

Чуйко О. С., Сілін Є. С.

ПЕДАГОГІЧНІ ВИМІРЮВАННЯ

рекомендовано до друку

Вченою радою Державного вищого навчального закладу

«Донбаський державний педагогічний університет»

як навчальний посібник

Слов'янськ

2019

УДК 37.091.212.2

Ч 87

Чуйко О. С., Сілін Є. С.

*Педагогічні вимірювання: Навчальний посібник – Слов'янськ:
Видавництво Б. І. Маторіна. – 2019. – 112 с.*

У посібнику викладено основні теми курсу дисципліни «Педагогічні вимірювання», які традиційно вивчаються в педагогічних закладах вищої освіти. Кожна тема містить стислі теоретичні відомості, типові приклади, питання для самостійної роботи.

Для студентів і викладачів закладів вищої освіти.

*Видання друкується за рішенням Вченої ради
Державного вищого навчального закладу
«Донбаський державний педагогічний університет»
Протокол № 1 від 24.09.2019*

Рецензенти:

декан фізико-математичного факультету, кандидат фізико-математичних наук, доцент кафедри математики та інформатики Кадубовський О. А.

учений секретар ІПММ НАН України, кандидат фізико-математичних наук, доцент Несмелова О. В.

Відповідальні за випуск:

старший викладач кафедри математики та інформатики *Дзюба М. В.*,

доцент кафедри математики та інформатики *Сілін Є. С.*

© *Чуйко О. С., Сілін Є. С.*, 2019

ЗМІСТ

ВСТУП	4
1. ПЕДАГОГІЧНА ДІАГНОСТИКА	7
1.1. ПОНЯТТЯ ТА КАТЕГОРІЇ ПЕДАГОГІЧНОЇ ДІАГНОСТИКИ	7
1.2. ТИПИ ВИМІРЮВАНЬ. КЛАСИФІКАЦІЯ ТЕСТІВ.....	14
2. ІСТОРІЯ РОЗВИТКУ ТЕСТУВАННЯ	22
3. ЗАГАЛЬНІ ВИМОГИ ДО ЯКОСТІ ТЕСТІВ	35
3.1. КРИТЕРІЇ ЯКОСТІ МЕТОДУ ВИМІРЮВАННЯ	35
3.2. РАНГОВА КОРЕЛЯЦІЯ	39
4. ЕТАПИ ТА ПРАВИЛА СТВОРЕННЯ ТЕСТІВ	44
4.1. ВИЗНАЧЕННЯ МЕТИ І ЗАВДАНЬ ТЕСТУВАННЯ.....	44
4.2. СКЛАДАННЯ ТЕСТОВИХ ЗАВДАНЬ.....	47
4.3. АПРОБАЦІЯ ТЕСТУ	52
4.4. ОБРОБКА ВІДПОВІДЕЙ.....	53
4.5. ФОРМУВАННЯ БАНКУ ТЕСТОВИХ ЗАВДАНЬ	55
5. РОЗРОБКА СПЕЦИФІКАЦІЇ ТЕСТУ	56
5.1. МАТРИЦЯ ТЕСТУ	56
5.2. КЛАСИФІКАЦІЯ НАВЧАЛЬНИХ ЦІЛЕЙ.....	58
6. ФОРМАТИ ТЕСТОВИХ ЗАВДАНЬ	61
6.1. ТЕСТОВІ ЗАВДАННЯ ЗАКРИТОЇ ФОРМИ.....	61
6.2. ТЕСТОВІ ЗАВДАННЯ ВІДКРИТОГО ТИПУ	75
7. АНАЛІЗ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ	80
7.1. МЕТОДИ ПІДРАХУНКУ БАЛІВ ЗА ЗАВДАННЯ ТЕСТУ	80
7.2. КОНВЕРТАЦІЯ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ У ШКАЛИ.....	84
7.3. СТАТИСТИЧНИЙ АНАЛІЗ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ.....	86
8. ПОМИЛКИ ТА ТЕХНІЧНІ ДЕФЕКТИ ТЕСТОВИХ ЗАВДАНЬ І ТЕСТУВАННЯ	92
8.1. ПОМИЛКИ ЗМІСТУ ТЕСТІВ	92
8.2. ПОМИЛКИ ОРГАНІЗАЦІЇ ТЕСТУВАННЯ	93
9. ПРИКЛАДИ ПОМИЛОК ТА ДЕФЕКТІВ ТЕСТОВИХ ЗАВДАНЬ	96
ВИСНОВКИ	99
ДОДАТОК 1. ПИТАННЯ ДО ЗАЛІКУ.....	101
ДОДАТОК 2. КОРОТКИЙ ТЛУМАЧНИЙ СЛОВНИК	103
ДОДАТОК 3. КРИТИЧНІ ЗНАЧЕННЯ КОЕФІЦІЄНТУ РАНГОВОЇ КОРЕЛЯЦІЇ.....	109
РЕКОМЕНДОВАНА ЛІТЕРАТУРА	110

ВСТУП

В умовах розбудови національної системи освіти, інтеграції в світову систему особливо актуальним стає забезпечення належного рівня фахової підготовки майбутнього вчителя. Важливою складовою професійної підготовки вчителя є навчальна дисципліна «Педагогічні вимірювання».

Представлений навчально-методичний посібник вміщує матеріали для організації аудиторної та самостійної роботи студентів. Тут наведено теоретичний та практичний матеріал, перелік основних питань курсу, питання для самоконтролю знань студентів, короткий тлумачний словник, список рекомендованої літератури.

Метою написання посібника є висвітлення проблем педагогічного тестування, його переваг та недоліків, технологій створення тестів; аналіз вимог до якості тестів та умов ефективного впровадження тестового контролю знань; розкриття класифікації тестових завдань, характеристика композиції завдань різних типів (формат, структура, правила оформлення, принципи оцінювання та інтерпретації результатів); аналіз помилок та технічних дефектів тестових завдань та процедур тестування.

Інтеграція України у міжнародне співтовариство вимагає співставлення освітніх рівнів, а також узгодження принципів підготовки спеціалістів, оскільки саме рівень професійних знань фахівців є одним із основних показників, що визначають добробут населення. Тому першочерговим завданням нині є запровадження сучасних технологій навчання й оцінювання.

Гостроту цієї проблеми зумовлено, з одного боку, появою приватних закладів освіти та можливістю навчання за кордоном, з іншого – застарілістю наукових основ дидактики контролю, недосконалістю системи оцінювання освітніх досягнень тих, хто навчається. А як відомо, саме контроль результатів забезпечує зворотний зв'язок у процесі навчання.

Вибір засобів педагогічної діагностики базується на відповідних дидактичних принципах вимірювання результатів навчання. Лише

підготовлений відповідним чином комплекс знань дає змогу з використанням певних діагностичних методів правильно оцінити рівень знань і вмінь суб'єктів навчання. Ось чому в педагогіці останнім часом виникла тенденція до використання кількісних методів педагогічного контролю. Серед засобів об'єктивного контролю найбільш науково обґрунтованим є метод тестування із залученням технічних засобів для сканування та обробки результатів.

У контексті реформування національної освіти набувають актуальності питання моніторингу якості вищої освіти. Нових перспектив набуває використання ефективної системи педагогічного оцінювання як інструменту виміру результатів педагогічного процесу та навчальних досягнень учнів.

Тестологія (наука про тести) – це напрям наукових досліджень у галузі теорії і практики вимірювання та оцінювання якісних характеристик, притаманних людині, де методом вимірювання є тестування. Цей метод передбачає, що інструментом вимірювання є тест, складений з тестових завдань; процедурою вимірювання є тестування; методом оцінювання є шкалювання.

Історія сучасної тестології почалася з появою перших зразків шкільних тестів успішності як інструменту індивідуальних вимірів ще у 1864 р. у Великобританії. 1917 – 1918 рр. у США були розроблені перші групові тести. З 60-х років ХХ ст. почалася нова хвиля використання тестів завдяки розробці педагогом Р. Тайлером програми National Assessment of Educational Progress (NAEP) – Державної програми педагогічних вимірів.

Фундаментальні сучасні дослідження тестування як методу педагогічної діагностики, висвітлені в роботах зарубіжних учених В. С. Аванесова, Т. Аджер, А. Анастасі, В. П. Беспалька, Л. Ф. Бурлачука, Дж. Гласс, Н. Гронланда, Л. Долінера, Р. Ібел, К. Інгенкампа, П. Клайна, А. Майорова, Р. Намбельтона, В. Переверзева, І. Раппопорта, Дж. Стенлі, М. Челишкової та ін.

У наш час практичне втілення ідеї тестування відбулося із введенням зовнішнього незалежного оцінювання (ЗНО) – стандартизованого тестування,

проходження якого є обов'язковим для всіх випускників українських шкіл, які бажають вступати до закладів вищої освіти України. Умови проведення всіх тестувань визначаються Українським центром оцінювання якості освіти – установою, що забезпечує організацію проведення ЗНО в Україні. Українським центром оцінювання розробляються завдання кожного тесту, встановлюється час на виконання цих завдань, правила виконання цих завдань та система їх оцінювання.

Теоретиками і практиками з тестології стали українські науковці Я. Болюбаш, І. Булах, В. Гогунський, Л. Гриневич, А. Єріна, Л. Коваленко, О. Локшина, К. Михайлов, М. Мруга, Л. Леонський, О. Ляшенко, М. Олійник, Л. Паращенко, Х. Порсева, Г. Терещук, Л. Фігурська, В. Хмельницький, В. Шпильовий, О. Яковенко та інші.

1. ПЕДАГОГІЧНА ДІАГНОСТИКА

1.1. ПОНЯТТЯ ТА КАТЕГОРІЇ ПЕДАГОГІЧНОЇ ДІАГНОСТИКИ

На думку провідного дослідника проблеми діагностування в освітньо-виховному процесі, німецького вченого К. Інгенкампа, педагогічна діагностика покликана, по-перше, оптимізувати процес індивідуального навчання, по-друге, в інтересах суспільства забезпечити правильне визначення результатів навчання і, по-третє, керуючись виробленими критеріями, звести до мінімуму помилки при переведенні учнів із однієї освітньої групи в іншу, при спрямуванні їх на різні курси у виборі спеціалізації навчання. Для досягнення подібної мети в процесі діагностичних процедур, з однієї сторони, визначаються передумови до навчання, які є в окремих індивідуумів і в представників освітньої групи загалом, а з другої, визначаються умови, необхідні для організації планомірного процесу освіти. За допомогою педагогічної діагностики аналізується освітній процес і визначаються результати навчання. При цьому під діагностичною діяльністю К. Інгенкамп розуміє процес, в ході якого (з використанням діагностичного інструментарію чи без нього) дотримуючись необхідних наукових критеріїв якості, вчитель спостерігає за учнями і проводить анкетування, обробляє дані спостережень і опитувань та повідомляє про отримані результати з метою характеристики поведінки, пояснення її мотивів чи передбачення поведінки в майбутньому.

Педагогічна діагностика розглядається також як процес розпізнавання явищ і визначення їх стану в певний момент на основі використання необхідних для цього параметрів. Таке трактування дозволяє виділити специфіку педагогічної діагностики як самостійного виду діяльності, розкрити її сутність, цілі і завдання.

Останнім часом поле діяльності педагогічної діагностики суттєво розширюється, вона стає все більш поліфункціональною, включає в себе такі функції як аналітична, оцінкова, функція контролю, корекційна, орієнтаційна, інформаційна, функції мотивації і стимулювання.

Аналітична функція – це психолого-педагогічний аналіз освітнього процесу на всіх рівнях його структурної організації як педагогічної системи. На відміну від контролюючої функції, спрямованої лише на виявлення недоліків, аналітична функція виявляє причинно-наслідкові зв'язки в освітньо-виховному процесі, між умовами і результатами навчання.

Власне-діагностична функція – це психолого-педагогічне вивчення засвоєних знань, вихованості і розвитку учня, а також рівня професійної компетентності вчителя.

Функція контролю – педагогічна діагностика дозволяє проводити оперативний контроль за процесом навчання та виховання учнів, оскільки містить інформацію про його стан.

Оцінювальна функція – це якісна і кількісна оцінка якості адміністрації школи, кожного вчителя і кожного учня.

Корекційна функція – це дидактична корекція освітньо-виховного процесу.

Інформаційна функція – це постійне інформування всіх учасників педагогічного процесу про результати педагогічної діагностики.

Діагностика педагогічна – це сукупність методів вимірювання й оцінювання кількісних показників успішності навчання, спрямовані на оптимізацію навчального процесу, диференціацію осіб, які навчаються, удосконалення навчальних програм.

Вимірювання – це формалізований процес оцінювання, який завершується кількісною оцінкою. В широкому розумінні, вимірювання – це процес надання чисельного значення певному показнику відповідно до його кількісного прояву із застосуванням чітко визначених правил вимірювання.

Вимірювання педагогічне – це спосіб упорядкування інформації про навчальні досягнення учня (чи студента), за якого систему чисел і відношень між ними ставлять відповідно до ряду вимірюваних факторів.

Оцінювання – формалізований або експертний процес визначення рівня навчальних досягнень учнів/студентів, який завершується виставленням балів.

Як один із методів вимірювання широко використовується тестування.

Тест – це сукупність тестових завдань, які укладені за певними правилами та процедурами для вимірювання певної властивості. Тест передбачає наявність стандартизованої процедури проведення, заздалегідь спроектованої технології обробки та аналізу результатів. Тест має відповідати таким характеристикам ефективності, як валідність і надійність. У педагогіці тестом називають специфічну уніфіковану форму контролю знань і навичок.

Термін «тест» почали використовувати в зарубіжній методиці для позначення будь-якого контрольного завдання як синонім поняття «контрольна робота», «опитування», «залік», «екзамен».

Американські тестологи визначають дві групи тестів: об'єктивні та суб'єктивні. В об'єктивних тестах правильність відповіді визначається за допомогою ключа відповідей; у суб'єктивних – оцінка впливає з оцінювальних суджень експертів. У вітчизняній педагогіці термін «об'єктивний тест» і «суб'єктивний тест» не використовують, а термін «тест» вживають у вузькому значенні – «розташований за певними вимогами комплекс завдань, які пройшли попередню апробацію, дають можливість виявити в тестованих ступінь компетенції, результати якої оцінюються відповідно до попередньо визначених критеріїв» (Б. А. Глухов).

Тестування – це метод вимірювання певних властивостей особи за допомогою тесту, спосіб визначення рівня знань і умінь студентів за допомогою спеціальних валідних тестових завдань.

Тестовий контроль – це метод емпіричного дослідження, який складається із спеціально підготовленого контрольного комплексу завдань, котрий дозволяє за порівняно короткі проміжки часу надійно і адекватно методом шкалювання визначити результативність пізнавальної діяльності студентів, тобто оцінити ступінь і якість досягнення кожним студентом цілей навчання.

Тестологія може бути психологічною, педагогічною, соціологічною, залежно від сфери застосування.

У педагогічних вимірах тестологія – це теоретико-методологічне і методичне обґрунтування процесів розробки і застосування педагогічних тестів. У вищій освіті педагогічне тестування є засобом вимірювання рівня знань, умінь і навичок студентів, а процес такого тестування називається педагогічним виміром навчальних досягнень. Педагогічне тестування стає особливо актуальним як реалізація потреб суспільства в одержанні об'єктивної інформації про якість отриманої і наданої освіти.

За предметом та метою вимірювання педагогічні тести відрізняються від психологічних, соціально-психологічних, соціологічних, культурологічних та інших. Власне психологічні тести (психодіагностика) мають на меті вивчення особистості студента (здібностей, інтелекту, емоційно-вольової сфери, динамічних властивостей тощо). Педагогічні тести (тести досягнень, тести успішності) мають на меті оцінку успішності оволодіння студентами конкретними знаннями і навіть окремими розділами навчальних дисциплін, визначення ефективності програм, підручників і методів навчання, особливостей роботи окремих викладачів, кафедр, педагогічних колективів тощо.

Отже, педагогічні тести як система взаємопов'язаних предметним змістом завдань зростаючої складності та специфічної форми дозволяють оцінювати структуру знань і вимірювати ступінь засвоєння учбового матеріалу, оволодіння необхідними вміннями й навиками, рівень учбових досягнень.

Завдання в тестовій формі – педагогічний засіб, який відповідає наступним вимогам:

- відповідність меті;
- лаконічність;
- логічна форма висловлювання;
- наявність місця для відповіді;
- однакові правила оцінювання відповідей;
- правильне розміщення елементів завдання;
- наявність однакової інструкції для всіх тестованих;

- адекватність інструкцій формі та змісту завдання.

Характеризується певним рівнем складності, після включення до тесту витирається з банку даних. Щоб уникнути можливої фальсифікації результатів, вибір завдань здійснюють автоматично за допомогою спеціальної комп'ютерної програми. Визначають кілька основних форм завдань. Завдання після проведення тестування аналізують не тільки щодо їх складності, а й у зв'язку між виконанням кожного конкретного завдання й загальним балом учня/студента за тест. У завданнях множинного вибору встановлюють можливість кожного варіанта відповіді.

Завдання тестові – складова частина тесту, що відповідає вимогам до завдань у тестовій формі та пройшла обов'язкову перевірку статистичних властивостей. Вимоги до тестових завдань:

- певна складність, що відповідає меті й рівню оцінювання;
- диференційна здатність (достатня варіативність тестових балів);
- позитивна кореляція балів завдань із балами всього тесту;
- відповідає вимогам чистоти форми й предметної чистоти змісту.

Переваги педагогічного тестування перед традиційними методами контролю знань:

– загальна позитивна орієнтація процесу оцінювання (тести спрямовані переважно на виявлення досягнень студентів, а не на виявлення помилок і недоліків);

– багатofункціональність: тести дозволяють враховувати особливості засвоєння кожного навчального предмету і окремих його розділів. Наявність кількісних показників для визначення повноти та глибини знань допомагає студенту адекватно й критично оцінити свої успіхи, підвищити рівень самостійності. Викладач може з'ясувати рівень й динаміку оволодіння студентами учбовим матеріалом відповідно до програми, організувати своєчасну індивідуальну допомогу, вчасно скорегувати власну педагогічну діяльність (зміст курсу);

– вища об'єктивність контролю за рахунок його відповідності заздалегідь визначеному стандарту відповідей, ліквідації підказок і списування, можливості автоматизації перевірки результатів тестування знань студентів, зменшення впливу суб'єктивного компоненту (настрою, рівня кваліфікації та особистісних характеристик викладача). Об'єктивність тестового контролю забезпечує єдині вимоги до всіх студентів, незалежно від їх попередніх навчальних досягнень;

– оперативність та економність контролю (економія часу за рахунок швидкості проведення й легкості обробки результатів; можливості зосередження уваги не на формуванні відповіді, а на осмисленні її суті; вчасне виявлення помилок та їхнє виправлення з мінімальними втратами);

– точність та диференційованість вимірювань учбових досягнень. Змістовна валідність тесту незрівнянно вища за будь-яку іншу форму діагностики. Результати тестування, завдяки особливій організації тестів, можуть бути представлені в шкалах, різних градаціях оцінок, балах, рейтингових позиціях кожного студента, а не лише чотирибальною шкалою («відмінно», «добре», «задовільно», «незадовільно»);

– вища ефективність та технологічність (можливість одночасного оцінювання успішності багатьох студентів; охоплення тестами значного навчального матеріалу; спрощення функцій експериментатора; можливість застосування під час самостійної роботи для контролю якості засвоєння матеріалу; можливість багаторазового повторення умов перевірки для з'ясування змін у рівні підготовки тощо);

– універсальність педагогічного тестування, охоплення ним усіх стадій процесу навчання, наявність однакових вимог для всіх студентів; можливість статистичного аналізу результатів тестування; сумісність тестових технологій з іншими сучасними освітніми технологіями;

– стимулювання позитивного ставлення студентів до систематичного вивчення курсу, кращого засвоєння матеріалу за рахунок глибшого розуміння структури матеріалу, акцентуації уваги на центральних позиціях, опорних поняттях, контрольних питаннях з тем;

– застосування тестової перевірки знань дає можливість формалізувати процес контролю, використовувати комп'ютерну техніку, сучасні педагогічні технології навчання, а також можливість поєднання тестування з різними дидактичними технологіями, методиками навчання та (або) його включення як складової до інших методів контролю;

– порівняно з іншими методами контролю тестова перевірка сприяє широкому охопленню всього складу студентів зі значним заощадженням навчального часу, тобто є реальна можливість збереження ресурсу (мінімізація часу випробування водночас із збільшенням кількості випробуваних, не обов'язкова присутність викладача, як безпосереднє, так і опосередковане пред'явлення тестів тощо);

– забезпечується широка можливість індивідуалізації та диференціації контролю з визначенням певного рівня навченості студентів; система тестових завдань сприяє здійсненню контролю самостійної роботи студентів, створює умови для самооцінювання досягнень;

– стандартизована процедура вимірювання знань та аналізу результатів сприяють об'єктивному та незаангажованому підходу до встановлення якості освіти та уможливають її моніторинг

Недоліки тестового контролю:

– складно, а часто й неможливо, простежити уміння студента будувати відповідь, логічно висловлювати свої думки та міркування мовою науки, обґрунтовувати судження та конкретизувати свою відповідь прикладами;

– тести успішності зазвичай виявляють лише кінцевий результат виконання завдання, тому утрудненою є перевірка глибинного розуміння предмету, оволодіння стилем мислення;

– тестування може призвести до певної стандартизації мислення, провокує студентів до поверхневого вивчення матеріалу, створює враження, що просте запам'ятовування («зубріння») формалізованих знань розв'язує проблеми навчання;

– відсутність безпосереднього контакту викладача зі студентами, складність виявлення їхніх індивідуальних відмінностей та потенційних можливостей;

– існує складність формування змісту тестових завдань гуманітарних дисциплін, коли не можна одержати однозначної відповіді на питання тесту (матеріал не піддається формалізації);

– досить велика ймовірність випадкового вибору правильної відповіді, коли деякі студенти обирають відповідь навмання або методом виключення;

– ситуація тестування може сприйматися студентами як гра-квест, єдиною метою якої є отримання «ключа» та передача його один одному;

– вірогідність впливу на результат випадкових чинників (тривожність, стомлення тощо);

– велика затрата часу на складання якісного тесту і необхідність спеціальної підготовки розробників тестів;

– недостатня теоретична та методична база педагогічних тестів і як наслідок - неточності у вимірюваннях або перекручування результатів;

– відсутність досвіду проходження тестування у випробуваних.

Отже, метод тестування передбачає, що інструментом вимірювання є тест, складений із тестових завдань, процедурою вимірювання є тестування, методом оцінювання є шкалювання. Шкала – форма фіксації сукупності ознак досліджуваного об'єкта з упорядкуванням їх у певну числову систему.

1.2. ТИПИ ВИМІРЮВАНЬ. КЛАСИФІКАЦІЯ ТЕСТІВ

Визначено два типи вимірювань: вимірювання, орієнтовані на норму, й вимірювання, орієнтовані на критерії. Такий поділ обумовлено різною метою вимірювання: для відбору або для встановлення рівня досягнень. Кожна мета вимірювання вимагає застосування різних типів вимірювань, розробки різних тестів.

Вимірювання з метою відбору (нормо-орієнтоване вимірювання).

Основне завдання: розрізнити екзаменованих за рівнем підготовки.

Вимірюються певні кваліфікаційні вимоги або вимоги щодо спроможності набуття певного освітнього чи кваліфікаційного рівня, наприклад, здатності навчатися у вищому навчальному закладі або отримати певну професійну кваліфікацію.

Тестові завдання: середньої складності (комплексний тест, що оцінює змістовний компонент початкових вимог до майбутнього фахівця з вищою освітою певного напрямку підготовки. Цим тестом визначається готовність абітурієнта до навчання у вищому навчальному закладі).

Приклад 1. У США, Канаді, Німеччині відбір до вищих навчальних закладів на медичні спеціальності здійснюється із застосуванням комплексного тесту, до якого входять три предмети: фізика, хімія і біологія. Але зміст цих дисциплін, який базується на шкільній програмі, зосереджується на тих розділах, що в майбутньому використовуватимуться під час вивчення об'єкту діяльності лікаря – людини та її здоров'я. Цей комплексний тест орієнтовано на визначення не лише програмних знань і вмінь абітурієнта, а й на визначення його готовності до навчання у вищому навчальному закладі за медичним фахом.

Вимірювання досягнень (критеріально-орієнтоване вимірювання)

Мета: визначити рівень досягнень особи, яка тестується, незалежно від результатів інших.

Тестові завдання: складність відповідає складності оцінюваного домену (рівень складності тестових завдань, а отже, тесту загалом, має відповідати рівню засвоєння матеріалу: знанням, розумінню, застосуванню, аналізу тощо).

Приклад 2. Найвідоміший американський тест SAT вимірює вербальні та математичні досягнення випускників середньої школи. За цим тестом визначають критерій їхньої готовності до навчання у вищому навчальному закладі.

Застосування тестування як методу вимірювання привело до появи та використання, особливо в розвинутих країнах, великої кількості різноманітних тестів. Це зумовило необхідність їх класифікації. Вона здійснювалася за

різними критеріями та принципами багатьма фахівцями, серед яких: І. Булах, А. Анастасі, В. Аванесов.

Наведемо узагальнену класифікацію, яка найбільш однорідно структурована.

Отже, педагогічні тести можуть бути класифіковані за:

- рівнем уніфікації (тести стандартизовані, не стандартизовані);
- рівнем упровадження (національні, навчального закладу, вчительські);
- статусом використання (обов'язкові, пілотні, дослідницькі);
- співвідношенням з нормами або критеріями (тести досягнень, тести порівняння або тести відбору);
- видом тестового завдання (тести з завданнями закритими і відкритими).

За рівнем уніфікації тести поділяються на стандартизовані та не стандартизовані.

Стандартизація (від англ. standart – типовий, нормальний) стосовно тестування означає приведення до єдиних норм процедури вимірювання та показників якості тесту, а саме:

- регламентація процедури проведення тестування;
- уніфікація інструкцій, бланків тестування;
- уніфікація засобів реєстрації результатів, їх обробки та збереження;
- характеристика контингенту, який тестується;
- термін і тривалість тестування.

Тест стандартизований – інструмент оцінювання, що вимірює стандартний набір широко визначених результатів навчання, використовує стандартні інструкції та стандартну методику оцінювання, а також дає змогу порівнювати результат окремого тестованого з результатами інших тестованих, які виконували той самий тест за аналогічних умов. Ці тести прив'язані переважно до норм, що призначені для порівняння рівня успішності тестованих певного навчального закладу з рівнем успішності репрезентативної вибірки групи тестованих (наприклад, на національному чи регіональному рівнях).

Раніше у світовій практиці такі тести ґрунтувалися на завданнях на вибір відповіді, передусім на завданнях з вибором однієї відповіді. Останнім часом тести модифікують, включаючи завдання на надання відповіді, а також групують у кластери для подальшої інтерпретації. Добре розроблені стандартизовані тести на перевірку успішності, як правило, мають такі характеристики:

- зміст тесту ґрунтується на матеріалах широко використовуваних підручників і методичних посібників.
- тестові завдання ґрунтуються на чіткому наборі специфікацій.
- тестові завдання апробовано, переглянуто, проаналізовано на предмет складності й дискримінаційної здатності, завдання з незадовільними статистичними характеристиками перероблено або вилучено з тесту.
- остаточний набір завдань відібрано на основі специфікації тесту.
- інструкції щодо проведення тесту й виставлення оцінок (балів навчальних досягнень) жорстко описані.
- тест проводять із вибраною групою тестованих, щоб встановити національні та/або регіональні норми для інтерпретації результатів тестів.
- остаточний варіант тесту публікують разом із методичним посібником, який описує статистичні характеристики тесту й методику його проведення й оцінювання, а також інтерпретацію та використання результатів.

Найбільш відомими у світі є стандартизовані тести SAT, GRE, GMAT, LSAT, MCAT, TOEFL тощо.

Стандартизації тесту передуює пілотне тестування. Воно здійснюється на репрезентативній вибірці з метою визначення параметрів тесту та уточнення процедур тестування і є обов'язковим етапом процесу стандартизації тесту. Тривалість цього процесу для відомих тестів коливається від 5 до 10 років, а вдосконалення та оновлення тесту здійснюється постійно.

За рівнем впровадження і статусом використання тести поділяються на загальнодержавні (національні), відомчі, або міністерські, тести навчального

закладу, кафедральні, особисті або неформальні. При цьому залежно від статусу використання тести можуть бути обов'язкові, пілотні й дослідницькі.

Означена класифікація тестів визначається різними вимогами до пілотних тестувань, які супроводжують стандартизацію тестів.

За співвідношенням із нормами або критеріями тести поділяються на орієнтовані на норму та орієнтовані на критерій.

Тести, орієнтовані на норму, використовують тестову стратегію типу «мету досягнуто – мету не досягнуто». Вони дають можливість, порівнюючи результати з деякою нормою, застосовувати їх на заліках та при інших випробуваннях, коли приймаються рішення альтернативного характеру: «зарахований – не зарахований», «прийнятий – не прийнятий».

Тести, орієнтовані на критерій, визначаючи індивідуальні результати, передбачають оцінювання взаємозв'язку цих результатів з попередньо встановленими критеріями.

За видом тестового завдання тести бувають з відкритими та закритими тестовими завданнями.

Тести з відкритими тестовими завданнями. У таких тестах відповіді до тестових завдань не надаються ані особі, яка тестується, ані особі, яка їх перевіряє, тобто це тестові завдання з вільною формою відповіді. Їх поділяють на завдання з пропусками, завдання на доповнення, завдання з короткою відповіддю та завдання у формі есе.

Завдання з короткою відповіддю можуть бути сформульовані у різній формі, але так, що запитання потребують короткої відповіді, яка є, як правило, результатом математичних розрахунків.

У тестах з напіввідкритими тестовими завданнями відповідь надається лише особі, яка їх перевіряє.

Тести з закритими тестовими завданнями передбачають, що варіанти відповідей до тестового завдання надаються також особі, яка тестується. Такі тести за формою відповіді можна розділити на альтернативні тестові завдання,

вибіркові тестові завдання, які передбачають вибір із кількох запропонованих відповідей, і тестові завдання на визначення відповідності.

Необхідно зауважити, що незалежно від типу розробленого тесту, класифікація може бути здійснена за методом тестування. За методом тестування розрізняють бланкове, комп'ютерне та комп'ютерне адаптоване.

Бланкове тестування передбачає, що экзаменовані відповідають на тестові завдання шляхом внесення олівцем або ручкою відповідей у спеціальні бланки відповідей.

Під час зовнішніх (масштабних) тестувань зчитування даних з бланків відповідей здійснюється переважно за допомогою комп'ютера.

Комп'ютерне тестування відрізняється від бланкового тим, що тест надається не у паперовому буклеті, а в базі даних комп'ютера. Тестові завдання відображаються на дисплеї, а відповіді вводяться экзаменованим безпосередньо з клавіатури комп'ютера для подальшого їх опрацювання на комп'ютері. Отже, для комп'ютерного тестування характерна автоматизація всього процесу вимірювання.

Метод комп'ютерного тестування має як певні переваги, так і певні недоліки порівняно з бланковим. До головної переваги можна віднести отримання результату тестування відразу після його закінчення. Основним недоліком є необхідність мати мінімальні навички роботи з комп'ютером, а також велика база тестових завдань як для забезпечення закритості экзаменаційного тесту, так і для забезпечення валідності змісту.

Під час сертифікаційних і ліцензійних тестувань дорослих тестові організації усе частіше використовують комп'ютерне тестування. Водночас вступні тестування абітурієнтів відбуваються у формі бланкових.

Комп'ютерне адаптивне тестування. Поява адаптивного тестування була викликана прагненням до підвищення ефективності педагогічних вимірювань, яка, як правило, пов'язувалася зі зменшенням числа завдань, часу, вартості тестування, а також із підвищенням точності оцінок учнів. В основі адаптивного підходу лежить індивідуалізація процедури відбору завдань тесту,

яка за рахунок оптимізації складності завдань відносно рівня підготовленості учнів забезпечує генерацію ефективних тестів.

Оптимізація складності завдань зазвичай проводиться покроково. Якщо учень виконує завдання вірно, то далі йому дається більш важке завдання. При неправильному виконанні завдання відбувається відхід назад до більш легкого завдання. Таким чином, у комп'ютерному адаптивному тестуванні число тестових завдань та їх рівень складності обираються індивідуально для кожного іспиту на підставі відповідей тестованого, а індивідуальна сукупність завдань утворює адаптивний тест. Адаптивні тести у групі випробовуваних складаються в основному з різних завдань і розрізняються за кількістю і складністю завдань тим сильніше, чим більша різниця серед випробуваних із рівня підготовленості.

Отримати одночасний приріст ефективності вимірювань за всіма критеріями неможливо, тому, зазвичай, при організації адаптивного тестування на перший план виходить один, у кращому разі, два критерії. Наприклад, в одних випадках при експрес-діагностиці в адаптивному режимі найбільша увага приділяється мінімізації часу випробування і кількості запропонованих завдань, а питання точності оцінок відходять на другий план. В інших випадках пріоритетною може бути точність вимірювання і тестування кожного випробуваного продовжується до тих пір, поки не досягається запланована мінімальна помилка вимірювання.

На довжині адаптивного тесту істотно позначається якість структури знань учнів. Зазвичай випробовувані з чіткою структурою знань виконують завдання з збільшенням рівня складності, уточнюючи з кожним черговим вірно виконаним завданням оцінку підготовленості. Вони виконують невелике число завдань адаптивного тесту і швидко доходять до порогу своєї компетентності. Учні з нечіткою структурою знань, у яких чергуються вірні і невірні відповіді, отримують коливання за рівнем складності завдань. Процес тестування затягується, оскільки не відбувається покрокового наростання точності

вимірювання і число завдань, адаптованих за рівнем складності, нерідко виявляється навіть більшим, ніж у звичайному, традиційному тесті.

Переваги адаптивного тестування. До числа важливих переваг комп'ютеризованого адаптивного тестування можна віднести:

- високу ефективність;
- високий рівень секретності;
- індивідуалізацію темпу виконання тесту;
- високий рівень мотивації до тестування у найбільш слабких учнів за рахунок виключення з процесу пред'явлення зайво важких завдань;
- повідомлення результату в інтервальної шкалою тестових балів кожному випробуваному негайно, відразу після закінчення його роботи над індивідуально підібраним набором завдань в адаптивному тесті.

2. ІСТОРІЯ РОЗВИТКУ ТЕСТУВАННЯ

У 1884 – 1885 роках англійський вчений Ф. Гальтон на Всесвітній виставці організував антропометричну лабораторію, де за три пенси відвідувачі могли перевірити свої фізичні якості (силу, швидкість реакції тощо), фізіологічні можливості організму і психічні властивості за 17 показниками (зріст, вага, ємність легень, станова сила, сила кісті й удару кулаком, запам'ятовування літер, гострота зору, розрізнення кольору тощо). За цією програмою було обстежено 9337 осіб. Послідовність завдань, які треба було виконувати для виявлення тих чи інших якостей особистості, одержала назву «тест». Ф. Гальтон також був основоположником використання шкал оцінювання, методів анкетування і методики вільних асоціацій, що використовуються донині. Вчений також, вибравши і спростивши статистичні формули, розроблені математиками, адаптував їх для використання під час оброблення результатів тестування. Тому саме його вважають автором методів математичної статистики у проведенні аналізу даних про індивідуальні особливості людини. В цьому напрямі пізніше працювало багато вчених, серед яких і учень Ф. Гальтона – К. Пірсон.

Американському психологу Дж. Кеттеллу також відводять особливе місце у розвитку сучасної тестології. У своїй лабораторії він розробив завдання, які назвав «розумовими тестами», їх призначенням було вивчення інтелектуального рівня студентів коледжів. Цей учений деякий час стажувався та працював в лабораторії Ф. Гальтона. Психолог вважав, що тест є засобом для проведення наукового експерименту, до якого повинні висуватись певні вимоги. Всі ідеї, запропоновані Дж. Кеттеллом у тестуванні, у наш час становлять основу сучасної тестології, а саме:

- рівність умов для всіх випробуваних;
- обмеженість у часі;
- відсутність сторонніх у лабораторії;
- надійність обладнання;

- забезпечення однакового інструментарію і чіткості умов завдань;
- оброблення результатів тестування статистичними методами.

Кеттелл запропонував як зразок 50 тестів, що включали різного роду виміру моторики, чутливості, часу реакції, часу, що витрачається на називання кольорів, кількості звуків, відтворених після однократного прослуховування, і ін. Він використовував ці тести у влаштованій ним при Колумбійському університеті лабораторії (1891 р.). Але досвід їх застосування в практичних цілях для прогнозу навченості студентів виявився невдалим: зв'язку між результатами тестування і успішністю навчання встановити не вдалося.

Варто зазначити, що Ф. Гальтон критикував Дж. Кеттелла за те, що той не вважав за необхідне порівнювати результати тестів із незалежними вимірюваннями тих самих змінних (говорячи сучасною психодіагностичною термінологією, Ф. Гальтон порушував проблему валідності тесту).

Учнем і колегою Ф. Гальтона був англійський учений-психолог К. Пірсон. Вивчивши дослідження Ф. Гальтона та Дж. Кеттелла, він розпочав власні дослідження в теорії тестології. Результатом його роботи було вдосконалення методів кореляційного, регресійного і факторного аналізу тестів. Створена ним теорія кореляції та введений коефіцієнт кореляції (1896 р.) почали активно використовувати для визначення надійності та валідності тестів. Важливим вкладом К. Пірсона у розвиток теорії тестів було визначення трьох основних принципів проведення дослідження за допомогою тестування, зокрема:

- використання серії однакових тестів для великої кількості людей;
- статистичне оброблення результатів;
- виділення еталону оцінки.

Новий крок у розвитку методу тестів був зроблений французьким лікарем і психологом А. Біне (1857 – 1911 рр.), творцем найпопулярнішою на початку ХХ ст. серії інтелектуальних тестів. До Біне тестувалися, як правило, відмінності в сенсомоторних якостях – чутливості, швидкості реакції тощо. Але практика вимагала інформації про вищі психічні функції, що позначаються

звичайно термінами «розум», «інтелект». Саме ці функції забезпечують придбання знань і успішне виконання складної пристосувальної діяльності.

Причиною, з якої А. Біне разом з Т. Симоном приступив до розробки першого в історії психодіагностики інтелектуального тесту, став практичний запит – створити методику, за допомогою якої можна було відокремити дітей, здатних до навчання, але ледачих і небажаючих вчитися, від дітей з особливими потребами, які не були здатні вчитися у звичайній школі. За ініціативи Міністерства народної освіти Франції на початку ХХ ст. стали створюватися спеціальні школи для розумово неповноцінних дітей, у зв'язку з чим треба було диференціювати дітей по їх можливостях вчитися ще до початку навчання, на порозі школи.

Біне і Симон провели серію досліджень пам'яті, уваги, мислення у дітей різного віку, починаючи з трьох років. Проведені на великій групі дітей експериментальні завдання були перевірені за статистичними критеріями і стали розглядатися як засіб виявлення рівня інтелектуального розвитку.

Перша серія тестів – Шкала інтелекту Біне-Симона (Binet-Simon Intelligence Development Echelle) з'явилася в 1905 році. Вона була призначена для дітей від трьох до 11 років і згодом кілька разів переглядалася авторами, які прагнули вилучити з неї всі завдання, що вимагають спеціального навчання. Розроблена шкала містила 30 завдань, розташованих по мірі зростання важкості. Наприклад, для дитини трьох років було потрібно:

- показати свої очі, ніс, рот;
- повторити речення довжиною до шести слів;
- повторити по пам'яті два числа;
- назвати намальовані предмети;
- назвати своє прізвище.

Якщо дитина вирішувала всі завдання, їй пропонувалися завдання більш старшої вікової ступені. Завдання вважалися відповідними до певної вікової ступені, якщо їх правильно виконувало більшість (80 – 90%) дітей даного віку.

Шкала Біне-Симона у своїх наступних редакціях (1908 р. і 1911 р.) була переведена на німецьку та англійську мови. Друга редакція (1908 р.) відрізнялася тим, що в ній був розширений віковий діапазон дітей – до 13 років, збільшено число завдань і введено поняття розумового віку. Ця шкала отримала саме широке поширення. Остання, третя, редакція шкали вийшла в рік смерті А. Біне і не мала суттєвих змін.

Завдання в шкалах Біне-Симона були згруповані по віках (від 3 до 13 років). Для кожного віку підбиралися певні тести шляхом дослідження великої групи дітей (300 осіб). Тести вважалися відповідними даної вікової щаблі, якщо їх вирішувало більшість дітей даного віку (80 – 90%). Показником інтелекту в шкалах Біне був розумовий вік, який міг розходитися з хронологічним. Розумовий вік визначався за рівнем тих завдань, які міг вирішити дитина. Якщо, наприклад, дитина, чий хронологічний вік дорівнює 3 років, вирішує всі завдання для 4-річних дітей, то розумовий вік цього 3-річну дитину зізнавався рівним 4 рокам. Розбіжність розумового і хронологічного віку вважалось показником або розумової відсталості (якщо розумовий вік нижче хронологічного), або обдарованості (якщо розумовий вік вище хронологічного).

Варто зазначити, що авторами проводився аналіз власних розроблених тестів. Для прикладу, якщо більшість дітей успішно виконували тест, він вважався надто легким; якщо ж мало – важкий. Іншими словами психологами вперше було порушено питання складності тестів.

Друга редакція шкали Біне послужила основою роботи з адаптації, перевірки та стандартизації, проведеної в Стенфордському університеті (США) групою психологів під керівництвом Л. М. Термена (1877 – 1956 рр.). Перший варіант адаптації тестової шкали Біне був запропонований в 1916 році та мав так багато серйозних змін у порівнянні з основним, що був названий Шкалою інтелекту Стенфорд-Біне (Stanford-Binet Intelligence Scale). Основних нововведень у порівнянні з тестами Біне було два: введення в якості показника по тесту коефіцієнта інтелектуальності (Intelligence Quotient – IQ), що виводиться з відношення між розумовою і хронологічним віками, і

застосування критерію оцінки тестування, для чого вводилося поняття статистичної норми.

Шкала Стенфорд-Біне отримала популярність у всьому світі. Вона мала кілька редакцій (1937 р., 1960 р., 1972 р., 1986 р., 1998 р.). В останній редакції вона застосовується і в даний час. Показник IQ, отримуваний за шкалою Стенфорд-Біне, на довгі роки став синонімом інтелекту. Новостворювані інтелектуальні тести стали перевірятися шляхом зіставлення з результатами Шкали Стенфорд-Біне.

Згодом було розроблено систему для колективного тестування. Цікаво, що її (так само як і шкалу Біне) також було розроблено на замовлення. Так, у 1917 році, за часів, коли армію США було задіяно у баталіях Першої світової війни, перед американськими психологами постало складне завдання: якнайшвидше відібрати і розподілити півтора мільйони рекрутів на різноманітні армійські служби. Чиновники зі спеціально створеного комітету доручили досліднику А. Отісу розробити для цього нові типи тестів. І такі тести було створено: «Тести Альфа» – тести для осіб, що володіли англійською мовою, і «Тести Бета» – тести для неписьменних та іноземців.

Застосування нових типів тестів – колективних, значно спрощувало проведення процедури відбору: адже відтепер тестування мали змогу проводити не лише кваліфіковані психологи, а й інші люди, що пройшли навчання з процедури проведення такого тестування.

Пізніше армійські тести значною мірою було вдосконалено, вони стали зразком для створення нових групових тестів інтелекту.

Подальшого розвитку теорія тестування набула у роботах Д. Векслера, який у 1939 році розробив шкалу розумового розвитку. Він вперше використав вербальні тести і тести на виконання (дію). Всього у тест Д. Векслера входило одинадцять субтестів. Ці тести є одними з найбільш популярних, що використовуються у світі.

На початку ХХ століття виокремився і педагогічний напрям у розвитку тестології. У 1864 році в Англії Дж. Фішер створив спеціальні книги («Scale

books»), в яких було подано запитання та варіанти відповідей для вибору для перевірки рівня знань учнів. Книги містили запитання з правопису, арифметики, читання, граматики, навігації. Ці праці можна вважати першими зразками шкільних тестів успішності, хоча теоретичні основи тестування було розроблено значно пізніше.

У 1884 році в США було опубліковано першу збірку тестів, що містила завдання з математики, історії, граматики, навігації та відповіді до них. Результати оцінювались за п'ятибальною шкалою. В цій книзі також подавалися зразки текстів творів і вказувались методи їх кількісної оцінки. Це був перший в історії випадок використання найпростіших статистичних розрахунків у педагогічній роботі. Така форма контролю швидко розповсюдилась у США; хоча вона мала не лише прихильників, а й противників, які дотримувались думки, що ніхто краще за вчителя не може оцінити здібності учня.

Початком систематичних вимірювань у педагогіці варто вважати 90-ті роки XIX століття: з 1894 року було розпочато систематичні групові тестові вимірювання у школі. Їх започаткував американський лікар і психолог Дж. М. Райс (1857 – 1934 рр.), який створив три різні тести для перевірки правопису. В 1902 році він склав серію арифметичних тестів; через рік було підготовлено мовний тест. Дж. Райс оцінював результати за створеною ним п'ятибальною шкалою. Та, незважаючи на новизну досліджень Дж. Райса, його роботи не були оцінені сучасниками належно.

Американець В. А. Макколл розділив тести на педагогічні (Educational Test) та психологічні (Intelligence Test). Основним завданням педагогічних тестів було вимірювання успішності учнів з різних шкільних предметів за певний період навчання, перевірка ефективності застосування окремих методів викладання та організації навчального процесу. В. Макколл обґрунтував мету використання педагогічних тестів так: для об'єднання учнів, що засвоюють однаковий об'єм матеріалу за визначений проміжок часу, у групі. Психологічні тести були спрямовані на визначення рівня інтелектуального розвитку.

Дж. Орлеанс та Г. Сілі у роботі «Об'єктивні тести» (1905 р.) наголосили на об'єктивному характері одержаних результатів і визначили такі головні цілі тестування:

– забезпечення викладачів інформацією як про рівень засвоєння вивченого матеріалу, так і про обсяг матеріалу, призначеного для подальшого вивчення;

– надання допомоги викладачам у виборі методів навчання.

Саме Дж. Орлеанс та Г. Сілі запровадили класифікацію тестів на індивідуальні та групові.

Засновником використання тестів у педагогіці США все ж вважають Е. Торндайка. Саме ним було запропоновано методичні аспекти розроблення тестів у педагогіці. В 1903 році В. Торндайк видав книгу «Психологія освіти» (Educational Psychology), в якій було описано ті види тестів, за допомогою яких можливо ефективно визначити успіхи у навчанні. У цих тестах використовувалися принципи вимірювання, розроблені у психологічних лабораторіях.

Одним з учнів В. Торндайка був С. Стоун, який розробив арифметичні тести під назвою «Стандартні тести Стоуна». Над удосконаленням арифметичних тестів працював ще один учень В. Торндайка – С. Куртіс. Результатом його роботи стала «Серія арифметичних тестів А» (1909 р.), а пізніше – удосконалена «Серія арифметичних тестів В» (1914 р.).

У 1910 – 1911 роках було проведено перевірку якості навчання у безкоштовних школах Нью-Йорка. Для цього було використано тести для вимірювання успішності учнів як засіб перевірки ефективності шкільної системи. Варто зазначити, що серед використовуваних тестів були і тести С. Куртіса.

Серед послідовників В. Торндайка варто згадати Б. Букінгема, який у 1913 році розробив шкалу правопису, яка була новим типом вимірювального інструмента. Б. Букінгем за теоретичну основу свого тесту прийняв твердження про те, що здатність до правопису може бути виміряна за тим рівнем

складності, яку може досягнути учень за розробленою шкалою. Принцип побудови шкали Б. Букінгема запозичили багато спеціалістів з педагогічного тестування для різних предметів, а саме: Л. П. Айрес використав його для своєї «Шкали правопису», С. Вуді – для тестів на перевірку математичних операцій, Х. Хотсом – для тестового вимірювання знань з алгебри, В. Хенмон та Х. Браун – для встановлення рівня засвоєння учнем латині, М. Трабу – для визначення мовознавчих знань і вмінь школярів, М. Дж. Ван-Вагенен – для перевірки виконання тестів з історії.

Після того, як Спілка американських шкільних рад у 1914 році затвердила тестування як об'єктивний метод вимірювання та оцінювання знань, у США розпочали активно застосовувати тести успішності.

У 1915 році Йеркс створив власну серію тестів, головною відмінністю яких була зміна системи підрахунку. Замість вікових часток, запропонованих А. Біне, досліджуваний отримував за кожну правильну відповідь певну кількість балів. Це зробило більш зручним підрахунок результатів тесту. Кількість балів переводилась за запропонованими стандартами, визначався коефіцієнт успішності.

Динамічний розвиток досліджень застосування тестів у педагогіці сприяв використанню статистичних методів на практиці. У «Дванадцятому щорічнику Національного суспільства з досліджень освіти» вперше було подано офіційні рекомендації з використання тестів у школі. В університетах було розроблено спеціальні курси підготовки майбутніх адміністраторів і технічних спеціалістів у галузі освіти з використання тестів. Основним підручником того часу з тестування було удосконалене В. Торндайком «Введення в теорію розумових та соціальних вимірювань».

У США допомогу вчителям у використанні тестів надавали створені в той час «Бюро освітніх досліджень» під керівництвом відомих фахівців з вимірювань. Ці установи забезпечували тестами школи, проводили стандартизацію уже існуючих тестів і складали нові, а також знайомили

вчителів та адміністрацію шкіл з тестами і методикою їх проведення. Використання тестів розпочалося, окрім США, у багатьох країнах світу.

Дещо пізніше почали з'являтися батареї тестів досягнень, початком чого стало видання в 1923 році першої редакції Стенфордського тесту досягнень (Stanford Achievement Test). Його авторами були кращі тогочасні фахівці в галузі технології тестів Т. Келлі, Дж. Рач та Л. Тьормен.

Важливе значення у розвитку тестології в Америці має робота Ф. Фрімена «Розумові тести. Їх історія, принципи та застосування» (1926 р.). У ній автор представив широкий огляд розвитку тестування у США, визначив головні його напрями: створення педагогічних тестів, основна мета яких – перевірка рівня знань на різних етапах навчання; розроблення розумових тестів, метою яких є визначення рівня загального розумового розвитку і спеціальних здібностей. У тому ж році було опубліковано монографію «Тести нового типу», в якій автор Ч. Грин спробував проаналізувати недоліки та переваги зазначених вище тестів, створив їх власну класифікацію і надав рекомендації щодо створення тестів, а саме:

- чітко визначити обсяг матеріалу, за яким проводиться тестування;
- визначити найкращу форму тесту для перевірки засвоєння певного матеріалу дослідницьким методом;
- чітко визначити час проведення тесту, експериментально визначивши середню швидкість відповідей учнів на кожне запитання;
- перевірити правильність та логічність включених до тестів тверджень;
- розташовувати запитання у порядку зростання складності, уникаючи закономірності у чергуванні правильних та неправильних відповідей.

Четверте десятиліття ХХ століття відзначилось застосуванням технічних засобів для підрахунку тестових показників.

В Європі розповсюдження тестів проходило не так активно. Особливому проходив розвиток методу тестування у вітчизняній педагогіці. Практичного значення тести набули після створення у 1925 році тестової

комісії при педагогічному відділі Інституту методів шкільної роботи, в завдання якої входило розроблення стандартизованих тестів для радянської школи, їх масове використання в закладах освіти. За допомогою таких тестів, що створювались на основі американських, можна було перевірити знання учнів з правопису, лічби, розв'язування задач, географії, природознавства і суспільствознавства. До тестів додавались інструкції і особові картки для врахування прогресу учня.

Прихильники наукового напрямку розвитку психології (Ф. Рибаків, Г. Россолімо та інші) активно сприяли розповсюдженню ідеї тестування, наслідуючи своїх західних колег. Пізніше тести для педагогічних цілей розробляли відомі радянські психологи і педагоги, серед яких – М. Басов, М. Берштейн, П. Блонський, С. Василейський та інші. В Україні свої новаторські дослідження з цього питання запропонували А. Мандирка (1931 р.) та М. Сиркін (1929 р.), приділяючи увагу теоретичним аспектам прикладної статистики у психологічному тестуванні, критеріям правильності тестових випробувань.

Зазначений час характеризується інтенсивним і неконтрольованим використанням тестів у системі народної освіти. Масові тестові дослідження не завжди підкріплювались серйозною перевіркою якості інструментарію. З огляду на поширення «тестоманії» та низки інших суб'єктивних причин було прийнято постанову «Про педологічні перекручення в системі Наркомпосу» від 4 липня 1936 року, в якій було заборонено використання тестів та анкет.

Загалом у постанові мова йшла не про дидактичні тести, а про тестування розумових здібностей. Але в атмосфері 30-х років ХХ століття термін «тест» став символом «педологічних перекручень». Ця постанова діяла до середини 50-х років ХХ століття, але й пізніше будь-які тести вважались «буржуазним засобом». Внаслідок цього навіть майбутні педагоги та психологи не вивчали тестології, не було професійної літератури, присвяченої проблемам конструювання та оцінювання якості тестів тощо.

Однак у розглядуваний час звучали виступи на користь тестів. Відомий психолог М. Басов зазначав, що «довга і гостра критика тестової методики... врешті-решт приведе не до повалення, не до скасування цієї методики, а навпаки, до її зміцнення і утвердження в тих межах, в яких вона, очевидно, має право на застосування та існування».

Хоча у 30-ті роки ХХ століття робота з тестами загальмувалась, наукове вивчення справжніх можливостей цього методу не припинялась; елементи методу тестів успішності (без згадування відповідного терміна) використовувались під час обліку успішності учнів у загальноосвітніх школах. Зразки тестових завдань можна знайти у збірниках вправ і задач, контрольних завдань з різних навчальних предметів.

Відкрито питання про можливість і доцільність використання тестів у радянській школі було знову поставлене лише в 60-х роках ХХ століття у зв'язку із виникненням і спробами реалізувати індивідуальний підхід у навчанні. В березні 1969 року на Центральній раді Спілки психологів СРСР психодіагностика була визнана однією з найменш розвинених галузей психології; було прийнято рішення про офіційний «дозвіл» на використання тестів.

У 70 – 80-х роках ХХ століття широко розповсюджувались збірники завдань і вправ, призначених для використання їх машинами типу «репетитор», і контрольні завдання-тести, призначені для використання з машинами типу «екзаменатор». Також було розроблено тестові завдання, що перевірялись за допомогою перфокарт.

У 70-х роках ХХ століття з'явилася низка розробок, в яких висвітлювався досвід і проблеми масових процедур контролю знань учнів і студентів за допомогою спеціально розроблених контрольних робіт, що мали ознаки класичного тестування. Їх авторами були С. Архангельський, В. Аванесов, Л. Болотник, Ю. Бєлий, Г. Батуріна, В. Беспалько, М. Бернштейн, Б. Володін, Л. Ітельсон, Є. Кульянов, Р. Касимов, В. Копилов, В. Короза, В. Мельникова, В. Мезинцев, В. Огорелков, Н. Тализіна. У цей період з'явилася значна

кількість робіт, присвячених математичним проблемам педагогічних вимірів. У них розглядалися питання розроблення шкал вимірювання результатів контрольних робіт, математичних моделей оброблення інформації у педагогічних дослідженнях, зокрема і з використанням ЕОМ, а також питання статистики.

Однак використання тестів багатьма не підтримувалося. Полеміка про правильність упровадження тестів періодично велася до середини 70-х років і майже повністю була припинена до початку 80-х років ХХ століття. У той час психодіагностика стала окремою галуззю досліджень і набула значення багатогалузевої науки – загальної психодіагностики. Спеціальними психодіагностиками зазначеного періоду можна вважати клінічну, професійну, спортивну та педагогічну. У цей час у психодіагностиці використовувалась обчислювальна техніка.

Наприкінці ХХ століття, після розпаду Радянського Союзу, розвиток психологічної діагностики, як й інших наук, певною мірою призупинився. Однак в основних психологічних центрах певна робота в галузі психодіагностики була продовжена.

1993 рік ознаменувався спробою створити тестування у школах, але вона виявилася невдалою. У 2002 році було проведено 200 пілотних тестувань по Україні, у яких взяли участь студенти перших курсів. У 2003 році тестування з історії та математики пройшов 3121 випускник 670 загальноосвітніх шкіл України. У 2004 році кількість охочих зросла трохи більше, ніж на тисячу, а в 31 вищому навчальному закладі були прийняті рішення про зарахування результатів ЗНО як вступних випробовувань.

У 2005 році Віктор Ющенко надав указ Міністерству освіти і науки України здійснити перехід до проведення вступних екзаменів для вступу у заклади вищої освіти шляхом складання ЗНО протягом 2005 – 2006 років. 31 грудня 2005 року Кабінет Міністрів створив Український центр оцінювання якості освіти й офіційно дозволив вступати до вишів за результатами тестування. У 2006 році ЗНО ввели на офіційному рівні. У Державному

бюджеті України вперше були передбачені фінанси для його запровадження і моніторингу освіти. Починає свою діяльність Український центр якості освіти, створюється 9 його регіональних центрів. Проводиться тестування 41818 випускників, де були задіяні 6300 інструкторів, 700 екзаменаторів.

У 2007 році всі освітні заклади, підпорядковані Міністерству освіти, зараховували абітурієнтів за сертифікатами ЗНО. Тестування проводилося лише з української мови, математики, історії (української та всесвітньої).

Предмет	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Українська мова	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Українська література	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓
Історія України	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Всесвітня історія	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓	✓
Математика	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Фізика	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓
Хімія	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Зарубіжна література (Світова література)	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	✓
Біологія	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓
Основи економіки	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗
Основи правознавства	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗
Географія	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓
Іноземна мова (англійська, німецька, французька або іспанська)	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓
Російська мова	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓
Примітки ✗ — не проводилося; ✓ — проводилося вибірково; ✓ — для всіх абітурієнтів											

3. ЗАГАЛЬНІ ВИМОГИ ДО ЯКОСТІ ТЕСТІВ

3.1. КРИТЕРІЇ ЯКОСТІ МЕТОДУ ВИМІРЮВАННЯ

1) **Валідність тесту** – характеризує адекватність і дієвість методів відповідно до конкретної ситуації застосування. Валідність фіксує, що оцінює тест і як добре він це робить (А. Анастасі).

Поняття валідності використовується для загальної характеристики тесту в аспекті відповідності одержаних результатів меті та вимогам оцінювання.

Тест не валідний. Це означає, що одержані результати не вимірюють того, що було визначено метою тестування. Наприклад, невідповідності інструменту вимірювання (тесту) меті оцінювання.

Приклад. Тест побудовано з тестових завдань на розуміння матеріалу, а навчання було спрямовано лише на знання базової інформації. Метою оцінювання є рівень виконання певних навичок, а тест базується на вимірюванні знань про виконання цих навичок.

Фактори, що знижують валідність результатів оцінювання:

- 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 містяться 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 тощо);
- 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;
- не враховано час, необхідний для виконання тесту;

– порушення процедур проведення тестування. Незадовільні умови контролю тощо.

Валідність змісту і відповідності може бути кількісно визначено через **коефіцієнт валідності** – розраховується як коефіцієнт кореляції між результатами тестування та результатами інших вимірювань, здійснених на тій самій групі з того самого предмета іншим методом (змінюється від -1 до 1). Загально прийнято, що результат тестування можна вважати валідним при коефіцієнті кореляції більшому ніж 0,6.

Якщо показник валідності від'ємний, тест вважають непридатним для використання, якщо низький (наближений до нуля), – доопрацьовують.

Коефіцієнт валідності дорівнює коефіцієнту кореляції між результатами, одержаними різними методами за однакових умов, і показує, наскільки збігаються результати вимірів.

2) Об'єктивність показує, наскільки мінімізовано вплив суб'єктивних факторів. Об'єктивність досягається шляхом стандартизації умов проведення вимірювання, умов оцінювання, умов аналізу результатів і забезпечується за допомогою об'єктивності проведення вимірювання, об'єктивності обробки даних та об'єктивності інтерпретації результатів.

Забезпеченню критеріїв валідності й об'єктивності сприяють певні регламентуючі документи, якими визначаються вимоги до побудови тестів, до процедур проведення тестування та вимоги до фахівців, які застосовують цей метод.

3) Надійність методу вимірювання визначається ступенем стійкості результатів. Перевірка надійності методу стосується насамперед співставлення результатів при повторних вимірах. Надійність тестових балів визначається за допомогою коефіцієнта надійності.

Коефіцієнт надійності дорівнює коефіцієнту кореляції між результатами, отриманими однаковим методом за однакових умов, і показує, наскільки збігаються результати вимірів. Коефіцієнт надійності змінюється від 0 до 1 і розраховується з використанням певних методик (найчастіше для його

розрахунку використовують рівняння Спірмена – Брауна, або коефіцієнта Кронбаха.

При створенні тестів, які використовуються під час широкомасштабних тестувань, розрахунок коефіцієнта надійності є обов'язковим. Під час тестувань, за результатами яких приймаються адміністративні висновки для екзаменованих, цей коефіцієнт має бути більшим ніж 0,9.

Характеристика тесту за коефіцієнтом надійності:

Інтервали коефіцієнта надійності	Характеристика тесту
0,90 – 0,99	Відмінний
0,85 – 0,89	Дуже хороший
0,80 – 0,84	Хороший
0,75 – 0,79	Задовільний
0,70 – 0,74	Мало задовільний

Чинники, що знижують надійність тестових балів:

- відсутність інструкцій або наявність нечітких і неповних інструкцій;
- мала кількість тестових завдань;
- неправильно розрахований час для проведення тестування;
- у тесті не враховано складність тестових завдань;
- неправильно визначені критерії оцінювання;
- допущені помилки в ключах відповідей до завдань на вибір правильної відповіді;
- порушено умови проведення тестування, існували чинники, які могли вплинути на результати;

Розроблено кілька методів визначення та підвищення надійності тесту:

Метод	Тип інформації
Метод повторного тестування	Стабільність балів тесту за певний період
Метод еквівалентних форм	Постійність балів тесту в різних формах тесту (тобто в різних завданнях)
Метод повторного тестування з еквівалентними формами	Постійність балів тесту за певний період і при різних формах тесту
Методи внутрішньої послідовності тесту	Постійність балів тесту в різних частинах

Кількість тестових завдань об'єднаних в одному тесті визначається поняттям **довжина тесту** або обсяг тесту. Для тематичної перевірки знань тестованих у процесі експрес-контролю можна використовувати тест довжиною в 15 – 20 тестових завдань; підсумкова перевірка знань та вмінь з того чи іншого навчального курсу вимагає включення до тесту 30 – 60 тестових завдань. У цілому банк тестових завдань з дисципліни повинен містити не менше 150 тестових завдань, адже деякі із завдань в процесі апробації можуть вилучатися. Фахівці підкреслюють, що надійність і об'єктивність тестової перевірки знань та вмінь зростає зі збільшенням довжини (обсягу) тесту.

Індекс складності (достатня складність завдання). Складність завдання дозволяє оцінити належний рівень знань і навичок при допустимому рівні (частці) помилок. Показником складності завдання є частка вибірки тестованих, які розв'язали чи не розв'язали дане завдання. У класичній теорії вимірів індекс складності завдання визначається у відсотках діленням кількості завдань з помилкою на загальну кількість запропонованих завдань, і вона варіює від 1 до 100%.

Інтервали індексу складності	Характеристика тесту
Понад 80	Дуже легкі
60 – 79	Легкі
40 – 59	Оптимальні
20 – 39	Складні
19 і нижче	Дуже складні

Наприклад, якщо лише 20 % випробуваних виконали завдання, його можна вважати складним для даної вибірки, якщо 80 % – легким. Підбір завдань за показниками складності важливий для успішного використання тестів. Так, при підборі надто складних тестових завдань валідність і надійність тесту різко зменшуються. Надто прості тестові завдання призведуть до одноманітності тесту та його неефективності. Надто легкі завдання, як і важкі, не дають змоги виявити творчі можливості студентів, рівень їхньої підготовки. Проте не варто вилучати з апробованих тестів усі завдання, які легко виконуються, оскільки тести успішності мають не лише виявляти здібних та обдарованих студентів, а й фіксувати результати навчання. Складні завдання

теж не має сенсу вилучати, оскільки в цьому випадку втрачається показник якісного рівня підготовки тієї чи іншої студентської групи (розумовий рівень студентів різних факультетів може бути різним).

Диференційна здатність (дискримінативність) – це розрізнявальна (розподільна) здатність тестових завдань, яка дозволяє достовірно розрізняти осіб тестування з різним рівнем знань. Якщо тест ефективний, то він відображує дійсний рівень знань. Якщо всі студенти показують однаковий результат (на завдання однаково добре відповідають як сильні, так і слабкі тестовані), то в тестовому завданні немає сенсу, оскільки він не дає інформації про відносний рівень результатів. Необхідно ретельно конструювати завдання, щоб забезпечити відповідний рівень дискримінативної здатності, тобто формулювати як саме запитання, так і можливі відповіді на нього бажано таким чином, щоб той студент, який дійсно має необхідні знання, міг відповісти правильно, а той студент, котрий не володіє необхідними знаннями, обрав би будь-яку з неправильних відповідей.

Індекс дискримінативності – це коефіцієнт кореляції між балом за весь тест і балом за дане тестове завдання. Він показує, наскільки добре дане тестове завдання розрізняє екзаменованих з високим і з низьким балами (коливається від -1 до +1).

Ефективність – критерій, який дозволяє порівнювати тести. Ефективним можна назвати тест, який краще, ніж інші тести, вимірює знання студентів потрібного рівня підготовки, з меншою кількістю завдань, якісніше, швидше, дешевше.

3.2. РАНГОВА КОРЕЛЯЦІЯ

На практиці часто виникає необхідність встановлення зв'язку між порядковими (якісними) змінними.

Якісна ознака – це ознака, яку неможливо виміряти точно, але вона дозволяє порівнювати об'єкти між собою. тобто можна встановити порядок, в

якому об'єкти розташовуються по ступеню прояви ознаки (тестові бали, екзаменаційні оцінки, якість житлових умов).

В таких випадках проблема оцінки тісноти зв'язку вирішується, якщо упорядкувати (ранжувати) об'єкти по ступеню прояви вимірюваних ознак. При цьому кожному об'єкту надається певний номер, називається **рангом**.

Якщо об'єкти ранжувати по двом ознакам, то маємо можливість оцінити тісноту зв'язку між ознаками, базуючись на рангах.

Коефіцієнт рангової кореляції Спірмена:

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^n (x_i - y_i)^2}{n^3 - n}, \quad |\rho| \leq 1,$$

де n – число пар спостережень (об'єм вибірки), x_i, y_i – ранги i -го об'єкту.

Коефіцієнт кореляції набуває значень у межах від -1 до $+1$. Чим ближче ρ до числа $+1$, тим більша узгодженість між результатами тестування. Від'ємне значення коефіцієнта кореляції свідчить про обернену (протилежну) узгодженість.

Алгоритм розрахунку коефіцієнта кореляції рангів:

1) замінюють кількісні (або напівкількісні, або якісні) ознаки ряду x і y на ранги, ранжуючи при цьому строго від меншої величини до більшої (або суворо від більшої до меншої на розсуд дослідника);

2) визначають умовні відхилення, тобто різницю рангів по кожному рядку;

3) зводять умовні відхилення в квадрат;

4) визначають суму квадратів умовних відхилень;

5) підставляють отримані дані у відому формулу і обчислюють коефіцієнт кореляції.

Оскільки за своєю природою якісні ознаки є випадковими величинами (наприклад, на остаточний бал учня за контрольну роботу впливають, зокрема, такі фактори: чи відволікали учнів під час написання роботи, чи мали вони можливість скористатися підказками), то і одержане значення коефіцієнта

рангової кореляції також є випадковою величиною. Постає питання, чи можна робити практичні висновки щодо наявності (відсутності) кореляційного зв'язку за одержаним значенням коефіцієнта рангової кореляції?

Для відповіді на поставлене питання необхідно перевірити статистичну значущість коефіцієнта кореляції. Для цього спочатку розраховують емпіричне значення t-критерію:

$$t_{емп} = \rho \times \frac{\sqrt{n-2}}{1-\rho^2}, \text{ де } \rho - \text{значення коефіцієнту кореляції Спірмена.}$$

Далі обирають певний рівень значущості α , найчастіше він дорівнює 0,05 чи 0,01. Рівень значущості α розуміють як ймовірність того, що наш висновок не підтвердиться. Для обраного рівня значущості та відомої кількості рангів n за спеціальною таблицею критичних значень коефіцієнта рангової кореляції Спірмена знаходять табличну величину $t_{табл}(\alpha, n)$ (Додаток 2).

Якщо фактичне значення $t_{емп}$ більше критичного, тобто $t_{емп} \geq t_{табл}(\alpha, n)$ – зв'язок визнається істотним. У протилежному випадку – $t_{емп} < t_{табл}(\alpha, n)$ зв'язок випадковий або істотність зв'язку не доведено.

Приклад. Знання десяти студентів перевірені по двом тестам А і В. Оцінки за стобальною шкалою виявилися такими (у першому рядку вказано кількість балів по тесту А, в другому – по тесту В):

Прізвище студента	Бали за тестом А	Бали за тестом В
Студент 1	90	92
Студент 2	86	93
Студент 3	95	83
Студент 4	84	80
Студент 5	70	55
Студент 6	75	60
Студент 7	62	45
Студент 8	50	72
Студент 9	57	62
Студент 10	60	70

Обчислити вибіркового коефіцієнт рангової кореляції Спірмена між балами по двом тестам.

Розв'язання.

Розв'язок прикладу зручно оформити за у вигляді таблиці:

Прізвище студента	Бали за тестом А	Ранг x_i	Бали за тестом В	Ранг y_i	Різниця рангів $x_i - y_i$	Квадрат різниці $(x_i - y_i)^2$
Студент 1	90	2	92	2	0	0
Студент 2	86	3	93	1	2	4
Студент 3	95	1	83	3	-2	4
Студент 4	84	4	80	4	0	0
Студент 5	70	6	55	9	-3	9
Студент 6	75	5	60	8	-3	9
Студент 7	62	7	45	10	-3	9
Студент 8	50	10	72	5	5	25
Студент 9	57	9	62	7	2	4
Студент 10	60	8	70	6	2	4
Разом						68

Обираємо принцип розташування рангів – за зростанням чи спаданням. Присвоюємо ранги x_i за виконання тесту А. Наприклад, ми обрали розташування рангів за спаданням, тому перший ранг $x_1 = 1$ присвоюємо студенту, який отримав найбільший бал за тест А (95 балів) – Студент 2. Другий ранг $x_2 = 2$ присвоюємо студенту, який отримав наступний за максимальним бал (90 балів) – Студент 1. Продовжуємо аналогічним чином міркувати, останній ранг $x_{10} = 10$ отримує студент із найменшим балом за тест А (50 балів) – Студент 8.

Аналогічно знаходимо ранги y_i за результатами тесту В. Ранг $y_1 = 1$ отримує студент із найбільшим балом (93) за тест В – Студент 2. Наступним буде Студент 1 у якого 92 бали. Тому Студент 1 отримує ранг 2.

Увага! Порядковий номер рангу НЕ обов'язково співпадає з номером студента у списку.

Далі знаходимо різниці рангів із відповідними номерами, квадрати цих різниць та суму. Обчислимо шуканий коефіцієнт рангової кореляції Спірмена, враховуючи, що $n=10$:

$$\rho = 1 - \frac{6 \cdot 68}{10^3 - 10} = 1 - \frac{408}{990} \approx 0,59.$$

Оскільки коефіцієнт рангової кореляції Спірмена є досить значним, то робимо висновок про існування зв'язку (кореляційної залежності) між результатами студентів за обома тестами.

Перевіримо, чи буде цей зв'язок статистично значущим. Розрахуємо емпіричне значення t-критерію:

$$t_{\text{емп}} = \rho \times \frac{\sqrt{n-2}}{1-\rho^2} = 0,59 \times \frac{\sqrt{10-2}}{1-0,59^2} = 2,56$$

Задамо рівень значущості $\alpha=0,05$. За допомогою Додатку 2 знаходимо значення $t_{\text{табл}}(\alpha, n) = t_{\text{табл}}(0,05; 10) = 2,23$. Оскільки $2,56 > 2,23$, то існуюча кореляційна залежність між результатами з тесту А та тесту В є статистично значущою й відображає реально існуючу закономірність (не є дією випадкових факторів).

Число ступенів свободи f	Рівень значимості α			
	0,10	0,05	0,01	0,001
1	6,31	12,70	63,70	637,00
2	2,92	4,30	9,92	31,60
3	2,35	3,18	5,84	12,90
4	2,13	2,78	4,60	8,61
5	2,01	2,57	4,03	6,86
6	1,94	2,45	3,71	5,96
7	1,89	2,36	3,50	5,40
8	1,86	2,31	3,36	5,04
9	1,83	2,26	3,25	4,78
10	1,81	2,23	3,17	4,59

4. ЕТАПИ ТА ПРАВИЛА СТВОРЕННЯ ТЕСТІВ

При складанні завдань тесту слід дотримуватися ряду правил та рекомендацій, необхідних для створення надійного, збалансованого інструмента оцінки успішності оволодіння студентами навчальною дисципліною.

4.1. ВИЗНАЧЕННЯ МЕТИ І ЗАВДАНЬ ТЕСТУВАННЯ

Зміст тесту має відповідати меті тестування. Тест – не простий набір, а система завдань, що володіє складом, цілісністю і структурою. До складу тесту входять як самі завдання, так і правила їх застосування і оцінювання, а крім того – дані щодо інтерпретації результатів тестування. Система тестування може застосовуватися як засіб усіх видів контролю: початкового (базового), поточного (тематичного), рубіжного (залікового), підсумкового (екзаменаційного) та самоконтролю якості засвоєння теоретичного матеріалу, відповідного рівня набутих знань і практичних навичок студентів. Мають бути чітко і однозначно усвідомлені всі знання, уміння і навички, володіння якими перевіряється за допомогою тесту, тестові завдання мають відповідати заздалегідь запланованим результатам навчання. Для цього необхідно здійснити аналіз змісту навчальної дисципліни. Необхідно провести аналіз контингенту для проведення тестування, час та умови використання тестів.

1) Аналіз змісту навчальної дисципліни в цілому, структурування навчального матеріалу. Структура тесту формується відповідно до цілей і завдань, які визначені в державному стандарті та програмі дисципліни. В основі конструкції тесту лежить структура базових тестових завдань, яка відповідає системі основних навчальних елементів окремої дисципліни. Вибір тестових завдань для поточного чи підсумкового контролю повинен здійснюватися за логікою навчального процесу, бути збалансованим та узгодженим зі змістом навчальної дисципліни. При відборі учбового матеріалу слід виділити круг тем кожного модуля, що включається в тест, і відносну кількість завдань, яким має

бути представлений кожен модуль курсу. Бажано, щоб тестові завдання для кожної окремої дисципліни (або однієї дисципліни для різних факультетів) мали свою специфіку. Контрольні питання за контролюючим розділом програми курсу перекладається у контролюючі завдання педагогічного тесту.

2) Підготовка специфікації тесту як документу, що визначає структуру й зміст контрольно-вимірювальних матеріалів із навчального предмета.

Після визначення цілей тестування і їх конкретизації розробляється план тесту. Цей крок ще часто називають розробкою специфікації тесту. Специфікація передбачає створення стислого змісту (плану) тесту (тестовий проект, матриця). Специфікація розподіляє завдання за змістом, формою та рівнем складності, визначає систему оцінювання окремих завдань і тесту в цілому. Проектування змісту тесту здійснюється на основі матеріалу, опрацювання якого передбачено навчальною програмою відповідного курсу та підручниками. Необхідно опрацювати створені завдання разом із тим теоретичним матеріалом, який надається студентам.

Бажано вказувати відповідність тесту певним джерелам (підручникам, посібникам). Дослідники рекомендують у таблицю специфікації включати наступні елементи:

- перелік форм тестових завдань;
- кількість завдань кожної форми в одному варіанті тесту;
- види вимірюваних інтелектуальних умінь;
- кодифікатор вибраних тематичних розділів змісту предмета;
- кількість завдань у кожному розділі тесту;
- рівень складності завдань і їх розподіл за розділами змісту та формою завдань.

3) Формулювання вимог до тесту – кількість тестових завдань та тривалість тестування.

Тест має бути повним та зваженим, представляти усі аспекти тем відповідної дисципліни без перекосів у бік тієї чи іншої теми та повторів

тестових завдань. Необхідно проаналізувати зміст завдань з позиції рівної представленості в тесті різних навчальних тем, понять тощо. Завдання не повинне містити чинників, що перешкоджають виконати його правильно.

Для важливих компонентів програми необхідно привести більше запитань, щоб можна було здійснювати вибір. Не варто перевантажувати тест завданнями, які легко формулюються, але не несуть інформації про знання студентів з теми. Кількість завдань у тесті має бути вказана у специфікації тесту і враховувати низку обмежень:

- вік тестованих;
- час, відведений на тестування;
- тип використовуваних тестових завдань;
- тип інтерпретації за результатами тестування.

Кількість тестових завдань не повинна бути занадто малою, тому що перевірка знань перетворюється на лотерею. Мінімальний кількісний склад контрольного тесту має становити 25 – 30 тестових завдань. Кількість тестових завдань не повинна бути занадто великою, тому що збільшується час тестування і накопичується втома, а в результаті оцінюється не рівень знань, а витривалість.

Приблизно половина питань, включених у підсумкове завдання, повинна бути на відтворення і формулювання визначень, правил, законів. Друга половина – містити такі завдання, щоб студенти могли продемонструвати своє вміння використовувати теоретичні знання в практичній діяльності або проілюструвати практичне застосування теорії. За предметну та лінгвістичну коректність тестових завдань несе відповідальність автор (лектор).

4) Визначення часу (часового інтервалу), який відводять для виконання тесту

Необхідно продумати визначену тривалість тестування відповідно до кількості тестових завдань. Обсяг завдання має бути розрахований так, щоб тестований міг його виконати за відведений проміжок часу. Для правильного розрахунку часу тестування враховують довжину та складність тесту. Доцільно

слідкувати за часом виконання тесту, щоб не перевантажувати студентів. Оптимальний час виконання тесту пропонують розраховувати за двома методами:

1. Тест пропонують виконати викладачеві з певного предмета. Час, за який він виконав тест, треба помножити на два. Це буде оптимальний час, за який тестовані зможуть виконати завдання.

2. Під час проведення пробного тестування визначити час, протягом якого 20% тестованих устигнуть виконати тест повністю.

5) Визначення моделі педагогічного тестування. Під час оцінювання у формі тестування найчастіше використовують класичну модель, класичну модель з урахуванням складності завдань, модель із зростанням складності, модель із розподілом завдань за рівнем засвоєння навчального матеріалу, модель з урахуванням часу виконання завдання, адаптивна модель та ін. Класична модель є найбільш простою, але має недоліки: випадковість при відборі завдань та неможливість при оцінюванні враховувати складність завдань. Тому класична модель має найнижчу надійність.

Класична модель з урахуванням складності завдань більш адекватна щодо оцінки знань тестованого, проте вона не дає можливості через випадковість вибірки заздалегідь визначити, які завдання за складністю отримає кожен студент. Модель із зростанням складності забезпечує можливість використання паралельних тестів, має вищу, ніж у попередніх моделях надійність результатів. Використання різнорівневих завдань дозволяє якісніше підійти до оцінки результатів навчальної діяльності. Найбільш прогресивною та надійною є адаптивна модель як вид комп'ютерного тестування.

4.2. СКЛАДАННЯ ТЕСТОВИХ ЗАВДАНЬ

Змісту тесту має відповідати найефективніша форма. Завдання, що входять в тест, підбираються так, щоб вони давали основу для перевірки деяких з таких категорій придбаних знань, як-от: назви, імена; факти; визначення; порівняння, зіставлення об'єктів; протилежності, суперечності; асоціації;

класифікації; причинно-наслідкові стосунки; алгоритми, процедури; технології і технологічні поняття; абстрактні поняття; методологія предмету тощо.

Завдання повинні бути реально наближені до предмета, з якого вимірюється рівень знань, мати конкретний зміст. Критеріями вибору виду завдання є: зрозумілість; попередження випадкового вгадування; компактність (стислість) і точність (чіткість). Чим різноманітнішою буде функціональна спрямованість тестів, тим об'єктивнішими будуть виміри результатів тестування студентів. Якщо кілька завдань є однаково ефективні, то треба вибрати той тип, який у тесті зустрічається рідше. Кожне тестове завдання повинно бути інформативним, предметним.

Різнманітність завдань поточної (повсякденної) корекції знань зробить тест менш монотонним й більш цікавим для студентів, що сприяє підвищенню мотивації навчання. Проте у підсумковому контролі рекомендується застосовувати моноформний тест, що містить завдання однакової форми. Завдання тесту мають бути сформульовані коротко, чітко, коректно й однозначно. Тест не повинний бути навантажений другорядними термінами, несуттєвими деталями з акцентом на механічну пам'ять. Якщо твердження складне, то формулювання необхідно скорочувати або поділити на декілька, оскільки непомірно довгі фрази ускладнюють розуміння завдання, викликають утому.

Інформація, що міститься в одному тестовому завданні, не повинна давати відповідь на інше. Великий об'єм фактичної та аналітичної інформації погано піддається формалізації, тому доцільно використовувати завдання у тестовій формі чи відкриті запитання.

1) Написання чітких інструкцій у формі настанови з описанням технології виконання завдання, способів позначення відповідей. До всього тесту або окремих його частин складаються чіткі інструкції. Інструкції мають бути адекватними формі та змісту завдання. Короткі стандартні інструкції зазвичай поміщаються перед завданням або групою завдань і по шрифтовому оформленню відрізняються від змістовної основи завдання і відповідей до

нього. Формулювати інструкцію слід якомога коротше, використовуючи прості речення без уточнень. Якщо завдання представлені однією формою, інструкція пишеться один раз для всього тесту. Якщо ж тест включає завдання різних форм, то перед кожною зміною форми завдання пишеться нова інструкція.

2) Добір варіантів відповідей (альтернатив) до тестових завдань.

Неправильні відповіді (дистрактори) до тестових завдань повинні бути правдоподібні, мати реальне відношення до предмета, що вивчається. Варіанти відповідей на кожне завдання повинні підбиратися таким чином, щоб виключалися можливості простого здогаду чи відкидання свідомо невідповідної відповіді. Правильна відповідь має стояти в завданнях так, щоб у тесті не спостерігалось певної закономірності. Можна використати такі прийоми формулювання дистракторів:

- використання правильних тверджень, які, однак, правильно не відповідають на запитання в основі;
- використання знайомих виразів і фраз. Оскільки їх часто використовують, такі відповіді можуть бути привабливими для тестованих, які мають поверхові знання з предмета.

3) Підбір тестових завдань відкритої форми з вільно конструйованою відповіддю.

Завдання вільної форми мають відповідати навчальним цілям, оцінювати різні рівні інтегрованих знань і вмінь, бути доступними. Завдання має передбачати відповідний метод оцінювання результатів виконання. Слід урахувати, що завдання відкритої форми з короткою відповіддю легко складати, проте не можливо застосовувати для оцінювання складних результатів навчання.

4) Упорядкування завдань у тесті, групування завдань за формою, підготовка їх до пробного тестування.

Загальний вигляд тесту має бути естетичним, привабливим. Тест складається з окремих тестових завдань, які є складовою частиною тесту. Всі елементи завдання розташовуються на визначених місцях, фіксованих у межах обраної форми. Тестові завдання мають бути розташовані у продуманому порядку, який залежить від цілей тестування,

тематики і складності завдань. Довільний порядок компоновки завдань теж може (і повинен) бути продуманим варіантом.

Нумерується кожне завдання, порядковий номер може змінюватися після об'єктивної оцінки труднощів завдання і вибору стратегії пред'явлення завдань тесту. Для кожного тесту повинен бути представлений блок основної інформації, що включає напрямок навчання, з яким пов'язаний тест; назву тесту, прохідний бал; час, відведений на виконання тесту; тип тесту (для самотестування, диференційований або ні); загальні рекомендації щодо роботи з тестом; ім'я студента, групу, дату тестування та іншу необхідну інформацію.

Кожне тестове завдання як складова частина тесту має нести своє смислове навантаження. Завдання кожного типу (однієї форми) можна розміщувати разом, об'єднавши в одну групу (це дає можливість кожному тестованому пристосуватись до даного типу завдань і, до того ж, для групи таких завдань можна використовувати одну інструкцію). Завдання підсумкового тестування мають бути послідовними у зростанні складності. Кожне завдання доцільно відділити одне від одного (бажано за допомогою вільного простору).

5) Оформлення тесту. Текст завдань тесту по можливості друкується на одній сторінці. Шрифтове оформлення завдання має бути таким, щоб суть завдання розумілася з одного погляду. Для полегшення візуального розрізнення частин завдання формат шрифту основної частини завдання вибирають відмінний від шрифту відповідей. Особливо важлива інформація позначається іншим шрифтом. Зазвичай текст основної частини пишеться прописними буквами, відповіді – рядковими, проте можна використовувати і різний розмір шрифту, наприклад, основну частину завдання надрукувати буквами розміром 14 пт, а відповіді – розміром 12 пт. Завдання тесту, згруповані в субтести (групи завдань, об'єднані однією інструкцією), варто помістити у спеціальному тестовому зошиті, а правильні відповіді відзначаються на окремих бланках. Кожну частину субтестів рекомендується починати з нової непарної сторінки, зазначивши на парній, що тест продовжено на наступній сторінці.

Варіанти відповідей розташовуються у стовпчик. Для виконання завдань з вільно конструйованою відповіддю має бути достатньо місця для записів. Місце для відповіді виділяється певним способом у тестовому зошиті або бланку відповідей по можливості однотипно.

б) Розроблення схеми оцінювання результатів тестування (правильні відповіді на завдання (ключі) та їх вага). Оскільки під час контрольних тестувань кінцева оцінка відіграє важливу роль, об'єктивність її виставлення повинна бути максимальною, а всі можливі варіанти фальсифікації результатів – виключені. При складанні плану майбутнього тесту слід звернути увагу на створення шкали оцінювання, тобто визначення ваги кожного тестового завдання. Для цього використовують критерії оцінювання навчальних досягнень студентів з даного предмету. Автор окремо прописує ключі – номери запитання і варіанти відповідей. Кожен збіг з ключем додає, як правило, опитуваному один бал. Завдання з вибором декількох правильних відповідей вважається за виконане правильно, якщо точно вибрані всі без виключення правильні відповіді. Як правило, один бал дається за правильно виконане завдання, нуль – за неправильно виконане. Можна використовувати і ширшу шкалу оцінок. Наприклад, за повністю виконане завдання давати три бали, при одній помилці (одна зайва відповідь або один пропущений) – два бали, і так далі, до нуля. Правила оцінки мають бути визначені заздалегідь і відомі тестованим перед початком тестування.

7) Розроблення критеріїв оцінювання завдань відкритої форми (на надання відповіді). Критерії мають чітко визначати найважливіші параметри виконання завдань. Критеріями оцінювання якості знань при виконанні завдань відкритої форми відповіді є: повнота, глибина, усвідомленість, конкретність, узагальненість, системність, ступінь суттєвості, ступінь інтеріоризованості. Необхідно перед педагогічним тестуванням ознайомлювати тестованих з критеріями оцінювання. Це сприяє усвідомленню ними мети, яку треба досягнути, ознайомлює учасників тестування з параметрами, яких слід дотримуватись для ефективного виконання завдання. Слід урахувати, що на

оцінювання результатів завдань відкритої форми може впливати рівень грамотності тестованого.

8) Проведення експертної оцінки (професійної й лінгвістичної) якості передтестових завдань. Група експертів оцінює конгруентність завдань визначеному для тестування навчальному матеріалу й меті тестування, тобто здійснює професійну, технологічну й лінгвістичну експертизу якості всієї множини тестових завдань, розроблених для конкретного тесту за визначеним обсягом матеріалу. Кожне завдання у тесті повинно бути індикатором відповідного знання, а досягнути цього можна лише за рахунок детальної експертизи цього тесту спеціалістами-предметниками.

Тестові завдання, як правило, створюються провідним лектором (викладачем вищої кваліфікації) або викладачами під його керівництвом. Як і для навчальних посібників, повинен діяти стандартний механізм рецензування та затвердження тестів. За результатами експертної оцінки відбувається доопрацювання форми й змісту завдань автором тестових завдань.

4.3. АПРОБАЦІЯ ТЕСТУ

Тест не можна відразу застосовувати для відповідальних оцінок (заліки, іспити), рекомендується проводити проміжний контроль і пілотні тестування. Бажано проведення тестування вибірки студентів більш ніж 100 осіб. Вибір оптимальної форми тестових завдань передбачає комплектування первинного, пробного, тесту. Перевірка й оцінювання підготовленого тесту має відбуватися за стандартизованими процедурами, прописаними в інструкціях. В інструкціях до процедури тестування вказуються послідовність проведення тестування, його тривалість, можливість користуватися додатковими оснащенням й розмірів кабінету тощо.

Тестування починається з того, що перед самим тестуванням інструктор оголошує правила поведінки під час тестування та технологію роботи з тестом (тестовим зошитом). Тестові завдання можуть бути заздалегідь віддруковані і роздані кожному студенту на листах паперу або, при комп'ютерному

тестуванні, видаватися на екран дисплея. Тестовані дають відповіді відповідно до інструкції – безпосередньо на листах паперу в місцях, спеціально передбачених для відповідей (бланках), або на дисплеї комп'ютера.

4.4. ОБРОБКА ВІДПОВІДЕЙ

Розробку тестів неможливо відокремлювати від математичної обробки результатів тестування та визначення якості розроблених тестів. Оцінювання результатів тестування здійснюється на основі критеріїв, які встановлюються робочою програмою навчальної дисципліни або програмою державної атестації. Тестування надає можливість представлення оцінки у наступних форматах: бал; відсоток від загального числа балів; літерна оцінка на основі отриманого відсотка. Перевірка, оцінювання і шкалювання результатів тестування може здійснюватися як тьютором, так і незалежним експертом. У найпростішому випадку оцінкою студента є сума балів за правильно виконані завдання. Підсумовування всіх оцінок за завдання визначає тестову оцінку випробовуваного, що показує ступінь оволодіння учбовим матеріалом.

Максимальний бал – це бал, який отримує студент, давши стовідсотково вірні відповіді на всі запропоновані питання. Максимальний бал рівний сумі балів (ваги, коефіцієнтів складності) всіх питань, заданих в ході тесту. Прохідний бал (критичний поріг оцінки) – це мінімальна частка від максимального бала, яку необхідно набрати для успішної здачі тесту. Для зручності оцінки прохідний бал тесту визначається у відсотках. Студенту досить набрати певний відсоток від максимального бала тесту, щоб тест вважався успішно пройденим.

1) Визначення рівня (коефіцієнту) складності тестових завдань.

Розподіл завдань тесту за складністю має бути збалансованим. Питанню тесту повинна завжди відповідати вага, або коефіцієнт складності. Традиційно вага всіх питань дорівнює одиниці. Однак викладач в процесі створення або зміни тесту може вказувати для кожного питання відповідну вагу, яка більш точно відображає рівень складності та знань студентів. З'ясування складності

тестових завдань здійснюється на основі аналізу статистичних показників кожного завдання. Показником складності тестового завдання є частка вибірки тестованих, які розв'язали чи не розв'язали дане завдання. Наприклад, якщо лише 20 % випробуваних виконали завдання, його можна вважати складним для даної вибірки, якщо 80 % – легким.

2) Інтерпретація результатів, вибраковування і коректування змісту тестових завдань, їх вдосконалення (остаточне коригування змісту тестових завдань). Аналітичні дослідження здійснюються у формі уточнення педагогічних характеристик як окремих тестових завдань, так і всього тесту в цілому, його валідності, надійності; пояснення, тлумачення результатів виконання завдань, обчислення тестологічних характеристик тестових завдань тощо. Корекція завдань та складання кінцевих тестів здійснюється у відповідності до алгоритму та вимог до тестів із застосування методів інтерпретації результатів: нормативно зорієнтованого або критеріально орієнтованого.

Метою аналізу тестових завдань викладачем (експертом) є не лише виявлення помилок у складанні тесту, а й недоліків при викладанні даного матеріалу. Після процедури тестування доцільно провести обговорення проблемних питань зі студентами та з'ясувати причини отриманих показників. Якщо те чи інше питання тесту було сформульовано коректно, то, можливо, студенти не зрозуміли саме зміст теми. У цьому випадку викладач може ще раз пояснити ті моменти, які були незрозумілими, відкоригувати методику викладання.

Результати тестування студентів можна вважати об'єктивними, якщо вони в основному корелюють з поточною успішністю. Зв'язок тестування з поточною успішністю повинен бути достатньо стійким. Тести вважаються надійними, коли коефіцієнт кореляції становить не менше 0,7 – 0,75.

3) Стандартизація тестів. Постановка завдань і опрацювання результатів тестування мають бути стандартизованими. Тест стає стандартизованим, тобто перетворюється в інструмент виміру знань лише після переробки або видалення

неякісних тестових завдань, проведення додаткового тестування, уточнення характеристик тесту, його специфікації (написання паспорту з нормами, умовами та інструкціями для його багаторазового використання в різних умовах). Щоб довести тест до повної готовності до застосування необхідно кілька років збирати статистичні дані з потоком 100–120 студентів. Стандартизація може обмежити діяльність викладачів, оскільки має вплив на не стандартизовані тести (викладацькі тести, які складаються самими викладачами для своїх студентів).

4.5. ФОРМУВАННЯ БАНКУ ТЕСТОВИХ ЗАВДАНЬ

Для визначення рівня сформованості знань та умінь з вузівських дисциплін, необхідно спроектувати фонд ефективних тестових завдань. Властивості завдань і тесту в цілому повинні забезпечувати якість і ефективність педагогічного вимірювання. Певний перелік тестових завдань повинен мати багаторівневу ієрархічну структуру: рівень навчальної теми дисципліни; рівень навчального модулю дисципліни; рівень дисципліни. Остаточний набір завдань komponується на основі специфікації тесту. Оновлення банку тестових завдань здійснюється щорічно з урахуванням результатів аналізу статистичних характеристик тестових завдань.

Мінімальний обсяг тестових завдань у банку тестових завдань визначається в залежності від обсягу годин, які заплановані студенту на вивчення матеріалу навчальної дисципліни і кількості студентів, які одночасно вивчають відповідну дисципліну. Якщо трудомісткість дисципліни становить 1 – 1,5 кредити, то мінімальна кількість тестових завдань на тему – 20; вибіркова кількість тестових завдань на одного студента у контрольному тесті – 15 – 25; загальна кількість тестових завдань з дисципліни – 100 – 160.

5. РОЗРОБКА СПЕЦИФІКАЦІЇ ТЕСТУ

5.1. МАТРИЦЯ ТЕСТУ

Перед створенням тесту з'ясовують його мету, розроблюють специфікацію.

Специфікація тесту – це опис поля успішності, яке потрібно виміряти. Зазвичай специфікацію готують у формі таблиці (матриці) – комбінаційного розподілу, що пов'язує результати навчання зі змістом і встановлює відносну значущість, яку треба приписати кожному з аспектів оцінювання.

У матриці тесту, зазвичай, по горизонталі розташовують змістові сфери, які передбачається виміряти, по вертикалі – їх маніфестації, тобто шляхи виявлення змістових сфер.

Побудова переліку вимірюваних властивостей залежить від мети тесту. При визначенні кількості завдань маніфестацій (виявів) важливо виокремити різні форми їх реалізації. Зазвичай на кожній осі розміщують від 4 до 7 категорій. Менша їх кількість звужує процес конструювання опитувальника, більша – обтяжує. У результаті створюють решітку, кількість осередків у якій дорівнюватиме кількості запланованих для вивчення параметрів, помноженій на кількість їх поведінкових маніфестацій:

Маніфестації	Змістовні сфери				

Розробляючи тести, як правило, створюють решітку розміром від 16 до 25 осередків (наприклад, 4x4, 4x5, 5x4 або 5x5), оскільки вона є ідеальною для тесту, котрий можна сконструювати, пред'явити і обробити. Потім визначають кількість завдань для кожного з осередків, враховуючи важливість одного із

параметрів щодо іншого або інших. Наприклад, змістовим сферам А і В слід приписати 40% значення, а С і D – 10%. Одночасно кожній маніфестації А, В, С і D приписують 25% значення. Загалом відсоткове значення всіх змістових сфер (по горизонталі) і всіх маніфестацій (по вертикалі) повинно дорівнювати 100%.

Розташування відсотків визначає кількість завдань, що належать до окремого осередку. Після цього з'ясовують кількість завдань, яку слід включити в тест. При цьому враховують розмір решітки і час, передбачений для розв'язання завдань. Дослідник повинен забезпечити надійність тесту, що вимагає збільшення кількості завдань, і мінімізувати кількість завдань для того, щоб досліджуваній зміг концентрувати увагу під час обстеження. Під час визначення кількості завдань тесту зважають на особливості досліджуваного контингенту.

Після визначення відсоткового значення кожного осередку матриці й встановлення кількості завдань тесту підраховують завдання для кожного осередку.

Маніфестації	Змістовні сфери				Кількість завдань
	А 40%	В 40%	С 10%	D 10%	
А 25%	8	8	2	2	20
В 25%	8	8	2	2	20
С 25%	8	8	2	2	20
D 25%	8	8	2	2	20
Кількість завдань	32	32	8	8	80

Кількість завдань для кожного осередку підраховують шляхом множення кількості усіх завдань, призначених для вимірювання певної властивості тестованого, на відсоткове значення її виявів. Наприклад, кількість завдань для лівого крайнього осередку решітки дорівнює 25% до 32 завдань, що становить 8 завдань – $25/100 \times 32 = 8$. Кількість завдань для кожного осередку округлюють.

5.2. КЛАСИФІКАЦІЯ НАВЧАЛЬНИХ ЦІЛЕЙ

Цілі навчання можуть бути виражені в таких поняттях і категоріях, як знання, уміння, навички, способи творчої діяльності й емоційно оцінні норми. Кожна з них, має не просте значення, однак у своїй сукупності вони дають повне уявлення про те, чому маємо вчити. Головна мета навчання і виховання полягає в передачі молодому поколінню нагромадженої попередніми поколіннями людської культури. Вимога чіткого та однозначного опису цілей є базовою для більш узагальненої вимоги діагностичності досягнення цілей навчання, яка, крім певного однозначного опису цілей навчання, означає також визначення методу вимірювання результатів навчання для оцінювання ступеня реалізації навчальних цілей.

Чіткому визначенню цілей навчання сприяє їх класифікація (таксономія). Розглянемо системи класифікації цілей навчання у трьох сферах (доменах): пізнавальній (Блум, 1956 р.), емоційній (Кратхволл, 1964 р.) і психомоторній (Сімпсон, 1966 р.).

1) Когнітивний домен, або пізнавальна сфера.

Класифікація пізнавальної сфери включає шість класів цілей, розміщених відповідно до складності:

- знання (найпростіший клас) – це здатність запам'ятовувати факти, принципи, процеси в різних предметних галузях;
- розуміння – здатність розуміти матеріал, що вивчається;
- застосування – здатність використовувати засвоєний матеріал у нових ситуаціях;
- аналіз – здатність структурувати навчальний матеріал так, щоб була зрозумілою загальна організаційна структура.
- синтез – здатність поєднати окремі частини знань з метою отримання цілого, що набуває нової якості.
- оцінювання – здатність давати оцінку даному навчальному матеріалу в рамках поставленої мети.

Класифікація рівнів пізнавальних процесів за Блумом:

Пізнавальний рівень (категорія)	Пізнавальний процес (інтелектуальне вміння)	Ключові слова для використання в основі тестових завдань
Знання	Упізнавання фактів, термінів, умов, понять, визначень, принципів	Визначте, перерахуйте, ідентифікуйте, назвіть. Хто? Коли? Де? Який?
Розуміння	Пояснення, Інтерпретація вивченого навчального матеріалу	Поясніть, передбачте, інтерпретуйте, виведіть, додайте, перетворіть, переведіть, наведіть приклад, обчисліть, перекажіть
Застосування	Використання понять або принципів, щоб розв'язати проблему (задачу) в типових і конкретних ситуаціях	Застосуйте, розв'яжіть, покажіть, використайте, змініть, продемонструйте, обчисліть
Аналіз	Деструктування системи на складові для виявлення відношень та ієрархії, організація зв'язків між частинами	Диференціюйте, порівняйте, відрізнити__від__, поєднайте з . чому працює ?
Синтез	Створення нового й оригінального зі складових (елементів)	Спроектуйте, сконструуйте, розробіть, сформулюйте, уявіть, створіть, змініть так, щоб... , напишіть есе (твір-мін іатюру)
Оцінювання	Формування судження, яке ґрунтується на заздалегідь визначених критеріях	Спроектуйте, сконструуйте, розробіть, сформулюйте, уявіть, створіть, змініть так, щоб.... Що було б ліпше...?

2) Особистісна, або емоційна сфера.

Емоційна сфера включає цілі, що стосуються емоцій, почуттів, цінностей, позицій та пов'язані зі змінами, які відбуваються в них.

Ця сфера поділяється на п'ять основних класів:

- сприймання – здатність бути уважним;
- реагування – ступінь активності;
- переконаність – готовність відстоювати власну точку зору, вміння вибирати цінності та ідеї;
- селективність – здатність виділяти головне;
- індивідуальність – вищий рівень сформованості особистості, що визначається індивідуальною лінією поведінки;
- психомоторний домен (сфера).

3) Психомоторна сфера пов'язана з рухами, маніпуляціями з матеріалом або об'єктами, а також з координацією. Ця сфера є найскладнішою для класифікації, оскільки за винятком найпростіших рефлексів, усі інші дії включають пізнавальні та емоційні компоненти.

Психомоторна сфера поділяється на п'ять класів відповідно до рівня розвитку психомоторних здібностей.

- перший клас: повторення дії внаслідок спостереження та наслідування – імітація;
- другий клас: дія за інструкцією, що зафіксована в усвідомленому вмінні – маніпуляція;
- третій клас: повноцінне, впевнене (без інструкцій) виконання дії, що контролюється свідомістю;
- четвертий клас: вміння виконувати узгоджено сукупність дій з усвідомленим контролем;
- п'ятий клас: вміння виконувати сукупність дій автоматично з повним засвоєнням знань.

6. ФОРМАТИ ТЕСТОВИХ ЗАВДАНЬ

У сучасній класичній – вітчизняній і зарубіжній – тестології існує загальноприйнята класифікація тестових завдань: завдання закритої форми; завдання на встановлення відповідності; завдання на встановлення послідовності; завдання відкритої форми.

6.1. ТЕСТОВІ ЗАВДАННЯ ЗАКРИТОЇ ФОРМИ

(завдання множинного вибору (multiple-choice) на вибір однієї або кількох правильних відповідей)

ФОРМАТ А (завдання з вибором однієї правильної відповіді).

Характеристика завдань формату А. Завдання з вибором правильної відповіді із запропонованих є найбільш поширеними та простими. Формат А діагностує знання, розуміння, застосування. Перевагами завдань із вибором правильної відповіді з кількох запропонованих варіантів є: стислість і зрозумілість; простота інтерпретації; висока технологічність й чітка структурованість; швидкість тестування, об'єктивність тощо. Недоліком завдань із вибором однієї правильної відповіді з кількох запропонованих, є великі часові витрати на складання якісного тесту, складність пошуку правдоподібних дистракторів, неефективність для оцінювання вміння розв'язувати проблеми, обмежена можливість студентів виразити себе. Тестові завдання альтернативної форми небажано використовувати у підсумкових тестах через досить високу ймовірність угадування.

Структурна схема завдання формату А. Тестове завдання з вибором однієї правильної відповіді складається з наступних компонентів:

– інструкція до виконання з пропозицією обрати правильну (або найточнішу) відповідь;

– основа (умова) завдання – текстова ситуативна задача, яка описує певну проблему, ставить запитання, формулює у формі розповідного речення твердження або незакінчене твердження);

– альтернатива: перелік варіантів відповідей (варіантів закінчення твердження чи альтернативи кількох тверджень), серед яких одна є правильною або найкращою, а інші – дистрактори – є неправильними відповідями;

– оцінка правильності виконання (ключ правильної відповіді, бал).

Завдань формату А можуть бути таких типів:

– завдання з вибором однієї правильної відповіді;

– завдання альтернативної форми (вибір відповідної альтернативи «правильно-неправильно»);

– завдання на вибір з кількох тверджень одного істинного (катафатичний тест);

– вибір зайвої (неправильної) відповіді (апофатичний тест);

– завдання на встановлення причиново-наслідкових відношень;

– завдання за принципом кумуляції.

Завдання з вибором однієї правильної відповіді – тестове завдання, яке має кілька варіантів відповідей, із яких тестований має вибрати одну правильну. Складається з основи та кількох альтернатив (варіантів відповідей).

Приклад 1. Вкажіть, у складі якої із поданих нижче груп наведені новоутворення ранньої юності

А) почуття дорослості; потреба у самоствердженні, особистісна рефлексія

Б) фізичний розквіт, пік активності, сили і витривалості

В) рефлексія, внутрішній план дії, уміння організовувати навчальну діяльність

Г) професійне і особистісне самовизначення, відкриття внутрішнього Я, світогляд

Приклад 2. Закінчіть речення: Рушійною силою розвитку особистості є

А) навчання

Б) реалізація вроджених потенцій

- В) протиріччя
- Г) задоволення потреб
- Д) саморозвиток

Приклад 3. Знайдіть помилку. Світоглядні позиції в ранній юності будуються на

- А) інтелектуальному розвитку
- Б) поведінкових реакціях
- В) рефлексії
- Г) набутій системі знань

Приклад 4. Серед пропонованих понять виберіть ті, що характеризують індивідуально-типологічні особливості мислення

- 1 самостійність
- 2 в'язкість
- 3 зрілість
- 4 критичність
- 5 образність
- 6 інтуїтивність
- 7 гнучкість
- 8 механістичність
- 9 швидкість розуму

Виберіть варіант правильної відповіді

- А) 3, 5, 7, 9
- Б) 1, 3, 6, 8
- В) 1, 4, 7, 9
- Г) 2, 6, 7, 8

Завдання альтернативної форми є модифікацією завдань з однією правильною відповіддю, застосовується для перевірки правильності констатації фактів. Завдання складається з основи – твердження або запитання, яке тестований має визначити як правильне чи неправильне – та двох альтернатив. Тестований має відповісти «так» або «ні», «правильно» чи «неправильно»,

«істинне твердження» чи «хибне» тощо. Перевагами завдань альтернативної форми є швидкість, простота, об'єктивність, економність. У поєднанні з вправами на інтерпретацію дають можливість оцінювати складні результати навчання. Недоліки: висока ймовірність угадування – 50%, важкість уникнути двозначності при складанні завдань, завдання не надають діагностичної інформації про прогалини в знаннях тестованого.

Приклад 1. Дивовижні образи роблять збережену інформацію більш яскравою, і тому її легше відтворити. Правильно чи ні?

А) правильно; Б) неправильно.

Приклад 2. Міжособистісна дистанція залежить від культурних традицій. Так чи ні?

А) так; Б) ні.

Приклад 3. Оцініть істинність твердження: “Порушення правильного дихання призводить до неухважності”

А) істинне твердження; Б) хибне.

Завдання катафатичної форми (вибір із кількох тверджень одного істинного) – інструмент оцінювання, у якому з альтернатив тільки одна сформульована як істинне твердження, решта варіантів відповідей – хибні.

Приклад. Визначте, яке з наведених тверджень є істинним

А) спадковість відіграє вирішальну, фатальну роль у розвитку здібностей, формуванні особистості

Б) спадковість відіграє домінуючу роль, її реалізація відбувається за наявності адекватного середовища та сильної волі

В) особистість за наявності сильної волі може виробити в собі будь-які здібності (навіть на рівні геніальності), незалежно від спадковості

Завдання апофатичної форми (вибір неправильної відповіді) – інструмент оцінювання, у якому із альтернатив тільки одна сформульована як хибне твердження, решта варіантів відповідей істинні. Такі тести рідко використовують у практиці тестування, хоча вони мають переваги перед катафатичними. Переваги: тестовані працюють з істинними твердженнями,

відкинувши незначний відсоток хибних; дає можливість включити в тест формулювання завдань із таким рівнем варіативності, вибіркості й ситуативності, які більш адекватно відображають матеріал; дає змогу включити завдання на узагальнення; скорочує довжину тесту (кількість тестових завдань), зберігаючи при цьому встановлений рівень змісту валідності й надійності педагогічних вимірювань.

Приклад 1. Знайдіть помилку. Психологічними характеристиками діяльності є

- А) урегульованість свідомою метою
- Б) суспільна природа
- В) випереджальне відображення реальності
- Г) спрямованість на перетворення життєвого середовища

Приклад 2. Знайдіть некоректне твердження. У чому виявляється ідентифікація особистості?

- А) у процесі виявлення подібності одного конкретного індивіда з іншим
- Б) у процесі бачення суб'єктом іншої людини як проекції себе
- В) у осмисленні людиною окремої дії, вчинку, сенсу буття
- Г) у механізмі перенесення себе в простір обставин іншої людини

Завдання за принципом кумуляції формулюється як накопичення елементів шляхом підсумування їх кількості в наступній групі до кількості в усіх попередніх. Завдання, побудовані за принципом кумуляції, цікаві тим, що не завжди остання відповідь є правильною.

Приклад 1. Виберіть правильну та найбільш повну відповідь Мислення дошкільника є

- А) наочно-дійовим
- Б) наочно-дійовим і образним
- В) наочно-дійовим, образним, логічним

Приклад 2. Визначте найбільш повну і правильну відповідь. Психологічна готовність дитини до шкільного навчання включає такі компоненти як

- А) розумовий і комунікативний
- Б) інтелектуальний, особистісний, спонукальний
- В) особистісний, інтелектуальний, соціально-психологічний
- Г) особистісний, інтелектуальний, соціально-психологічний, ресурсний

Правила оформлення тестового завдання формату А. Одне тестове завдання не повинно містити підказку відповіді на інше. До кожного тесту повинна бути складена інструкція. Інструкція повинна бути зрозумілою, в ній повинно вказуватись, як вибирати відповіді і яким чином вписувати (відмічати) їх. В інструкції може міститись додаткова інформація, яка необхідна для тестованих. Інструкція поміщається перед завданням, вона друкується шрифтом, що відрізняється від шрифту самого завдання.

Основу завдання доцільно формулювати переважно стверджувальним реченням чітко, однозначно й лаконічно. Основа тестового завдання повинна містити достатню кількість інформації, не можна опускати суттєві деталі, без яких вибір правильної відповіді може бути досить складним. Основа, як правило, формулюється позитивно (частка НЕ погано сприймається у тестовому завданні). Якщо основу сформульовано заперечним реченням (не більше ніж 5% завдань у тесті), треба підкреслити заперечну частку НЕ або виділити її будь-яким іншим способом.

Змістовну частину завдання не перенавантажують другорядними деталями, вона включає мінімум ключових слів (речення до 8 слів із простою синтаксичною конструкцією), необхідних для правильного розуміння завдання.

Формулювання запитання в основі доцільно починати з дієслова. Для кращого розуміння сенсу завдання сильні слова «самий», «найбільший», «найменший», «якнайкращий» доцільно ставити на самому початку.

Ймовірність вгадування правильної відповіді для формату А найбільша серед всіх форматів тестових завдань, тому рекомендується використовувати 3 – 5 варіантів відповідей. Відповіді мають бути змістовними і, по можливості, короткими.

Дистрактори у тестовому завданні можуть бути або абсолютно неправильними, або правильними лише частково. Усі дистрактори повинні бути вірогідними. Для забезпечення надійності тестів всі дистрактори мають бути діючими й однаково привабливими, тобто явно неправильні чи взагалі неприйнятні відповіді необхідно вилучити як форму «замаскованої» підказки. Вибір будь-якого дистрактора, як і правильної відповіді, в ідеалі повинні бути рівно ймовірні. Таке правило відповідає вимозі дискримінативності тестових завдань. Дистрактори повинні мати малу змістовну та формальну контрастність. При розв'язуванні завдань із малою змістовною контрастністю студентові необхідно диференціювати тонкі розбіжності між варіантами відповідей. Формальна контрастність передбачає, що форма правильної відповіді не повинна значно відрізняється від форми неправильних відповідей (довжина, повнота).

По можливості в число невірних відповідей включаються типові помилки тестованих, зафіксовані з практики навчання. Жоден із дистракторів не повинен бути частково правильною відповіддю, яка за певних умов може розглядатися як правильна відповідь. Варто використовувати чіткі, зрозумілі формулювання, застосовувати лише відомі тестованим терміни без двозначності, рідко вживаних слів, сленгу, діалектизмів. У відповідях не рекомендується використовувати слова «все», «жоден», «ніколи», «завжди», «всі названі», «з названих жоден», тому що такі слова/словосполучення допомагають угадати правильну відповідь. Слід виключити застосування у якості правильної відповіді чи дистрактора таких фраз, як «правильної відповіді немає», «всі відповіді правильні», «все з вищевказаного» або «всі відповіді неправильні», «нічого з вищевказаного». Якщо варіанти відповідей починаються з одного слова (кількох однакових слів), їх потрібно перенести в основу завдання. Усі варіанти відповідей мають бути граматично узгодженими з основою завдання.

В кінці змістовної частини завдання, а також в кінці варіантів відповідей не рекомендується ставити розділові знаки. Розташовувати кожен варіант

відповіді доцільно з нового рядка (у стовпчик, але не більше двох колонок). Букви правильної відповіді необхідно варіювати.

Методи підрахунку балів завдань формату А. Основна форма контролю – це виконання або невиконання тесту. У найпростішому випадку оцінкою студента є сума балів за правильно виконані завдання.

ФОРМАТ Х (завдання множинного вибору з невказаною кількістю правильних відповідей)

Характеристика завдань формату Х. Формат Х накладає жорсткі умови на дистрактори. На відміну від формату А дистрактори повинні бути абсолютно неправильними відповідями. Оскільки кількість правильних відповідей невідома, то частково правильні дистрактори завжди будуть викликати суперечки, чи є вони правильними відповідями, чи ні. Імовірність вгадування правильної відповіді значно менше, ніж у форматі А.

Структурна схема завдання формату Х. Тестове завдання багатовибіркового типу складається з наступних компонентів:

- інструкція до виконання, у якій зазначається, що відповідей може бути кілька;
- основа завдання – змістовна ситуативна задача, яка описує певну проблему, ставить запитання, формулює у формі розповідного речення твердження або незакінчене твердження;
- альтернатива: перелік декількох варіантів відповідей (від 5 до 12), будь-яка кількість з яких може бути правильною у форматі класифікацій, а інші є неправильними відповідями;
- оцінка правильності виконання (ключ правильних відповідей, бали).

Приклад. Завдання складається з кількох варіантів відповідей, із яких кілька правильних. Виберіть УСІ правильні, на Вашу думку, відповіді й позначте їх у бланку. Для того, щоб справити сприятливе враження на співрозмовника, доцільно

А) демонструвати йому власну значимість

- Б) виявляти доброзичливе ставлення
- В) виявляти зацікавленість у ньому
- Г) дотримувати дистанції у спілкуванні
- Д) щиро посміхатися

ФОРМАТ N (завдання із задалегідь указаною кількістю правильних відповідей)

Характеристика завдань формату N. Формат N за формою такий же, як і формат X, однак у його запитанні має бути вказана кількість правильних відповідей. Це дозволяє пом'якшити вимоги до дистракторів. Вони, як і для формату A, можуть бути частково правильними. Правильною відповіддю на запитання тестового завдання вважається зазначена в запитанні кількість найбільш правильних дистракторів з усього набору. Якщо потрібно вказати не одну вірну відповідь, а декілька з великого числа варіантів відповідей, то без знання учбового матеріалу вгадати відповідь неможливо, таким чином інформаційна цінність завдань цього формату, вища, ніж у завдань з вибором однієї правильної відповіді.

Структурна схема завдання формату N. Тестове завдання багатовибіркового типу складається з наступних компонентів:

- інструкція до виконання (обов'язково вказується скільки правильних відповідей має завдання);
- основа завдання – змістовна ситуативна задача, яка описує певну проблему, ставить запитання, формулює у формі розповідного речення твердження або незакінчене твердження);
- альтернатива: перелік декількох варіантів відповідей (може бути до 30), будь-яка кількість з яких – N – може бути правильною (2, 3, 4, ... 12);
- оцінка правильності виконання (ключ правильних відповідей, бали).

Приклад 1. Із запропонованого списку психологічних понять виберіть ЧОТИРИ, які відносяться до характеристик вольової поведінки. Виберіть правильні, на Вашу думку, відповіді й позначте їх у бланку.

- А) витримка
- Б) наполегливість
- В) самооцінка
- Г) рефлексія
- Д) енергійність
- Е) дисциплінованість
- Ж) агресивність
- З) авантюризм

Приклад 2. У завданні серед запропонованих варіантів відповідей знайдіть ДВІ ПРАВИЛЬНІ з погляду соціальної психології. Виберіть правильні, на Вашу думку, відповіді й позначте їх у бланку. Лідер може бути

- А) тимчасовим
- Б) постійним
- В) офіційним
- Г) шанованим
- Д) неофіційним
- Е) не шанованим

Методи підрахунку балів завдань формату Х та N. Завдання з вибором декількох правильних відповідей вважається за виконане правильно, якщо точно вибрані всі без виключення правильні відповіді. Це оцінювання здійснюється за методом дихотомічного оцінювання або оцінювання з ваговим коефіцієнтом. При дихотомічному оцінюванні в інструкції вказано, що тестований має вибрати в кожному завданні конкретну кількість правильних відповідей із запропонованих варіантів (наприклад, три із шести). Якщо тестований вибрав всі правильні відповіді, йому ставлять один бал (або інший за ваговим коефіцієнтом), якщо ж він хоч в одній відповіді помилився, – отримує нуль балів.

Можна використовувати і ширшу шкалу оцінок із застосуванням методу часткового бала з введенням порогу оцінювання. Метод часткового бала використовують тоді, коли тестованим заздалегідь не повідомляють кількість

правильних відповідей, а лише вказують, що їх кілька. Якщо, наприклад, завдання має шість варіантів відповідей, із яких три правильні, тоді тестований, який відповів правильно на всі варіанти, отримує три бали, якщо вибрав дві правильні відповіді – два бали, одну – один бал. Якщо тестований вибрав будь-який інший набір відповідей, де поряд із правильними є неправильні, за виконання завдання він отримує нуль балів. Правила оцінки мають бути визначені заздалегідь і відомі учасникам тестування перед його початком.

ФОРМАТ R (завдання на встановлення відповідності з використанням категорії логічних пар або розширеного вибору)

Характеристика завдань формату R. Формат R передбачає наявність у завданні переліку вихідних умов у колонці ліворуч (слів, словосполучень, речень, дат, формул, термінів тощо), до яких необхідно підібрати логічні відповідності з деякого набору (список відповідей у колонці праворуч). У завданнях цього типу слід встановити відповідність між твердженнями з категорії A та твердженнями з категорії 1 (або твердженнями з категорії 1 та твердженнями з категорії A). Тестований має порівняти матеріал лівої й правої колонок і утворити правильні логічні пари.

Кількість відповідностей, як правило, перевищує кількість тверджень. Деякі відповідності можуть одночасно підходити для кількох тверджень, тобто у одних завданнях відповідь можна вибирати тільки один раз, у інших – кілька разів. Порядок тверджень не має значення.

Варто звернути увагу, що завдання цього формату мають значну чутливість до невідповідних ключів.

Варіанти завдань формату R:

- завдання на встановлення відповідностей;
- завдання на встановлення зв'язку;
- завдання перехресного вибору;
- завдання з форматом логічних пар або встановлення аналогій.

Структурна схема завдання формату R. Тестове завдання на відповідність складається з наступних компонентів:

- інструкція до виконання (може бути ще й додаткова інструкція);
- основа завдання або спільне запитання;
- альтернатива: перелік декількох варіантів висхідних умов (від 4 до 20), до яких потрібно встановити відповідність із запропонованим форматом відповідей (кожен варіант відповіді може використовуватися один раз, більше одного разу або зовсім не використовуватися);
- оцінка правильності виконання (ключ правильних відповідей, бали).

Приклад 1. Доберіть до кожного виду провідної діяльності відповідний період вікового розвитку, якому ця діяльність властива

Види провідної діяльності	Віковий період
А) емоційно-безпосереднє спілкування	1. дошкільнята
Б) предметно-маніпулятивна діяльність	2. старший шкільний вік
В) сюжетно-рольова гра	3. підлітки
Г) навчальна діяльність	4. ранній вік
Д) інтимно-особистісне спілкування	5. немовля
Е) професійно-навчальна діяльність	6. молодший школяр
	7. вік дорослості
	8. молодість

Відповідь:

А __, Б __, В __, Г __, Д __, Е __.

Приклад 2. Віднайдіть адекватні приклади кількісних та якісних змін у розвитку

Види змін	Приклади кількісних та якісних змін у розвитку
А Кількісні зміни	1. Збільшення ваги тіла
	2. Поява новоутворень
Б Якісні зміни	3. Збільшення словникового запасу
	4. Оволодіння знаками і мовленням
	5. Збільшення кількості рухів
	6. Поява нових механізмів, процесів, структур

Відповідь:

А __, __, __.

Б __, __, __.

Методи підрахунку балів за завдання на встановлення відповідності формату R. Оцінювання відповідей на завдання формату R можна здійснювати за оцінкою суми правильних логічних пар. Існує два методи, які застосовують залежно від мети тестування й підходів до інтерпретації результатів. Перший – дихотомічна оцінка або оцінка з ваговим коефіцієнтом. Тестованому в інструкції повідомляють, що він має встановити відповідність між матеріалом двох колонок, причому тільки за умови, що всі логічні пари буде утворено правильно, він отримає один бал (або встановлений за завдання ваговий коефіцієнт).

Другий спосіб – застосування часткового бала. В інструкції вказують, що за кожну правильну логічну пару тестований отримає один бал. Тоді максимальну кількість балів тестований отримає за умови, що він установив правильно всі відповідності, якщо він утворив правильно дві пари – отримає два бали, одну – відповідно один бал. Якщо він не вибрав жодної правильної відповіді, отримує нуль. За такого способу оцінювання підвищується диференційна здатність завдання, тому що тестовані з частковим знанням мають вищу оцінку, ніж тестовані з повним незнанням предмета (теми).

ФОРМАТ R 1 (завдання на встановлення правильної послідовності)

Характеристика завдань формату R1. Формат R1 – це різновид формату R, в якому порядок тверджень має значення (подія А передує події Б). Завдання добре структуроване й чітке. Як правило, запитання тестового завдання цього формату вимагає розташувати по порядку набір тверджень, що описують будь-які дії або відсортувати твердження за будь-якою ознакою. Завдання на встановлення правильної послідовності перевіряє сформованість алгоритмічного мислення.

Варіанти завдань формату R1 :

- завдання на встановлення правильної послідовності, порядку (упорядкування елементів послідовності);
- завдання на систематизацію, перестановку;
- завдання на групування.

Приклад 1 (завдання на встановлення правильної послідовності).

Завдання на відтворення ієрархічних зв'язків. Побудуйте правильну ієрархію потреб за Маслоу

- А) потреби самозбереження
- Б) потреби самоактуалізації
- В) потреби у любові та шануванні
- Г) соціальні потреби
- Д) фізіологічні потреби

Відповідь: __, __, __, __, __ .

Приклад 2 (завдання на відтворення вірної послідовності). Прийняття смерті – це неперервний процес, що поділяється на декілька фаз. Встановіть правильну послідовність, встановивши нумерацію цифрами

- ___ депресія
- ___ прийняття
- ___ торги
- ___ заперечення
- ___ гнів

Структурна схема завдання формату R1. Тестове на встановлення правильної послідовності складається з наступних компонентів:

- інструкція до виконання завдання (встановлення правильної послідовності, порядку; здійснення систематизації, перестановки, групування);
- основа – формулювання завдання (проблеми);
- альтернатива: пронумерований список елементів;
- оцінка правильності виконання (підрахунок правильних пар подій або послідовностей).

Методи підрахунку балів завдань формату R1 на встановлення правильної послідовності. Самою природою питання для його оцінювання завдань формату R1 вимагається підрахунок послідовностей, які утворені з пар подій, що вибрані учасником тестування. Тому при оцінюванні відповіді достатньо обмежитися підрахунком правильних пар подій. Таким чином, якщо для послідовності подій, вказаної учасником тестування, утворити всі можливі пари, то кількість пар, що співпали з парами, які утворені для правильної послідовності, може виступати оцінкою відповіді.

Офіційна схема оцінювання формулюється у термінах пар подій з уточненням, що мова йде про послідовні пари. Градування питань за складністю можна ввести відповідно до мінімальної довжини послідовностей, котрі беруться до розгляду.

6.2. ТЕСТОВІ ЗАВДАННЯ ВІДКРИТОГО ТИПУ

ФОРМАТ W (завдання з вільно конструйованою відповіддю).

Характеристика завдань формату W. Формат W – це відкритий формат, який передбачає наявність запитання і відповіді тестованого у довільній формі без указаних можливих варіантів відповідей. Варіанти відповідей в цьому форматі не використовуються. Відкритий формат є потужним засобом перевірки знань, оскільки учень повинен самостійно, без наведених варіантів відповідей правильно відповісти на запитання завдання.

Формат W передбачає обов'язкову участь тьютора (експертів) та критеріїв оцінки відповідей.

Дослідники зазначають низку недоліків тестових завдань відкритого типу:

- оцінювання результатів є трудомістким і потребує багато часу;
- оцінювання таких завдань є найбільш суб'єктивним порівняно з іншими формами завдань;
- складно розрахувати час, необхідний для виконання завдання;
- складно передбачити результати тестованих через високий ступінь свободи щодо структури твору (есе), стилю викладу та глибини розкриття теми;
- каліграфічний почерк, задовільне естетичне оформлення роботи й фальсифікації під час проведення тестування (списування, порушення процедур тощо) підвищують оцінки, а поганий почерк, орфографічні й пунктуаційні помилки їх знижують.

Варіанти завдань відкритого типу з короткою текстовою відповіддю формату W:

- завдання з пропусками частин речення у запропонованому тексті;
- завдання з підкресленням частин речення;
- завдання з короткою числовою відповіддю або розрахунком за формулою.

Варіанти завдань відкритого типу з розгорнутою текстовою відповіддю формату W:

- завдання з довільною розгорнутою відповіддю;
- завдання на завершення відповіді, її доповнення;
- завдання у формі мікротвору, есе.

Структурна схема завдання формату W. Тестове завдання відкритого типу складається з наступних компонентів:

- інструкція до виконання;
- основа завдання – запитання, твердження, речення;

- альтернатива: виконання тестованим завдання згідно з власним баченням, яке має являти собою твердження з невідомою змінною;
- оцінка правильності виконання (критерії оцінки, бали).

Приклад 1. Підтримайте або спростуйте думку: «Усі обдаровані діти стають згодом талановитими дорослими».

Приклад 2. Мимовільна увага пов'язана з інстинктивною поведінкою. Так чи ні? Сформулюйте тезу, наведіть два – три переконливі аргументи, які найкраще підтвердять Ваші міркування.

Приклад 3 (завдання на заповнення прогалін у тексті). Доповніть список перелічених видів пам'яті: довільна, образна, рухова, словесно-логічна, довготривала, короткочасна, _____, _____, _____.

Методи підрахунку балів завдань формату W. Для оцінювання завдань з вільно конструйованою відповіддю важливими є якісні характеристики, певні критерії, які використовують експерти для винесення максимально об'єктивного судження. Критерії чітко визначають найважливіші параметри виконання завдань. Критеріями оцінювання якості знань при виконанні завдань відкритої форми відповіді є: повнота, глибина, усвідомленість, конкретність, узагальненість, системність, ступінь суттєвості, ступінь інтеріоризованості.

Завдання з короткою відповіддю відзначаються високим ступенем регулювання відповіді. Оцінювання тестових завдань відкритого типу з короткою текстовою відповіддю реалізовується як посимвольне порівняння рядка відповіді з еталонним зразком і присвоєння відповіді наперед заданої оцінки. Завдання з короткою відповіддю спроможні оцінювати різноманітні результати навчання на рівні знань, фактів, правил, термінів, тобто здатність відтворення відповіді по пам'яті.

Завдання відкритої форми з короткою відповіддю на практиці оцінюються у 3 бали за наступними критеріями:

- 1 бал зараховується за роботу на 3 – 4 речення загального змісту без суттєвої конкретизації;

– у 2 бали оцінюється відповідь, у якій вища ступінь суттєвості, глибини, наведено конкретні факти, зроблено певні узагальнення;

– 3 бали зараховуються за стислу, але глибоку відповідь, коли правильно охарактеризовано психічне явище, наведені приклади, спостерігається високий рівень суттєвості та інтеріоризованості знань.

Окрім того, береться до уваги логічність та послідовність викладу, загальний рівень мовного оформлення (прийнятний/ неприйнятний).

Завдання відкритої форми з розгорнутою відповіддю передбачає побудову певного висловлювання. Виконання завдання відкритої форми з розгорнутою відповіддю свідчить про те, як тестовані вміють створювати власні міркування, формулювати думки, ідеї, дотримувати певної структури тексту тощо.

Завдання з розгорнутою відповіддю характеризує високий ступінь свободи у відповідях, вони оцінюють результати навчання найвищого рівня (уміння аналізувати, синтезувати, інтегрувати й застосовувати ідеї, трактувати, формулювати власне ставлення щодо запропонованої теми). Повнота та правильність виконання завдання визначається експертами, які перевіряють завдання за авторськими схемами оцінювання.

Завдання відкритої форми не мають формалізованого ключа. Здебільшого вони оцінюються у 5 балів з наступними критеріями:

– 5 балів зараховується за відповідь, у якій студент легко вживає лексику загальнонаукового та психологічного характеру; демонструє уміння самостійно аналізувати, оцінювати, порівнювати, узагальнювати опанований матеріал; застосовує основні положення теорії для вирішення стандартних та нестандартних завдань, робить правильні висновки; вміє інтерпретувати інформацію з різних джерел, виокремлювати головні ідеї, конкретизувати свою відповідь доречними прикладами, встановлювати міжпредметні зв'язки;

– 4 бали зараховується за відповідь, у якій студент дає ґрунтовні відповіді на поставлені запитання, володіє понятійним апаратом, однак допускає

незначні неточності, не грубі фактичні помилки; робить висновки з несуттєвими неточностями;

– 3 бали зараховується за відповідь, у якій необхідні навички роботи з засвоєним матеріалом сформовані недостатньо. Студент допускає значні неточності, фактичні помилки;

– 2 бали зараховується за відповідь, у якій студент демонструє інертність мислення, схильність до шаблону; спостерігається формальне засвоєння знань, розумова інертність, слабка усвідомлюваність знань. Студент допускає значні неточності, грубі фактичні помилки;

– 1 бал зараховується за відповідь, у якій спостерігається фрагментарність, поверховість міркувань, наявне змішування понять; наявні труднощі у формуванні змістових узагальнень.

Орієнтовний обсяг роботи завдання з довільною розгорнутою відповіддю – 1 сторінка (200 – 250 слів).

7. АНАЛІЗ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ

7.1. МЕТОДИ ПІДРАХУНКУ БАЛІВ ЗА ЗАВДАННЯ ТЕСТУ

1) **Методи підрахунку балів за завдання з однією правильною відповіддю.** У практиці педагогічного тестування застосовують наступні методи підрахунку тестових балів:

а) Традиційний метод «Сума всіх балів» (дихотомічне оцінювання).

Формула підрахунку балів:

$$X = \sum_{i=1}^n x_i.$$

Основні переваги: простота у використанні. Можливість швидкого «ручного» підрахунку. Широко використовують для бланкового тестування й «ручної» перевірки.

Основні недоліки: неможливість корекції на вгадування.

б) Корекція на вгадування «Правильний мінус неправильний». Формула підрахунку балів:

$$X = R_i - \frac{W_i}{k-1},$$

де R_i – кількість правильних відповідей; W_i – кількість неправильних відповідей; k – кількість варіантів відповідей у кожному завданні.

Основні переваги: можливість корекції на випадкове вгадування відповідей. Використовують під час масового бланкового тестування й автоматизованого зчитування результатів шляхом сканування. Забезпечує високу надійність тестування.

Основні недоліки: складність масового опрацювання результатів при бланковому тестуванні.

в) Корекція на вгадування «Правильний плюс пропуск». Формула підрахунку балів:

$$X = R_i + \frac{H_i}{k},$$

де R_i – кількість правильних відповідей; H_i – кількість пропущених завдань; k – кількість варіантів відповідей у кожному завданні.

Основні переваги: можливість корекції при небажанні тестованих угадувати відповіді. Використовують під час масового бланкового тестування й автоматизованого зчитування результатів шляхом сканування. Забезпечує високу надійність тестування.

Основні недоліки: складність масового опрацювання результатів при бланковому тестуванні.

2) Методи підрахунку балів за завдання множинного вибору з кількома правильними відповідями.

У практиці педагогічного тестування використовують два методи:

- дихотомічне оцінювання або оцінювання з ваговим коефіцієнтом;
- застосування методу часткового бала.

Вибір кожного з них залежить від мети тестування й підходів до інтерпретації його результатів.

При **дихотомічному оцінюванні** в інструкції вказано, що тестований має вибрати в кожному завданні конкретну кількість правильних відповідей із запропонованих варіантів (наприклад, три із шести). Якщо тестований вибрав всі правильні відповіді, йому ставлять один бал (або інший за ваговим коефіцієнтом), якщо ж він хоч в одній відповіді помилився, – отримує нуль балів.

Недолік цього способу – низька диференційна здатність завдання, тому що відсутня інформація про рівень знань учнів/студентів із частковими знаннями і з повним незнанням теми. Доцільно застосовувати в жорстких критеріально зорієнтованих тестах та для професійної атестації.

Метод часткового бала використовують тоді, коли тестованим заздалегідь не повідомляють кількість правильних відповідей, а лише вказують, що їх кілька. Якщо, наприклад, завдання має шість варіантів відповідей, із яких три правильні, тоді тестований, який відповів правильно на всі варіанти, отримує три бали, якщо вибрав дві правильні відповіді – два бали, одну – один

бал. Якщо тестований вибрав будь-який інший набір відповідей, де поряд із правильними є неправильні, за виконання завдання він отримує нуль балів.

3) Методи підрахунку балів за завдання на встановлення відповідності.

Існує два методи, які застосовують залежно від мети тестування й підходів до інтерпретації результатів.

Перший – дихотомічна оцінка або оцінка з ваговим коефіцієнтом. Тестованому в інструкції повідомляють, що він має встановити відповідність між матеріалом двох колонок, причому тільки за умови, що всі логічні пари буде утворено правильно, він отримує один бал (або встановлений за завдання ваговий коефіцієнт).

Недолік – низька диференційна здатність завдання, тому що воно не відрізняє учнів з повним незнанням від учнів з частковим знанням теми.

Другий спосіб – застосування часткового бала. В інструкції вказують, що за кожну правильну логічну пару тестований отримує один бал. Тоді максимальну кількість балів тестований отримує за умови, що він встановив правильно всі відповідності, якщо він утворив правильно дві пари – отримує два бали, одну – відповідно один бал. Якщо він не вибрав жодної правильної відповіді, отримує нуль. За такого способу оцінювання підвищується диференційна здатність завдання, тому що тестовані з частковим знанням мають вищу оцінку, ніж тестовані з повним незнанням предмета(теми).

Орієнтовна шкала дихотомічного оцінювання та визначення ваги тестового завдання відповідного формату

Форми (формат) завдань	Час виконання одного завдання у хвиликах	Критерії оцінювання завдань (кількість балів за правильну відповідь)
Завдання з вибором однієї правильної відповіді (завдання з множинним вибором)	1 – 1,5	1 бал
Завдання на перевірку знання психологічних понять	1	2 бали

Форми (формат) завдань	Час виконання одного завдання у хвиликах	Критерії оцінювання завдань (кількість балів за правильну відповідь)
Завдання на встановлення хибності або істинності твердження («Вірно – невірно» або «Так-ні»)	1	1 бал
Завдань із завершенням відповіді	1	1 бал
Завдання на відтворення вірної послідовності	2	2 бали або за методом часткового балу з введенням порогу оцінювання
Завдання з вибором кількох правильних відповідей	2	2 бали або за методом часткового балу з введенням порогу оцінювання
Завдання на встановлення хибності або істинності одного з кількох наведених понять	2	3 бали або за методом часткового балу з введенням порогу оцінювання
Завдання на встановлення логічної відповідності	2	2 бали або за методом часткового балу з введенням порогу оцінювання
Завдання з підбором відповіді	2	3 бали
Завдання на визначення групи правильних відповідей	3	3 – 4 бали
Завдання на встановлення хронологічної послідовності	3	3 бали або за методом часткового балу з введенням порогу оцінювання
Завдання на перевірку вміння встановлювати причинно-наслідкові зв'язки	5	3 бали або за методом часткового балу з введенням порогу оцінювання
Завдання на класифікацію інформації за визначеним критерієм (завдання у форматі класифікації)	5	3 бали

Форми (формат) завдань	Час виконання одного завдання у хвиликах	Критерії оцінювання завдань (кількість балів за правильну відповідь)
Завдання відкритої форми з короткою відповіддю	5	3 бали
Завдання на завершення за власним вибором, заповнення пробілу	5	3 бали
Завдання для аналізу психологічного документа або фрагмента першоджерела	10	3 бали
Завдання на перевірку вміння формулювати та аргументувати власну думку (творче есе)	10	5 балів
Завдання відкритої форми з розгорнутою відповіддю (наприклад, проблемне питання, психологічна задача тощо)	10	5 балів

7.2. КОНВЕРТАЦІЯ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ У ШКАЛИ

Розрахунок результату – остаточне оцінювання – це один із завершальних етапів вимірювання. Процедура оцінювання полягає в конвертації одержаного при тестуванні первинного результату(бала) у відповідне число(оцінку) за певною шкалою.

Процедура конвертації має назву **шкалювання**. Питання шкалювання є центральною теоретичною та методичною проблемою педагогічної діагностики, при вирішенні якої необхідно визначитися з вибором шкали, початку відліку, масштабу, інтервалу змін та норми.

Номинальна шкала – це найнижчий рівень вимірювання, за яким визначається наявність або відсутність вимірюваної ознаки без встановлення градацій за ступенем її прояву (наприклад, «залік – незалік»).

Порядкова (рангова) шкала. При вимірюванні на цьому рівні визначається не лише наявність певної ознаки, а й ступінь її прояву (наприклад, «погано», «середньо», «добре», «дуже добре», «відмінно»). Для кожного опитуваного визначається місце, яке він посідає згідно з цією шкалою залежно від ступеня вияву ознаки, за якою проходить вимірювання.

Шкала оцінок успішності, яка була прийнята у школах України, є порядковою шкалою (інтервали між оцінками при 12-бальній системі оцінювання не відповідають дійсному розриву між реальними результатами учнів).

Шкала інтервалів відрізняється від рангової тим, що однаковим відстаням у градації вимірюваного параметра відповідають однакові інтервали на цій шкалі (на шкалі є нуль – точка відліку, одиниця виміру і напрям). За цією шкалою підлягають конвертуванню лише ознаки, розподілені за нормальним законом.

Конвертація результатів тестування у бали. При конвертації результатів тестування використовуються два підходи: критеріальний і статистичний.

Критеріальний підхід. Приклад. Максимальна кількість балів за тест – 50. Бали конвертуються у 12-бальну шкалу так: «12» – 50, «11» – 45, «10» – 40, «9» – 35, «8» – 30, «7» – 25, «6» – 20, «5» – 15, «4» – 10.

Статистичний підхід. Для визначення критеріїв конвертації первинних тестових балів у оцінки підраховується середній показник (M) та стандартне відхилення (s) результатів тестування. Якщо критерій склав $A\%$, то:

«2» – нижче від $A\%$;

«3» – від $A\%$ до значення $(M - s)$;

«4» – від значення $(M - s)$ до значення $(M + s)$;

«5» – від значення $(M + s)$ та вище.

Приклад 1. Конвертація результатів тестування у чотирибальну шкалу при використанні статистичного підходу:

критерій склав – 40,5%. Статистичні дані: $M=70$, $s = 10$.

«2» – нижче ніж 40,5%;

«3» – від 40,5% до 60% ($M - s$);

«4» – від 61% ($M - s$) до 80% ($M + s$);

«5» – від 81% ($M + s$) та вище.

7.3. СТАТИСТИЧНИЙ АНАЛІЗ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ

Після тестування обов'язковим є аналіз його результатів. Аналіз необхідний для того, щоб тестований міг перевірити, наскільки адекватно він оцінює свої знання та скоректувати свою підготовку. Викладач не лише фіксує факт помилок та оголошує вірну відповідь, але й ґрунтовно пояснює випробуваним причини помилкових дій. Випробуваним, які припустилися помилок, надається можливість у письмовому вигляді або усно відповісти на питання: «Чому обрана відповідь не вірна?» «Наведіть докази на користь вірної відповіді».

Робота над помилками, осмислення помилки, її причини, з'ясування помилкової ланки в ланцюзі розумових дій значно зменшують вірогідність запам'ятовування помилкових знань, «витіснення» вірних відповідей. Таким чином, виключається практика «вгадувань» тестованими вірного рішення наведеної в тесті задачі.

Класична теорія тестів ґрунтується на статистичних методах аналізу результатів тестування. Розглянемо найпростіші і необхідні процедури статистичного опрацювання результатів тестування знань і методи оцінки якості тесту. Позначимо через x_{ik} кількісну оцінку успішності виконання завдання № k i -им тестованим. Результати тестування звичайно представляються у вигляді таблиці (матриці) з n рядками та m стовпцями, де n – кількість осіб, m – кількість тестових завдань. Матриця тестових результатів показує результат виконання всіх завдань учасниками тестування.

№ з/п	ПІБ	Тестові завдання				
		1	2	3	...	m
1.	Особа 1	x_{11}	x_{12}	x_{13}	...	x_{1m}
2.	Особа 2	x_{21}	x_{22}	x_{23}	...	x_{2m}
...
n	Особа n	x_{n1}	x_{n2}	x_{n3}	...	x_{nm}

Подальша робота поділяється на дві частини: аналіз досягнень випробуваних осіб та аналіз тестових завдань.

1) Аналіз досягнень випробуваних осіб. На практиці прийнято, як правило, використовувати дихотомічну шкалу оцінок результатів. У результаті правильного виконання завдання тестований отримує один бал, $x_{ik} = 1$, в протилежному випадку – нуль балів, $x_{ik} = 0$. Результат можна оцінювати не лише нулем чи одиницею, але й присвоювати певний ваговий коефіцієнт, що відповідає складності завдання.

Процес статистичного опрацювання матриці результатів тестування будемо розглядати послідовно. На першому кроці обчислюємо індивідуальні початкові бали всіх учнів (студентів) як арифметичну суму всіх одержаних балів:

$$y_i, i = 1, 2, 3, \dots, n, y_i = \sum_{k=1}^n x_{ik}, \text{ де } y_i \text{ – це результат (індивідуальний бал) } i\text{-го}$$

тестованого після проходження тесту.

Далі обчислюємо середній результат \bar{y} сумарних балів учасників тестування

$$\bar{y} = \frac{\sum_{i=1}^n x_{ik}}{n}$$

та встановлюємо середні структурні до яких відносяться такі величини – мода і медіана.

Мода (Mo) (модальна величина) ряду – це така величина, яка найбільш часто зустрічається в даному розподілі.

Медіаною (Me) називається така величина, що займає серединне положення у варіаційному ряду, в якому варіанти розташовані в зростаючому або спадаючому порядку.

Наступним кроком є дослідження питання про відхилення сумарних балів учасників тестування від середнього балу за тест. Для цього застосовують такі поняття математичної статистики як дисперсія та стандартне відхилення. Більші значення дисперсії чи стандартного відхилення свідчать про більші відхилення сумарних балів тестованих від середнього балу всієї групи.

Дисперсія D обчислюється за формулою:

$$D = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1},$$

а стандартне відхилення S знаходиться як корінь квадратний із дисперсії:

$$S = \sqrt{D}.$$

2) Аналіз тестових завдань. Важливою вимогою до тестових завдань є їх об'єктивний рівень складності. В тесті немає місця завданням з невідомою мірою складності. Завдання стають тестовими лише після емпіричної перевірки міри їх складності. Складність завдань можна визначати двома способами :

– на основі оцінювання передбачуваної кількості і характеру розумових операцій, необхідних для вдалого виконання завдань;

– на основі емпіричної перевірки завдань, з підрахунком частки неправильних відповідей. У класичній теорії тестів багато років розглядалися тільки емпіричні показники складності. У сучасних теоріях навчальних тестів, які використовуються в дистанційному навчанні, більше уваги приділяється

характеру розумової діяльності у процесі виконання тестових завдань різних форм.

Емпірично складність завдання визначається додаванням елементів матриці по стовпчиках і дорівнює числу правильних відповідей, отриманих за кожним завданням – R_k . Чим більше правильних відповідей на завдання, тим воно легше для даної групи тестованих. У міру простоти показник R_k зручний, але доти, поки не з'являться інші групи з іншою кількістю тестованих. Тому для одержання об'єктивних характеристик величину R_k ділять на кількість тестованих у кожній групі (обсяг вибірки):

$$p_k = \frac{R_k}{n}.$$

У результаті отримуємо нормований статистичний показник – частка правильних відповідей, p_k . Статистика p_k довго використовувалася як показник рівня складності завдання в класичній теорії тестів. Пізніше була усвідомлена певна її неточність: адже збільшення значення p_k означає не зростання складності завдання, а, навпаки, зростання легкості. Тому з показником складності завдань стали асоціювати протилежну статистику – частку неправильних відповідей, q_k . Вона обчислюється як відношення числа неправильних відповідей w_k (від англ. wrong – неправильний) до кількості учасників тестування n :

$$q_k = \frac{w_k}{n}, p_k + q_k = 1.$$

Зі збірника тестових завдань викидаємо завдання, що не володіють дискримінативністю: $p_k > 0.9$ (надто легкі), $p_k < 0.2$ (надто важкі).

Наступною вимогою до тестових завдань є варіація балів. Якщо на деяке завдання правильно відповідають всі випробувані, то таке завдання стає не тестовим. Учасники тестування відповідають на нього однаково: між ними немає варіації. Відповідно з даного завдання в матриці будуть стояти лише

одиниці. Нетестовим вважається завдання, на яке немає жодної правильної відповіді. Варіація по ньому теж дорівнює нулеві. Нульова варіація означає практичну потребу видалення завдання із тесту.

Зручною мірою варіації є значення дисперсії D_k результатів випробуваних по k -ому завданню:

$$D_k = \frac{\sum_{i=1}^n (x_{ik} - \bar{x})^2}{n-1} \text{ де } k = 1, 2, \dots, m; \bar{x} = \frac{\sum_{k=1}^m x_{ik}}{n}.$$

Обчисливши дисперсію, можна знайти і стандартне відхилення $S_k = \sqrt{D_k}$.

Приклад 2. Десять учнів пройшли тестування за 6 тестовими завданнями (кількості учнів та завдань спеціально обрано невеликими, для спрощення розрахунків). Результати наведено у таблиці. Завдання 1 – 4 оцінюються дихотомічно – 0 чи 1 бал, завдання 5 та 6 – методом часткового балу від 1 до 3.

1) Аналіз результатів учнів. Середній бал цієї групи складає

$$\bar{y} = \frac{\sum_{i=1}^n x_{ik}}{n} = \frac{7+5+6+3+9+2+8+10+9+4}{10} = 6.3 \text{ балів. Медіана – 6}$$

балів, мода – 9 балів. Розрахунок дисперсії проведемо у таблиці (останні два стовпчики).

$$D = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = \frac{68.1}{9} = 7.57$$

. Далі знайдемо стандартне відхилення $S = \sqrt{D} = \sqrt{7.57} = 2.75$.

2) Аналіз тестових завдань. Число правильних відповідей на кожне тестове завдання наведено у таблиці. Тестове завдання №2, для якого частка правильних відповідей $p_2 = 1$ необхідно видалити з банку завдань. Для завдання №6 такий висновок не зовсім вірний, оскільки завдання оцінюється

методом часткового балу. В силу того, що $p_4 = 0.3$ необхідно звернути увагу на тестове завдання №4.

ПІБ учнів	Тестові завдання						Сума балів, y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
	№1	№2	№3	№4	№5	№6			
Учень 1	1	1	0	1	2	2	7	0.7	0.49
Учень 2	0	1	0	0	3	1	5	-1.3	1.69
Учень 3	1	1	1	0	2	1	6	-0.3	0.09
Учень 4	0	1	0	0	1	1	3	-3.3	10.89
Учень 5	1	1	1	1	3	2	9	2.7	7.29
Учень 6	0	1	0	0	0	1	2	-4.3	18.49
Учень 7	1	1	1	0	2	3	8	1.7	2.89
Учень 8	1	1	1	1	3	3	10	3.7	13.69
Учень 9	1	1	1	0	3	3	9	2.7	7.29
Учень 10	1	1	0	0	1	1	4	-2.3	5.29
Число правильних відповідей, R_k	7	10	5	3	9	10			
Частка правильних відповідей, p_k	0.7	1	0.5	0.3	0.9	1			
Частка неправиль- них відповідей, q_k	0.3	0	0.5	0.7	0.1	0			

8. ПОМИЛКИ ТА ТЕХНІЧНІ ДЕФЕКТИ ТЕСТОВИХ ЗАВДАНЬ І ТЕСТУВАННЯ

8.1. ПОМИЛКИ ЗМІСТУ ТЕСТІВ

1. Завдання тесту не відповідають цілям тестування.
2. Недостатня кількість завдань, яка не охоплює змісту дисципліни або теми, яку перевіряють тестами.
3. Неякісний тест, граматична невідповідність між умовою та варіантами відповідей. Нечіткість формулювання, що породжує неоднозначність, а, отже, і дезорганізує навчальну роботу студентів.
4. Підбір завдань та використання тестів без урахування їх складності.
5. Завдання у тесті не упорядковані (не представлений блок основної інформації, завдання не згруповано за формою тощо).
6. Відсутність інструкцій або їхня двозначність, що знижує предметну чистоту змісту завдання, створює перешкоди для його розуміння тестованими.
7. Частина варіантів відповіді вичерпує всі можливі варіанти, що вимагає введення пункту «усі відповіді правильні» і «усі відповіді невірні», «нічого з вищезазначеного», які трактуються як логічні підказки.
8. Надмірність інформації в тексті завдання, багатослівність відповідей та дистракторів. Надмірність вихідних даних (особливо у підсумкових тестах) об'єктивно заважає студентові сприйняти суть завдання в умовах психологічної напруги, пов'язаної з процесом тестування. Намагання автора тестових завдань вкласти у тест якнайбільше корисної інформації виправдано лише у навчальних завданнях тестової форми для самопідготовки та самоперевірки засвоєння знань студентами. Згодом розгорнуті завдання скорочуються до ключових формулювань і набувають вигляду контролюючих тестів. Чим коротше відповіді та дистрактори, тим швидше студент, який знає матеріал, зробить правильний вибір.
9. Пропозиції виконання невластивих тестовому завданню дій – пояснити відповідь, якщо це не стосується завдань у формі тестів, коли

викладач, використовуючи письмові тести, дає студентам можливість усно обґрунтовувати свої відповіді у формі дискусій під час практичних (семінарських) занять.

10. Виключне застосування небажаної питальної форми основної частини завдання.

11. Використання дистрактора, що «не працює». Якщо частота вибору дистрактора дорівнює нулю, тобто дистрактор жоден зі студентів не вибрав жодного разу, то це означає, що його «неправильність» явно впадає в око навіть дуже слабким студентам. Він зайвий. Замінивши такий дистрактор, ми підвищуємо якість тестового завдання.

12. До правильної відповіді включені слова чи фрази з умови, що підвищує ймовірність відгадування студентами правильної відповіді.

13. Правильна відповідь найдовша та найбільш деталізована, тобто спостерігається стилістична неоднорідність варіантів відповідей або нелогічний порядок цих варіантів.

14. Одноманітність побудови завдань (форм); зорієнтованість на неістотне, другорядне.

15. Цифрові показники подані безсистемно (не в єдиному форматі).

16. Умови тестового завдання складні або заплутані.

17. Допущені помилки в ключах.

8.2. ПОМИЛКИ ОРГАНІЗАЦІЇ ТЕСТУВАННЯ

1. Відсутність процедур кваліфікованої експертної перевірки тестових матеріалів і технологій тестування.

2. Стихійність, відсутність систематичності поточного контролю знань через порушення наступності між тестами поточного й підсумкового контролю; нераціональне використання ефективних методів і форм контролю.

3. Відсутність дидактичної цілеспрямованості під час організації навчального тестування (завдання для тренування, самопідготовки та самоперевірки).

4. Невміння викладача, як автора тестових завдань, передбачити (надалі врахувати) характер типових помилок, прогалин у знаннях студентів, а також порушення раціональної послідовності розумових операцій, які можуть призвести до помилок при відповідях на те чи інше завдання.

5. Неправильно розрахований час проведення тестування, незадовільний контроль з боку інструктора.

6. Помилки при обчисленні балів у завданнях на вибір відповіді.

7. Відсутність подальшого доопрацювання створених тестів викладачами. Недостатність внутрішньої мотивації авторів щодо підвищення валідності, надійності, дискримінативності завдань, а також проведення за результатами тестування подальшої корекції знань студентів.

8. Неврахування психологічної готовності студентів до тестування (поточного, підсумкового), узагальнених та специфічних умов роботи в аудиторії під час проведення контрольних заходів.

9. Неврахування основних чинників, які впливають на результати контролю та оцінки знань: загальна і спеціальна підготовка викладача та особисті його якості (принциповість, почуття відповідальності).

10. Значний суб'єктивізм у формуванні змісту тестів, методик проведення тестового контролю; розбіжність у вимогах різних викладачів щодо оцінки одних і тих самих тестових завдань.

Наслідки помилок:

– некоректна підготовка тестів чи організація тестування стає дуже небезпечним інструментом, який дозволяє тиражувати винятково репродуктивні знання;

– створюється небезпека неякісним тестуванням одержати недостовірні дані;

– виявлення недостатньої уваги до творчої діяльності, вмінь і навичок студентів;

- орієнтація більшості тестів на репродукцію знань призводить до навантаження лише пам'яті, а не мислення, внаслідок чого тестування втрачає свою прогностичну здатність;
- нівелювання інтересу студентів до виконання тестових завдань, зниження авторитету викладача, який пропонує неякісні тестові завдання.

9. ПРИКЛАДИ ПОМИЛОК ТА ДЕФЕКТІВ ТЕСТОВИХ ЗАВДАНЬ

1) Нечіткість формулювання, що породжує неоднозначність і дезорганізує навчальну роботу студентів; граматичні й стилістичні дефекти; одноманітність шрифтового оформлення.

Приклад 1. Якщо спостерігач є екстерналії, то причини поведінки індивіда, якого він спостерігає будуть бачитися йому:

- а) в особистісних особливостях що спостерігається;
- б) індивідуальні особливості, що спостерігається;
- в) індивідуально -психологічних особливо спостерігаємого;
- г) в особливостях що спостерігається як суб'єкта діяльності;
- д) в особливостях що спостерігається як соціального суб'єкта;
- е) у зовнішніх обставин.

2) Два запропоновані варіанти вичерпують усі можливі варіанти, одноманітність шрифтового оформлення.

Приклад 2. Спроможність дитини самостійно розв'язувати завдання позначається терміном:

- а) зона найближчого розвитку;
- б) актуальний рівень розвитку.

3) Частина варіантів відповіді вичерпує всі можливі варіанти, що вимагає введення пункту «обидві відповіді вірні», «обидві відповіді невірні», які трактуються як логічні підказки.

Приклад 3. Відчуття і сприймання належать до рівня:

- а) чуттєвого пізнання;
- б) абстрактного пізнання;
- в) обидві відповіді вірні;
- г) обидві відповіді невірні.

4) Неправильне розташування альтернатив, запропоновані варіанти вичерпують усі можливі варіанти, граматичні помилки, одноманітність шрифтового оформлення.

Приклад 4. Одноманітна і монотонна діяльність на уроці може стати причиною розвитку таких несприятливих психічних станів: а) зниження психічної активності, ослаблення уваги, втрата інтересу до діяльності; б) посилення психічного збудження, відраза до діяльності, дратівливість, озлобленість; в) обидві відповіді вірні; г) обидві відповіді не вірні.

5) Логічна помилка у основі і ключі.

Приклад 5. До структури учбової діяльності відносяться компоненти:

- а) мотивація, учбові задачі, учбові дії;
- б) мотивація, учбові задачі, учбові дії, контроль;
- в) мотивація, учбові задачі, учбові дії, контроль, оцінка;
- г) усі відповіді вірні;
- д) усі відповіді невірні.

В основі цього завдання не вказано, що потрібно обрати найбільш повну відповідь, а у ключі відповідей правильною подається відповідь «в», хоча за елементами кожна відповідь є правильною, у тому числі і «г».

б) Стилiстична неоднорiднiсть варiантiв вiдповiдей (правильна вiдповiдь є найдовшою).

Приклад 6. Несвiдоме представляє собою:

- а) форму вiдображення, при якiй певнi компоненти психiчної дiяльностi людини не усвiдомлюються;
- б) найвищу форму вiдображення;
- в) переживання людини;
- г) вci вiдповiдi неврiнi.

7) Варіанти відповідей граматично не узгоджені з основою завдання.

Приклад 7. В якому з рядків перераховано види уваги за участю волі?

- а) наочно-дійовий, наочно-образний, абстрактний;
- б) рухова, емоційна, образна, словесно-логічна;
- в) мимовільна, довільна, післядовільна;
- г) екстероцептивні, інтероцептивні, пропріоцептивні.

Приклад 8. Провідна діяльність – це:

- а) діяльність, якою людина займається постійно;
- б) діяльність, яка є найскладнішою і найактуальнішою для людини;
- в) яка сприяє виникненню новоутворень;
- г) усі відповіді вірні;
- д) усі відповіді невірні.

Приклад 9. Виберіть правильну відповідь.

Які з перерахованих характеристик мислення не властиві молодшому шкільному віку?

А. Егоцентричне мислення.

Б. Неможливість утримувати свою розумову активність на об'єктах, не пов'язаних з конкретними життєвими ситуаціями, з емпіричним досвідом.

В. Мислення стає центром свідомої діяльності дитини.

Г. Підвищується питома вага та взаємозв'язок мисленневих і мовних компонентів.

ВИСНОВКИ

Застосування тестової перевірки знань студентів підвищує ефективність педагогічного контролю з будь-якої дисципліни, дозволяє більш об'єктивно й надійно оцінювати рівень засвоєння знань, набуття умінь і навичок, сприяє підвищенню мотивації навчання, формує адекватне самооцінювання студентами результатів власної діяльності.

Розробка системи тестування – це складний процес, який вимагає високого рівня професійної підготовки в галузі теорії тестових завдань (Item Response Theory) та спільної роботи багатьох авторів-розробників, запровадження надійного програмного забезпечення, використання різноманітних науково обґрунтованих та ефективних технологій контролюючої (моніторингової) діяльності, врахування психологічно-педагогічних аспектів процесу модернізації університетського навчання.

Мистецтво конструювання тестових завдань вимагає, крім знання основних принципів тестології досконалого знання свого навчального предмета самим викладачем. Відповідальність за семантичну коректність (коректність питань, варіантів відповідей, логічну пов'язаність питань модуля) покладається на провідного викладача (лектора).

Реалізацію дидактичного потенціалу тестових методик у вивченні психолого- педагогічних дисциплін можливо забезпечити за умов:

- уніфікації змісту дисциплін гуманітарного циклу вищої школи, систематизації основних категорій наук, відповідно до Галузевих стандартів вищої освіти психолого-педагогічної підготовки фахівців педагогічного профілю;

- поєднання методично вивіренних традиційних підходів та сучасних технологій оцінювання рівня засвоєння знань, набуття умінь і навичок, тому що навіть великого обсягу якісних тестових завдань недостатньо для повноцінного засвоєння та перевірки знань з курсу;

- удосконалення професійної готовності викладача до методичної діяльності використання тестових технологій як засобу контролю фахової підготовки майбутніх спеціалістів;
- забезпечення викладачів методичною літературою (дефіцит якої впливає на якість розроблених тестів та організацію тестування), яка б складала комплекс відповідних теоретичних і практичних знань щодо тестології;
- перетворення тестування на робочий повсякденний інструмент високої роздільної здатності завдяки конструюванню якісних (вілідних, надійних) тестів, адекватних сучасним цілям модернізації навчання й освіти;
- забезпечення кваліфікованої експертної перевірки тестових матеріалів й процедур тестування;
- урізноманітнення форм (типів) тестових завдань та їх складності;
- формування психологічної готовності (мотиваційного, когнітивного, операційного, особистісного компонентів) до випробування учасників тестування;
- розв'язання актуальних проблем методичної підготовки майбутніх учителів до використання тестів у професійній діяльності.

ДОДАТОК 1. ПИТАННЯ ДО ЗАЛІКУ

1. Історія розвитку тестування. Розвиток тестології в Україні.
2. Основні поняття та категорії педагогічної діагностики.
3. Нормо-орієнтовані та критеріально-орієнтовані педагогічні тести.
4. Поняття завдання в тестовій формі і тестового завдання, вимоги до них.
5. Критерії якості методу вимірювання (поняття валідності, об'єктивності та надійності тестів).
6. Стандартизовані тести.
7. Класифікація тестів (за рівнем уніфікації, за рівнем упровадження, за статусом використання).
8. Класифікація тестів (за видом тестового завдання, за методом тестування).
9. Комп'ютерне тестування. Комп'ютерне адаптивне тестування.
10. Довжина тесту. Матриця змісту тесту. Якість тесту.
11. Класифікація навчальних цілей. Таксономія Блума. Інші види таксономій.
12. Завдання з вибором однієї правильної відповіді (приклади завдань, рекомендації до їх змісту, принципи формулювання та добору варіантів відповідей, рекомендації до оцінювання завдань).
13. Завдання з вибором декількох правильних відповідей (приклади завдань, рекомендації до їх змісту, принципи формулювання та добору варіантів відповідей, рекомендації до оцінювання завдань).
14. Завдання на встановлення відповідності (логічні пари) (приклади завдань, рекомендації до їх змісту, принципи формулювання та добору варіантів відповідей, рекомендації до оцінювання завдань).
15. Завдання на встановлення правильної послідовності (приклади завдань, рекомендації до їх змісту, принципи формулювання та добору варіантів відповідей, рекомендації до оцінювання завдань).
16. Завдання відкритої форми (з короткою та з розгорнутою відповіддю) (приклади завдань, рекомендації до їх змісту, рекомендації до оцінювання завдань).
17. Технологія розроблення тестових завдань різних форм (правила написання умови, варіантів відповідей).

18. Технологічний цикл розроблення валідного педагогічного тесту.

19. Конвертація результатів тестування у бали. Шкали оцінювання.

20. Обчислювання балів за формулою для тестів множинного вибору (коригування здогадки).

21. Зовнішнє незалежне оцінювання (основні нормативні документи, структура тестових завдань з різних предметів, оцінювання і шкалювання результатів).

22. Моніторинг в освіті (завдання, функції, історія розвитку моніторингу в освіті, види моніторингових досліджень, особливості педагогічного моніторингу).

23. Міжнародні моніторингові порівняльні дослідження якості освіти (мета проведення, основні міжнародні порівняльні моніторингові дослідження).

ДОДАТКОВІ ПИТАННЯ:

1. Який тип оцінювальної діяльності використовується у сфері освіти?
2. Яку інформацію ви хочете отримати у результаті оцінювання?
3. Хто отримує користь від результатів оцінювання?
4. Наскільки сильно оцінювання впливає на схильність індивідуума до навчання?
5. Чи відрізняється навчання та оцінювання в школах від навчання та оцінювання в університетах?
6. Що таке якість освіти?
7. Чи можна взагалі оцінювати якість освіти та проводити моніторинг освітніх реформ?
8. Навіщо вимірювати і оцінювати якість освіти та проводити моніторинг освітніх реформ?
9. За якими показниками і критеріями можна достатньо повно та об'єктивно охарактеризувати наявний стан освіти і засвідчити досягнення (чи недосягнення) освітніх цілей?
10. Хто визначає достатній рівень якості освіти?
11. Які фактори (чинники) впливають на якість освіти? Як виміряти й оцінити ступінь їхнього впливу (суттєвий або несуттєвий)?

ДОДАТОК 2. КОРОТКИЙ ТЛУМАЧНИЙ СЛОВНИК

АЛЬТЕРНАТИВИ – варіанти відповідей на запропоноване запитання чи твердження, з яких одна правильна (у завданнях з однією правильною відповіддю), а решта — правдоподібні, але неправильні (дистрактори).

АЛЬТЕРНАТИВНЕ ОЦІНЮВАННЯ – тип оцінювання результатів навчально-пізнавальної діяльності студентів методом, що є альтернативним до традиційного.

АНАЛІЗ ТЕСТОВИХ ЗАВДАНЬ – метод, що дає можливість визначити якість тестових завдань та їх внесок у мету педагогічного оцінювання, якої має досягнути тест. При аналізі спираються на статистичні показники результатів виконання тестових завдань, зокрема на їх складність і дискримінаційну здатність.

АПРОБАЦІЯ ТЕСТУ – пробне педагогічне тестування на певній вибірці учнів/студентів, яке проводять перед контрольним (сертифікаційним) тестуванням для усунення неточностей у завданнях, оцінки складності тесту та його окремих частин.

БАНК ТЕСТОВИХ ЗАВДАНЬ – певний перелік тестових завдань з багаторівневою ієрархічною структурою (рівень навчальної теми дисципліни; рівень навчального модулю дисципліни; рівень дисципліни).

ВАЛІДНІСТЬ методу – це комплексна характеристика, яка визначається як параметрами засобів і процедури вимірювання, так і властивостями ознаки, яка досліджується. Термін «валідність результатів» означає придатність їх для тієї мети, ради якої проводилося тестування.

ДИСКРИМІНАТИВНІСТЬ – розрізнявальна (розподільна) здатність тестових завдань, яка дозволяє достовірно розрізняти осіб тестування з різним рівнем знань.

ДИСТРАКТОРИ (англ. distractors) – у тестових завданнях, неточні, неповні, неправильні, але найбільш правдоподібні варіанти відповідей завдань з вибором однієї або кількох правильних відповідей.

ЕФЕКТИВНІСТЬ ТЕСТІВ – порівняльний критерій, який дозволяє порівняти тести. Ефективним можна назвати тест, який краще, ніж інші тести, вимірює знання студентів потрібного рівня підготовки, з меншою кількістю завдань, якісніше, швидше, дешевше, і все це за можливістю одночасно.

ЗАВДАННЯ ТЕСТОВЕ – складова частина тесту, що відповідає вимогам до завдань у тестовій формі та пройшла обов'язкову перевірку статистичних властивостей.

ЗАВДАННЯ ТЕСТОВЕ ЗАКРИТОЇ ФОРМИ – завдання тесту з вибором відповіді з декількох запропонованих варіантів.

ЗАВДАННЯ ТЕСТОВЕ ВІДКРИТОЇ ФОРМИ – завдання тесту, в якому студент повинен самостійно формулювати свою відповідь.

ЗАВДАННЯ З ОДНІЄЮ ПРАВИЛЬНОЮ ВІДПОВІДДЮ – тестове завдання, яке має кілька варіантів відповідей, із яких студент має вибирати одну правильну. Складається з основи та кількох альтернатив (варіантів відповідей).

ЗАВДАННЯ ТЕСТОВЕ НА ВСТАНОВЛЕННЯ ВІДПОВІДНОСТІ – тестове завдання на вибір відповіді, що складається з основи та низки альтернатив варіантів відповідей розміщених у дві колонки. Одна колонка (ліворуч) це перелік вихідних умов (слів, словосполучень, речень, дат, формул, термінів, тощо), до яких тестований має відшукати відповідну відповідь у другій колонці (праворуч), яку називають списком відповідей. Тестований має порівняти матеріал лівої і правої колонок і утворити правильні логічні пари.

ЗАВДАННЯ ТЕСТОВЕ НА ВСТАНОВЛЕННЯ ПРАВИЛЬНОЇ ПОСЛІДОВНОСТІ – тестове завдання, у якому тестований має встановити правильну послідовність елементів, дій, подій, операцій тощо.

ЗАВДАННЯ ТЕСТОВЕ НА НАДАННЯ ВІДПОВІДІ – тестові завдання без указаних можливих варіантів відповідей. Вимагають від тестованого самостійно сформулювати правильну відповідь на поставлене запитання та відповідною її оформити (записати одним словом, цифрою, буквою, словосполученням або підготувати розгорнуту відповідь на кілька сторінок).

ЗАВДАННЯ З КОРОТКОЮ ВІЛЬНОЮ ВІДПОВІДДЮ – це тестові завдання, на які тестований повинен написати відповідь словом, словосполученням або числом. Є окремим випадком завдань відкритих.

ЗАВДАННЯ З РОЗГОРНУТОЮ ВІДПОВІДДЮ – тестові завдання, на які тестований повинен записати відповідь у вигляді одного або декількох речень чи формул. Є окремим випадком завдань відкритих.

ЗАВДАННЯ ПЕРЕСТАНОВКИ – тестові завдання, в яких потрібно переставити елементи списку в потрібній послідовності. Наприклад, розставити події у хронологічному порядку.

ЗАВДАННЯ ОБ'ЄКТИВНЕ – завдання, яке можна оцінити без того, щоб експерт висловлював власне судження щодо якості відповіді. До таких завдань належить: завдання з вибором однієї або кількох правильних відповідей, завдання на встановлення відповідності, завдання на встановлення правильної послідовності.

КЛЮЧ ДО ТЕСТУ – сукупність ключів до всіх завдань тесту.

КРИТЕРІЙ – показник, з яким порівнюють результати тестування для визначення валідності тесту; набір кількісних і якісних характеристик, які використовують для винесення судження щодо якості виконання завдання.

МАКСИМАЛЬНИЙ БАЛ – це бал, який отримує студент, давши стовідсотково вірні відповіді на всі запропоновані питання. Максимальний бал рівний сумі балів (ваги, коефіцієнтів складності) всіх питань, заданих в ході тесту.

МОДЕЛЬ ПЕДАГОГІЧНОГО ТЕСТУВАННЯ – метод, відповідно до якого учню/студенту пропонують тестові завдання й оцінюють результати тестування.

МОНІТОРИНГ – використання засобів і методів контролю засвоєння учбового матеріалу.

НАДІЙНІСТЬ ТЕСТУ – це показник, який характеризує точність і стійкість результатів вимірювання за допомогою тесту при його багатократному застосуванні.

ОБСЯГ ТЕСТУ – кількість завдань, із яких складається тест.

ОСНОВА ЗАВДАННЯ – одна зі складових завдань з вибором однієї або кількох правильних відповідей, що ставить запитання, формулює проблему чи описує ситуацію, до якої запропоновано кілька альтернатив. Основою може бути запитання або незакінчене твердження.

ОЦІНЮВАННЯ – формалізований або експертний процес визначення рівня навчальних досягнень студентів, який завершається виставленням балів. Результати оцінювання навчально-пізнавальної діяльності студентів виражаються в оцінці, яку ставлять відповідно до розроблених критеріїв.

ПЕРЕДТЕСТ – тест, результати якого опрацьовують, інтерпретують, а завдання коригують для укладання остаточного варіанта тесту.

ПОМИЛКА – результат хибних висновків і суджень, типове відхилення від норм і правил. Помилка відображає певну стадію засвоєння матеріалу.

ПРОВЕДЕННЯ ТЕСТУВАННЯ – алгоритм виконання послідовних дій під час тестування. Процедура проведення описується в детальних інструкціях. У них повно і послідовно описується процедура тестування, її тривалість, санітарні умови, характеристики приміщень, права та обов'язки учасників тестування тощо.

ПРОЕКТ ТЕСТУ – вихідний тестовий матеріал, який у подальшому буде трансформовано у систему тестових завдань.

ПРОХІДНИЙ БАЛ – це мінімальна частка від максимального бала, яку необхідно набрати для успішної здачі тесту. Для зручності оцінки прохідний бал тесту визначається у відсотках.

СКЛАДНІСТЬ ЗАВДАННЯ – здатність завдання оцінити належний рівень знань і навичок при допустимому рівні (частці) помилок.

СПЕЦИФІКАЦІЯ ЗАВДАННЯ – спеціалізований вид документа, який використовують для складання певної групи тестових завдань. Це формальні системоутворювальні вказівки від організатора педагогічного тесту для авторів завдань. Може включати таку інформацію, як необхідні форми завдань, види інструкцій, типи поставлених задач і проблем, обмеження для основи,

характеристики варіантів відповідей, особливості правильної відповіді й дистракторів. Формулює правила створення завдань. Мета тестування визначає якість інформації, описану в специфікації завдання.

СТАНДАРТИЗАЦІЯ – уніфіковані процедури планування, оцінювання й виставлення балів.

СУБТЕСТ – група завдань у тесті з чітко вираженими спільними характеристиками і функцією. Зазвичай за кожен субтест ставлять окрему оцінку, що вказує експерту на профіль виконання. Педагогічний тест може складатися з кількох субтестів.

ТЕСТ – сукупність запитань, які переважно вимагають однозначної відповіді, укладений за певними правилами та процедурами, передбачає попередню експериментальну перевірку й відповідає таким характеристикам ефективності, як валідність і надійність. Тест у вузькому значенні – розташований за певними вимогами комплекс завдань, які пройшли попередню апробацію, дають можливість виявити в тестованих ступінь компетенції, результати якої оцінюються відповідно до попередньо визначених критеріїв.

ТЕСТ ПЕДАГОГІЧНИЙ – система взаємопов'язаних предметним змістом завдань специфічної форми, що дозволяють оцінювати структуру і вимірювати рівень знань.

ТЕСТИ ДОСЯГНЕНЬ КРИТЕРІАЛЬНО-ЗОРІЄНТОВАНІ – тип тестів, що призначені для визначення рівня індивідуальних досягнень відносно певного критерію на підставі логіко-функціонального аналізу змісту завдань. В якості критерію (еталону) звичайно розглядаються конкретні знання, уміння навички, що необхідні для успішного виконання конкретного завдання.

ТЕСТИ ДОСЯГНЕНЬ НОРМАТИВНО-ЗОРІЄНТОВАНІ – нормована методика, що дозволяє оцінити відносну якість досягнень студента порівняно з іншими.

ТЕСТИ ПРОЦЕСУАЛЬНІ – тестованим надається можливість продемонструвати когнітивні здібності, що оцінюються на практиці у природній ситуації.

ПЕРВИННИЙ БАЛ (він же **СИРИЙ БАЛ**) – перший результат тестування, що отримується в результаті виконання тесту. Він свідчить про те, скільки завдань тестований розв’язав правильно.

ПСИХОЛОГІЧНА ГОТОВНІСТЬ СТУДЕНТІВ ДО ВИКОНАННЯ ТЕСТІВ – це комплекс мотивів, знань, умінь і навичок, особистісних якостей, які забезпечують успішне здійснення тестування.

ТЕСТОВИЙ БАЛ (він же **СТАНДАРТИЗОВАНИЙ БАЛ**) – визначається в процесі обробки первинних, або сирих, балів. Тестовий бал враховує не тільки кількість правильно вирішених завдань, але і всю сукупність характеристик тесту і тестових завдань.

ТЕСТОВЕ ЗАВДАННЯ – складова частина тесту, що відповідає вимогам до завдань у тестовій формі та пройшла обов’язкову перевірку статистичних властивостей.

ТЕСТОЛОГІЯ – наука про створення і використання тестів. У галузі педагогічних вимірювань тестологія – це теоретико-методологічне й методичне обґрунтування процесів розробки й застосування педагогічних тестів.

ТЕСТУВАННЯ ПІЛОТНЕ – порівняно широкомасштабне «пробне» проведення оцінювання для виявлення будь-яких недоліків до того, як систему оцінювання вважатимуть завершеною й почнуть її запровадження.

СТАНДАРТИЗОВАНИЙ ТЕСТ – це тест, який пройшов апробацію на досить великій кількості учасників тестування і має стабільні показники якості, а також специфікацію – паспорт з нормами, умовами та інструкціями для його багаторазового використання в різних умовах.

ФОРМА ЗАВДАННЯ – різновид тестових завдань певної формальної структури, спосіб організації, упорядкування та існування змісту тесту. Завдання, які використовують у тестах на перевірку успішності, мають такі форми: завдання з вибором однієї або кількох правильних відповідей; завдання на встановлення відповідності, завдання на встановлення правильної послідовності, відкриті завдання з короткою відповіддю, відкриті завдання з розгорнутою відповіддю.

ДОДАТОК 3. КРИТИЧНІ ЗНАЧЕННЯ КОЕФІЦІЄНТУ РАНГОВОЇ КОРЕЛЯЦІЇ

Число ступенів свободи n	Рівень значимості α			
	0,10	0,05	0,01	0,001
1	6,31	12,70	63,70	637,00
2	2,92	4,30	9,92	31,60
3	2,35	3,18	5,84	12,90
4	2,13	2,78	4,60	8,61
5	2,01	2,57	4,03	6,86
6	1,94	2,45	3,71	5,96
7	1,89	2,36	3,50	5,40
8	1,86	2,31	3,36	5,04
9	1,83	2,26	3,25	4,78
10	1,81	2,23	3,17	4,59
11	1,80	2,20	3,11	4,44
12	1,78	2,18	3,05	4,32
13	1,77	2,16	3,01	4,22
14	1,76	2,14	2,98	4,14
15	1,75	2,13	2,95	4,07
16	1,75	2,12	2,92	4,01
17	1,74	2,11	2,90	3,96
18	1,73	2,10	2,88	3,92
19	1,73	2,09	2,86	3,88
20	1,73	2,09	2,85	3,85
21	1,72	2,08	2,83	3,82
22	1,72	2,07	2,82	3,79
23	1,71	2,07	2,81	3,77
24	1,71	2,06	2,80	3,74
25	1,71	2,06	2,79	3,72
26	1,71	2,06	2,78	3,71
27	1,71	2,05	2,77	3,69
28	1,70	2,05	2,76	3,66
29	1,70	2,05	2,76	3,66
30	1,70	2,04	2,75	3,65
40	1,68	2,02	2,70	3,55
60	1,67	2,00	2,66	3,46
120	1,66	1,98	2,62	3,37
∞	1,64	1,96	2,58	3,29

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Авраменко О. В. Вимірювання в освіті: Підручник / За редакцією О. В. Авраменко. – Кіровоград: Лисенко В. Ф., 2011. – 360 с.
2. Булах І. Є. Створюємо якісний тест: навч. посіб. / І. Є. Булах, М. Р. Мруга. – К. : Майстер-клас, 2006. – 160 с.
3. Жерар Франсуа-Марі, Ксав'є Роеж'єр. Як розробляти та оцінювати шкільні підручники / Пер. з франц. М. Марченко. – К.: К.І.С., 2001. – 352 с.
4. Короткий тестологічний словник-довідник / Упорядник Л.Т.Коваленко. К.: Грамота, 2008. – 160 с. (Серія «Словник»).
5. Кухар Л. О., Сергієнко В. П. Конструювання тестів. Курс лекцій: навч. посіб. / Л. О. Кухар, В. П. Сергієнко. – Луцьк, 2010. – 182 с.
6. Лукіна Т. О. Педагогічна діагностика: завдання, методи, інструменти: навчально-методичні матеріали до модуля / Т. О. Лукіна. – К.: Проект «Рівний доступ до якісної освіти в Україні», 2007. – 59 с.
7. Майоров А. Н. Теорія і практика створення тестів для системи освіти / А. Н. Майоров – Львів, 2001. – 296 с.
8. Методичні рекомендації з організації тестового контролю освітньо-професійної підготовки вчителя. – Тернопіль: видавництво ТНПУ ім. В. Гнатюка, 2004. – 100 с.
9. П. Клайн. Справочное руководство по конструированию тестов: Введение в психометрическое проектирование: Перевод с английского / Под ред. Л. Ф. Бурлачука. Киев: ПАН Лтд., 1994. – 288 с.
10. Педагогічне оцінювання і тестування. Правила, стандарти, відповідність. Наукове видання / [Болюбаш Я. Я., Булах І. Є., Мруга М. Р., Філончук І. В.] – К.: Майстер-клас, 2007. – 272 с.
11. Удосконалення змісту й технологій оцінювання якості підготовки майбутніх фахівців відповідно до вимог Європейської асоціації якості освіти: матеріали регіонального науково-практичного семінару / За ред. Г. В. Теращука. – Тернопіль: Вид-во ТНПУ імені В. Гнатюка, 2007. – 160 с.

12. Ярощук Л. Г. Основи педагогічних вимірювань та моніторингу якості освіти: навчальний посібник / Л. Г. Ярощук – Бердянськ: Видавець Ткачук О. В., 2010. – 248 с.

ЕЛЕКТРОННІ РЕСУРСИ

- 1.** <http://testportal.gov.ua/> – Український центр оцінювання якості освіти.
- 2.** <https://mon.gov.ua/ua/tag/zno-2019> – Міністерство освіти і науки України, сторінка щодо ЗНО 2019.
- 3.** www.nbuv.gov.ua – веб-сторінка бібліотеки ім. В. І. Вернадського.
- 4.** <http://osvita.ua/> – Український освітній портал.

Педагогічні вимірювання

Навчальний посібник

Автори

кандидат фізико-математичних наук, доцент кафедри математики та інформатики **Чуйко О. С.**,

кандидат фізико-математичних наук, доцент кафедри математики та інформатики **Сілін Є. С.**

Підписано до друку 12.09.2019

Формат 60x84 1/16. Ум. др. арк. 1025.

Наклад 100 прим. Зам. № 1101.

Видавництво Б. І. Маторіна

84116, м. Слов'янськ, вул. Г. Батюка, 19.

Тел.: +38 06262 3-20-99; +38 050 518 88 99. E-mail: matorinb@ukr.net

Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру видавців, виготівників і розповсюджувачів видавничої продукції ДК №3141, видане Державним комітетом телебачення та радіомовлення України від 24.03.2008 р.
