_1 і H_2 – кількість студентів сильної та слабкої груп відповідно, що правильно розв'язали завдання; N_1 і N_2 – загальна кількість студентів у сильній та слабкій групах.

Тестові завдання з низьким індексом диференційної здатності ($I_D < 0,3$) підлягають перегляду (замінюються або видозмінюються). На думку О. Майорова, типовими недоліками низькодискримінативних завдань є надмірна складність, незрозумілість формулювання; неоднозначність умов; очевидність розв'язку; абсурдність й нереальність варіантів відповіді; існування декількох правильних відповідей, що не обумовлено у завданні [3].

Усунення означених недоліків на етапі апробації тесту дає змогу підвищити дискримінативність завдань й забезпечити достатньо малу ймовірність їх успішного розв'язання (нерозв'язання) студентами з різним рівнем інженерно-графічної підготовки, що сприяє підвищенню об'єктивності результатів тестування.

Важливою характеристикою тесту є його надійність, що визначає ступінь відтворюваності результатів тестування, їх точність [2].

Встановлення надійності тестів для вхідного й підсумкового діагностування рівня інженерно-графічної підготовки студентів здійснювалося з допомогою коефіцієнта кореляції Пірсона на основі індивідуальних балів студентів, одержаних за тестування.

Для визначення надійності тестів використовувався метод розщеплювання (поділу) одержаних результатів за парними й непарними завданнями з наступним обчисленням відповідної кореляції: чим вища кореляція, тим вища надійність тесту. Такий метод зручний, оскільки передбачає лише одну сесію тестування.

Підрахунок коефіцієнта надійності тестів методом розщеплювання здійснювався за формулою [2]:

$$r_{XY} = \frac{N \sum_{i=1}^N X_i Y_i - \left(\sum_{i=1}^N X_i \right) \cdot \left(\sum_{i=1}^N Y_i \right)}{\sqrt{\left(N \sum_{i=1}^N X_i^2 - \left(\sum_{i=1}^N X_i \right)^2 \right) \cdot \left(N \sum_{i=1}^N Y_i^2 - \left(\sum_{i=1}^N Y_i \right)^2 \right)}}$$

де r_{XY} – коефіцієнт надійності тесту за методом розщеплювання;

X_i та Y_i – індивідуальні бали i -го студента за виконання парного і непарного завдань тесту відповідно;
 N – кількість студентів, що брали участь в апробації тесту ($N = 150$).

Таким чином, у процесі обчислень було встановлено коефіцієнти надійності (кореляції Пірсона) тестів для вхідного та підсумкового тестування рівня інженерно-графічної підготовки студентів, які склали 0,78 та 0,73 відповідно. Проте, одержані результати відображають надійність тестів лише за половиною завдань, тому для отримання загальних (повних) коефіцієнтів надійності тестів необхідно скористатися формулою Спірмена-Брауна [2]:

$$r_n = \frac{2 \cdot r_{XY}}{1 + r_{XY}}$$

Остаточний коефіцієнт надійності вхідного тесту склав 0,87, а підсумкового – 0,85. Оскільки одержані значення коефіцієнтів надійності тестів виявилися більшими за 0,7, можна стверджувати про високу надійність й узгодженість одержаних результатів вимірювання [2].

Забезпеченню прийнятної надійності тестів сприяло комплексне врахування негативних чинників, які знижують об'єктивність результатів тестування [5]:

1. *Суб'єктивізм в оцінюванні завдань тесту*, що проявляється здебільшого при оцінюванні відкритих завдань з розгорнутою (вільною) відповіддю. Для усунення цього фактору пропонувалися тестові завдання здебільшого закритої форми, або відкритої з короткою (однозначною) відповіддю.

2. *Можливість вгадування правильної відповіді*, особливо студентами з низьким рівнем інженерно-графічної підготовки. Для мінімізації цього чинника, при формуванні неправильних відповідей (дистракторів) закритих завдань враховувалися такі рекомендації:

- усі варіанти відповіді повинні бути однаково правдоподібними;
- кількість дистракторів у завданнях з однією правильною відповіддю повинна бути не меншою 4; у завданнях з декількома правильними відповідями – не меншою 6;
- усі варіанти відповіді мають бути чіткими, лаконічними, виражати єдину думку й, по можливості, – однакові за розміром;
- дистрактори мають розташовуватися у строго визначеній послідовності (в алфавітному порядку, за величиною, за ступенем зростання чи спадання ключової ознаки та ін.);
- правильна відповідь у різних завданнях не повинна періодично повторюватися або постійно бути під однаковим номером;
- умова завдання не повинна містити будь-які вербальні асоціації, що сприяють вибору правильного варіанта відповіді шляхом здогадування;
- слова, що часто повторюються, виключаються з відповідей й вводяться в основний текст завдання;
- у дистракторах не рекомендується використовувати слова, які сприяють вгадуванню правильної відповіді (наприклад, "всі", "ніколи", "жоден" та ін.);
- необхідно уникати відповідей, що взаємозумовлюються.

3. *Відсутність логічно коректного формулювання умови завдання*, що ускладнює його розуміння й сприяє, здебільшого, несвідомому вибору правильної відповіді.

4. *Необігрунтований вибір вагових коефіцієнтів тестових завдань*. Кількість балів за оцінювання кожного тестового завдання встановлювалася на етапі апробації, залежно від коефіцієнту трудності. Завдання з низькою трудністю (легші) характеризувалися меншим ваговим коефіцієнтом; з високою (важчі) – більшим.

5. *Довжина тесту*. Збільшення довжини тесту призводить до підвищення його надійності, проте зростає ймовірність досягнення порогу втомлюваності студентів при роботі із занадто довгими тестами

й, як наслідок, зростання кількості випадкових помилок. Враховуючи результати досліджень в галузі тестології [1; 3; 5], нами прийнято оптимальну довжину тестів для вхідного і проміжного діагностування – 45 завдань.

6. *Відсутність стандартної інструкції до тесту.* Розроблені тести для вхідного й проміжного діагностування рівня інженерно-графічної підготовки майбутніх учителів технологій містили вичерпні відомості щодо процедури тестування, способів вибору правильного варіанта відповіді, принципу нарахування балів й оцінювання студентів.

Означені вище чинники виявлялися й усувалися на етапі апробації тестів та їх застосування.

Висока надійність тесту є необхідною, проте не цілком достатньою умовою для обґрунтування його якості. Тому педагогічні тести піддають додатковій перевірці на валідність результатів.

На думку В. Аванесова, валідність визначає придатність тестових результатів для досягнення цілей тестування [1], а О. Майоров переконаний, що валідність характеризує спроможність тесту адекватно відобразити ті якості досліджуваних, які піддаються вимірюванню [3].

Розрізняють три види валідності тестів – змістову, конструктну і критеріальну (емпіричну). Змістова валідність характеризує тест за ступенем його відповідності предметній області (галузі знань, змісту навчального предмета та ін.); конструктна – визначається у випадку представлення вимірювальних властивостей у формі абстрактного образу (концептуальної моделі); критеріальна – передбачає наявність зовнішнього критерію для кореляції результатів тестування [2].

У межах наукового дослідження було обрано критеріальну валідність тестів, при цьому зовнішнім критерієм виступали результати виконання студентами підсумкової контрольної (графічної) роботи з креслення.

Валідність педагогічного тесту характеризується величиною коефіцієнта валідності, який обчислювався як коефіцієнт кореляції Пірсона між результатами тестування та успішністю виконання підсумкової контрольної (графічної) роботи з креслення за формулою [5]:

$$r_{xy} = \frac{\sum_{i=1}^N (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{N \cdot \sqrt{S_x^2 \cdot S_y^2}},$$

де X_i – кількість балів i -го студента за тестування;

Y_i – кількість балів i -го студента, отримана за виконання підсумкової контрольної (графічної) роботи з креслення;

\bar{X} та \bar{Y} – середні значення результатів тестування та виконання підсумкової контрольної (графічної) роботи з креслення;

N – кількість студентів, що брали участь в апробації тесту ($N = 150$).

S_x^2 і S_y^2 – дисперсії балів студентів за тестування і виконання підсумкової контрольної (графічної) роботи з креслення відповідно.

Коефіцієнт валідності тесту для вхідного діагностування рівня інженерно-графічної підготовки майбутніх учителів технологій становив 0,79, для підсумкового діагностування – 0,77. Оскільки одержані значення коефіцієнтів валідності тестів виявилися більшими за 0,6, можна стверджувати про їх високу валідність й одержання достовірних результатів вимірювання [4].

Важливим на етапі апробації педагогічного тесту є встановлення оптимальної тривалості процедури тестування, що визначається кількістю і складністю представлених завдань. Відомий дослідник О. Майоров зазначає, що кожен тест має характеризуватися встановленим часом виконання, зменшення або збільшення якого знижує кількісні показники тесту [3]. Отже, некоректне встановлення тривалості тестування унеможливує досягнення його мети й одержання об'єктивних результатів.

На думку В. Аванесова, визначення часу тестування є важливим системоутворюючим чинником при розробці тесту. При цьому вчений вважає, що оптимальна тривалість тестування – це час від початку процедури вимірювання до моменту настання втомлюваності, який можна встановити емпіричним шляхом, відслідковуючи момент досягнення максимуму дисперсії тестових результатів [1].

Отже, при визначенні тривалості тестування мають враховуватися такі основні положення: 1) тривалість тестування встановлюється емпірично на основі одержаних результатів апробації тесту; 2) час виконання тесту доцільно розрахувати, керуючись величиною дисперсії (появою максимального значення) для серії тестувань різних, проте якісно однорідних вибірок студентів; 3) тривалість тестування не має перевищувати 90 хв.; 4) довжина тесту не повинна перевищувати 60 – 70 завдань [2].

Таким чином, у процесі апробації тестів для вхідного і підсумкового діагностування рівня інженерно-графічної підготовки студентів було встановлено оптимальну тривалість процедури вимірювання – 90 хв.

Необхідним етапом у процесі конструювання й апробації тестів є встановлення ваги кожного завдання, тобто оптимального вагового коефіцієнта, який враховується при визначенні кількості балів.

Вага тестового завдання встановлюється, залежно від його складності, тобто характеру навчально-пізнавальної діяльності студентів у процесі розв'язання, а також з урахуванням показника трудності завдання, одержаного емпіричним шляхом.

Заключний етап апробації розроблених тестів полягав у встановленні шкали оцінювання результатів тестування, яка б уможливила об'єктивне визначення відповідного рівня інженерно-графічної підготовки майбутніх учителів технологій. Можливість розв'язання тестових завдань відповідної групи (рівня) складності безпосередньо залежить від характеру навчально-пізнавальної діяльності студентів у процесі знаходження правильної відповіді, сформованості інженерно-графічних умінь і навичок, а також розвитку індивідуальних особливостей особистості (мислення, уяви, пам'яті та ін.). Відповідно до цього можна стверджувати, що успішність розв'язання студентами тестових завдань певного рівня складності може слугувати об'єктивним показником їхньої інженерно-графічної підготовки.

Висновки. Узагальнюючи вище зазначене, можна зробити висновок, що ефективність тестових технологій й об'єктивність одержаних результатів педагогічного вимірювання залежить від чіткості дотримання процедури конструювання й апробації тестових завдань. Особливу увагу у процесі апробації тесту необхідно приділяти визначенню трудності і дискримінативності завдань, а також формі їх представлення; визначенню науково-обґрунтованих критеріїв якості тесту (надійності, валідності); встановленню часових норм проведення тестування; визначенню вагових коефіцієнтів тестових завдань; інтерпретації і шкалюванню одержаних результатів.

Використані джерела

1. Аванесов В.С. Композиция тестовых заданий / В.С. Аванесов. – М. : Адепт, 1998. – 272 с.
2. Ким В.С. Тестирование учебных достижений: [монография] / В.С. Ким. – Уссурийск : Из-во УГПИ, 2007. – 214 с.
3. Майоров А.Н. Теория и практика создания тестов для системы образования. (Как выбирать, создавать и использовать тесты для целей образования) / А.Н. Майоров. – М. : "Интеллект-центр", 2001. – 296 с.
4. Парашенко Л.І. Тестові технології у навчальному закладі : [метод. пос.] / Л.І. Парашенко, В.Д. Леонський, Г.І. Леонська / наук. ред. О.І. Ляшенко. – К. : ТОВ "Майстерня книги", 2006. – 217 с. : іл.
5. Чельшкова М.Б. Теория и практика конструирования педагогических тестов : [учеб. пос.] / М.Б. Чельшкова. – М. : Логос, 2002. – 432 с. : ил.

Nyshchak I.

METHOD OF CONSTRUCTION OF TEST FOR LEVEL MEASUREMENT OF ENGINEERING-GRAPHIC PREPARATION OF FUTURE TEACHERS OF TECHNOLOGY

In the context of scientific research aimed at checking the effectiveness of methodical system of teaching engineering-graphics disciplines of future teachers of technology, testing is the main method of scientific and pedagogical research.

Pedagogical test – a system of parallel tasks of increasing difficulty and a specific form that allows you to qualitatively and effectively measure the level and structure of preparedness of investigated. Thanks to the standardized procedure and pre-designed technology of processing and analysis of test results it is possible to obtain objective information about the course of the experimental work, establishing the objective level of engineering-graphic preparation of students at various stages of training. The process of developing of pedagogical tests should be carried out in accordance with the main of scientific statements and provide design, expertise and implementation of test methods, establishing procedures for testing and processing, analysis and evaluation of the results (determination of the level of engineering-graphics knowledge and skills).

In the process of shaping the content of tests must take into account the following requirements: correct location of tasks (conditions, answers, additional information); the logical form presentation of the conditions tasks, its concise and unambiguous; correct presentation of the task; the optimal number of answers; rational methods of evaluation. The process of creating pedagogical test to identify primary and final level of engineering-graphic preparation of future teachers of technology should include preliminary testing, necessary for correcting some tasks (substitutions, additions, modifications, etc.), and verification of scientifically based criteria of quality of the tasks (reliability, validity) to ensure optimum effectiveness of pedagogical measurement.

Stage of verification of the test involves definition of difficulties and differential ability of tasks; identify tasks that need to be updated (submitted with errors, inaccuracies, and controversial interpretations etc.); definition of scientifically based criteria of quality of the test (reliability, validity); setting of the time necessary to perform a specific task, and the test as a whole; the definition of weighting coefficients of tasks; scaling and interpretation of the test results.

Key words: *engineering-graphics preparation, teacher of technology, educational measurement, pedagogical test.*

Стаття надійшла до редакції 01.03.2016 р.