

*European-Ukrainian project TRAST “**TR**ue **Asses**ment”
European Commission’s TEMPUS/TACIS program
Центр тестування при МОЗ України*

І. Є. Булах, М. Р. Мруга

СТВОРЮЄМО ЯКІСНИЙ ТЕСТ

Навчальний посібник для викладачів,
учителів і методистів

Київ
МАЙСТЕР-КЛАС
2006

ББК 74.04

Б 90

Рецензенти:

О. І. Ляшенко — доктор педагогічних наук, професор, головний вчений секретар АПН України, дійсний член АПН України

В. В. Олійник — доктор педагогічних наук, професор, ректор Центрального інституту післядипломної педагогічної освіти України

Розроблено за підтримки Програми TEMPUS/TACIS
Європейської Комісії в межах проекту "Справедливе оцінювання".

Адреса інтернет-порталу Проекту www.testportal.org.ua.

Проект TRAST "Справедливе оцінювання" фінансується
програмою TEMPUS/TACIS Європейської комісії, 2004—2007.

Жодна частина цієї публікації не може бути відтворена у будь-якому вигляді
і будь-яким чином без попередньої згоди Центру тестування при МОЗ України
та Програми TEMPUS/TACIS Європейської Комісії.

Б90 Булах І. Є., Мруга М. Р.

Створюємо якісний тест: Навч. посіб. — К.: Майстер-клас, — 2006 — 160 с.
ISBN 966-444-005-1

У посібнику обговорюється вітчизняний і закордонний досвід впровадження у практику педагогічної діяльності сучасних методів та технологій педагогічного оцінювання. У першій частині розглянуто основні принципи та правила побудови тесту і тестових завдань. У другій — етапи розробки валідного тесту а також приклади стандартизованих тестів як закордонних, так і вітчизняних.

У посібнику наведено також приклади предметних тестів, розроблених групами викладачів під час проведення тренінгів у міжнародній літній школі у 2005 р. Навчання проводилося у межах міжнародного проекту "Справедливе оцінювання — мережа дистанційного навчання для вчителів — TRAST". Цей проект передбачає поширення серед освітян (вчителів, викладачів та керівників навчальних закладів) нових знань та підходів у педагогічному оцінюванні.

Книга призначена для науковців, педагогів-практиків, працівників системи підвищення кваліфікації вчителів, студентів педагогічних ВНЗ.

ББК 74.04

ISBN 966-444-005-1

© І. Є. Булах, М. Р. Мруга

© Центр тестування при МОЗ України

ЗМІСТ

ВСТУП	5
ЧАСТИНА 1. ОСНОВИ ПЕДАГОГІЧНОГО ТЕСТУВАННЯ	7
1.1. Основні поняття та категорії тестування	8
<i>Метод вимірювання</i>	8
Типи вимірювань	9
Критерії якості методу вимірювання	13
<i>Інструмент вимірювання</i>	16
Тести як інструмент вимірювання	16
Якість тесту	23
<i>Технологічний цикл тестового екзамену</i>	25
Процедури вимірювання й оцінювання	26
<i>Критерії аналізу об'єктивності результатів</i>	28
1.2. Основні принципи та правила побудови тесту	30
<i>Класифікація навчальних цілей</i>	30
Когнітивний домен, або пізнавальна сфера	30
Особистісна, або емоційна сфера	31
Психомоторний домен (сфера)	32
<i>Формулювання цілей навчання</i>	32
Правила формулювання цілей	33
<i>Домени компетентності</i>	34
1.3. Створення тестових завдань	39
<i>Правила створення тестових завдань</i>	39
Правила написання умови	40
Правила написання варіантів відповідей	42
<i>Формати тестових завдань</i>	49
<i>Технологія створення тестових завдань</i>	59
Технічні дефекти тестових завдань	60
<i>Створення тестових завдань різних когнітивних рівнів</i>	62
1.4. Обробка та представлення результатів тестування	68
<i>Ідентифікація учасників тестування та їхніх відповідей</i>	68
<i>Розрахунок результатів</i>	70
Конвертація результатів тестування у шкали	71
Шкали оцінювання	71
Конвертація результатів тестування у бали	72
<i>Контроль можливих порушень</i>	73
<i>Представлення результатів</i>	75
Форми індивідуальних результатів	75
Форми порівняльного аналізу	75
Технологічні форми	76
1.5. МЕТОДИ ВИЗНАЧЕННЯ КРИТЕРІЮ “СКЛАВ – НЕ СКЛАВ”	78
<i>Методи встановлення критерію</i>	79
Основні зауваження	82
ЧАСТИНА 2. ПРОЕКТУВАННЯ СТАНДАРТИЗОВАНОГО ТЕСТУ	83
2.1. Етапи розробки валідного тесту	84
<i>Проектування тесту</i>	84

Мета проведення іспиту	85
Опис домену, що діагностується	86
Вибір методу вимірювання	87
Матриця змісту тесту	87
Вибір форматів тестових завдань	88
Формування тесту та його специфікація	89
Докази валідності та надійності результатів	89
Вибір раціонального та аргументованого прохідного бала	90
Урівнювання тесту	90
Фальсифікація	91
2.2. ПРИКЛАДИ СТАНДАРТИЗОВАНИХ ТЕСТІВ	92
<i>Тест перевірки загальних академічних здібностей</i>	93
Мета та метод оцінювання	94
Домен змісту, що оцінюється, та навчальні цілі	95
Структура (матриця) тесту	96
Формати тестових завдань	103
Висновки	106
<i>Медичні ліцензійні тестові іспити</i>	107
Мета ліцензійних інтегрованих тестових екзаменів	108
Форми представлення результатів тестування	111
Критерії “склав – не склав”	111
Матеріали ліцензійних інтегрованих тестових екзаменів “Крок” (прикладі)	113
2.3. ПРИКЛАДИ ПРЕДМЕТНИХ ТЕСТІВ	118
<i>Суспільствознавство</i>	118
Мета оцінювання	118
Матриця змісту, що оцінюється	118
Навчальні цілі	119
Матриця тесту	119
Зразки тестових завдань	120
<i>Фізика</i>	125
Матриця змісту, що оцінюється	125
Матриця тесту	126
Зразки тестових завдань	127
ЧАСТИНА 3. ДОДАТКОВІ МАТЕРІАЛИ	131
3.1. ІНФОРМАЦІЙНІ РЕСУРСИ	132
3.2. ІСТОРІЯ РОЗВИТКУ ТЕСТУВАННЯ	134
3.3. КРИТЕРІЇ ВАЛІДНОСТІ РЕЗУЛЬТАТІВ ВИМІРЮВАННЯ	143
3.4. ОСНОВНІ ЗАКОНИ ТА ЗАКОНОМІРНОСТІ	146
ЛІТЕРАТУРА	151

Вступ

*Вимірювання – права рука навчання.
Парнелл Д., 1973*

Приєднавшись до Болонської декларації у травні 2005 року, Україна стала 44-м учасником Болонського процесу. Однак для того, щоб ця участь не була формальною, необхідно докорінно змінити ставлення до сфери освіти і до процесів, які в ній відбуваються. 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 завдань 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.

Метод тестування широко використовується в європейських країнах: Німеччині, Швеції, Норвегії, Великій Британії тощо. У Франції, наприклад, уже давно однією зі складових системи відбору до вищих навчальних закладів є використання ранньої, починаючи ще зі старших класів середньої школи, диференціації учнів відповідно до їхніх природних здібностей. За основу беруться показники успішності з різних дисциплін шкільної програми, які визначаються за результатами систематичного тестування академічних та інтелектуальних здібностей.

Вимірювання й оцінювання результату навчання може здійснюватися за допомогою засобів педагогічної діагностики, створення яких потребує спеціальних знань у галузі стандартизованих методів контролю якості освіти. Засобами діагностики є професійно орієнтовані тести, впровадженню яких має

передувати період їх стандартизації з обов'язковим визначенням їх психометричних показників.

Отже, йдеться про відтворення в Україні технології стандартизованого тестування, яка використовується в усіх професійних тестуваннях у більшості розвинених країн. Нині тестування вийшло далеко за межі системи освіти, його застосовують під час масових іспитів на професійну придатність, при відборі до вищих навчальних закладів, масових призовів до армії, комплектації спеціальних підрозділів поліції тощо. Серед причин, що обумовили широке використання тестування, можна вказати такі:

- бізнесовий інтерес до найефективнішого використання людських ресурсів;
- прагнення винагороджувати талант незалежно від його соціального походження.

Проблемою, що нині постала перед українським суспільством, є відсутність об'єктивних критеріїв, які забезпечували б порівняння фахівців, закладів, послуг, програм тощо і створювали б основу для конкуренції, тим самим стимулюючи підвищення якості освіти. Ця ситуація певною мірою є наслідком того, що формування фахівців (початкова, середня, вища освіта) здійснюється за відсутності об'єктивного оцінювання їхніх особистих здібностей і професійної компетентності. Причиною цього є відсутність викладацьких кадрів, здатних застосовувати уніфіковані для всієї країни методи об'єктивного контролю під час виховання та навчання. Суб'єктивізм при оцінюванні учня сприяє скоріше протекціонізму, ніж конкуренції. Тому формування нового покоління через педагогічний ланцюг “новий викладач – новий випускник – новий фахівець” сприятиме підвищенню індексу людських ресурсів України, а отже, і зростанню її економічного потенціалу. Загальнонаціональна перевірка успішності навчання, рівня засвоєння змісту освіти — найважливіше завдання на сьогодні.

Пропонована книга складається з трьох частин. У першій частині розглянуто основи педагогічного тестування, зокрема основні поняття та категорії тестування, наведено основні принципи і правила побудови тесту, а також правила створення тестових завдань, порядок обробки і представлення результатів тестування. У другій частині розглянуто основи і етапи проектування стандартизованого тесту, шляхи його валідазації, наведено приклади стандартизованих і предметних тестів. Третя частина – додатки. В ній подано основні інформаційні ресурси, наведено коротку історію розвитку тестування, критерії валідності результатів вимірювання, основні закони й закономірності методу тестування, а також методи визначення критерію “склав – не склав”.

Частина 1.

Основи педагогічного тестування

Визначення основних понять педагогічного тестування

Інтерпретація основних характеристик методу тестування

Етапи створення якісних тестів

1.1. Основні поняття та категорії тестування

Метод вимірювання

Інструмент вимірювання

Технологічний цикл тестового екзамену

Критерії аналізу об'єктивності результатів

Перш ніж застосовувати на практиці методи педагогічного тестування, потрібно усвідомити, що поняття “тестування” може використовуватися, по-перше, як метод вимірювання, по-друге, як узагальнена назва процесу вимірювання.

У першому випадку, зокрема, ми маємо чітко виокремлювати такі поняття, як метод тестування та інструмент, яким здійснюється тестування. У випадку опису процесу вимірювання ми маємо розрізнити технологію тестування і процедури, які забезпечують здійснення процесу вимірювання.

Зупинимося окремо на кожному з основних понять і категорій тестування: методі вимірювання, інструменті тестування, технології та процедурі тестування.

Метод вимірювання

Даючи визначення методу вимірювання, з'ясуємо передусім, що таке вимірювання в загальному розумінні та у контексті педагогічних вимірювань. Вимірювання (measurement) – це *процес* надання чисельного значення певному показнику відповідно до його кількісного прояву із застосуванням чітких визначених правил вимірювання. Тоді як метод вимірювання – це *спосіб*, за допомогою якого здійснюється надання кількісного значення показнику, який вимірюється.

Якщо як метод вимірювання використовується тестування, то воно розглядається як процес вимірювання кількісних показників за допомогою тесту.

Тестування (testing) як метод – це процес вимірювання кількісних показників за допомогою тесту

Тест (test) – це сукупність тестових завдань (item), підібраних за певними правилами для вимірювання певного кількісного показника

Отже, метод тестування передбачає, що інструментом вимірювання є тест, складений із тестових завдань, процедурою вимірювання є тестування, методом оцінювання є шкалювання.

У випадку педагогічного вимірювання, тестування — це метод вимірювання певного рівня знань екзаменованого. Тест складається із сукупності тестових завдань, підібраних за певними правилами, про що йтиметься далі.

У науковій літературі використовується багато визначень поняття “тестування”, запропонованих, як правило, психологами. В Європі класичним є визначення Г. Лінерта: “Тестування – це звичайний науковий метод дослідження однієї або кількох ознак особи, які емпірично розрізняються, мета якого – визначити відносну ступінь прояву ознаки особи на основі максимального використання кількісних показників” [1]. У педагогічних дослідженнях не можна задовольнитися цим визначенням, оскільки воно не включає критеріально-орієнтоване тестування. Більш влучними є визначення В. Візерковського: “Тестуванням можна вважати загалом всі контрольовані ситуації, в яких, по-перше, діагностично-релевантна поведінка викликається стандартними подразниками і для яких, по-друге, розроблено правила інтерпретації, що дозволяє на підставі спостережуваної поведінки зробити висновок про наявність та ступінь прояву здібностей” [2] та К. Інгенкампа: “Тестування – це метод педагогічної діагностики, за допомогою якого вибір поведінки, що репрезентує передумови або результати навчального процесу, повинен максимально відповідати принципам “співставлення”, об’єктивності, надійності та валідності вимірів. Він повинен пройти обробку й інтерпретацію і бути прийнятним для застосування у педагогічній практиці” [3].

Типи вимірювань

Вимірювання з метою відбору

Вимірювання досягнень

Визначено два типи вимірювань: вимірювання, орієнтовані на норму, й вимірювання, орієнтовані на критерії. Такий поділ обумовлено різною метою вимірю-

вання: для відбору або для встановлення рівня досягнень [4]. Кожна мета вимірювання вимагає застосування різних типів вимірювань, розробки різних тестів.

Тести, орієнтовані на норму, мають тестову стратегію типу “мету досягнуто – мету не досягнуто”. Порівнюючи результати з деякою нормою, їх можна застосовувати на іспитах та при інших випробуваннях, коли приймаються рішення *альтернативного* характеру: “зарахований – не зарахований”, “прийнятий – не прийнятий”.

Тести, орієнтовані на співвідносну групу, передбачають співставлення індивідуальних результатів тестування з результатами, отриманими на репрезентативній вибірці. Тести, орієнтовані на критерій, визначаючи індивідуальні результати, передбачають оцінювання взаємозв’язку цих результатів з попередньо встановленими критеріями.

Тести, класифіковані за цією ознакою, репрезентують два напрями у теорії вимірювання якісних ознак: методи, орієнтовані на критерії, та нормативно-групові методи.

Таким чином, коли ми починаємо розробляти тест, то чітко маємо усвідомлювати, для якого вимірювання він створюється.

Вимірювання з метою відбору

За типом вимірювання з метою відбору належить до нормо-орієнтованого, основне завдання якого – розрізнити екзаменованих за рівнем підготовки. При цьому змістовим компонентом, що оцінюється (доменом оцінювання), виступає не лише зміст тієї чи іншої програми з дисципліни, а й зміст, яким забезпечуються 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.

Під час нормо-орієнтованого вимірювання тест, як правило, складається з тестових завдань середньої складності. Якщо ви хочете відібрати найкращих абітурієнтів для навчання у вищому навчальному закладі певного напрямку підготовки, то, спираючись на досвід країн, де тестування застосовується вже десятки років, слід вибирати не предметні тести, а комплексний тест. Останнім оцінюється змістовий компонент початкових вимог до майбутнього фахівця з вищою освітою певного напрямку підготовки. Цим тестом визначається готовність абітурієнта до навчання у вищому навчальному закладі.

Узагальнену інформацію щодо нормо-орієнтованого вимірювання наведено на рис. 1.

Нормо-орієнтоване вимірювання:

Основне завдання: розрізнити (диференціювати) екзаменованих

Домен: великий; вимірюється в "ширину", містить інформацію про загальні риси оволодіння змістом

Тестові завдання: середньої складності

Результат: відносний

Сфера використання: відбір, ранжування, порівняння індивідуумів, оцінювання програм, порівняння освітніх інституцій

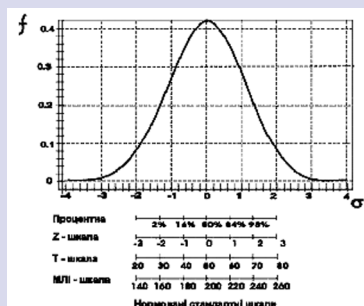


Рис. 1. Вимірювання з метою відбору

Розглянемо для прикладу відбір до вищих навчальних закладів на *медичні спеціальності* [3]. У таких країнах, як США, Канада, Німеччина цей відбір 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. Аналогічна ситуація має місце під час відбору абітурієнтів за педагогічною, інженерною, юридичною професіями тощо.

Вимірювання досягнень

Ще один тип вимірювання – вимірювання досягнень. Його мета – визначити рівень досягнень особи, яка тестується, незалежно від результатів інших. У цьому випадку домен оцінювання (змістовий компонент) вибирається не широкий, а з певним ступенем деталізації програмного матеріалу, який оцінюється. При цьому рівень складності тестових завдань, а отже, тесту загалом, має відповідати рівню засвоєння матеріалу: знанням, розумінню, застосуванню, аналізу тощо.

Найвідоміший американський тест SAT вимірює вербальні та математичні досягнення випускників середньої школи [5]. За цим тестом визначають критерій (критерії) їхньої готовності до навчання у вищому навчальному закладі.

Узагальнену інформацію щодо критеріально-орієнтованого вимірювання наведено на рис. 2.

Критеріально-орієнтоване вимірювання (домено-орієнтоване вимірювання):

Основне завдання: визначити абсолютний рівень досягнень індивідуума (незалежний від інших)

Домен: вузький; вимірюється в "глибину", містить інформацію про оволодіння конкретними елементами змісту.

Тестові завдання: складність відповідає складності оцінюваного домену

Результат: абсолютний (порівняння з критерієм, стандартом)

Сфера використання: сертифікація, ліцензування, оцінка мінімальної компетентності

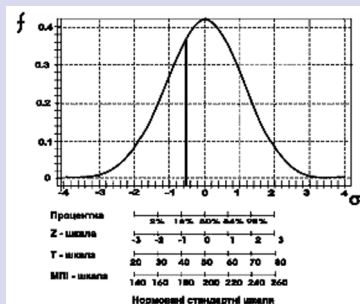


Рис. 2. Вимірювання з метою сертифікації

Вимоги до випускника середньої школи в Україні визначають державні стандарти, до випускника середньої школи в країнах Європейського Союзу — відповідними вимогами до ключових компетентностей. Якщо ви ставите завдання визначити досягнення учнем рівня, передбаченого державними стандартами для випускника середньої школи, необхідно створити комплексний тест досягнень, який встановлюватиме, відповідає учень рівню стандарту чи ні. Якщо потрібно визначити, чи засвоїв учень програму з певної дисципліни, можна створити тест досягнень, але вже не комплексний, а предметний.

Таким чином, у сучасній тестології сформувалося дві стратегії вимірювання. Це привело до диференціації не лише тестів, а й тактики тестування: вимірювання, орієнтоване на норми, та вимірювання орієнтоване на критерії. Крім відмінностей щодо стратегії та тактики процесу вимірювання, вони мають суттєву різницю щодо закономірностей, яким підпорядковуються результати вимірювання. Так, вимірювання, орієнтовані на норми, ґрунтуються на теоретичних положеннях, основними з яких є такі:

- здібності людини є генетичними даними;
- за рівнем здібностей люди розподіляються згідно з нормальним законом розподілу;
- тестування дає змогу визначити місце людини відносно математичного очікування на кривій Гауса.

Водночас при вимірюваннях, орієнтованих на критерії, форма розподілу не має вирішального значення. При таких вимірюваннях, залежно від рівня прийняття рішення щодо досягнення мети, використовуються різні моделі

обробки результатів, зокрема й такі, що не базуються на нормальному законі розподілу.

Критерії якості методу вимірювання

Валідність
Об'єктивність
Надійність
Точність

Вибір того чи іншого методу вимірювання знань ґрунтується на критеріях, які визначають їх якість. Найважливішими критеріями є валідність, об'єктивність, надійність і точність [4]. Дамо їх визначення та означимо зміст кожного з них.

Валідність

Найбільш фундаментальним є критерій *валідності*. Цей термін походить від англійського “valid” і означає обґрунтований, дійсний, придатний, той, що має силу. Найбільш вдале визначення валідності педагогічного тесту дав А. Анастасі [6].

Валідність тесту – це “поняття, яке визначає, що вимірює тест і наскільки якісно це здійснюється” (А. Анастасі)

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

Ми часто можемо чути, що тест, за яким проводилося тестування, не валідний. Що це означає? А те, що одержані результати не вимірюють того, що було визначено метою тестування. Це може бути наслідком дії кількох чинників. Наприклад, невідповідності інструменту вимірювання (тесту) меті оцінювання. Наведемо декілька прикладів. 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. Завдання, які входять до тесту, не є стандартизованими, отже, невалідність результатів може бути наслідком неякісних тестових завдань.

У стандартних вимогах до педагогічних тестів їх валідність має визначатися як комплекс інформації про характеристики тесту, процедури тестування та методики оцінювання.

Ми пропонуємо підхід до аналізу якості тестів і тестових завдань на основі визначення їх головних характеристик і параметрів, коли загальне поняття валідності щодо процесу вимірювання і оцінювання рівня знань диференціюється за функціональною ознакою таким чином:

- валідність методу (валідність змісту, відповідності, прогнозу);
- валідність тесту (валідність тестових завдань, процедури тестування, процедури оцінювання).

Валідність – це комплексна характеристика, що визначається як параметрами засобу та процедури вимірювання, так і властивостями досліджуваної ознаки

Валідність методу – це відповідність того, що вимірюється цим методом, тому, що він має вимірювати. Отже, критерій валідності встановлює сферу дійсності, для якої метод дає статистично достовірні результати. Валідність методу при вимірюванні успішності навчання можна класифікувати за такими критеріями:

- валідність змісту;
- валідність відповідності;
- валідність прогнозу.

Валідність *змісту* – це відповідність вимог до змісту.

Порушення паритету між вимогами та змістом навчання призводить до порушення достовірності результатів вимірювання, навіть якщо вибраний метод відповідатиме іншим критеріям (об'єктивності та надійності)

Валідність *відповідності* – це відповідність результатів вимірювання та оцінювання, одержаних різними методами.

Валідність змісту і відповідності може бути кількісно визначено через *коефіцієнт валідності*. Але необхідно зауважити, що розраховується коефіцієнт валідності опосередковано – як коефіцієнт кореляції між результатами тестування та результатами інших вимірювань, здійснених на тій самій групі з того самого предмета іншим методом. Результат тестування можна вважати валідним при коефіцієнті кореляції більшому ніж 0,6.

Отже, можна дати наступне визначення коефіцієнта валідності.

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

Об'єктивність

Ще один критерій методу вимірювання – *об'єктивність*. Він 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.

Критерії валідності й об'єктивності фактично є первинними, оскільки порушення одного з них призводить до порушення критеріїв надійності та точності, які, незважаючи на їх важливість, виступають як похідні.

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. Детально про правила, стандарти і відповідальність під час тестування піде мова у наступних книгах цієї серії.

Надійність

Надійність методу вимірювання визначається ступенем стійкості результатів. Цей показник впливає на точність, з якою можна виміряти ту чи іншу конкретну ознаку. Перевірка надійності методу стосується насамперед співставлення результатів при повторних вимірах.

Надійність методу залежить від:

- об'єктивності методу (об'єктивності процедур тестування);
- параметрів інструменту вимірювання (якості тесту);
- стабільності характеристики, що вимірюється.

Ступінь надійності методу визначається за допомогою коефіцієнта надійності.

Коефіцієнт надійності (R) 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 коефіцієнт α -Кронбаха.

При створенні тестів, які використовуються під час широкомасштабних тестувань (а до них належать передусім національні та галузеві тестування), 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.

Точність чи помилка вимірювання

Коефіцієнт надійності тісно пов'язаний з помилкою вимірювання, яка визначається *точністю* методу.

Точність методу визначає мінімальну або систематичну похибку, з якою можна провести вимірювання даним методом.

Чим вища точність методу, тим менша помилка під час вимірювання, тим більша надійність цього методу. З теорії похибок випливає таке твердження: при усуненні інших систематичних похибок зміни у результатах вимірювання підпорядковуються статистичним закономірностям. Існують певні формули, які дають змогу після вимірювання визначити коефіцієнт точності (**E**). Це, зокрема, дозволяє кількісно розрахувати міру точності та врахувати її при визначенні кінцевих результатів.

Інструмент вимірювання

Ще одне поняття, яке ми розглянемо, — інструмент, за допомогою якого здійснюється вимірювання. Інструмент вимірювання — це категорія, що визначає засіб, за допомогою якого здійснюється вимірювання. Наприклад, під час усного опитування і письмової роботи інструментом вимірювання є *білет або структуровані завдання*; під час тестування — *тест*, складений з *тестових завдань*.

Тести як інструмент вимірювання

Види тестів *Якість тесту*

Нагадаємо, що тест (test) — це сукупність тестових завдань (item), підібраних за певними правилами для вимірювання певної властивості.

Зауважимо, що з середини 50-х рр. ХХ ст. педагогічне вимірювання та педагогічне тестування відокремилися від психологічних досліджень і сформувалися в окрему науку — тестологію, яка має свої закони і правила, що дають можливість визначати та науково обґрунтовувати показники якості тесту [7]. Детальніше історію розвитку тестування наведено у Частині 3. Додаткові матеріали.

Класифікація тестів

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: К. Інгенкамп [3], І. Булах [4], А. Анастасі [6], В. Аванесов [8, 9], В. Беспалько [10], О. Киверялг [11], М. Розенберг [12].

Наведемо узагальнену класифікацію, яка, на думку авторів, найбільш однорідно структурована і за якою класи тестів згруповано за однією певною ознакою.

Отже, педагогічні тести можуть бути класифіковані за:

- рівнем уніфікації (тести стандартизовані, нестандартизовані);
- 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) [13].

За рівнем уніфікації тести поділяються на стандартизовані та нестандартизовані.

Розглянемо тести, класифіковані за цією характеристикою, більш детально. *Стандартизація* (від *англ.* standard – типовий, нормальний) стосовно тестування означає уніфікацію, тобто приведення до єдиних норм процедури вимірювання та показників якості тесту. Завдяки стандартизації методики вимірювання забезпечується можливість зіставлення результатів тестування, визначення тестових оцінок у відносних стандартизованих показниках, а також зіставлення тестових оцінок, одержаних за різними тестовими методиками.

Як зазначено в роботі [14], у психологічній діагностиці під стандартизацією розуміємо два різних поняття.

У першому випадку мається на увазі стандартизація процесу вимірювання та оцінювання, 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.

У другому випадку стандартизація використовується у більш вузькому значенні, визначаючи лише процес конвертації результатів тестування у шкалу оцінок.

Застосуємо термін “стандартизація” у першому розумінні до тесту.

Стандартизованим називається тест, що має комплексну характеристику, яка визначається його властивостями, процедурою вимірювання і процедурою шкалювання, а також чіткою регламентацією процедури та логістики (організації) процесу тестування. Порушення будь-якого з елементів зводить нанівець усю роботу зі створення стандартизованого тесту, що значно обмежує сферу його застосування та скасовує наслідки за одержаними результатами. Якщо ви підготували якісний тест, але не забезпечили відповідної регламентованої процедури вимірювання, то отримаєте результат, який взагалі нічого не відбиває. Доречно навести думку Р. Ебеля про те, що найбільшої шкоди навчальному процесу завдають не диференційовані за співвідношенням із нормою або критерієм тести, а власне нестандартизовані тести та такі, що проведені з порушенням процедур [15].

Стандартизації тесту передують *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.

Серед основних характеристик стандартизованого тесту нижче розглянемо його довжину і психометричний комплекс параметрів якості, що утворюють фактично паспорт тесту, за яким визначається його якість.

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. Так, тестуванню за загальнонаціональними тестами має передувати ретельне, неодноразове пілотне тестування на репрезентативній “національній” вибірці, оскільки результати цих тестувань здебільшого мають *адміністративні* наслідки.

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

Тести, орієнтовані на норму, використовують тестову стратегію типу “мету досягнуто – мету не досягнуто”. Таким чином, 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 співвідносну групу, передбачають зіставлення індивідуальних результатів тестування з результатами, отриманими на репрезентативній

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

Тести, класифіковані за цією ознакою, як визначалося раніше, репрезентують два напрями у теорії вимірювання якісних ознак: методи, орієнтовані на критерій, та нормативно-групові методи. Але, на думку провідних фахівців у галузі вимірювання Дж. Макка, В. Ханта, Р. Ебеля та Р. Фріке, існує тісна взаємозалежність цих методів один від одного.

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 структурованого есе.

Роз'яснень потребують завдання з короткою (продукованою) відповіддю. Вони можуть бути сформульовані у різній формі подання знань, 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.

Окремо необхідно зауважити, що, незалежно від типу розробленого тесту, *класифікація* може бути здійснена *за методом тестування* залежно від *процесу, процедури та технології тестування*.

За методом тестування ми маємо розрізнити тестування *бланкове, комп'ютерне* та *комп'ютерне адаптоване*.

Бланкове тестування, яке ще називається pencil-pen testing, передбачає, що екзаменовані відповідають на тестові завдання шляхом внесення олівцем або ручкою відповідей у спеціальні бланки відповідей.

Під час зовнішніх, а отже, масштабних тестувань, зчитування даних з бланків відповідей здійснюється переважно за допомогою комп'ютера. Існує два типи сканерів: спеціалізовані й універсальні. Спеціалізовані mark reader створено спеціально для зчитування бланків відповідей (кожної відповіді окремо).

Універсальні сканери мають широкий спектр застосування і зчитують бланк як “картинку”. Перші сканери вважаються точнішими. Але останнім часом потужний розвиток універсальних сканерів щодо швидкості та обсягів зчитування робить їх конкурентними із спеціалізованими сканерами.

Комп’ютерне тестування, яке в зарубіжній літературі отримало назву computer-based testing (СВТ), 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 екзаменованим 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 порівнянно 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 та в інтерфейсному середовищі тестової програми, а також велика база якісних тестових завдань як для забезпечення закритості екзаменаційного тесту, так і для забезпечення валідності змісту. Остання вимога потребує формування тесту не довільно (наприклад, рандомізованим методом), а відповідно до матриці змісту тесту (детально про матрицю змісту тесту далі).

Під час сертифікаційних і ліцензійних тестувань дорослих тестові організації усе частіше використовують комп’ютерне тестування. Водночас вступні тестування абітурієнтів у країнах, де такі тестування здійснюються, відбуваються у формі бланкових.

Окреме місце серед двох названих методів займає адаптивне тестування. Адаптивне тестування, або більш загальноприйнятий термін – **комп’ютерне адаптивне тестування** (САТ), являє собою тестування за допомогою комп’ютера, під час якого надання тестових завдань залежить від відповідей екзаменованого на попередні запитання. Звичайно, якщо екзаменована особа відповідає на запитання правильно, то наступне запитання буде складнішим. І навпаки, якщо екзаменований відповідає неправильно, то наступне запитання буде легшим. Отже, тест “орієнтується” на рівень здібностей екзаменованого [16]. Тому різні екзаменовані, що тестуються за системою САТ, найімовірніше, складатимуть різні форми тесту. Компетентний екзаменований повинен мати мало (або зовсім не мати) однакових завдань порівняно з тим, чия компетентність значно нижча. Такого ніколи не було в традиційному тестуванні, де кожен отримував такі самі завдання, як і інші. У традиційному тесті вимір “кількість правильних відповідей” працює як головний показник результату тестування. В САТ це не працюватиме, тому що всі екзаменовані приблизно на половину запитань дадуть правильні

відповіді. Більш компетентні екзаменовані дадуть правильні відповіді на половину запитань зі складного набору. Менш компетентні дадуть правильні відповіді на половину запитань зі свого більш легкого набору. Об'єднуюча основа, яка утримує всі ці тести разом, – особливий тип психометричних теорій, так звана Теорія Відповідей на Тестові Завдання (IRT). У ній порівнюється компетентність екзаменованого зі складністю завдання і передбачається ймовірність, з якою ця особа дасть правильну відповідь на наступне запитання. Якщо рівень компетентності особи набагато вищий від рівня складності завдання, то така ймовірність буде високою. Якщо рівень складності завдання набагато вищий за рівень компетентності особи, то така ймовірність буде низькою. Ми дізнаємося найбільше тоді, коли ця ймовірність складе приблизно 50 на 50 ($p = 0,5$). Алгоритм вибору завдань полягає у відборі завдань, які є найбільш інформативними і водночас відповідають специфікаціям змісту, критичним для належного тесту. Компетентність екзаменованого розраховується за складністю завдань, які він отримав і на які дав відповідь.

Теоретичні та технологічні засади САТ лише розробляються і, як видно з наведеного, потребують не лише потужної бази каліброваних тестових завдань, а й чіткого алгоритму процесу тестування.

Довжина тесту

Тест складається з певної кількості тестових завдань. Виникає запитання: яким чином ці тестові завдання підібрати і в якій кількості?

Кількість завдань, з якої формується екзаменаційний тест, розраховується за допомогою статистичних методів для забезпечення достовірності висновків і називається довжиною тесту. Певну залежність [17] довжини тесту від точності вимірювання наведено нижче.

Довжина тесту, тестових завдань	Точність вимірювання, %
400	5
100	10
25	25

Якщо треба виміряти рівень знань більш точно, у вас має бути довший тест. Безумовно, довжина тесту обмежується певними психофізіологічними характеристиками контингенту, який тестується, а тут наведено лише статистичні (ідеальні) показники.

Матриця змісту тесту

Інший чинник, що впливає на довжину тесту, — це зміст, який у даному випадку має бути репрезентований певною кількістю тестових завдань, що виступають узагальненням змісту, який оцінюється відповідно до мети. Поданий таким чином зміст, який оцінюється за допомогою тесту, отримав назву **матриці тесту** (рис. 3). *Це зовсім нове поняття, але дуже важливе для побудови якісного тесту. Необхідно підкреслити, що наявність матриці тесту є вихідною необхідною умовою, без виконання якої набрані тестові завдання не можуть називатися інструментом вимірювання*, а отже, не можуть створити якісний тест. Вони залишаються тільки набором окремих тестових завдань навіть тоді, коли ці тестові завдання якісні.

Як видно з рис. 3, матриця тесту може бути структурованою не лише за змістом, а й за рівнями і часом засвоєння матеріалу, форматами тестових завдань, за видами діяльності тощо. Отже, це багатовимірна матриця, і чим детальніше її побудуємо, тим більш точні, сфокусовані вимірювання одержимо. Тоді ці вимірювання не лише констатуватимуть загальний результат, а й стануть реальним інструментом управління якістю підготовки через потужний зворотний зв'язок із процесом навчання.

Під час проектування матриці слід діяти в такій послідовності:

- сформулювати мету оцінювання;
- описати домен, що діагностується;
- створити N-вимірну матрицю змісту тесту, яка може складатися, наприклад, зі змісту, рівнів засвоєння знань, навичок тощо;

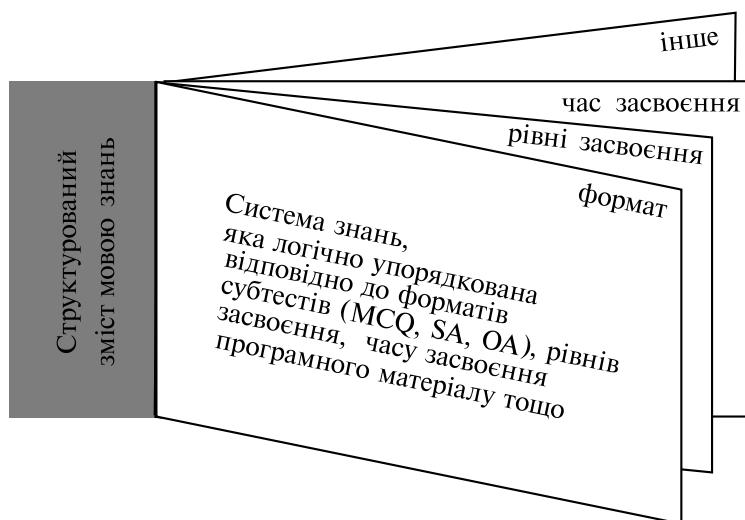


Рис. 3. Структура змісту тесту у вигляді матриці тесту

- визначити співвідношення окремих елементів структури змісту тесту;
- визначити специфікацію тестових завдань.

Приклади матриць тесту наведено у розділах 2.2 «Приклади стандартизованих тестів» і 2.3 «Приклади предметних тестів».

Якість тесту

Якість тесту – це окрема категорія, що базується на певних теоріях і закономірностях. Вона, безумовно, залежить від того, на який тип вимірювання ми орієнтувалися і який статус використання має тест: обов’язковий, пілотний або дослідницький.

Визначення якості тесту базується на певних теоретико-методологічних засадах та методах, що використовуються під час аналізу результатів тестування.

Методи, що використовуються під час аналізу, поділяються на аналітичні (експертні) та емпіричні (статистичні, математичні). Самостійну функцію статистичного аналізу тесту виконує *психометричний* аналіз, що являє собою сукупність емпіричних методів визначення якості інструменту вимірювання, які базуються на даних самого вимірювання (на відміну від оцінки його якості без урахування даних вимірювання). В Україні цей метод відносять до *кваліметрії*.

Сучасна теоретико-методологічна база тестування спирається на такі теорії тестування:

- Класична теорія тестування (J. Laserfeld, R. Thorndike, A. Anastasi, L. Cronbach, R. Brennan);
- Теорія “запитання – відповідь” – IRT – Item Response Theory (G. Rash, M. Birenbaum, R. Hambleton);
- ТМППТ – Теорія моделювання та параметризації педагогічних тестів.

Не зупиняючись на відмінностях названих теорій тестування, підкреслимо, що всі вони визначають певні показники, за якими встановлюється якість тесту. Ці показники формують так званий *психометричний комплекс тесту* та *тестових завдань*, обов’язкові компоненти якого наведено нижче.

Основні характеристики психометричного комплексу тесту

Наведемо головні характеристики психометричного комплексу тесту, які є обов’язковими для стандартизованого тесту, тобто такого, що використовується під час широкомасштабних національних тестувань, і за якими визначається якість тесту та процедур тестування:

- середнє (M), мінімальне та максимальне значення;

- середнє квадратичне відхилення σ ;
- коефіцієнт надійності R (α -Кронбаха, Кьюдер-Ричардсона);
- помилка вимірювання – E ;
- коефіцієнт валідності (за наявності);
- контингент екзаменованих (кількість, структура).

Додаткові характеристики (відхилення від нормального розподілу):

- коефіцієнт асиметрії A_s (асиметричність кривої);
- коефіцієнт ексцесу E_x (зміна гостроти кривої біля вершини).

Психометричні характеристики тестового завдання

На відміну від психометричних характеристик тесту, психометричні характеристики тестового завдання стосуються його складності та ступеня, за якими це тестове завдання розрізняє екзаменованих за вимірюваною ознакою, наприклад за рівнем знань. Основні психометричні характеристики тестового завдання такі:

- складність тестового завдання (індекс складності P_{diff} або P -величина);
- розподільна здатність тестового завдання (індекс дискримінації ID).

Складність тестового завдання (індекс складності P_{diff} або P -величина) визначає, наскільки конкретне тестове завдання було складним для осіб, які тестувалися.

Що таке складність тестового завдання? Те, що є складним для учня першого класу, не є складним для учня десятого. Тобто тестове завдання є складним лише стосовно конкретної групи учнів, конкретного рівня, класу, школи тощо. Індекс складності дорівнює відсотку студентів, які правильно відповіли на дане тестове завдання. Він коливається від 0 до +1,0. Отже, чим вищим є цей показник, тим легшим є тестове завдання: якщо 100% учнів відповіли правильно, то отримуємо показник 1. Залежно від значення складності тестове завдання кваліфікується таким чином:

- $0,20 < P_{diff} \leq 0,36$ – надто складне;
- $0,36 < P_{diff} \leq 0,84$ – середньої складності;
- $0,84 < P_{diff}$ – надто легке.

За цим показником просто диференціювати легке, середньої складності, складне завдання або відгадане.

Розподільна здатність тестового завдання, індекс дискримінації – 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. Індекс може коливатися від -1 до $+1$. Якщо $ID = 0$, то всі екзаменовані відповідали однаково (добре або погано).

Найпростіший метод розрахунку (“ручний”) такий. За результатами цього тестування необхідно визначити різницю між складністю тестового завдання для

групи сильних (Hi) і для групи слабких (Lo) екзаменованих. Залежно від кількості тестованих осіб, групи слабких і сильних утворюються шляхом розподілу кількості осіб, які тестувалися, навпіл або на чотири чверті. Формула розрахунку має такий вигляд: $ID = P_{diff}(Hi) - P_{diff}(Lo)$.

Тестове завдання має достатню розподільну здатність при $ID \geq 0,2$.

Якщо $ID \leq 0,0$, сильні студенти відповідають гірше за слабких, і це тестове завдання відбраковується. При $0 < ID \leq 0,2$ його розподільна здатність мала і воно недостатньо (або погано) розподіляє екзаменованих.

Технологічний цикл тестового екзамену

Під час створення тестових екзаменів використовуються відповідні технології, які мають охоплювати чотири основних компоненти: *банк тестових завдань, тест, тестування, аналіз*. Здебільшого тестові організації створюють спеціалізований професійний комплекс, що забезпечує реалізацію більшості потенційних можливостей, які надає метод тестування, та конкретні завдання, визначені специфікою того чи іншого тестового екзамену. Так, для проведення медичних ліцензійних тестових іспитів як засобу діагностики якості підготовки фахівців у галузі охорони здоров'я в Україні було розроблено комплекс комп'ютерних програм підтримки технології стандартизованого тестування, що отримав назву Standard Test.

Система Standart Test призначена для організації і супроводу процесу тестування, що здійснюється у бланковій формі з комп'ютерним скануванням результатів та обробкою у нормативні терміни даних в обсязі галузевого рівня. Система має розвинуті засоби обробки й аналізу результатів тестування і може бути задіяна на всіх етапах тестування. Основні режими роботи системи:

- формування банку тестових завдань;
- експертиза тестових завдань;
- створення тесту для іспиту;
- формування банку студентів, що проходять тестування;
- первинна обробка результатів тестування студентів;
- аналіз результатів тестування;
- розрахунок психометричних показників тесту і тестових завдань, а також підготовка статистичного сертифіката тесту;
- підготовка і друк звітів за результатами тестування тощо.

Робота системи здійснюється у вигляді замкненого технологічного циклу, наведеного на рис. 4.

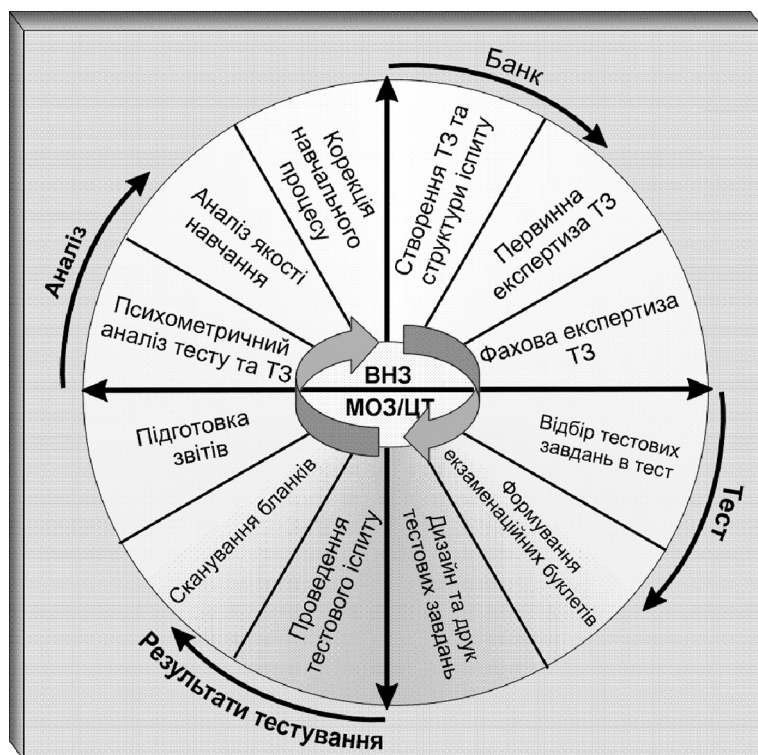


Рис. 4. Технологічний цикл стандартизованого тестування

Якщо ви бажаєте створити якісний інструмент вимірювання – якісний тест, що працюватиме, технологічний цикл має бути обов’язково замкнений. Спочатку необхідно створити банк тестових завдань відповідно до змісту, який діагностується. Отже, до банку ввійдуть тестові завдання з певною специфікацією, узгодженою зі змістовим доменом. Далі відповідно до матриці тесту формується тест із чітко визначеної кількості тестових завдань.

Отже, формування тесту відповідно до матриці змісту унеможливить введення до тесту неважливих за змістом тестових завдань.

Процедури вимірювання й оцінювання

Звертаємо увагу знову на цей момент, оскільки процедура вимірювання, безумовно, впливає на отримувані результати. Можна створити ідеальний якісний інструмент, але не досягти мети тестування через порушення процедури вимірювання (як це було, наприклад, в Україні при першій спробі шкільного атестаційного тестування у 1993–1994 рр., коли варіанти екзаменаційних буклетів

ще до екзамену опинилися в руках екзаменованих). Було також порушено й процедуру самого процесу тестування, оскільки в більшості шкіл, де воно відбувалося, викладачі допомагали учням. Отже, лише при дотриманні регламенту та процедур вимірювання можна одержати валідні результати.

Конкретизуємо означені вище поняття.

Процедура вимірювання – це категорія, що показує, яким чином і за яких умов здійснюється вимірювання. Вона визначає, як саме відбувається вимірювання. Під час тестування – це *процедура тестування*, що визначається умовами та вимогами вимірювання.

Процедура шкалювання визначає, як виміряна величина набуває кількісного визначення, складовими якого є такі елементи: *первинний бал, конвертація, шкала, оцінка*.

Для стандартизованих тестувань є певні обов'язкові вимоги щодо їх проведення. Уявіть ситуацію. Тестологи працюють над створенням якісного тесту. Але він – лише інструмент вимірювання. Необхідно також процедурно правильно усе виміряти й оцінити. Якщо процедура вимірювання порушуватиметься, то не буде отримано валідного результату.

У медичних вищих навчальних закладах України тестування відбувається одночасно у багатьох аудиторіях. Для проведення тестування залучають велику кількість людей – прокторів (представників Центру тестування), які мають здійснювати його у вищих навчальних закладах за однаковими вимогами. Їм відомо, що від дотримання процедури тестування залежить річна робота цілої організації.

Для контролю за процедурою тестування використовуються певні методики. Серед них – сканування результатів окремо по кожній аудиторії та застосування спеціальних статистичних критеріїв перевірки валідності результатів тестування [18]. Це дає змогу незалежно проконтролювати дотримання процедур тестування у кожній аудиторії, де відбувалося тестування, а отже, роботу кожного представника Центру тестування. Якщо за результатами аналізу статистичних показників, звітів і протоколів буде виявлено, що процедуру тестування порушено, то результати вважатимуться невалідними. Але під час проведення обов'язкових тестувань це має негативні наслідки лише для проктора та іміджу тестової організації.

В основу названих критеріїв покладено аналіз результатів за основними статистичними показниками. Для прикладу наведено висновки, зроблені за критеріями валідності результатів.

Критерії аналізу об'єктивності результатів

Аналіз валідності результатів базується на спеціально розроблених методиках зіставлення з критеріями аналізу об'єктивності результатів, в основі яких – оцінка міри відхилення від нормального закону результатів тестування та процедура їх нормалізації [17]. Ступінь відхилення від нормального закону визначається за такими параметрами, як асиметрія A_s і ексцес E_x , та встановленням їх статистичної достовірності за t -критерієм Стьюдента. Залежно від значення та знаку асиметрії або ексцесу результати можна поділити на п'ять загальних груп, що дають змогу визначити критерії прийняття рішення щодо валідності процесу вимірювання та наявності або відсутності статистично достовірної дії факторів, що впливають на результати вимірювання.

Критерії кількісного аналізу

Критерій I	“результати <u>об'єктивні</u> ” – валідний тест
Критерій II	“результати <u>завищені</u> – дії з боку викладачів або студентів”
Критерій III	“результати <u>завищені</u> ”
Критерій IV	“результати <u>занижені</u> ”
Критерій V	“результати <u>занижені</u> ” – порушення умов тестування

Критерій I

Якщо дані аналізу засвідчують, що $A_s = 0$, $E_x = 0$, то можна зробити такі **висновки**: результати вимірювання об'єктивні, а отже, і процес вимірювання об'єктивний, що є необхідною умовою його валідності.

Критерій II

Якщо дані аналізу засвідчують наявність від'ємної асиметрії та додатного ексцесу $A_s = “-”$, $E_x = “+”$, то можна зробити такі **висновки**: результати вимірювання є завищеними, тобто процес вимірювання є невалідним.

Причиною цього може бути невалідність тесту та моделі тестування. Така ситуація, крім названих випадків, найчастіше має місце в разі втручання екзаменаторів у процес тестування або за наявності в осіб, що тестуються, правильних відповідей або інших допоміжних матеріалів.

Критерій III

Якщо дані аналізу засвідчують наявність від'ємних асиметрії та ексцесу $A_s = “-”$, $E_x = “-”$, то можна зробити такі **висновки**: результати є завищеними за рахунок невалідності тесту.

Критерій IV

Якщо дані аналізу свідчать про наявність додатної асиметрії та від'ємного ексцесу: $A_s = "+"$, $E_x = "-"$, то можна зробити такі **висновки**: результати є заниженими за рахунок невалідності тесту.

Критерій V

Якщо дані аналізу засвідчують, що $A_s = "+"$, $E_x = "+"$, то можна зробити такі **висновки**: результати вимірювання є заниженими за рахунок невалідності тесту, форми діалогу, умов тестування.

Детальніше про критерії валідності результатів вимірювання див. у додатку 3.3.

1.2. Основні принципи та правила побудови тесту

*Класифікація навчальних цілей
Формулювання цілей навчання
Домени компетентності*

Класифікація навчальних цілей

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 результату навчання для оцінювання ступеня реалізації навчальних цілей.

Чіткому визначенню цілей навчання сприяє їх класифікація (таксономія). Американські педагоги та психологи створювали її впродовж двадцяти років, починаючи з кінця 50-х рр. минулого століття. Системи класифікації цілей навчання вимагали розробки цілей у трьох сферах (доменах): пізнавальній, емоційній і психомоторній [19]. Першою було розроблено систему класифікації у пізнавальній сфері (Bloom, 1956 р.). Наступним кроком було створення системи класифікації в емоційній сфері (Krathwohl, 1964 р.) і, нарешті, Simpson (1966 р.), Kibler (1970 р.) і Harrow (1972 р.) розробили систему класифікації цілей навчання в психомоторній області. Роботу було завершено на початку 80-х рр. Таксономія цілей навчання являє собою опис цілей навчання у термінах професійної діяльності (поведінки).

Ця загальноприйнята міжнародною освітянською спільнотою система цілей навчання є основою вибору цілей оцінювання, а отже, і методів педагогічної діагностики.

Когнітивний домен, або пізнавальна сфера

Пізнавальна сфера містить цілі, що описують знання та інтелектуальні здібності екзаменованих [19].

Класифікація пізнавальної сфери включає **шість класів цілей**, розміщених відповідно до складності: знання, розуміння, застосування, аналіз, синтез та оцінювання знань.

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.



Особистісна, або емоційна сфера

Емоційна сфера включає цілі, що стосуються емоцій, почуттів, цінностей, позицій та пов'язані зі змінами, які відбуваються в них [20].

Ця сфера поділяється на *п'ять основних класів*: сприймання, реагування, переконаність, селективність, індивідуальність.

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)

Психомоторна сфера пов'язана з рухами, маніпуляціями з матеріалом або об'єктом, а також з координацією [21].

Ця сфера є найскладнішою для класифікації, оскільки, за винятком найпростіших рефлексів, усі інші дії включають пізнавальні та емоційні компоненти. Безумовно, класифікація психомоторної сфери особливо корисна для викладачів професійних шкіл (медичної зокрема), де формування професійних вмінь та практичних навичок є необхідною умовою для успішної роботи у сфері професійної діяльності. 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.

Наведена класифікація цілей успішно використовується в усіх розвинених країнах, що дозволяє застосовувати сучасні методи вимірювання успішності навчання, які дидактично узгоджені з метою навчання та контролем знань.

Формулювання цілей навчання

Розглянемо більш детально основні вербальні моделі, що використовуються для опису і формулювання цілей освіти та навчання різного рівня [15].

Опис кінцевих цілей навчання в термінах діяльності ґрунтується на чіткому визначенні рівня сформованості та на дотриманні вимоги узгодженості між

Когнітивний домен (сфера) – рівні класифікації

Таксономічний рівень	Класифікація
1.00	Знання (запам'ятовування попередньо вивченого матеріалу)
2.00	Розуміння (розуміння змісту вивченого матеріалу)
3.00	Застосування (вміння використати вивчений матеріал у нових ситуаціях)
4.00	Аналіз (уміння розчленувати ціле на складові елементи)
5.00	Синтез (уміння створити ціле з частин)
6.00	Оцінювання (вміння визначити цінність та придатність певних засобів для досягнення певної мети)

предметом діяльності і дієсловом, яким сама дія визначається. Наведемо нижче терміни (дієслово і пряме доповнення – дія і предмет дії), які відповідають прийнятій таксономії цілей навчання.

Пізнавальний (когнітивний) домен має шість таксономічних рівнів. Під час створення галузевих стандартів як системи ієрархічно підпорядкованих нормативних документів чітко класифікуються загальні (кінцеві) та специфічні (проміжні або предметні) цілі навчання. Це, зокрема, передбачає, що рівень сформованості загальних (кінцевих) цілей має починатися з 2.00 – розуміння вивченого матеріалу, тоді як рівень 1.00 – запам'ятовування вивченого матеріалу – є базисним при формулюванні предметних цілей.

Використовуючи відповідні вербальні принципи та моделі при формулюванні кінцевих і предметних цілей навчання, до уваги треба брати ключові слова. При формулюванні цілей навчання необхідно, по-перше, чітко класифікувати таксономічний рівень, по-друге, визначити предмет дії, по-третє, вибрати відповідні дієслова і пряме доповнення.

Правила формулювання цілей

Кожна мета (принаймні вузькоспеціальна) може бути сформульована у певних межах деякої діяльності (поведінки). Цілі діяльності визначають, які саме дії повинен виконувати учень (студент) і що саме контролюється під час оцінювання досягнення поставленої ним мети. Слід зазначити, що вираз “цілі поведінки або діяльності” студентів не слід плутати з “поведінкою” з погляду психології (до цього призводить широке тлумачення терміна “виконання” замість “поведінка”).

При застосуванні суб'єктно-діяльного підходу, на якому ґрунтується система стандартів вищої освіти, для оцінювання професійних вмінь студентів та визначення стратегії навчання необхідним є формулювання цілей з точки зору “поведінки або професійної діяльності”. Це дає змогу сформулювати цілі більш

точно, включаючи умови, необхідні для процесу формування певної поведінки, та критерії, що слід враховувати під час оцінювання знань студентів.

Прихильники терміна “цілі поведінки або професійної діяльності” стверджують, що його використання є дієвою допомогою під час розробки навчальних планів; оптимізує процес навчання та сприяє досягненню студентами поставленої мети. Викладач краще розуміє цілі, сформульовані в рамках професійної діяльності, і тому знає, як краще організувати процес навчання. На цих принципах побудовано педагогічні системи професійної медичної освіти більшості розвинених країн. Тому ми маємо будувати сучасну, інтегровану в світову, педагогічну систему освіти саме на таких методологічних засадах, дотримуючись саме таких дидактичних принципів та закономірностей.

Зупинимося на п’яти *основних правилах* формулювання цілей освіти і навчання.

Правило 1. Правильно вибирати дієслово для опису професійної поведінки. Крім того, допустимим є використання наказової форми цих дієслів.

Правило 2. Вказувати на умови, за яких студент демонструватиме здатність виконати завдання, а також визначати критерії або стандарти, згідно з якими здійснюватиметься оцінювання знань.

Правило 3. Цілі необхідно формулювати за допомогою однозначних термінів.

Правило 4. Формулювання цілей має бути чітким, не переважаним зайвими словами.

Правило 5. Цілі мають бути унітарними, тобто кожне положення повинно належати до одного процесу [15].

Домени компетентності

Ефективне досягнення кінцевого результату можливе лише за умови чіткого його визначення, опису вимог та орієнтації навчального процесу на досягнення визначених вимог. Одним із сучасних підходів до встановлення освітніх вимог є визначення доменів компетентності, які є кінцевими результатами навчання, та їх подальша специфікація у вигляді конкретних освітніх цілей, що є кінцевими або проміжними цілями навчання.

Термін “компетентність” як наукове та практичне поняття увійшов до фахового словника педагогів в усьому світі порівняно недавно та досі потребує конкретизації визначення. Протягом короткого часу він змінювався від надзвичайно вузького розуміння у вигляді переліку конкретних вмінь та навичок, які можна продемонструвати та перевірити, до широкого розуміння

підготовленості учня до життя, а студента – до роботи. Від використання терміна компетентності відмовлялися, але зараз до нього повертаються знов.

На сучасному етапі новим поштовхом до поширення та уточнення змісту терміна “компетентність” у Європі та Америці стало визнання неготовності учнів до активного життя в умовах інформаційного суспільства. Таке суспільство потребує все більшого обсягу інформації, яку опанувати в межах шкільної програми вже неможливо, і тому в учнів необхідно формувати не лише знання, а й ключові (основні, базові) компетентності.

Сучасне визначення поняття “компетентність” дали два окремих європейських проекти у сфері ключових компетентностей.

Завданням проекту Організації економічного співробітництва і розвитку (OECD) з визначення ключових компетентностей було обґрунтування нормативних, теоретичних і концептуальних основ для визначення та відбору ключових компетентностей [22]. У межах цього проекту було прийнято такі визначення.

Компетентність – це здатність успішно задовольняти особисті чи суспільні вимоги і потреби або виконувати завдання та проводити діяльність. Компетентність складається з когнітивних і некогнітивних компонентів.

Ключові компетентності – це компетентності, які є важливими для багатьох сфер життя, вони є складовою успішного життя особистості і добре функціонуючого суспільства. Вибір та визначення ключових компетентностей залежать від суспільних цінностей.

Таке визначення компетентності ґрунтується на функціональному підході, за якого людина є компетентною не сама по собі, а по відношенню до реалізації зовнішніх функцій. Зовнішній прояв залежить від внутрішньої індивідуальної структури компетентності, яка є конгломератом знань, когнітивних вмінь, маніпуляційних навичок, мотивації, моральних і етичних цінностей, інтересів, ставлень, емоцій та інших соціально-психологічних компонентів, що мобілізуються разом для ефективною дії. Тобто внутрішня структура містить когнітивні, психомоторні, емоційні, мотиваційні, соціальні, етичні та поведінкові компоненти.

Проект Європейської комісії дав таке структурне визначення компетентності та структурно-функціональне визначення ключових компетентностей [23].

Компетентність – це гібридна характеристика, що складається з комбінації знань, вмінь і ставлень. Компетентність може набуватися за різних умов: формально і неформально, цілеспрямовано або ненавмисно.

Ключові компетентності – це багатофункціональна сукупність знань, вмінь і ставлень. Вона може бути застосована в різних умовах і її потребують всі індивідууми для особистої діяльності (розвитку) та інтеграції у суспільство,

а також для працевлаштування. Її необхідно набути до моменту завершення обов’язкової шкільної освіти або навчання, і вона є основою для навчання протягом життя.

Мета освіти у вигляді досягнення компетентності відрізняється від мети освіти у вигляді оволодіння знаннями та навичками, оскільки компетентність не дорівнює *окремим* знанням, вмінням і навичкам, а є значно ширшим поняттям. Водночас компетентність не означає відмови від опанування знаннями або вміннями чи формування певних ставлень, а означає їх підпорядкованість більш широкій, глобальній меті.

Співвідношення внутрішньої структури компетентності та її орієнтованість на зовнішні вимоги на тлі конкретного контексту або умов добре проілюстровано на рис. 5 [22].

Властивості та характеристики людини, такі як знання, когнітивні вміння, практичні навички, ставлення, емоції, цінності, етичні погляди, мотивації тощо, є основою **потенціалу компетентності**. Залежно від характеру зовнішніх вимог і контексту частина цього потенціалу компетентності людини мобілізується для активної дії – виникає **проявлена компетентність**. За різних умов, у різному контексті у людини мобілізуються, активізуються різні властивості. Навпаки, для реалізації однакових вимог у різних однаково компетентних індивідуумів можуть мобілізуватися різні “набори” властивостей для реалізації компетентності.

З наведеним визначенням пов’язано поняття “домен компетентності”.

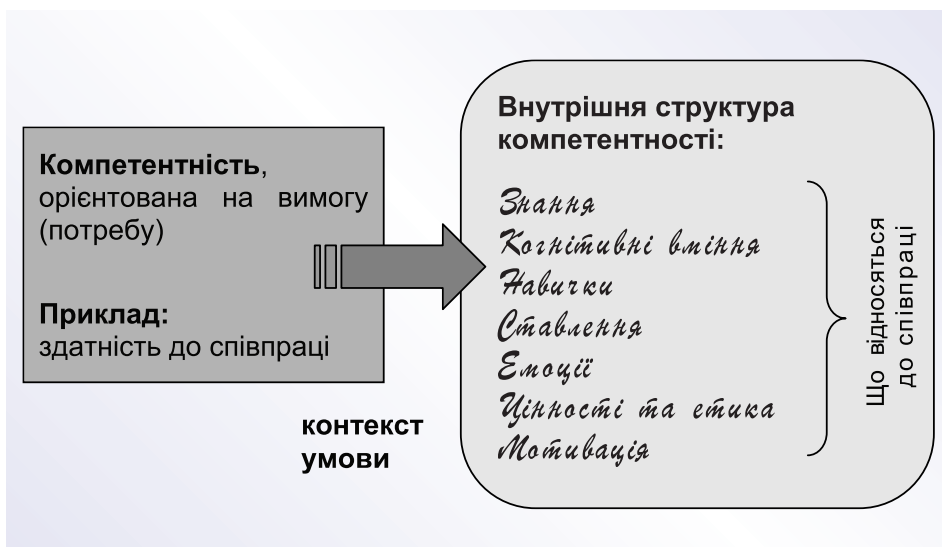


Рис. 5. Внутрішня структура компетентності

Домен компетентності – це сукупність порівняно однорідних компетентностей, спрямованих на розв’язання порівняно однорідних задач.

Домени компетентності можуть бути визначені та структуровані по-різному: за однорідністю завдань діяльності, за однорідністю компетентностей, за однорідністю предметної сфери тощо. Раніше традиційним підходом було структурування за предметами, галузями знань (предметні домени). Але у такий спосіб складно сформувати міжпредметні вміння і знання, тому нині все частіше як кінцеві цілі навчання визначаються інтегровані компетентності, що мають набуватися під час опанування різних предметів та поза формальним навчанням.

Прикладом інтегрованих міжпредметних domenів компетентності для обов’язкової шкільної освіти є переліки *ключових компетентностей, що пропонуються європейськими проектами*.

Ключові компетентності проекту Європейської комісії з освіти:

- спілкування рідною мовою;
- спілкування іноземною мовою;
- володіння інформаційно-комп’ютерними технологіями;
- здатність до кількісного мислення та компетентності з математики, природничих наук і технологій;
- підприємництво;
- міжособистісні та громадянські компетентності;
- вміння вчитися (навчання тому, як вчитися);
- володіння загальною культурою на необхідному рівні.

Ключові компетентності для успішного життя та добре функціонуючого суспільства, що пропонуються проектом OECD, можна подати у такому вигляді.

Автономні дії передбачають:

- здатність захищати та відстоювати власні права, інтереси та потреби;
- здатність до відповідальності та обмеження;
- здатність будувати та реалізовувати життєві плани та особисті проекти;
- здатність діяти в певних умовах і бачити більш широку картину.

Використання інтерактивних засобів (тобто засобів взаємодії зі світом) передбачає здатність користуватися:

- мовою, символами та текстом інтерактивно;
- знаннями та інформацією інтерактивно;
- новими технологіями інтерактивно.

Функціонування в соціально-гетерогенних спільнотах передбачає здатність:

- добре ставитися до інших;
- до співпраці;
- керувати та вирішувати конфлікти.

Як правило, домени компетентності визначають експерти. При визначенні доменів компетентності для професійної освіти враховуються структура професійної діяльності, перспективи розвитку відповідної науки, аналізуються статистика галузі, типові помилки тощо.

Домени компетентності визначаються кінцевими цілями навчання на певному етапі (наприклад, на етапі закінчення набуття середньої або вищої освіти).

Доцільно розрізняти домен компетентності, що є очікуваним результатом освіти, та домен навчання (або домен навчального плану), що є змістом освіти, вже відібраним і включеним до програми навчання.

Для ефективного опанування доменами компетентності вони мають бути чітко візуалізовані для учнів та вчителів. Навколо них мають будуватися навчання та оцінювання, має відслідковуватися також прогрес у їх досягненні. В умовах перевантаження інформацією при відборі змісту освіти (знання та інформація, що вивчаються, вміння і ставлення, що формуються, тощо) пріоритети мають віддаватися змісту, що сприяє загальній меті опанування доменами компетентності.

Відібраний зміст освіти структурується у термінах навчальних цілей відповідно до **таксономії освітніх цілей**. Засоби педагогічного оцінювання розробляються відповідно до структурованого та визначеного в термінах навчальних цілей змісту освіти.

На сучасному етапі розроблено різноманітні ефективні та надійні засоби педагогічного оцінювання інформаційного змісту освіти та *когнітивних* вмінь (для будь-якого рівня таксономії Блума), порівняно ефективні – для оцінки *маніпуляційних* і комплексних навичок, значно менш ефективні та надійні методи – для оцінювання *емоційного, мотиваційного, вольового та ціннісного* доменів. Класифікацію для останніх зазначених доменів на сьогодні не розроблено.

1.3. Створення тестових завдань

Правила створення тестових завдань

Формати тестових завдань

Технологія створення тестових завдань

Створення тестових завдань різних когнітивних рівнів

З точки зору вимірювання тестові завдання можна поділити на дві загальні категорії: ті, в яких екзаменовані *вибирають* правильні чи найкращі відповіді, базуючись на інформації, поданій у тестовому завданні, та ті, в яких вони мають *надати (створити, згенерувати)* відповіді самостійно. Когнітивні здатності, що потрібні для відповіді на завдання з двох зазначених категорій, є різними незалежно від змісту завдання. Завдання з першої категорії називаються запитаннями багатовибіркового типу (MCQ — multiple-choice question). До другої категорії належать завдання з короткою або довгою відповіддю, есе.

У цьому розділі детально розглянуто формати тестових завдань багатовибіркового типу, які протягом більш ніж сторічної практики застосування методу тестування набули найбільш широкого використання. Найпоширеніші завдання складаються з двох частин:

- **умови**, що описує певну проблему і ставить завдання перед екзаменованим;
- **списку варіантів відповідей**, серед яких як мінімум один є правильною чи найкращою відповіддю, а решта — дистрактори — є неправильними.

Основні питання, що взаємопов'язані між собою і безпосередньо впливають на якість тестових завдань, — це принципи або правила створення тестових завдань, формати завдань, технологія створення та технічні дефекти. Зупинимося на кожному з цих питань окремо.

Правила створення тестових завдань

Перш ніж розглядати формати тестових завдань, зупинимося на основних правилах, якими необхідно керуватися під час створення тестових завдань. Наведемо десять основних.

1. Кожне тестове завдання має оцінювати досягнення важливої та суттєвої освітньої цілі. Слід уникати перевірки тривіальних або надмірно вузько-спеціальних знань.
2. Кожне тестове завдання має перевіряти відповідний рівень засвоєння знань, в тому числі вищі когнітивні рівні.

3. Умова має містити чітко сформульоване завдання. Завдання має фокусуватися на одній проблемі.
4. Варіанти відповідей мають бути гомогенними (однорідними).
5. Усі дистрактори мають бути вірогідними (правдоподібними).
6. Інформація, що міститься в одному тестовому завданні, не повинна давати відповідь на інше тестове завдання.
7. Не рекомендується використовувати як правильну відповідь чи дистрактор фразу “все з вищевказаного”.
8. Не рекомендується використовувати як правильну відповідь чи дистрактор фразу “нічого з вищевказаного”.
9. Умова по можливості має бути сформульована позитивно.
10. Необхідно уникати при формулюванні умов підказок типу:
 - граматична невідповідність між умовою та варіантами відповідей;
 - повторення у правильній відповіді слів з умови;
 - використання прикладів з підручника чи лекції як тестових завдань;
 - найдовша правильна відповідь;
 - найдетальніша правильна відповідь;
 - дистрактори, що виключають один одного;
 - ситуації, коли одне тестове завдання є підказкою для другого.

Правила написання умови

Умова — це стимул для відповіді, яка описує певну проблему і ставить завдання перед екзаменованим. Умова повинна допомогти екзаменованому чітко уявити, поставлену перед ними проблему. Умова може містити лише завдання або складатися із вступної інформації та запитання, пов’язаного з наведеною інформацією.

Умова може подаватися у формі запитання, у наказовій формі або у формі незавершеного твердження. **Рекомендується використовувати форму запитання або наказову форму**, які є легшими для екзаменованих і ставлять перед ними більш чітке завдання. При незавершених твердженнях автори тяжіють до написання дистракторів, що не належать до навчальної цілі, на яку спрямовано завдання.

Невдало	Краще
Увертюра — це...	З якою метою використовується увертюра в опері?

Якщо все ж таки використовується формат незавершеного твердження, пропуск в останньому не повинен бути на початку або в середині, його треба розмістити в кінці фрази. Крім того, навіть у незавершеному твердженні умова має бути “завершена” з точки зору змісту, щоб на неї можна було відповісти, не дивлячись на список варіантів відповідей.

Невдало	Краще
У теорії ймовірностей _____ – це будь-який факт, що внаслідок досліджу може відбутися або не відбутися. <i>Правильна відповідь: подія</i>	Як у теорії ймовірностей називають будь-який факт, що внаслідок досліджу може відбутися або не відбутися?
	АБО
	У теорії ймовірностей будь-який факт, що внаслідок досліджу може відбутися або не відбутися, називають _____. <i>Правильна відповідь: подія</i>

Включайте в умову центральну ідею завдання та максимум інформації, що міститься у тестовому завданні. Умова має бути відносно довгою, а відповіді – короткими. В іншому випадку для здійснення вибору екзаменовані повинні декілька разів перечитувати довгі відповіді, що ускладнює та сповільнює процес. Необхідно, щоб на тестове завдання можна було відповісти, не читаючи варіантів відповідей (правило “закритих варіантів відповідей”).

Невдало	Краще
Великі – це підприємства, в яких: <i>Правильна відповідь: середньо-облікова чисельність працюючих за звітний період перевищує 1000 осіб, а обсяг валового доходу від реалізації продукції за рік перевищує суму еквівалента EUR 5 млн. [інші відповіді за такою ж структурою описують середнє та мале підприємства].</i>	Вкажіть, до якої категорії за розміром згідно із законом належать підприємства, в яких середньооблікова чисельність працюючих за звітний період перевищує 1000 осіб, а обсяг валового доходу від реалізації продукції за рік перевищує суму еквівалента EUR 5 млн. <i>Правильна відповідь: великі.</i>

Умова може бути досить детальною, але необхідно уникати багатослів'я (зайвих слів), “мішури” (стороннього матеріалу, що не стосується проблеми) та каверз (інформації, що свідомо заплутує екзаменованих).

Невдало	Краще
Приклад штучного використання складної умови	
Хворий скаржиться на біль у горлі, підвищення температури, слабкість. При огляді мигдалики вкриті рясним світлим нальотом. Який збудник викликає захворювання ангіною?	Який збудник викликає захворювання ангіною? <i>[текст на початку не потрібний для відповіді на запитання]</i>

Штучне ускладнення умов слід відрізнити від ситуацій, коли довгі та складні умови є доречними. Такими ситуаціями є: тестування з метою оцінки навичок читання; тестові завдання, в яких відповідь базується на описі ситуації, як у медицині, завдання, що перевіряють уміння екзаменованого виділяти суттєву інформацію, тощо.

Приклад природного використання складної умови
Хворий скаржиться на біль у горлі, підвищення температури, слабкість. При огляді мигдалики вкриті рясним світлим нальотом. Поставте найбільш імовірний діагноз

Формулюйте умови позитивно. Уникайте негативних формулювань, які вимагають протилежної, порівняно з більшістю тестових завдань, дії (вибір неправильного, гіршого), та є складними для розуміння екзаменованими. Особливо заплутує застосування негативної умови у форматі TF.

Невдало	Краще
Визначте, що НЕ належить до переваг формату А	Вкажіть недолік формату А
Основним поняттям аксіологічного аналізу НЕ є поняття “цінності”. <i>Правильна відповідь: ні</i> [правильність відповіді спірна]	Основним поняттям аксіологічного аналізу є поняття “цінності”. <i>Правильна відповідь: так</i>

Не базуйте тестове завдання закритого типу на думці, погляді, суб’єктивній оцінці. Такі завдання можуть мати декілька відповідей, які можна захистити і не мати однозначної правильної відповіді. У завданні мають бути певні критерії, що дають змогу екзаменованому зробити обґрунтований вибір.

Невдало	Краще
Хто є найвидатнішою людиною у світі?	Хто, за рейтингом журналу People, є найвидатнішою людиною у світі?
Політичну проблему НЕ СЛІД плутати з: <i>Правильні відповіді: подією, причинами, обставинами, симптомами, наслідками</i> Неправильні відповіді: ЗМІ, політичним рішенням, мат. моделлю	[Запитання некоректне. Наприклад, хіба проблему слід плутати з іншими поняттями?]

В умові не повинно бути граматичних та інших підказок для вибору або виключення будь-якої відповіді. Використання абсолютних термінів (“ніколи”, “завжди”) не рекомендується.

Правила написання варіантів відповідей

Написання *варіантів відповідей* є найскладнішим завданням під час створення тестового завдання. Серед відповідей мінімум одна є правильною, або найкращою, решта – дистрактори – є неправильними.

Доцільно створювати щонайменше 4–5 варіантів відповідей. Усі дистрактори мають бути правдоподібними і однорідними. Дистрактори, які не є вірогідними і однорідними, не працюють, збивають екзаменованих; їх не треба включати до складу завдання. Не варто штучно збільшувати кількість дистракторів за рахунок невірогідних.

При підборі дистракторів доцільно використовувати поширені помилки, хибні уявлення, об’єкти, що відповідають лише частині характеристик, наведених в умові, тощо. Водночас у дистракторах **не повинно бути каверзної, фальшивої та хибної інформації**. Використовуйте правильні твердження, але такі, що не нале-

жать до даного контексту. Наприклад, при завданні на встановлення наслідків конкретної ситуації може бути чотири типи відповідей за ступенем правильності наслідку та його зв'язку з наведеною конкретною ситуацією:

- 1) реальний наслідок, пов'язаний з конкретною ситуацією (правильна відповідь);
- 2) реальний наслідок, не пов'язаний з ситуацією;
- 3) нереальний або неправильно описаний наслідок, пов'язаний з конкретною ситуацією;
- 4) нереальний або неправильно описаний наслідок, не пов'язаний з конкретною ситуацією.

Використовуйте дистрактори з тим самим ступенем “технічності”, “науковості”, що й правильна відповідь. Для екзаменованих із недостатніми знаннями такі дистрактори виглядають однаково привабливими.

Уникайте фразування правильної відповіді цитатою з підручника або стереотипним виразом. Оскільки дистрактори автор має вигадати самотужки, фразування відрізняється, й екзаменовані можуть впізнати правильну відповідь зовнішньо.

При однакових запитаннях від варіантів відповідей залежить складність тестового завдання.

Надзвичайно складне суб'єктивне завдання	Легке завдання
<p><i>Вкажіть колір океану на мапі:</i></p> <p>A Блакитний B Лазурний C Синій D Синьо-зелений E Ультрамариновий</p>	<p><i>Вкажіть колір океану на мапі:</i></p> <p>A Жовтий B Синій C Червоний</p>

Якщо можливо, **варіанти відповідей треба розміщувати системно** (в логічному, в алфавітному порядку, у порядку збільшення або зменшення чисел тощо).

Невдало	Краще
<p><i>Дні циклу, в які запліднення є найбільш імовірним:</i></p> <p>A 11–17-й B 25–29-й C 3–7-й D 8–11-й</p>	<p><i>Дні циклу, в які запліднення є найбільш імовірним:</i></p> <p>A 3–7-й B 8–11-й C 12–17-й D 25–29-й</p>
<p><i>Розщеплення поживних речовин у травному тракті людини завершується у:</i></p> <p>A Дванадцятипалій кишці B Тонкому кишечнику C Товстому кишечнику D Шлунку</p>	<p><i>Розщеплення поживних речовин у травному тракті людини завершується у:</i></p> <p>A Шлунку B Дванадцятипалій кишці C Тонкому кишечнику D Товстому кишечнику</p>
<p><i>Вказати, яка температура поверхні характерна для червоних зірок (°C):</i></p> <p>A 2 000–3 000 B 6 000–7 000 C 12 000 D 20 000</p>	<p><i>Вказати, яка температура поверхні характерна для червоних зірок (°C):</i></p> <p>A 2 000–3 000 B 6 000–7 000 C 12 000–13 000 D 20 000–25 000</p>

Відповіді мають бути **незалежними одна від одної** та не перетинатися між собою. Відповіді, що перетинаються, рідко бувають правильними, і це слугує підказкою для екзаменованих.

Невдало	Краще
<p><i>Вкажіть правильні критерії адекватності структури проекту і системи контролю його реалізації (знайти 2 правильні відповіді):</i></p> <p>A Ступінь задоволеності керівника проекту системою планування і контролю</p> <p>B Ступінь незадоволеності керівника проекту системою планування і контролю</p> <p>C Ступінь задоволеності керівника проекту системою мотивації персоналу</p> <p>D Ступінь незадоволеності керівника проекту системою мотивації персоналу</p> <p>E Ступінь задоволеності робочої групи організаційною структурою</p> <p>F Ступінь незадоволеності робочої групи організаційною структурою</p> <p>G Ступінь задоволеності робочої групи результатами роботи менеджера проекту</p> <p>H Ступінь незадоволеності робочої групи результатами роботи менеджера проекту</p> <p>[кожна пара дистракторів є різними ступенями прояву одного й того ж явища, зробити між ними вибір неможливо]</p>	<p><i>Вкажіть правильні критерії адекватності структури проекту і системи контролю його реалізації (знайти 2 правильні відповіді):</i></p> <p>A Задоволеність керівника проекту системою планування і контролю</p> <p>B Задоволеність керівника проекту системою мотивації персоналу</p> <p>C Задоволеність робочої групи організаційною структурою</p> <p>D Задоволеність робочої групи результатами роботи менеджера проекту</p>
<p><i>Унаслідок зміни розміру державних витрат (G) і кількості грошей в обігу (M) відсоткова ставка (i) залишилася незмінною, а дохід (Y) зменшився. Можна визначити, що:</i></p> <p>A (G) та (M) зросли</p> <p>B (G) зменшилося, (M) зросло</p> <p>C (G) зросло, (M) зменшилося</p> <p>D (G) та (M) або одночасно зросли, або одночасно зменшилися</p> <p><i>Правильна відповідь: B</i> [дистрактор D є більш повним, ніж дистрактор A, тобто вони <i>перетинаються</i>]</p>	<p><i>Унаслідок зміни розміру державних витрат (G) і кількості грошей в обігу (M) відсоткова ставка (i) залишилася незмінною, а дохід (Y) зменшився. Як змінилися розміри державних витрат (G) і кількості грошей в обігу (M)?</i></p> <p>A (G) та (M) зросли</p> <p>B (G) зменшилося, (M) зросло</p> <p>C (G) зросло, (M) зменшилося</p> <p>D (G) та (M) зменшилися</p> <p><i>Правильна відповідь: B</i></p>

Відповіді мають бути **однорідними за змістом**. Усі відповіді мають належати до однієї навчальної мети. Якщо перед екзаменованим стоїть завдання вибрати *кращу* відповідь або *кращі* відповіді, необхідно ці відповіді порівняти між собою. Це можливо лише тоді, коли вони належать до однієї категорії. Не слід додавати штучних відповідей. Якщо частина відповідей охоплює всі можливі варіанти, а решта додається для збільшення кількості, ці додані відповіді відкидаються екзаменованими відразу. Така ситуація є логічною підказкою, яка зменшує кількість відповідей, серед яких екзаменованим вибирає.

Невдало	Краще
<p>Доповніть фразу: "Україна займає площу 603,7 тис. кв. км, і..."</p> <p>A Є однією з найбільших за розмірами держав у Європі</p> <p>B Займає перше за розмірами місце в Європі</p> <p>C Займає 5,7% території Європи</p> <p>D Перевищує за розмірами Францію</p>	<p>Враховуючи, що Україна займає площу 603,7 тис. кв. км, укажіть її місце за розмірами серед країн Європи.</p> <p>A Перше</p> <p>B Друге</p> <p>C Третє</p> <p>D Четверте</p> <p>E П'яте</p>
<p>Кодекс законів про працю регулює:</p> <p>A Трудові відносини підприємств державної власності</p> <p>B Трудові відносини підприємств недержавної власності</p> <p>C Трудові відносини на підприємствах регулюються постановами Кабінету Міністрів</p> <p>D Трудові відносини працівників усіх підприємств незалежно від форми власності</p> <p>[неоднорідна відповідь С; логічна неузгодженість варіанта С з умовою; дієслова; частину інформації з дистракторів можна перенести в умову]</p>	<p>Трудові відносини працівників підприємств якої форми власності регулює Кодекс законів про працю?</p> <p>A Державної</p> <p>B Недержавної</p> <p>C Будь-якої</p>

Відповіді мають бути **короткими та простими за структурою**. Слова, що повторюються в усіх відповідях, слід додавати до умови. До відповідей бажано не додавати дієслів, що ускладнюють сприйняття та розуміння. Оптимально, щоб відповіді містили іменники, прикметники, числівники як окреме слово, словосполучення, просту комбінацію слів. Використовувати речення у відповідях не рекомендується.

Невдало	Краще
<p>Вирізнити причини більшої суворості клімату Антарктиди порівнянно з Арктикою:</p> <p>A В Арктиці, на відміну від Антарктиди, клімат пом'якшує Північний Льодовитий океан</p> <p>B Абсолютні висоти території в Арктиці більші, ніж в Антарктиді</p> <p>C У європейському секторі Арктики пом'якшує клімат Північноатлантична течія</p> <p>D В Антарктиді, на відміну від Арктики, дують стокові вітри, які формують клімат цієї південної області</p>	<p>Вкажіть причини більшої суворого клімату Антарктиди порівнянно з Арктикою:</p> <p>A Відсутність Північного Льодовитого океану</p> <p>B Менші абсолютні висоти території</p> <p>C Відсутність Північноатлантичної течії</p> <p>D Наявність стокових вітрів</p>
<p>Галузева угода укладається між:</p> <p>A Відповідним галузевим міністерством та керівниками підприємств галузі</p> <p>B Відповідним галузевим міністерством та Кабінетом Міністрів України</p> <p>C Відповідним галузевим міністерством та Центральною радою профспілок України</p> <p>D Відповідним галузевим міністерством та Центральною радою профспілки галузі</p> <p>Правильна відповідь: D</p>	<p>З ким відповідне галузеве міністерство укладає галузеву угоду?</p> <p>A З керівниками підприємств галузі</p> <p>B З Кабінетом Міністрів України</p> <p>C З Центральною радою профспілок України</p> <p>D З Центральною радою профспілки галузі</p>

Невдало	Краще
<p><i>Під оптимістичним сценарієм розуміють:</i></p> <p>A Прогноз, що передбачає розвиток фірми, якщо ситуація на ринку складеться найбільш несприятливим чином, що, звісно, призведе до зниження прибутків</p> <p>B Прогноз, що передбачає найбільш вірогідне прогнозування зміни ситуації на ринку і відповідний найбільш реальний розвиток фірми</p> <p>C Прогноз, що передбачає розвиток бізнесу, якщо ринкова ситуація складеться найбільш сприятливим чином і фірма отримає найбільший прибуток</p> <p><i>Правильна відповідь: C</i></p>	<p><i>Яку ринкову ситуацію називають оптимістичним сценарієм щодо прогнозу розвитку фірми?</i></p> <p>A При найбільш несприятливій ситуації і відповідному зниженні прибутків</p> <p>B При найбільш імовірній ситуації і відповідному найбільш реальному розвитку</p> <p>C При найсприятливішій ситуації і відповідному найбільшому прибутку</p> <p><i>Правильна відповідь: C</i></p>

Відповіді мають бути **подібними за зовнішніми ознаками**, структурою, стилістикою тощо, зокрема подібною має бути довжина відповідей. Типовою помилкою є створення найдовшої, детальної, конкретної, найповнішої правильної відповіді, до якої додаються коротші дистрактори. Інколи довжина правильної відповіді становить абзац, тоді як дистрактори містять одне-два слова.

Невдало	Краще
<p><i>Політична система – це:</i></p> <p>A Сукупність відповідних інститутів, що беруть участь у формуванні, розподілі та здійсненні державної влади з метою управління суспільними процесами та представництва інтересів соціальних груп населення</p> <p>B Специфічна модель побудови суспільства, що володіє монополією на владу в масштабах усього суспільства</p> <p>C Організація, що забезпечує політичне і адміністративно-державне управління суспільними процесами</p> <p><i>Правильна відповідь: B</i></p>	<p><i>Що таке сукупність відповідних інститутів, що беруть участь у формуванні, розподілі та здійсненні державної влади з метою управління суспільними процесами та представництва інтересів соціальних груп населення?</i></p> <p>A Державні установи</p> <p>B Органи виконавчої влади</p> <p>C Політична система</p> <p><i>Правильна відповідь: C</i></p>
<p><i>Систему судів загальної юрисдикції складають:</i></p> <p>A Місцеві, апеляційні, вищі спеціалізовані суди</p> <p>B Верховний суд, Конституційний суд, Апеляційний суд</p> <p>C Касаційний суд, місцеві, апеляційні, вищі спеціалізовані суди, Верховний суд, Конституційний суд, Апеляційний суд</p> <p><i>Правильна відповідь: C</i></p>	<p>[Якщо зміст такого тестового завдання є важливим, доцільно замінити формат на формат X]</p>
<p><i>Якщо текст документа друкується з одного боку аркуша, номери проставляються:</i></p> <p>A Посередині верхнього поля аркуша арабськими цифрами на відстані не менше ніж 10 мм від краю</p> <p>B У правому верхньому кутку</p> <p>C У лівому верхньому кутку</p> <p><i>Правильна відповідь: A</i></p>	<p><i>Якщо текст документа друкується з одного боку аркуша, номери проставляються:</i></p> <p>A Посередині верхнього поля</p> <p>B У правому верхньому кутку</p> <p>C У лівому верхньому кутку</p>

Відповіді мають **відповідати умові граматично, стилістично та логічно**. Невідповідність найчастіше спостерігається при застосуванні незавершеної форми твердження в умові.

Невдало	Краще
<p><i>Обігові активи використовуються підприємством протягом:</i></p> <p>A Тривалого періоду B Залежно від періоду обіговості C Залежно від складу обігових активів D Одного року або одного операційного циклу</p> <p><i>Правильна відповідь: D</i> [граматична неузгодженість варіантів B та C; подвійна і найбільш конкретна правильна відповідь]</p>	<p><i>Як довго підприємство використовує обігові активи?</i></p> <p>A Тривалий період B Залежно від періоду обіговості C Залежно від складу обігових активів D Впродовж року або одного операційного циклу</p> <p><i>Правильна відповідь: D</i></p>
<p><i>Комісійна форма оплати праці здійснюється:</i></p> <p>A Встановлення комісією розміру премії до тарифної ставки B Встановлення комісією розміру премії до посадового окладу C Після пред'явлення товарів на комісію D Залежно від вартості здійсненої операції через попередньо визначений відсоток</p> <p><i>Правильна відповідь: D</i> [стилістична неузгодженість варіантів A і B; найдовша правильна відповідь]</p>	<p><i>У чому полягає сутність комісійної форми оплати праці?</i></p> <p>A У встановленні комісією розміру премії до тарифної ставки B У встановленні комісією розміру премії до посадового окладу C У сплаті при пред'явленні товарів на комісію D У сплаті попередньо визначеного відсотка від вартості здійсненої операції</p> <p><i>Правильна відповідь: D</i></p>
<p><i>Як скорочуються назви наукових ступенів?</i></p> <p>1. Кандидат фізико-математичних наук A К. ф.-м. н. B К. ф. н. 2. Кандидат філологічних наук C К. філ. н. 3. Кандидат філософських наук [логічна підказка]</p>	

Уникайте використання фрази “**все з вищезазначеного**” або її еквівалентів, що автоматично призводить до появи відповідей, що перетинаються. Ця фраза водночас є підказкою, якщо екзаменований визначить хоча б 2 відповіді, як правильні, і причиною для неправильної відповіді, якщо екзаменований зупиниться на першій правильній відповіді, яку побачить. **Уникайте фрази “нічого з вищезазначеного”**, правильність якої завжди можна відстояти. Обидві фрази, по суті, змінюють формат А (формат найкращої або правильної відповіді) на формат Х (будь-яка кількість правильних відповідей або їх відсутність). **Уникайте абсолютних слів “завжди”, “ніколи” та розпливчастих “інколи”, “ймовірно”**. Абсолютні терміни, звичайно, пов'язані з неправильними дистракторами, й екзаменовані їх уникають, вибираючи менш категоричні відповіді. Розпливчасті терміни щодо частоти явищ навіть експертами розуміються по-різному, тому відповідь може суттєво залежати від суб'єктивного розуміння екзаменованого, наприклад, що треба розуміти під словом “часто”.

Невдало	Краще
<p><i>Прискорена амортизація має:</i></p> <p>A Тільки великі переваги B Тільки великі недоліки C Є рядовим методом, не має ні переваг, ні недоліків D Як переваги, так і недоліки.</p> <p><i>Правильна відповідь: D</i> [абсолютний термін “тільки”, граматична неузгодженість варіанта C; різна довжина і структура відповідей]</p>	<p><i>Які переваги має прискорена амортизація?</i></p> <p>A B C D</p>
<p><i>Інфляція може бути пов’язана:</i></p> <p>A Тільки зі зростанням ВВП B Зі зростанням або зменшенням ВВП залежно від джерел її розвитку C Тільки зі зменшенням ВВП D Зі зростанням ВВП, коли економіка розвивається за умов повної зайнятості E Усі відповіді правильні.</p> <p><i>Правильна відповідь: B</i> [абсолютний термін “тільки”; граматична неузгодженість варіанта E; фраза “усі відповіді правильні”; алогічна відповідь E (відповіді A і C одночасно неможливі); перетин відповідей B і D, де B є більш повною відповіддю]</p>	
<p><i>Якщо оптимальний вибір для даного споживача складається тільки з одного блага, то можна стверджувати, що:</i></p> <p>A Ціни обох благ однакові B Гранічна норма заміщення дорівнює відношенню цін даних благ C Споживач максимізує граничну норму заміщення D Усі попередні відповіді помилкові</p> <p><i>Правильна відповідь: D</i> [відповідь “нічого з вищезазначеного”; різна довжина і структура відповідей; дієслова у відповідях]</p>	

Змістові слова з умови не повинні повторюватися у відповідях. Це є підказкою.

Невдало	Краще
<p><i>Соціологія.</i></p> <p>A Наука, що вивчає систему отримання інформації про соціальні об’єкти шляхом споглядання чи експерименту і переведення її на рівень фактів B Наука про систему знань щодо процесу дослідження раніше проведеного аналізу інформації про об’єкт дослідження</p> <p><i>Правильна відповідь: A</i> [повтор частини слова з умови “соціальний”]</p>	<p><i>Як називається наука, що вивчає систему отримання інформації про соціальні об’єкти шляхом споглядання чи експерименту і переведення її на рівень фактів?</i></p> <p>A Соціальна статистика B Соціологія C Соціометрія</p>
<p><i>Чи є авторитарними такі режими: традиційний, змагальна олігархія, авторитарно-бюрократичний, військовий, популістський, егалітарно-авторитарний, комуністичний, авторитарно-егалітарний</i></p> <p>A Так B Ні</p> <p><i>Правильна відповідь: A</i> [повтор частини слова з умови “авторитарний”]</p>	<p>[Якщо зміст цього тестового завдання є важливим, доцільно змінити формат A на формат X]</p>

Як і умови, варіанти відповідей слід формулювати позитивно.

Формати тестових завдань

Формати: А, ситуаційний кластер, Х, N, R, В, TF, D, К, С, Е, Н, I

За період від 50-х рр. минулого століття, коли вперше було використано тестові завдання багатовибіркового типу, розроблялося й апробувалося більше 40 різних форматів тестових завдань. Одні з них підтвердили свої діагностичні властивості, інші – ні. Перший апробований формат, – це формат А. Він і сьогодні є найбільш уживаним, оскільки, по-перше, дає можливість перевірити знання на різних рівнях засвоєння. По-друге, при створенні тестового завдання формату А автори тестових завдань, дотримуючись правил, роблять менше помилок. Отже, перейдемо до розгляду найбільш поширених форматів тестових завдань. У пропонованій книзі застосовується позначення форматів буквами (автори Case & Swanson [24], які понад 20 років займалися практичним дослідженням властивостей різних форматів тестових завдань).

Формат А

Тестові завдання з *однією* найкращою відповіддю.

Тестові завдання з вибором однієї найкращої або правильної відповіді є такими, що вживаються найчастіше. Структура цього тестового завдання така: умова (завдання), запитання за цією умовою та чотири-п'ять варіантів відповідей, один з яких є найбільш правильним.

Неправильні варіанти відповіді можуть бути не абсолютно неправильними, але вони все ж таки менш правильні, ніж еталонна відповідь. У такій ситуації варіанти відповіді можна умовно розташувати так:

D	C	A	E	V
найменш правильний				найбільш правильний

Форматом А при відповідній зміні умови і запитання можна перевірити з першого по третій рівні когнітивного домену: знання, розуміння і застосування знань.

Приклад тестового завдання

Формат А. Правильна відповідь	
З якою швидкістю має рухатись автомобіль, щоб за 120 хв проїхати 20 км?	
A	5 км/год
B	10 км/год
C	20 км/год
D	40 км/год
E	60 км/год
Правильна відповідь: B	

Приклад тестового завдання

Формат А. Найкраща відповідь	
Який з нижченаведених географічних факторів виявився найбільш важливим для розвитку ранньої єгипетської цивілізації?	
A	Доступ до Середземного моря
B	Захищеність Лівійською пустелею
C	Розташування у долині ріки Ніл
D	Помірний клімат
Правильна відповідь: C	

Ситуаційний кластер

Хоча кластер і не вважається окремим форматом тестового завдання у чистому вигляді, він широко використовується при тестуванні з різних дисциплін. Кластер складається з ввідного сценарію (опису ситуації) з наступними завданнями, що прямо стосуються цього сценарію.

До кожного сценарію ставиться два і більше запитань. Характер цих запитань визначається змістом та елементами компетентності, вказаними в домені. При створенні кластерів необхідно стежити, щоб запитання були пов'язані зі сценарієм та не були “зчеплені” між собою. Це означає, що відповіді на запитання мають бути незалежними одна від одної. Тобто відповідь на одне запитання не повинна бути підказкою для відповіді на інші запитання з цього ж кластера і не повинна впливати на них. Цим форматом також діагностуються з першого по третій рівні когнітивного домену: знання, розуміння і засвоєння знань.

Приклад тестового завдання. Формат “Кластер”. Завдання 1–2

У великій популяції лабораторних тварин протягом декількох поколінь відбувалося випадкове схрещування. Відсоток тварин, у яких спостерігалася рецесивна риса (bb), становив 49 як перед початком випадкового схрещування, так і наприкінці. У решти тварин спостерігався домінуючий фенотип, до того ж зовнішньої різниці між гетерозиготами і гомозиготами не було.	
1. Яку частоту має алель b у генофонді?	
A 0,70	D 0,30
B 0,51	E 0,07
C 0,49	
Правильна відповідь: A	
2. Яка доля гетерозигот (Bb) у популяції?	
A 0,51	D 0,09
B 0,42	E 0,07
C 0,21	
Правильна відповідь: B	
3. Який найвірогідніший висновок можна зробити з факту, що частота алелі b не змінилася протягом часу?	
A	У популяції відбувається генетичний зсув
B	Два фенотипи рівною мірою адаптуються до лабораторних умов
C	Генотип BB є летальним
D	Має місце високий рівень мутації алелі B в алель b
E	Має місце статевий відбір, що сприяє алелі b
Правильна відповідь: B	

Формат Х

Тестові завдання множинного вибору “правильно – неправильно”.

Зовні цей формат ідентичний формату А, проте у ньому може бути будь-яка кількість правильних відповідей.

Під час розгляду завдань форматів А та Х постає проблема порівняння завдань, які ці формати ставлять перед екзаменованим. На думку Case & Swanson [25], ці формати ставлять неідентичні завдання. Відповідаючи на завдання формату Х, екзаменовані мають вирішувати, наскільки правильним є кожний запропонований варіант для того, щоб бути вибраним як правильна відповідь. Водночас, якщо в завданнях формату А дистрактори не є абсолютно неправильними, в завданнях формату Х кожний варіант відповіді має бути або абсолютно правильним, або ж абсолютно неправильним, з тим, щоб була виключена неоднозначність. Для виконання цієї умови автори тестових завдань формату Х, на відміну від тестових завдань формату А, найчастіше перевіряють лише засвоєння першого рівня когнітивного домену – знання фактів та концепцій, тобто декларативних, а не процедурних знань.

Приклад тестового завдання. Формат Х

Визначити серед поданих прізвищ осіб, які стали в історії України всенародно обраними президентами

- A** Леонід Кравчук
- B** Михайло Грушевський
- C** Леонід Кучма
- D** Михайло Горбачов
- E** Віктор Ющенко

Правильна відповідь: А, С, Е

Формат N

Тестові завдання з декількома найкращими відповідями.

Як альтернативу формату Х (множинний вибір “правильно – неправильно”) та негативно сформульованим тестовим завданням (Що з вказаного ..., ОКРІМ?; Що з вказаного НЕ ...?) деякі автори [26] пропонують використовувати формат N. У тестових завданнях такого формату екзаменованій має вибрати певну кількість (наприклад, дві, три чи чотири) відповідей із запропонованого списку варіантів. Перелік варіантів відповідей здебільшого досить довгий (до 30 варіантів).

Формат N доцільно використовувати у таких ситуаціях, як, наприклад, опис висновків, дослідження, рекомендації, в яких доречно дати більше однієї відповіді.

За діагностичними властивостями тестове завдання формату N, так само як і формату А, діагностує знання, розуміння і застосування.

Приклад тестового завдання. Формат N

Вкажіть, які три з наведених прикладів можна вважати словосполученнями:	
A Яскраве сонце	E Брат і сестра
B Звучить музика	F Плисти за течією
C Сміятися весело	G Нехай приходить
D Хата сниться	
Правильна відповідь: A, C, F	

Формат R

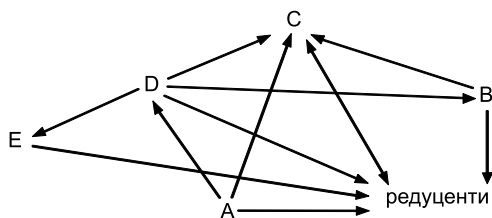
Тестові завдання розширеного вибору.

Цей формат називається форматом розширеного вибору. Кожний його блок складається з трьох-чотирьох умов і спільного для них переліку варіантів відповідей, а інколи, — зі вставного запитання [25]. Кількість варіантів відповідей може коливатися від 4 до 20. Як правило, варіанти відповідей розміщено на початку блоку, перед умовами.

Формат R діагностує з першого по третій рівні когнітивного домену. Його структура така: ми маємо, наприклад, п'ять варіантів відповідей, які можуть використовуватися лише раз, або не використовуватися жодного разу, одне запитання до умов і три-чотири умови.

Приклад тестового завдання. Формат R. Завдання 1–4

На рис. зображено ланцюги живлення у певній екосистемі. Кожна буква позначає певний вид. Стрілками позначено напрями руху енергії.



1. Який вид, найвірогідніше, репрезентує людей, якщо вони є частиною даної екосистеми?
2. Яким видом представлені фотосинтетичні організми?
3. Особини якого виду, найімовірніше, є травоядними?
4. Особини якого виду, найімовірніше, є всеїдними?

Правильні відповіді: 1 – C, 2 – A, 3 – D, 4 – C

Формат B

Тестові завдання логічних пар або визначення відповідності.

Тестові завдання формату B належать до категорії логічних пар, що складаються з 3–5 визначень або цифрових значень, які є варіантами відповідей та переліку слів чи фраз, що містять завдання. Варіанти відповідей позначено буквами, а слова чи фрази, що містять завдання, — пронумеровано. Екзаменований має

підібрати до кожного пронумерованого завдання один найбільш відповідний варіант відповіді, позначений буквою.

Оскільки кожний варіант відповіді може використовуватися більше одного разу або не використовуватися взагалі, тестові завдання формату В неможливо вирішити методом виключення.

Передбачалося, що тестові завдання цього формату розширять використання завдань багатовибіркового типу, даючи змогу оцінити кілька взаємопов'язаних предметів в одному блоці. На відміну від сучасних форматів категорії логічних пар, завдання формату В часто не містили вступного запитання, внаслідок чого були досить неоднозначними. Загалом цей формат успішно використовується дотепер, втім, у наш час поступово витісняється форматом R.

Приклад тестового завдання. Формат В

Визначте відповідність між назвою речовини і класом (групою) речовин, до якого(ої) вона належить.	
<p><i>РЕЧОВИНА</i></p> <ol style="list-style-type: none"> Хром (II) гідроксид Кальцій гідросульфат Калій хлорат Хром (III) гідроксид 	<p><i>КЛАС (ГРУПА) РЕЧОВИН</i></p> <ol style="list-style-type: none"> Кислота Кисла сіль Амфотерний гідроксид Середня сіль Нерозчинна основа
Правильна відповідь: 1 – E, 2 – B, 3 – D, 4 – C	

Формат TF (не рекомендується для застосування)

Тестові завдання TF (True/False, Правильно/Неправильно, Так/Ні, Істинне/Хибне тощо) – спрямовані на перевірку вміння екзаменованих визначати правильність або неправильність твердження. Цей формат широко застосовується для поточного оцінювання, неформального тестування або самоконтролю, але майже ніколи не застосовується у великих програмах стандартизованого тестування.

Формат має суттєві вади. По-перше, високий ступінь вгадування правильної відповіді (50%). По-друге, він сприяє перевірці тривіальних знань і спрямованих на перевірку запам'ятовування фактів. По-третє, твердження у завданні має бути абсолютно правильним/неправильним, оскільки за наявності певного ступеню правильності екзаменованому важко вирішити, наскільки правильним має бути твердження, щоб вважатися правильним. Але існує й серйозніша проблема, обумовлена психологічними особливостями екзаменованих, що мають різні психологічні типи. Дослідження показали, що характер помилкових відповідей у завданнях певною мірою залежить від формулювання запитання в негативному чи позитивному вигляді; частина екзаменованих має схильність до стверджувальних відповідей (“Так”, “Правильно”), частина – до відхилення їх (“Ні”, “Неправильно”).

Тому в цілому тести з такими тестовими завданнями є менш надійними, не рекомендується застосовувати їх з метою контролю під час важливих тестувань.

Приклад тестового завдання. Формат TF

Темперамент – це природні, генетично обумовлені особливості індивіда, зумовлені інтенсивністю і динамікою психологічних процесів, зокрема ступенем урівноваженості, емоційності, загальної та мімічної рухливості.

- A** Так
B Ні

Правильна відповідь: A

Як альтернатива тестовому завданню формату X інколи використовується тематична група завдань формату TF. Такою групою оцінюються ті самі знання, що й у форматі X, але, на відміну від останнього, формат TF дає змогу диференціювати знання кожного факту окремо, що підвищує точність та розподільну здатність вимірювання.

Приклад такої групи наведено нижче. Однак групування не пов'язаних між собою тестових завдань формату TF під загальним запитанням “Які з наступних тверджень є правильними?” категорично не рекомендується внаслідок психологічних факторів, наведених вище, оскільки вони суттєво впливають на результати побудованого таким чином тесту.

Приклад тестового завдання. Група 4x T3 формату TF

Для кожного тестового завдання дайте відповідь А, якщо воно правильне, і В, якщо неправильне. Які фактори впливають на надійність тесту?

1. Довжина тесту
2. Розмір тестового завдання
3. Діапазон тестових балів
4. Формат тестових завдань

Правильна відповідь: 1 – А, 2 – В, 3 – А, 4 – А

Формат альтернативного вибору

Формат альтернативного вибору є скороченою формою запитання багатовибіркового типу, в якому є лише дві відповіді, серед яких одна правильна або краща. Одні автори відносять до цієї групи вищенаведений формат TF, водночас інші відстоюють протилежну позицію, оскільки у запитаннях альтернативного вибору екзаменованій має порівняти два варіанти, тоді як формат TF не є порівнянням у чистому вигляді: в цьому випадку екзаменованій має подумки створити протилежний приклад і зробити вибір між вигаданим і наведеним твердженнями. Серед тестологів існують як противники, так і активні прихильники використання цього формату.

До позитивних рис формату можна віднести легкість створення, порівняно з традиційними MCQ, можливість поставити більше запитань протягом одиниці тестового часу, що може забезпечити краще охоплення змісту та зробити тест більш надійним. Цей формат, зрозуміло, використовується у ситуаціях, коли існує лише дві альтернативи. Інша цікава думка – тестові завдання альтернативного вибору є гарним інструментом для сильних екзаменованих, для яких складно підібрати достатню кількість дистракторів, що працюють.

Утім, внаслідок дуже високого ступеня вгадування правильної відповіді (50%), широке застосування тестових завдань такого формату під час відповідальних тестувань не рекомендується.

Приклад тестового завдання

Формат альтернативного вибору	
Вкажіть найбільш ефективний спосіб мотивації дитини	
A Покарання	
B Похвала	
Правильна відповідь: B	

Формат D (не рекомендується до застосування)

Тестові завдання формату D є комплексними завданнями з категорії логічних пар. Кожне складається з трьох категорій, позначених буквами, і п'яти пронумерованих ситуацій. Екзаменований має виконати два завдання:

- 1) визначити категорію, з якою пов'язано чотири з п'яти ситуацій;
- 2) визначити ситуацію, що не належить до тієї ж категорії, що решта чотири.

До завдання додається детальна Інструкція.

Приклад тестового завдання. Формат D

<p>Інструкція. На запитання треба дати дві відповіді. У списку ліворуч наведено три категорії, позначені буквами. Чотири з п'яти пронумерованих тестових завдань зі списку праворуч тим чи іншим чином стосуються <i>ОДНІЄЇ</i> з цих категорій. Дайте відповіді на такі запитання.</p> <ol style="list-style-type: none"> 1. Зі списку ліворуч замалюйте букву, якій відповідають чотири тестових завдання. 2. Зі списку праворуч замалюйте цифру завдання, яке <i>НЕ</i> стосується тієї ж категорії, що й інші чотири. 	
<p>Вирізнити політику радянської влади, з якою пов'язано чотири з п'яти ознак, а також ознаку, не пов'язану з цією політикою.</p>	
<p>КАТЕГОРІЯ</p> <p>A "Військовий комунізм"</p> <p>B Нова економічна політика</p> <p>C Продовольча програма</p>	<p>ЗАВДАННЯ</p> <ol style="list-style-type: none"> 1. Продподаток 2. Госпрозрахунок 3. Відміна карткової системи 4. Товарно-грошові відносини 5. Приватна власність
Правильна відповідь: B. 2	

Передбачалося, що такі завдання допоможуть виявити розуміння відмінностей між рядом подібних факторів. Проте цей формат виявився складним для написання, а інструкції до його застосування – досить заплутаними. Крім того, формат D має низьку розподільну здатність.

Формат K (не рекомендується до застосування)

Формат K найбільш поширений з категорії “правильно – неправильно”. Інша назва – комплексне запитання багатовибіркового типу (complex MCQ). Завдання цього формату складаються з умови та чотирьох варіантів відповідей, один чи більше з яких є правильними. Завдання екзаменованого полягає у виборі правильної комбінації варіантів відповідей.

Вважалося, що тестові завдання формату K мають перевіряти глибокі знання та розуміння різних аспектів, процесів чи методів і вимагати від екзаменованого знання різних фактів за даною темою. Втім, виявилось, що такі тестові завдання є надмірно складними, вимагають від екзаменованого запам’ятовування правильної комбінації та відповідної їй букви. Крім того, можлива комбінація відповідей є підказкою, що значно знижує розподільну здатність тестових завдань і надійність тесту. Створити якісні однозначні завдання цього формату важко, оскільки вони можуть охоплювати лише абсолютно правильні або абсолютно неправильні факти. Ці завдання виявилися менш ефективними, ніж завдання інших форматів багатовибіркового типу, і мали меншу відносну достовірність на одиницю часу тестування.

Одна з причин розробки такого формату – технічне забезпечення вибору декількох відповідей, що водночас можна відсканувати оптичним сканером, налаштованим на одну відповідь. З розробкою інших програм і пристроїв сканування, переходом на комп’ютерне тестування, а також у зв’язку з наявністю глибоких структурних дефектів від цього формату відмовилися, замінивши його на простий формат X (невизначена кількість правильних відповідей) або на формат N (визначена кількість правильних відповідей).

Приклад тестового завдання. Формат K

Зазначити, до яких держав належали західноукраїнські землі напередодні Другої світової війни:
1. Молдова 2. Польща 3. Росія 4. Румунія A Тільки 1, 2, 3 B Тільки 1, 3 C Тільки 2, 4 D Тільки 4 E Усі правильні
Правильна відповідь: C

Формат С (не рекомендується до застосування)

Тестові завдання формату С зовні схожі на завдання формату В, проте екзаменований повинен визначити “правильні – неправильні” відповіді. Тестове завдання цього формату складається з переліку визначень, які є варіантами відповідей, та переліку слів чи фраз, що містять завдання. Варіанти відповідей позначено буквами, а слова чи фрази, що містять завдання, – пронумеровано. Екзаменований має вирішити щодо кожного пронумерованого завдання, правильна відповідь А чи відповідь В, чи правильні обидві відповіді (варіант С), або одна з них є неправильною (варіант D). При цьому кожний варіант відповіді може використовуватися більше одного разу або не використовуватися взагалі.

Цей тип тестових завдань використовувався для порівняння та протиставлення двох певних фактів. За рівнем складності завдання формату С відповідають завданням формату К. Основна проблема формату С полягає у тому, що екзаменований мусить визначити, наскільки правильною є кожна відповідь, щоб бути вибраною як еталон. Наприклад, якщо завдання асоціюється з варіантами відповідей А і В, але більше відповідає варіанту А, то екзаменований має вирішити: вибрати відповідь А чи обидві відповіді – А і В. Якщо завдання погано асоціюється з варіантами відповідей А і В, то екзаменований повинен визначити ступінь асоціації або вибрати варіант D (всі відповіді неправильні). Судження такого типу не пов’язані зі знаннями, але примушують екзаменованого здогадуватися про те, що саме автор запитання мав на увазі.

Приклад тестового завдання. Формат С

Інструкція. Кожний блок запитань відповідності складається зі списку 4-х позначених буквами варіантів відповідей, за якими розташовано декілька пронумерованих тестових завдань. Для кожного пронумерованого завдання виберіть *ОДИН* варіант відповіді, що найбільш точно відповідає завданню. Кожен варіант відповіді може бути вибраний один, декілька разів або жодного разу.

- A** Гепатит А
B Гепатит В
C Обидва захворювання
D Жодне із захворювань
1. Антибіотики – засіб лікування гострого гепатиту
 2. Клінічні ознаки захворювання виникають через 3 тижні після контакту з носієм
 3. Гострий гепатит часто переходить у хронічну форму
 4. Захворювання можна попередити за допомогою щеплення

Правильна відповідь: 1 – D, 2 – А, 3 – В, 4 – В

Формат Е (не рекомендується до застосування)

Тестові завдання формату Е з множинними правильними чи неправильними відповідями базуються на аналізі взаємозв’язків. Завдання цього формату складається з речення з двома основними частинами: твердження і причини цього твердження. Відповідаючи на такі завдання, екзаменований мусить вибрати варіант відповіді А, якщо обидві частини – істинні твердження і причина є пра-

вильним поясненням першого твердження; В – якщо обидві частини є істинними твердженнями, але причина не є правильним поясненням першого твердження; С – якщо перша частина – істинне твердження, а друга – хибне; D – якщо перша частина – хибне твердження, а друга – істинне; Е – якщо обидві частини – хибні твердження.

Вважалося, що для правильної відповіді на такі тестові завдання необхідно критично мислити і розуміти основні принципи. Проте тестові завдання формату Е складні для написання і заплутані для екзаменованих.

Приклад тестового завдання. Формат Е

ТВЕРДЖЕННЯ		ПРИЧИНА
1	Довжина світового дня змінюється протягом року	ОСКИЇКИ Земля обертається навколо своєї осі
2	Грози найчастіше спостерігаються восени	ОСКИЇКИ Грози виникають при зниженні температури
A	Істинне	Істинне Причина правильно пояснює твердження Хибне Причина НЕправильно пояснює
B	Істинне	
C	Істинне	
D	Хибне	
E	Хибне	

Правильна відповідь: 1 – В, 2 – Е

Формат Н (не рекомендується до застосування)

Тестові завдання формату Н є завданнями на кількісне порівняння. Ці завдання складаються зі спарених тверджень, що описують два порівнюваних об'єкти, і трьох варіантів відповідей: А – якщо А більше, ніж В; В – якщо В більше, ніж А; С – якщо обидва приблизно однакові.

Загальновизнаним є те, що слід обмежувати тестові завдання, які вимагають запам'ятовування абсолютних кількісних величин, проте вважається, що тестові завдання формату Н корисні у випадках, коли знання кількісної інформації є важливим. Проблема для екзаменованих у вирішенні таких завдань полягає у визначенні того, наскільки великою має бути різниця, щоб бути значущою.

Приклад тестового завдання. Формат Н

Наведені спарені твердження описують два об'єкти, які необхідно порівняти кількісно. Відзначте необхідну літеру:
A Якщо (А) більше, ніж (В) B Якщо (В) більше, ніж (А)
1. (А) Валентність S у формулі SO ₃ (В) Валентність S у формулі SO ₂
Правильна відповідь: А
2. (А) Середня тривалість життя у Лаосі (В) Середня тривалість життя у Канаді
Правильна відповідь: В

Формат І (не рекомендується до застосування)

Тестові завдання формату І подібні до формату Н. Вони містять пари фраз, які описують певні параметри, що можуть змінюватися один відносно одного. Екзаменованій має вибрати варіант відповіді А, якщо обидва параметри мають прямо пропорційний зв'язок (зростання параметра з першої фрази супроводжується зростанням параметра з другої фрази); варіант відповіді В, — якщо параметри мають обернено пропорційний зв'язок (зростання параметра з першої фрази супроводжується зменшенням параметра з другої фрази, і навпаки), варіант відповіді С, якщо параметри змінюються незалежно один від одного.

Тестові завдання формату Н та формату І не набули поширення. Через меншу кількість варіантів відповідей, порівняно з іншими форматами тестових завдань, імовірність вгадування правильної відповіді у них є дуже великою. Крім того, такі завдання концентруються на окремих деталях, а не на наукових положеннях.

Приклад тестового завдання. Формат І

Наведені спарені твердження описують умови або кількості, які пов'язані або не пов'язані між собою. Відзначте необхідну букву:

- А** Якщо збільшення параметра (А) супроводжується збільшенням параметра (В)
В Якщо збільшення параметра (А) супроводжується зменшенням параметра (В)
С Якщо зміна параметра (А) не обов'язково супроводжується зміною параметра (В)
- (А) Концентрація речовини у літрі води
(В) Маса речовини у літрі води

Правильна відповідь: А

- (А) Концентрація речовини
(В) Маса речовини

Правильна відповідь: С

Технологія створення тестових завдань

Під час створення тестових завдань необхідно:

- Визначитися з завданням (проблемою), яке має бути продіагностоване під час тестування. Воно має відповідати програмним або кваліфікаційним вимогам і бути чітко сформульованим.
- Уявити умови завдання і варіанти відповідей у вигляді мініатюрної моделі знань. Вибрати певні дані чи факти, що мають бути перевірені. Продумати варіанти відповідей на поставлені запитання і записати.

Що можна використовувати як варіант відповіді? Педагоги, працюючи зі студентами (учнями), знають типові помилки останніх і можуть використовувати ці помилки як варіанти відповідей.

Технічні дефекти тестових завдань

Технічні тестові дефекти — це недоліки в тестових завданнях, що погіршують його вимірювальні якості і можуть давати перевагу певним екзаменованим (або навпаки), незалежно від рівня навчальних досягнень цих екзаменованих. Тестові дефекти, здебільшого пов'язані не зі змістом тестових завдань, а з їх формулюванням та побудовою.

Технічні дефекти, що зустрічаються при написанні тестових завдань, можна поділити на три класи:

- 1) Дефекти, пов'язані з досвідом у тестуванні, які дають перевагу досвідченим у тестуванні студентам (на виявлення цих дефектів екзаменованими, крім того, спрямовано зусилля репетиторів). Такі дефекти полегшують тестове завдання.
- 2) Дефекти, пов'язані з надмірною складністю тестового завдання.
- 3) Структурні дефекти, що виникають внаслідок недотримання правил побудови тестового завдання. У більшості випадків структурні дефекти супроводжуються технічними тестовими дефектами.

Дефекти, пов'язані з досвідом у тестуванні, поділяються на такі категорії:

- граматичні підказки — один чи більше дистракторів не відповідають умові тестового завдання граматично або стилістично;
- логічні підказки — частина варіантів відповіді вичерпує всі можливі варіанти або наявність змістовно алогічних дистракторів;
- абсолютні терміни — використання у варіантах відповідей слів “завжди” або “ніколи”;
- найдовша, найдетальніша правильна відповідь — правильна відповідь більш конкретна чи більш повна, ніж інші варіанти відповіді;
- повторення слів — слово чи фразу з умови долучено до правильної відповіді;
- конвергенція — найчастіше згадування правильних компонентів відповіді у неправильних відповідях (зустрічається у багатокомпонентних відповідях).

Дефекти, пов'язані з надмірною складністю

Складність здебільшого полягає у намаганні створити тестове завдання, що діагностує більш високий рівень засвоєння знань. Під час складання тестового завдання замість чіткого формулювання припускаються дефектів, що називаються дефектами з надмірною складністю. Найчастіше це трапляється у випадках, коли:

- варіанти відповіді довгі, складні або подвійні;
- цифрові показники наведено безсистемно (не в єдиному форматі);

- варіанти відповідей стилістично або змістовно неоднорідні, їх порядок нелогічний;
- як варіант відповіді використано фразу “нічого з вищезазначеного”;
- наявні розпливчасті терміни (“можливо”, “інколи” тощо);
- умови тестового завдання складні або заплутані.

Структурні дефекти (помилки у побудові завдань)

Прикладами таких дефектів є:

- недотримання правила “закритих варіантів відповідей” – ситуація, коли на тестове завдання неможливо відповісти після прочитання його умови та вступного запитання, а необхідно бачити варіанти відповідей;
- коротка умова – довгі відповіді (умова має містити максимум інформації, а відповіді – мінімум);
- застосування дієслів у відповідях (якщо можливо, відповіді не повинні містити дієслова, оскільки вони ускладнюють розуміння тексту);
- відсутність чіткого завдання (тестове завдання має ставити чітке запитання перед студентом, а не вимагати від нього зрозуміти або здогадатися, що саме хотів автор тестового завдання);
- дистрактори (неправильні варіанти) неправдоподібні, неоднорідні або різної довжини порівняно з правильною відповіддю;
- декілька задач (“елементів знань”, “різні когнітивні рівні” тощо) в одному тестовому завданні;
- тестові завдання з негативними формулюваннями, наприклад, “Вибрати **НЕПРАВИЛЬНУ ВІДПОВІДЬ**”, “Що правильно, **КРИМ**”, “Що **НЕ МОЖНА** робити” тощо (структурування завдань у такій формі слід уникати; якщо це абсолютно неможливо, необхідно виділити негативні слова, щоб екзаменованій звернув увагу, що перед ним стоїть інше завдання – знайти гірше або неправильне – порівняно зі звичайним завданням “знайти краще або правильне”).

З досвіду можна рекомендувати такий порядок створення тестового завдання:

1. Визначте зміст та когнітивний рівень, на який створюватимете тестове завдання.
2. Напишіть умову тестового завдання.
3. Продумайте відповіді до нього.
4. Покажіть колезі лише умову завдання.

Якщо його відповідь буде збігатися з вашою, яку ви йому не показували, можете вносити це тестове завдання на пілотну пробацію.

Діагностичну якість тестового завдання визначить психометрія, отримана після тестування.

Створення тестових завдань різних когнітивних рівнів

Під час розробки інструменту оцінювання у фахівця часто виникає потреба створити тестові завдання, спрямовані на оцінювання різних рівнів когнітивного домену, що здебільшого викладено у меті оцінювання.

Під час розробки тестових завдань найпростіше створювати завдання на перевірку засвоєння фактових знань та їх розуміння. Тому за відсутності цілеспрямованих зусиль фахівця щодо напрацювання тестових завдань, які оцінюють вищі когнітивні рівні (застосування, аналіз, синтез і оцінювання), виникає тест, який оцінює лише нижчі когнітивні рівні (знання, в кращому випадку знання і розуміння). Оцінювання, в яких використовуються лише такі тестові завдання, здебільшого піддаються ґрунтовній критиці з боку колег, батьків та екзаменованих. Щоб уникнути цієї ситуації, необхідно мати структуру змісту іспиту (матрицю тесту) і створювати завдання саме того когнітивного рівня та змісту, який заплановано.

Таксономія Блума має певну ієрархію. Очевидно, що для застосування певних знань спочатку слід їх набути, запам'ятати і зрозуміти, а також зуміти викликати з пам'яті. Цілком зрозуміло, що неможливо засвоїти весь обсяг знань, а потім намагатися їх зрозуміти, чи спочатку зрозуміти весь фактовий матеріал, а потім його аналізувати. Тобто набуття різних когнітивних умінь відбувається по певній спіралі: ми розуміємо і запам'ятовуємо знання, вчимося з окремих вивчених елементів складати ціле (синтез) або застосовувати їх, а потім повертаємося до вивчення нових фрагментів знань. Тому потреба створювати завдання на вимірювання різних когнітивних рівнів не залежить від освітнього етапу чи рівня навчальних досягнень екзаменованого, а твердження про те, що в молодшій школі треба перевіряти лише фактові знання, а в старшій – лише вміння аналізувати і синтезувати, є некоректним.

Які когнітивні вміння задіяні під час відповіді на запитання, можна визначити як за поставленим завданням, так і за особливостями екзаменованого. Якщо новачку необхідно задіяти свої аналітичні здібності, щоб дійти певного висновку, досвідченому фахівцеві достатньо буде згадати аналогічний випадок зі своєї ж практики. Тому одне і те ж тестове завдання може оцінювати різні когнітивні рівні у різних екзаменованих, і встановити абсолютно чітку межу між тестовими завданнями «сусідніх» когнітивних рівнів не завжди можливо.

Створення тестових завдань на вимірювання вищих когнітивних рівнів є складним процесом, що потребує багато часу. Автор має проявити творчий підхід для створення привабливих та якісних завдань. Можна дати такі поради для створення якісних завдань.

1. Використовуйте в умові практичні або реальні ситуації. Наприклад, в одному параграфі наведіть проблему з життєвої ситуації. Запитання може

- бути пов'язане із застосуванням принципів для вирішення проблеми або порівнянням і оцінюванням кількох альтернатив з метою вибору кращої за певним критерієм.
- Надайте схему обладнання апарата та поставте завдання проаналізувати або оцінити його. Наприклад, «Що відбудеться в точці А, якщо ...?» або «Як частина А співвідноситься з частиною В», «Як слід змінити розташування ..., щоб...?», «Якої частини не вистачає для того, щоб...».
 - Використовуйте реальні цитати з газет, книг тощо або вигадані. Поставте завдання щодо інтерпретації або оцінювання цих цитат.
 - Використовуйте графічний матеріал із завданням інтерпретувати його.
 - Використовуйте діаграми, таблиці, рисунки, що потребують не лише впізнання, а й інтерпретації чи аналізу.

Наведемо декілька прикладів тестових завдань різних когнітивних рівнів.

Рівень знань. Вважається найнижчим рівнем когнітивного домену. Тестові завдання цього рівня, переважно, вимагають згадати раніше засвоєний матеріал. Більшість професійних екзаменаційних агентств докладають зусиль до мінімізації в тесті кількості тестових завдань рівня знань. Тестові завдання перевіряють знання термінології, фактів, визначень, класифікацій, правил, методів, критеріїв, принципів, теорій тощо.

Приклади тестових завдань. Когнітивний рівень. Знання фактів

Хто є автором теорії еволюції?	
A Р. Гук	D К. Лінней
B Ч. Дарвін	E Г. Мендель
C А. Левенгук	
Виберіть серед наведених тверджень формулювання, що належить закону сталості складу речовин.	
A Сума мас реагентів дорівнює сумі мас продуктів реакції	
B Маса речовини є визначальною кількісною характеристикою цієї речовини	
C Відносна атомна маса елемента є важливою властивістю цього елемента	
D Склад речовини залишається постійним незалежно від місця та способу її добування	
Що є предметом вивчення в теорії імовірностей?	
A Закономірності випадкових величин	
B Імовірності подій	
C Числові характеристики випадкових величин	
Як у комбінаториці називають сполучення, що включають усі елементи заданої множини і відрізняються один від одного тільки порядком їх розміщення?	
A Перестановки	
B Розміщення	
C Сполучення	

Рівень розуміння. Розуміння вимагає інтерпретації матеріалу. Припускається, що екзаменований володіє певним набором знань, а екзаменатор оцінює розуміння цих знань. Діями, що відображають розуміння, є, наприклад, трансформація текстового матеріалу мовою рівнянь, інтерпретація графіків і рисунків тощо.

Приклади тестових завдань. Когнітивний рівень. Розуміння

Якби живі клітини, схожі на земні клітини, було знайдено на іншій планеті, де відсутній кисень, яка частина клітини, найімовірніше, була б відсутня?

- A Клітинна мембрана
- B Ядро
- C Мітохондрія
- D Рибосома
- E Хромосоми

Розташуйте події у тому порядку, в якому вони відбувалися на Землі:

- A Заселення суші хребтовими
- B Вимирання мамонтів
- C Накопичення в атмосфері вільного кисню
- D Виникнення фотосинтезу
- E Розквіт епохи іхтіозаврів та плезіозаврів

На сторонах куба позначено цифри від 1 до 6. Яка ймовірність випадіння парного числа?

- A 1/6
- B 3/6
- C 2/6
- D 0

Яким виразом можна записати задачу: «В класі вдвічі більше хлопчиків, ніж дівчаток. Якщо в класі 10 хлопчиків, скільки дівчаток навчається у класі?»

- A $x = 10 : 2$
- B $x = 10 * 2$
- C $x = 10 + 2$
- D $x = 10 - 2$

Серед наведених простою речовиною є:

- A Водень
- B Вода
- C Хлороводень
- D Натрій хлорид

У хворій 36 років після фізичного навантаження з'явилася ядуха, кашель з виділенням харкотиння. Вона страждає на ревматизм та мітральну ваду серця. Аускультативно над серцем та нижніми відділами легенів дрібно- та середньопухирчасті вологі хрипи. Який механізм погіршення стану хворої є головним?

- A Зниження лімфатичного відтоку
- B Зниження онкотичного тиску крові
- C Підвищення агрегації еритроцитів та тромбоцитів
- D Підвищення гідростатичного тиску крові
- E Підвищення кінцево-діастолічного тиску в лівому шлуночку

З чим з нижченаведеного пов'язаний архітектурний стиль будівлі на рисунку.

- A Візантійською імперією
- B Династією Мін
- C Стародавньою Грецією
- D Стародавнім Римом



Рівень застосування. Застосування характеризує здатність використовувати вивчений матеріал у новій ситуації. До цієї когнітивної категорії належить застосування законів, теорій, концепцій у практичних ситуаціях, розрахунки із застосуванням вивчених формул тощо.

Приклади тестових завдань. Когнітивний рівень. Застосування

Розрахуйте масову частку розчину, якщо у 100 г води розчинили 20 г солі	
A 0,17	D 5,00
B 0,20	E 6,00
C 1,00	
У дворічної дитини різко підвищилася температура (до 37,2°C). Об'єктивно: на незміненому тлі шкіри дрібноплямистий, блідо-рожевий висип, розташований переважно на розгинальній поверхні кінцівок, сідницях і спині; незначна гіперемія глотки, нежить, збільшення потиличних і задньошийних лімфатичних вузлів. Який найбільш імовірний діагноз?	
A Ентеровірусна інфекція	
B Кіп	
C Краснуха	
D Псевдотуберкульоз	
E Скарлатина	

Рівень аналізу. Аналіз – це здатність розкласти ціле на частини. Прикладами завдань, що стосуються аналітичних здібностей, можуть бути такі, що потребують зробити невизначені припущення, розрізнити факти та наслідки, вибрати доречні дані. Аналіз вимагає не лише володіння знаннями, а й розуміння взаємозв'язків між різними типами інформації. Аналіз вважається вищим когнітивним рівнем.

Приклади тестових завдань. Когнітивний рівень. Аналіз

Які зміни умов наведеної рівноважної реакції призведуть до її зсуву ЛІВОРУЧ? $C + O_2 \rightarrow CO_2(g)$, $\Delta H = -394$ кДж	
A Збільшення температури реакції вдвічі	
B Зменшення температури реакції вдвічі	
C Збільшення кількості C вдвічі	
D Зменшення кількості C вдвічі	
Мітохондрії вважаються електростанціями клітини, оскільки вони надають енергію для клітинного метаболізму. Який із фактів є НАЙБІЛЬШ переконливим свідченням цієї функції мітохондрій?	
A Велика кількість мітохондрій у м'язовій тканині	
B Виявлення мітохондрій у більшості типів клітин рослин і тварин	
C Наявність у мітохондріях ферментів циклу Кребса	
D Наявність подвійної мембрани у мітохондріях	
E Утворення у мітохондріях АТФ	
Розгляньте таблицю і вкажіть, до якої комбінації країн, поданих у рядках А, В і С, може належати наведена статистика.	
1. A – Південна Корея; B – Кенія; C – Канада	
2. A – Шрі Ланка; B – Німеччина; C – Таїланд	
3. A – Шрі Ланка; B – Таїланд; C – Швеція	
4. A – Намібія; B – Португалія; C – Ботсвана	

	ВВП (подушний) 1991	Зростання ВВП (подушного) 1980–1991	Збільшення населення 1980–1991	Структура працевлаштування 1980–1985 (%)		
				Сільське господарство	Промисловість	Послуги
A	500	2,5%	1,5%	51	20	29
B	1570	5,8%	1,6%	74	8	8
C	25110	1,7%	0,3%	6	32	32

Водій вантажівки їхав з м. Петрівці до м. Колиньки. Невдовзі після проїзду через м. Оселів він зупинився випити кави в м. Ставки, подолавши половину шляху. Відстань між якими містечками є найбільшою:

- A** Петрівці – Оселів
- B** Оселів – Колиньки
- C** Ставки – Колиньки
- D** Оселів – Ставки

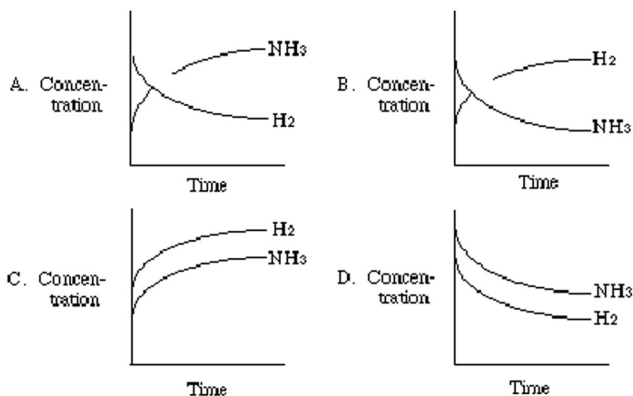
Рівень синтезу. Синтез – це процес складання цілого з частин. При цьому перевага надається творчій, креативній поведінці. Прикладами завдань, що стосуються здатності до синтезу, можуть бути такі, що потребують складання плану дій або експерименту в конкретній ситуації, створення нових прикладів, формулювання класифікаційних принципів. Тестові завдання рівня аналізу і синтезу тісно пов'язані між собою.

Приклади тестових завдань. Когнітивний рівень. Синтез

Ірина та Михайло досліджують вплив солі на ріст рослин. Для цього вони щоденно поливають рослини розчином солі. Яким чином вони можуть найкраще вдосконалити своє дослідження?

- A** Щоденно збільшувати кількість солі у воді
- B** Охолоджувати сольовий розчин перед поливом
- C** Додати фарбу, щоб побачити, як глибоко проникає сольовий розчин у рослину
- D** Поливати частину рослин сольовим, а решту – безсольовим розчином
- E** Чергувати полив рослин сольовим та безсольовим розчинами

Який із наведених графіків відображає зміну концентрації H_2 та NH_3 , у якому випадку з цих елементів утворюється аміак?



Рівень оцінювання. Оцінювання стосується здатності давати оцінку, зокрема критичну, процедурам, висновкам, доказам, перспективам. Створення тестових завдань на рівень оцінювання є складним. Важко розрізнити тестові завдання рівня оцінювання та рівня застосування знань. Тестові завдання рівня оцінювання не можуть мати лише одну правильну відповідь. Як правило, такі завдання базуються на порівнянні, виборі за певним критерієм (кращий, гірший, ефективний, дешевий, доступний тощо), вони є об'єктивнішими, коли таке порівняння можливе. Прикладами завдань можна буде оцінити, чи підтверджено висновки доказами, чи виконано завдання, чи відповідають аргументи наведеним даним тощо.

Приклади тестових завдань. Когнітивний рівень. Оцінювання

Незалежно від ступеня реалістичності виконання наступних процедур, який напрям наукових досліджень, найімовірніше, буде найбільш обґрунтованим і слугуватиме прямим доказом еволюційних зв'язків між різними видами?

- A** Аналіз хімічного складу накопичених запасів живлення в жіночих гаметах
- B** Аналіз ферментів циклу Кребса
- C** Спостереження за формою та розташуванням ЕПС
- D** Порівняння деталей молекулярної структури ДНК
- E** Визначення загального відсотка білків у клітинах

Яка з наведених властивостей електромагнітної радіації підтверджує висновок, що світло – це хвилі?

- A** Інтенсивність варіює з квадратом відстані від джерела
- B** Металеві поверхні випускають електрони при опроміненні їх світлом
- C** Світло огинає дуже маленькі об'єкти на своєму шляху
- D** Швидкість світла у вакуумі є постійною

1.4. Обробка та представлення результатів тестування

Ідентифікація учасників тестування

Розрахунок результатів

Контроль можливих порушень

Представлення результатів

У цьому розділі ми маємо визначити основні етапи обробки результатів тестування та встановити способи представлення даних залежно від мети оцінювання.

Розглянемо кожний етап окремо.

Ідентифікація учасників тестування та їхніх відповідей

Найголовнішими в обробці результатів є ідентифікація учасників тестування (учнів, студентів) та верифікація їхніх відповідей. Ця копітка робота займає багато часу і має проводитися дуже ретельно. Наше завдання – впевнитись у тому, що всіх, хто тестувався, зафіксовано, і бланки їхніх відповідей ототожнено.

На сьогодні на вступних іспитах до вищих навчальних закладів письмові роботи кодуються. Оскільки бланки відповідей також кодуються, ми маємо бути впевнені, що після перекодування ідентичність особи, яка тестувалася, зберігається. Кодування бланків відповідей обумовлено кількома причинами. По-перше, це внутрішній обов'язковий технологічний етап для роботи з базами даних осіб, які тестувалися. По-друге, це певним чином забезпечує “закритість” результатів. Про яку ж корупцію тоді може йти мова, якщо все кодується? Втім реально певні фальсифікації можливі на кожному етапі тестування. Яким же чином можна їх уникнути при “закритості” результатів? Для цього слід реалізувати систему тристороннього контролю: оригінал відповідей має знаходитися у Центрі тестування, а дві копії – в особи, яка тестувалася (абітурієнта, студента), та в адміністратора, відповідального за тестування (ректора, директора). За такої системи можна дозволити *екзаменованому* заповнювати бланк олівцем, що дає змогу виправляти свої помилки під час екзамену, а *тестовій організації* у відведений регламентом термін – здійснити ретельну верифікацію результатів. Лише після цього можна провести обробку результатів і представити їх у затвердженій формі.

Окремо розглянемо питання про сканування результатів безпосередньо після екзамену. Все залежить від мети. Якщо мета полягає в тому щоб показати як працює сканер, або як вивести результати на дисплей, тобто продемонструвати технологічний бік тестування, тоді цю процедуру можна застосовувати, але бажано її демонструвати на етапі реєстрації екзаменованих. Якщо ж публічне сканування результатів після екзамену здійснюється з метою демонстрації відкритості, то це підміна понять об'єктивності і технологічності, оскільки будь-яка технічна проблема буде сприйнята як фальсифікація результатів. Саме тому провідні міжнародні тестові центри під час проведення масштабних тестувань здійснюють цю процедуру у спеціальних технологічних приміщеннях, витрачаючи на сканування, обробку і представлення результатів від одного тижня до одного місяця. Продовження терміну оприлюднення результатів і унеможливлення фальсифікацій – моменти, що мають бути відкриті і процедурно задокументовані. При скануванні можуть бути присутні окремі представники від екзаменованих, які фактично забезпечують громадський контроль за обробкою результатів.

На рис. 6 наведено приклад ідентифікації осіб, які тестувалися, а також методику верифікації їхніх відповідей, яка використовується у Центрі тестування при МОЗ України.

Сканирование форм									
База данных									
\\Server\test\STTest\Tests\E04_pretest_2003\E04\05U10_ukr.unrv									
Сканировать					Посчитать текущего		Рассчитать всех		
1	000000143	Шарипов	МО	2	5	3	4	32	52
2	000000129	Терещенко	ВО	97	6	0	1	66	53
3	000000108	Радченко	ЖГ	42	1	0	14	65,5	54
4	000000036	Жидченко	ОО	91	1	2	6	68,5	55
5	000000027	Дементьева	ТВ	47	3	0	3	80,5	56
6	000000092	Онищенко	ГП	21	3	0	7	65,5	57
7	000000047	Цюковська	ЖЮ	65	5	22	7	56,5	58
8	000000103	Плохенко	ВВ	28	7	6	5	51,5	59
9	000000003	Амбарцумян	АО	53	8	2	2	56	60
10	000000054	Колоша	АВ	23	6	6	3	54	61
11	000000140	Целуйко	АС	43	4	0	4	63	62
12	000000141	Целуйко	ОБ	54	9	2	4	59	63
13	000000021	Гула	ЛП	22	9	1	3	57	64
14	000000009	Батрак	НЮ	35	9	0	0	73	65
15	000000148	Яременко	ОВ	30	5	1	0	67	66
16	000000130	Титаренко	ДА	96	4	24	0	61,5	67
17	000000118	Середа	АВ	3	7	1	1	37	68
18	000000033	Дубровін	ОВ	58	1	0	1	61,5	69
19	000000132	Ткаченко	РО	80	7	3	6	66,5	70
20	000000043	Ігнатенко	НО	35	9	1	8	58,5	71
21	000000031	Доценко	ОС	62	7	1	4	67	72
22	000000042	Івашенко	ЮМ	62	7	0	2	57	73
23	000000091	Огородник	МВ	93	3	0	13	70,5	74
24	000000136	Толстикова-Ц	ОА	66	1	0	1	78,5	75

Рис. 6. Ідентифікація учасників тестування та верифікація правильності сканування

У більшості випадків, особливо при зовнішніх, а отже масштабних тестуваннях, зчитування даних із бланків відповідей здійснюється за допомогою сканера. Сьогодні існує два типи сканерів: спеціалізовані й універсальні.

Спеціалізовані “марк рідери” (OCR), створено спеціально для зчитування бланків відповідей, коли зчитується кожна відповідь окремо. Універсальні сканери зчитують бланк як “картинку”. Перші є точнішими. Але останнім часом удосконалення універсальних сканерів щодо швидкості зчитування робить їх конкурентоспроможними порівняно з спеціалізованими сканерами.

Розрахунок результатів

Маємо декілька підходів до розрахунку результатів. Наведемо три з них:

1. За кількістю правильних відповідей:

$$X_i = \frac{N_e}{N} * 100\%$$

2. З урахуванням неправильних відповідей:

$$X_i = \frac{N_e - \frac{N_{не}}{m-1}}{N} * 100\%$$

3. Зі штрафом за вгадування:

$$X_i = \frac{N_e - \frac{N_{не}}{m-1} * W}{N} * 100\%$$

Перший підхід – простий розрахунок, за допомогою якого вираховується кількість правильних відповідей, що ділиться на загальну кількість відповідей, і визначається відсоток.

При *другому* підході за неправильні відповіді екзаменований штрафується. Коефіцієнт може коливатися від 1/4 до 1.

При *третьому* підході вводиться поправка на вгадування.

Головним при виборі методики розрахунку бала є те, що методика має бути чіткою і відкрито прописаною у процедурі підрахунку результатів. Особи, які тестуються, мають заздалегідь знати, яким чином буде здійснюватися підрахунок результату, оскільки це впливає на тактику поведінки особи, під час іспиту. Наприклад, за відсутності штрафу за неправильні відповіді екзаменований може закреслювати навіть такі відповіді, в яких невпевнений. В іншому випадку така тактика є неправильною.

Конвертація результатів тестування у шкали

Розрахунок результату – остаточне оцінювання – це один із завершальних етапів вимірювання. Процедура та методика розрахунку суттєво впливають на остаточні результати, на можливість аналізу та на статистичну достовірність висновків, тому визначення валідності оцінювання є надзвичайно важливим етапом у процесі вимірювання.

Процедура оцінювання полягає в конвертації одержаного при тестуванні первинного результату (бала) у відповідне число (оцінку) за певною шкалою.

Методика оцінювання полягає у визначенні алгоритму виконання зазначеної процедури.

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*.

Шкали оцінювання

Залежно від рівня вимірювання характеристики існують чотири шкали: номінальна, порядкова (рангова), шкала інтервалів і шкала відношень.

Номінальна шкала – це найнижчий рівень вимірювання, за яким визначається наявність або відсутність вимірюваної ознаки без встановлення градацій за ступенем її прояву (наприклад, “залік – незалік”). Шляхом цих вимірювань можна визначити частоту або ймовірність, з якою ця ознака є наявною або відсутньою в експериментальній групі.

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”). Шкала поділяється за кількістю рівнів, градацій, ознак. 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 нині прийнята у школах та вищих навчальних закладах України, є порядковою шкалою, оскільки інтервали між оцінками, наприклад при чотирибальній системі оцінювання, не відповідають дійсному розриву між реальними результатами.

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, але для кожної зі шкал

вони вибираються довільно. Саме тому не можемо робити висновки про співвідношення між різними значеннями шкали. Але вона називається *метричною*, і з величинами, визначеними за нею, можна проводити адитивні дії.

Слід зауважити, що за цією шкалою підлягають конвертуванню лише ознаки, розподілені за нормальним законом.

Шкала відношень. Це шкала, на якій початок відліку є не довільним, а чітко визначеним (наприклад, шкала довжини, ваги). Серед соціальних наук ознаки, які можна було б виміряти за цією шкалою, ще не відомі.

Таким чином, шкалювання – це процес конвертації у відповідне число за певною шкалою та за наявності певної норми, виміряних шляхом тестування результатів. При конвертації використовують два підходи, які визначають відповідні одиниці конвертованої величини:

- визначення рангового місця;
- визначення положення відносно норми величини, яка вимірюється.

Це вимагає введення двох одиниць виміру: **відносної й абсолютної**.

Відносна одиниця вимірюється у відсотках або місцях, абсолютна – у відстані до норми або в балах.

Реалізація першого підходу, за якого визначається рангове місце за рівнем відповідної ознаки, не дає можливості визначити, чи досягнуто мету навчання. Для з'ясування цього надзвичайно важливого для процесу навчання питання застосовуються різні критеріальні норми і конвертація в абсолютні одиниці за нормованою шкалою.

Наведемо приклади різних підходів, що застосовуються при конвертації результатів.

Конвертація результатів тестування у бали

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

Критеріальний підхід

Наведемо приклад. Правильна відповідь на кожне тестове завдання оцінюється в один бал. Результат за тест складається з суми балів, набраних абітурієнтом за кожний модуль. Максимальна кількість балів за тест – 50. Бали за тест 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

Наведемо приклад. Точний результат переведення набраних балів у шкільну оцінку буде опубліковано після проведення тестування і аналізу його результатів. За результатом тестування складається рейтинговий список абітурієнтів, порядковий номер у цьому списку визначає рейтинг абітурієнта з тесту. Якщо кілька абітурієнтів набрали однакову кількість балів, то їхній рейтинг є однаковим і обчислюється як середньоарифметичне відповідних місць у списку.

Для конвертації результатів тестових екзаменів в оцінки “2”, “3”, “4”, “5” застосовуються нормо-орієнтовані критерії, що визначаються за допомогою статистичних методів. 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 $(M - \sigma)$;

“4” – від значення $(M - \sigma)$ до значення $(M + \sigma)$;

“5” – від значення $(M + \sigma)$ та вище.

Приклад конвертації результатів тестування у чотирибальну шкалу при використанні статистичного підходу:

Критерій склав – $40,5\%$. Статистичні дані: $M = 70$, $\sigma = 10$;

“2” – менше ніж $40,5\%$;

“3” – від $40,5\%$ до 60% ($M - \sigma$);

“4” – від 61% ($M - \sigma$) до 80% ($M + \sigma$);

“5” – від 81% ($M + \sigma$) та вище.

Контроль можливих порушень

Розглянемо обов’язкову складову обробки результатів – контроль можливих порушень. Контроль можливих порушень здійснюється паралельно з розрахунками індивідуальних результатів тестування і базується на результатах статистичного та психометричного аналізу. Саме він забезпечує обґрунтованість валідності одержаних результатів.

У Центрі тестування при МОЗ України, як уже зазначалося, розроблено програмно-технічний комплекс, що забезпечує реалізацію всіх обов’язкових етапів технології тестування відповідно до міжнародних вимог. Приклад одного з “технологічних вікон” наведено на рис 7.

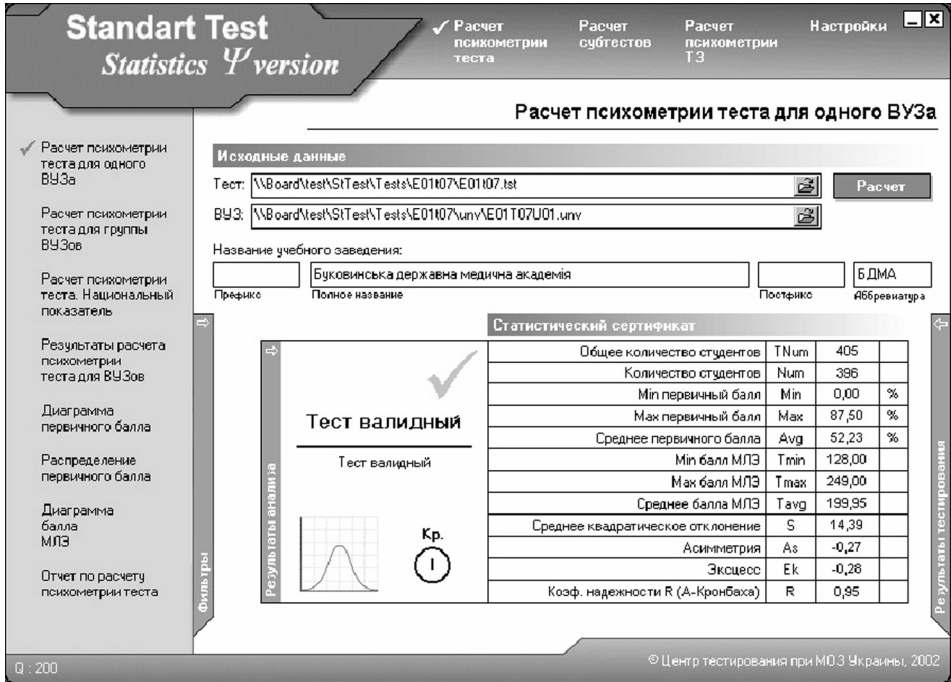


Рис. 7. Розрахунок психометрії тесту

Як зазначалося вище, психометричний аналіз тесту включає визначення таких характеристик тесту: середнє (M), мінімальне та максимальне значення; середнє квадратичне відхилення (s), коефіцієнт надійності (R) (α -Кронбаха, Кьюдер-Ричардсона); похибка вимірювання (E); коефіцієнт валідності (якщо наявний); контингент екзаменованих (кількість, структура).

Статистичний аналіз базується на визначенні ступеня відхилення результатів тестування від закону нормального розподілу за двома характеристиками: коефіцієнтами асиметрії та екссесу (див. рис. 8, 9).

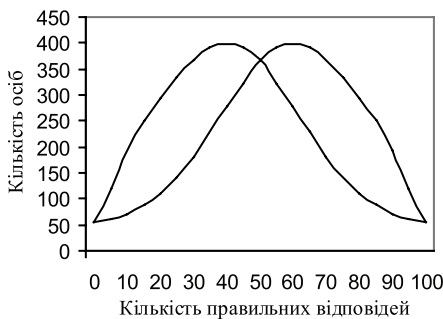


Рис. 8. Крива розподілу з $As \neq 0$, $Ex = 0$

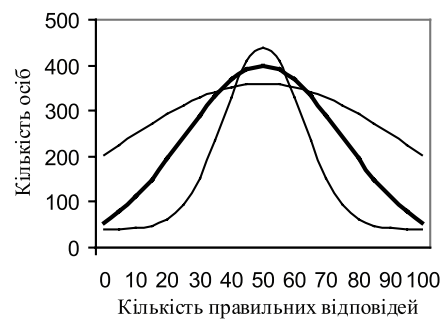


Рис. 9. Крива розподілу з $Ex \neq 0$, $As = 0$

Коефіцієнт асиметрії A_s – асиметричність кривої – (див. рис. 8) показує, наскільки результати тестування зміщуються ліворуч або праворуч відносно кривої нормального розподілу. Зміщення праворуч, як правило, свідчать про те, що результати є завишеними. Зміщення ліворуч – заниженими.

Коефіцієнт ексцесу E_x (див. рис. 9) характеризує зміна гостроти кривої біля вершини.

Детальніше порядок застосування отриманих даних для аналізу тестових завдань розглядається у розділі 3.3.

Представлення результатів

Представлення результатів є завершальним етапом педагогічного вимірювання та оцінювання. Форми звітів за результатами оцінювання, залежно від функцій, які вони виконують, можна поділити на індивідуальні, форми порівняльного аналізу і технологічні форми.

Форми індивідуальних результатів

Індивідуальні результати можуть наводитися у вигляді:

- відсотка правильних відповідей за тест або відповідного нормованого бала;
- відсотка правильних відповідей або балів за субтестами, відповідно до структури змісту тесту та специфікації тестових завдань;
- результату за 12-бальною або 4-бальною шкалою;
- рейтингового місця або перцентіля.

Наступним комплектом звітів, які можна формувати на основі тестування, є форми для порівняльного аналізу.

Форми порівняльного аналізу

Форми порівняльного аналізу являють собою результати тестування, згруповані за *статистичними характеристиками* осіб, які тестувалися, та за *їхніми адміністративними даними*.

Ці форми створюються на етапі реєстрації учасників тестування на основі статистичних характеристик і адміністративних баз даних.

Для отримання детальної інформації щодо успішності за типом навчальних закладів, формою навчання, гендерним фактором або за будь-якими іншими показниками, слід збирати ці дані на етапі реєстрації.

Технологічні форми

Технологічні форми можна визначити за такими основними типами щодо психометричного аналізу:

- психометрія тесту (за всією вибіркою, за навчальним закладом, за аудиторією тощо);
- валідність і надійність тесту;
- валідність процедури тестування;
- психометрія тестових завдань за тестом (за всією вибіркою, за навчальним закладом тощо).

Останнім етапом є оприлюднення результатів тестування, зокрема індивідуальних результатів.

Відомо, що коли тестування є *добровільним*, наприклад, при вступі до вищих навчальних закладів США, абітурієнт, який складає SAT, може не повідомляти про його результати до школи, в якій він навчався. Він отримує результати тестування особисто. І лише за його бажанням атестаційна організація ETS, яка здійснює тестування, надсилає результати у навчальні заклади, вказані абітурієнтом у анкеті реєстрації. Таким чином забезпечується право абітурієнта на закритість інформації.

Результати *обов'язкових* тестувань, які є одним з етапів навчання, завжди оприлюднюються, а отже, є повністю відкритими.

Способи представлення результатів можуть мати будь-яку форму. Як правило, це форма таблиці, стовпчаста або кругова діаграма. Інколи для візуалізації наводять гістограми первинного бала або вже конвертованого результату. Але, на нашу думку, якщо наводяться конкретні коефіцієнти асиметрії та ексцесу, то в цьому немає потреби.

Крім гістограми, можна представити результати тестування як функцію розподілу. Така форма дає можливість встановити перцентиль особи, яка тестувалася. Оскільки це не дуже часто вживане поняття, дамо йому пояснення. На рис. 10 наведено одні й ті ж результати тестування у формі *функції щільності розподілу* — ■ — (аналог гістограми або полігону частот) та у формі *функції розподілу* — ▲ — (аналог перцентильної шкали). Графік функції розподілу у тестології називається *характеристичною кривою тесту*.

На цьому рисунку на горизонтальній осі результати тестування – відсоток правильних відповідей за тест. А на вертикальній осі відкладено частоту або відсоток учнів, які одержали той чи інший результат. Застосуємо результати статистичного аналізу (криву функції щільності розподілу і криву функції розподілу) до визначення рейтингового місця особи, яка тестується.

Виберемо, наприклад, результат тесту – 60% (позначено вертикальною пунктирною лінією). За кривою функції щільності розподілу – (■) – визначимо, що кількість учнів, які отримали такий результат, становить 24% (нижня штрихпунктирна горизонтальна лінія). Скористаємося тепер перцентильною шкалою – (▲) – для встановлення рейтингу учня. За даною перцентильною шкалою учень, який набрав 60% правильних відповідей, має перцентиль 48 (верхня пунктирна горизонтальна лінія). Це означає, що 48% учнів набрали бал нижчий за 60% правильних відповідей. Перцентиль дає можливість визначити рейтингове місце. У нашому прикладі рейтингове місце учня, який має перцентиль 48, дорівнює 52 ($100 - 48 = 52$).

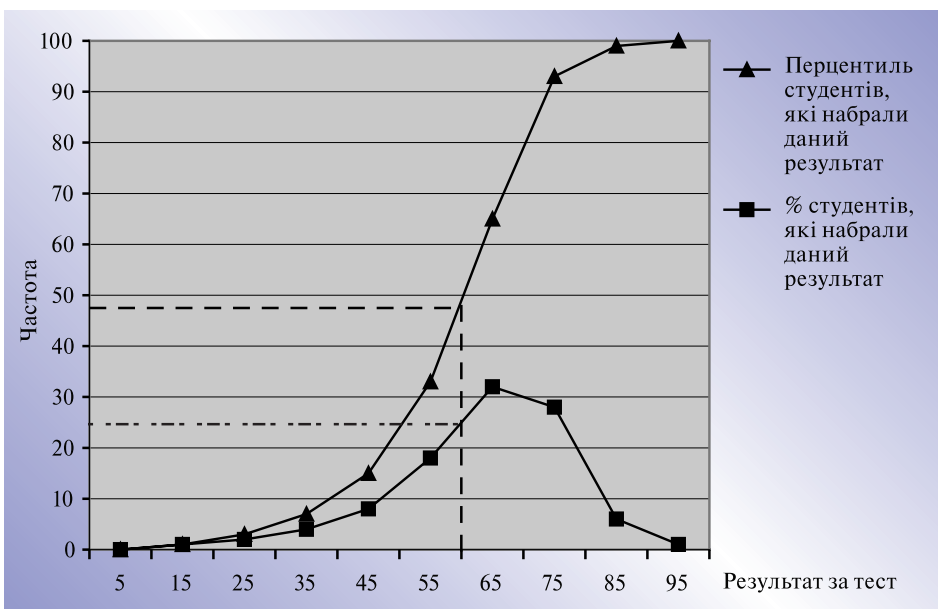


Рис. 10. Закон розподілу і функція розподілу первинного бала

Ще один приклад: 70% – результат учня за тест. Визначимо за графіком відсоток учнів, які досягли такого ж результату. Він дорівнює 30. Перцентиль учня дорівнює 80, а рейтингове місце, відповідно, 20.

Чому ми користуємося перцентильною шкалою? Тому що вона дає змогу визначити місце особи, яка тестувалася, незалежно від року тестування та складності тесту.

1.5. Методи визначення критерію “склав – не склав”

Типи критеріїв

Методи встановлення критерію

Критерій (стандарт) відбиває мінімально достатній рівень якості, необхідний для задовільної оцінки. Стандарти якості знаходять застосування у багатьох сферах життя, в тому числі й у сфері освіти. Вибір критерію для визначення якості знань та вмінь є вирішальним у навчальному процесі і потребує надзвичайно відповідального підходу. Рішення щодо визначення критерію оцінки знань виносяться на основі екзаменаційного бала та іншої інформації. Найпоширенішою практикою є прийняття рішення щодо кожного екзаменованого в індивідуальному порядку.

У цьому випадку екзаменатор враховує екзаменаційний бал та іншу інформацію про екзаменованого, яку він вважає важливою, робить оцінювання та виносить рішення. Такий спосіб має свої переваги: оскільки немає необхідності завчасно визначати критерії прийняття рішення, екзаменатор може використовувати будь-яку інформацію про екзаменованого на власний розсуд; індивідуальне оцінювання дає шанс кожному екзаменованому бути оціненим як особистість. Але цей спосіб має і серйозні вади: він надзвичайно суб'єктивний, результат оцінювання часто залежить від враження, яке екзаменований справив на екзаменатора; крім того, екзаменатор має можливість використовувати інформацію, яка не відповідає цілям і завданням екзамену (наприклад, якщо екзамен має на меті визначити рівень знань або вмінь у певній сфері, недоцільно для прийняття рішення використовувати інформацію про відвідування занять, старанність, сумлінність тощо). Крім того, за відсутності чітко сформульованих критеріїв прийняття рішення екзаменатор може бути звинувачений в упередженості чи непослідовності. Тому в інтересах як екзаменованих, так і екзаменаторів використовувати **правило прийняття рішень**, яке буде застосовуватися до всіх екзаменованих однаково. Це правило визначить, яка саме інформація і у який спосіб буде використовуватися для прийняття рішення щодо кожного екзаменованого.

Найпростішим та найпоширенішим правилом прийняття рішень є **поділ екзаменованих на дві групи**: групу з вищим балом і групу з нижчим балом. При цьому група з вищим балом вважається такою, що склала екзамен, а група з нижчим – такою, що не склала. Щоб використати результати екзамену в такий спосіб, необхідно визначити бал, за яким відбуватиметься поділ на групи. Бал, що розділяє екзаменованих, буде **прохідним** балом або критерієм.

Абсолютні та відносні критерії

Критерії можуть бути *відносними* чи *абсолютними*. **Відносний критерій** залежить від порівняння між собою балів екзаменованих. На запитання “Який результат є задовільним?” можна відповісти, використовуючи результати інших екзаменованих. Наприклад, якщо бал є вищим за середній результат групи, то це є задовільний результат; якщо бал не знаходиться серед 20% гірших результатів групи, то це є задовільний результат.

Кожне з цих тверджень є відносним критерієм, залежним від результатів інших екзаменованих. Таким чином, бал кожного окремого екзаменованого буде порівнюватися з критерієм, який залежить від результатів інших екзаменованих, при цьому чим кращими будуть результати екзаменованих, тим вищим буде критерій. Вибір прохідного бала за відносним критерієм не викликає труднощів. Потрібно просто вибрати такий бал, що пропускає певну кількість (чи відсоток) екзаменованих.

Абсолютний критерій на відміну від відносного не залежить від результатів екзаменованих, яких будуть оцінювати за даним критерієм. Для людини, яка складає екзамен за абсолютним критерієм, не має значення, які результати отримають інші, оскільки отриманий ними бал не впливатиме на критерій. Для визначення абсолютного критерію існують різні методи:

- методи, що базуються на судженнях про екзаменаційні завдання;
- методи, що базуються на судженнях про окремих екзаменованих;
- методи, що базуються на судженнях про групи екзаменованих.

Методи встановлення критерію

Будь-який критерій – чи то абсолютний, чи то відносний – базується на певному судженні. Критерій дає відповідь на запитання “Який результат є задовільним?” Відповісти на це запитання можна лише згідно з певним судженням. Для різних методів визначення критерію виносяться різні типи суджень, але незалежно від типу судження, всі вони мають бути винесені експертами з урахуванням мети екзамену.

Методи, що базуються на судженнях про екзаменаційні завдання

Три методи вибору критерію, що належать до цієї групи, ґрунтуються на концепції “маргінального” (“граничного”) екзаменованого, тобто такого, рівень знань та вмінь якого знаходиться на межі груп з високим і низьким балом. Таким чином, прохідним має бути бал, що очікується від екзаменованого, рівень знань і вмінь якого знаходиться на межі цих двох груп. Такі методи вимагають суджень, винесених шляхом постановки певних запитань на екзамені. Їх названо на честь авторів: метод Недельського [27], метод Ангоффа [28] та метод Ебеля [29]. Кожен з зазначених методів вимагає від експертів *суджень різного типу*.

Метод Недельського. Запропонований Л. Недельським у 1954 р. метод може використовуватися лише до екзаменаційних завдань багатовибіркового типу,

оскільки вимагає *судження щодо кожної можливої неправильної відповіді*. Завдання експерта полягає в тому, щоб, розглядаючи завдання, ідентифікувати ті неправильні відповіді, які маргінальний екзаменований зміг визнати явно неправильними, тобто не кращими з наведених відповідей.

Метод Ангоффа. Метод, запропонований У. Ангоффом у 1971 р., схожий на метод Недельського, але, на відміну від останнього, може застосовуватися не лише до багатовибіркових запитань. Прохідний бал за методом Ангоффа вираховується також за очікуваними результатами для кожного екзаменаційного завдання, проте ця процедура не потребує розгляду кожної неправильної відповіді.

Метод Ебеля. На відміну від двох попередніх методів складається з двох етапів: класифікації тестових завдань за групами і винесення судження у числовому виразі для кожної групи. Класифікація тестових завдань здійснюється за двома категоріями: складністю і важливістю. Ебель запропонував три рівня складності: “легкий”, “середній”, “складний”, та чотири рівня важливості: “суттєвий”, “важливий”, “прийнятний” і “сумнівний”. Після класифікації завдань експерти визначали ймовірний процент правильних відповідей маргінального екзаменованого для кожної з 12 груп тестових завдань. Процедура роботи експертів є аналогічною тій, що рекомендована для методів Недельського і Ангоффа, проте її варто застосовувати окремо для кожного етапу методу Ебеля.

Методи, що базуються на судженнях про окремих екзаменованих

Методи даної групи було розроблено у роботах S. A. Livingston і M. J. Zieky [30] та R. A. Berk [31]. Їх застосування вимагає двох типів інформації про кожного екзаменованого: його екзаменаційний бал і рівень його компетентності. Ця група методів включає:

- метод протилежних груп;
- метод маргінальної групи.

Основна перевага цих методів полягає у тому, що рішення приймається на основі оцінки компетентності людей як достатньої чи недостатньої для досягнення певної мети. Тому винесення суджень цього типу базуватиметься на даних принаймні двох незалежних експертних вимірювань.

Метод протилежних груп. При застосуванні методу протилежних груп, для того, щоб розподілити екзаменованих на дві групи, експерти повинні мати можливість оцінити компетентність кожного екзаменованого і мати чітке уявлення про рівень компетентності, необхідний для успішного складання тесту.

Близьким до методу протилежних груп, розробленому Livingston та Zieky [30], є метод, запропонований Berk [31]. За цим методом прохідний бал встановлюється на підставі даних, отриманих від двох груп екзаменованих: групи, що пройшла навчання в сфері, оцінюваній тестом, і групи, що такого навчання не проходила. Прикладами можуть бути група випускників і група першокурс-

ників. Обидві групи екзаменованих складають один і той же тест, за результатами цього складання будуються криві розподілу екзаменаційних балів. Якщо обидві криві повністю збігаються, то жодного рішення щодо прохідного бала прийняти неможливо, оскільки в ідеалі ці криві не повинні мати жодної точки перетину. В дійсності найбільш типовою картиною буде крива розподілу з від'ємною асиметрією для групи, що пройшла навчання, з частковим збігом кривих обох груп. *У точці перетину цих кривих і буде знаходитися прохідний бал.*

Метод маргінальної групи. Додатковим до методу протилежних груп є метод маргінальної групи. Він базується на ідеї, що прохідний бал має бути очікуваним від маргінального екзаменованого. У цьому аспекті метод маргінальної групи схожий на методи, що базуються на судженнях щодо екзаменаційних завдань. Різниця лише в тому, що експертам не потрібно оцінювати дії маргінального екзаменованого, а достатньо лише визначити фактичних екзаменованих як маргінальних у сфері компетентності, що оцінюється тестом. Прохідний бал буде середнім балом (медіаною), який ділитиме групу точно навпіл.

Основна перевага цього методу — простота. Однак його суттєвим недоліком є те, що маргінальні екзаменовані складають незначний відсоток від усіх екзаменованих, і експертам може бути складно їх однозначно ідентифікувати.

Методи, що базуються на судженнях щодо групи екзаменованих

Дані методи базуються на судженнях щодо групи екзаменованих (бажано великої групи), яка часто називається контрольною групою. За найпростішим з цих методів вибирається такий прохідний бал, який пропустив би певний відсоток екзаменованих у контрольній групі. Наприклад, якщо існує аргументоване твердження, що 85% торішніх екзаменованих були адекватно компетентними (це судження не базується на екзаменаційних балах), то вибирається бал, що пропустив би 85% торішніх екзаменованих, і такий бал використовується як прохідний для тесту поточного року. Якщо тест щорічно змінюється, то необхідно проводити процедуру урівнювання тесту поточного року та тесту, який складала контрольна група. Основний недолік цього методу полягає у тому, що досить складно оцінити точну кількість чи відсоток компетентних студентів серед торішніх екзаменованих. Проте якщо таке судження можна винести з прийнятним ступенем точності, то цей метод може слугувати своєрідним реальним контролем для методів, що базуються на судженнях щодо тестових завдань.

Компромісні методи

Методи визначення прохідного бала, які поєднують нормо-орієнтовані та критеріально-орієнтовані підходи, називаються компромісними. Процедури, розроблені Hofstee [32], Beuk [33] та De Gruijter [34], різняться у деталях, проте всі вони вимагають від експертів прийняття рішень щодо прохідного бала та відсотка екзаменованих, які успішно складуть тест. Для прийняття цих рішень експерти повинні знати зміст тестового іспиту і рівень компетентності екзаменованих.

Компромісні методи використовуються як запасний варіант при застосуванні будь-якого іншого методу визначення прохідного бала. Після того як експерти визначили прохідний бал методом, наприклад, Недельського, їм пропонується дати відповідь на такі запитання:

Яким, на вашу думку, буде відсоток екзаменованих, які вперше складають цей іспит і успішно його складуть? (Коротко обґрунтувати рішення);

Який, на вашу думку, прохідний бал, прийнятний для цього іспиту? (Коротко обґрунтувати рішення).

Суть *методу Hofstee* полягає у перегляді експертами тестових завдань іспиту; після ознайомлення зі змістом іспиту експерти встановлюють такі значення, що використовуються при встановленні критерію:

- мінімальний прийнятний відсоток екзаменованих, які не складуть іспит;
- максимальний прийнятний відсоток екзаменованих, які не складуть іспит;
- мінімальний прохідний бал;
- максимальний прохідний бал.

Після тестування будується графік залежності відсотка екзаменованих, які не склали іспит, від прохідного бала (процентної кількості правильних відповідей). Вказані вище чотири отриманих значення формують прямокутник на цьому графіку. Далі проводиться діагональ від верхнього лівого кута прямокутника до нижнього правого. Точка перетину діагоналі з кривою визначає числове значення прохідного бала.

Розглянуті критерії свідчать, з одного боку, про важливість здійснення процедур встановлення реальних критеріїв прийняття рішення відносно достатнього рівня професійної компетентності, з іншого – про необхідність залучення кваліфікованих експертів і додаткових джерел для прийняття такого рішення.

Основні зауваження

1. Звітні форми за результатами тестування мають містити настільки повну інформацію, наскільки це забезпечується можливостями методу тестування. Інформація може бути повною мірою використана лише тоді, коли при створенні тесту реалізовано усі етапи його проектування: визначено домен оцінювання, побудовано матрицю тесту, створено тестові завдання і проведено відповідну специфікацію.
2. Звітні форми мають забезпечувати реалізацію зворотного зв'язку з навчальним процесом. Отже, ця інформація дає можливість після її детального аналізу впливати на навчальний процес. В іншому випадку вона не є такою, що формує, вона лише констатує.
3. Дані у звітних формах за інформативністю мають забезпечити здійснення моніторингу та управління якістю підготовки. Отже, будуючи систему моніторингу, слід знати, якими даними ми можемо і повинні користуватися, а вже тоді – як ці дані отримати, тобто виміряти.

Частина 2. Проектування стандартизованого тесту

Засвоєння основних етапів розробки тесту

*Створення якісних тестів відповідно до
етапів розробки*

Визначення якісних показників тесту

*Встановлення ступеня валідності
результатів тестування*

2.1. Етапи розробки валідного тесту

Проектування тесту

Використання тестових іспитів дає можливість провести педагогічні виміри, що сприяють прийняттю рішень з мінімальною похибкою. Те, якою мірою вдасться скоротити кількість похибок і, відповідно, збільшити кількість правильних рішень, прямо залежить від якості тесту. Створення тестового інструменту з добрими психометричними показниками потребує проходження певних етапів процесу розробки:

1. Визначення мети проведення іспиту.
2. Опис домену, що діагностується.
3. Вибір методу вимірювання, який забезпечуватиме валідність.
4. Проектування матриці змісту тесту.
5. Вибір форматів тестових завдань.
6. Формування бази даних тестових завдань.
7. Формування тесту та його специфікація.
8. Визначення логістики тестування.
9. Вибір аргументованого прохідного бала.
10. Проведення тестування.
11. Обробка та аналіз результатів.
12. Встановлення валідності методу оцінювання.

Ці етапи запропонували і структурували Hambleton і Zaal [35], Impara [36], узгодивши з технічними стандартами для освітніх і психологічних тестів, створених спільно Американською асоціацією освітніх досліджень, Американською психологічною асоціацією і Національною радою США з вимірювання в освіті (1999).

Розглянемо докладніше основні проблеми, що постають перед розробниками ліцензійних і сертифікаційних тестових екзаменів, від успішного вирішення яких залежить успіх самого екзамену.

Мета проведення іспиту

Дуже часто (причому зовсім необґрунтовано) ми починаємо розробляти тест і створювати тестові завдання, не визначивши попередньо мету тестування. Тому, перш ніж починати розробку тесту, необхідно визначити області, які будуть оцінюватися (області прийняття рішень або домени) та типи висновків, які ми при цьому бажаємо зробити.

Існують різні класифікації цілей проведення іспитів. Gronlund [37] запропонував таку:

- з метою розподілу екзаменованих на групи (рівні);
- з метою поточного контролю;
- з метою діагностики;
- з метою підсумкового контролю.

Mehrens і Lehman [38] відзначають, що іспити проводяться також для:

- прийняття освітніх рішень, таких як оцінка успішності та діагностика;
- прийняття консультативних рішень у професійній, освітній та особистісній сферах;
- прийняття адміністративних рішень, таких як відбір, розподіл кандидатів на групи (рівні), оцінювання навчального плану;
- прийняття дослідницьких рішень.

На різних етапах навчання ми можемо робити різні висновки щодо *статусу екзаменованого*. Іспити, що проводяться до навчання, мають на меті розділити екзаменованих на групи (рівні). Таке тестування виявляє рівень підготовленості екзаменованого і допомагає йому бути обраним на відповідну програму навчання. Прийняття рішень щодо *рівня досягнень* вимагає існування попередньо визначеного *критерію*. На етапі до навчання результати такого іспиту сприяють прийняттю рішення щодо відбору кандидатів. У процесі освоєння змісту навчального плану для визначення сильних і слабких сторін студентів (учнів) використовуються діагностичні тести. Іспити, які проводяться по завершенні навчання, мають оцінити успішність студентів (учнів).

Тестування під час навчання проводиться ще й з метою *корекції*. Такий іспит може не відрізнитися від діагностичного іспиту, оскільки вдало розроблений тестовий екзамен може виконувати як діагностичну, так і коригувальну функцію. Тобто результати діагностичного тесту мають виявляти як сильні сторони екзаменованих, так і труднощі, які у них виникають. На базі отриманих результатів можна розробити відповідну коригувальну програму. По закінченні навчання на підставі екзаменаційних балів приймається рішення про переведення студентів (учнів) на наступний навчальний семестр (рік).

Висновки щодо успішності групи чи системи переважно мають *дослідницький* та оцінювальний характер. Наприклад, ми можемо використати показники успішності групи чи класу для визначення переваг двох різних навчальних планів. Слід зауважити, що іспити, які проводяться протягом навчання, а також по закінченні його з метою переведення (на наступний курс), оцінювання успішності та звітування, теж використовуються для зворотного зв'язку. Крім того, студенти (учні) хочуть дізнатися про свій прогрес, а також про власні сильні та слабкі сторони; викладачі ж хочуть оцінити успішність своїх студентів (учнів) та якість викладання змісту і набутих когнітивних умінь.

Іспити можуть також проводитися для *прогнозування успішності* екзаменованого при складанні певного критеріально-орієнтованого тесту. В сфері професійної орієнтації тести використовуються для передбачення, в якій професії екзаменований найбільш імовірно досягне успіху. При застосуванні прохідного бала можна зробити висновки щодо відбору і/або набору.

Більшість тестів спрямовано на оцінювання доменів навчального плану та когнітивного. Для сертифікаційного іспиту наші судження переважно стосуються статусу екзаменованого щодо досягнень успішності. Тобто на підставі екзаменаційного бала (балів), отриманого за тест, ми визначаємо статус екзаменованого в аспекті знань, умінь та професійного ставлення, що описують домен. Проте якби перед нами стояло завдання відібрати найкращих студентів для програми навчання, мета проведення іспиту і висновки, зроблені за результатами його проведення, були б іншими.

Встановлення мети проведення іспиту та типів висновків, які буде зроблено, визначають решту процесу розробки тесту.

Опис домену, що діагностується

Першочерговим завданням є чітке визначення домену, який буде оцінюватися екзаменом.

Залежно від мети іспиту (оцінювання) домен може включати: зміст навчального плану та програм навчання, вимоги освітнього стандарту, зміст компетентностей тощо. Існують різні підходи до аналізу та опису домену відповідно до типу тесту.

Для предметного тесту – це відповідність кінцевим цілям вивчення предмета, відповідність етапу навчальної програми, залежність від форми оцінювання та висновків на основі отриманих результатів.

Для атестаційного, сертифікаційного тесту – це відповідність освітнім стандартам, вимогам щодо прогнозування успішного засвоєння професії (вміння, що вимагаються від фахівця), “європейським компетентностям”.

Вибір методу вимірювання

Залежно від мети оцінювання, домену, необхідно вибрати відповідний метод вимірювання.

Так, під час оцінювання когнітивного домену може бути використано метод тестування. А залежно від мети оцінювання може бути розроблено або нормо-орієнтований, або критеріально-орієнтований тести.

Нагадаємо, що тест, який визначає статус особи відносно результатів інших осіб за цей самий тест, називається *нормо-орієнтованим*. Значимість результату екзаменованого виявляється у порівнянні з іншим екзаменованим або певною нормативною групою. Цей тест застосовується при відборі. З іншого боку, нормо-орієнтовані тести є несумісними з сучасною тенденцією забезпечення компетентності. Результати таких тестів не виявляють наявності певних умінь в екзаменованих; вони просто сортують екзаменованих за рівнем їхнього оволодіння набором знань і вмінь.

Критеріально-орієнтовані тести, на відміну від нормо-орієнтованих, вимірюють ті елементи компетентності, якими оволоділи (чи не оволоділи) екзаменовані. Критеріально-орієнтований тест встановлює статус екзаменованого відносно детально визначеного домену діяльності. Відмінною рисою критеріально-орієнтованих тестів є їх здатність давати чіткий опис вимірюваних умінь. Оскільки першочерговим завданням ліцензійних чи сертифікаційних іспитів є захист громадськості від некомпетентності фахівців, модель вимірювання мусить бути критеріально-орієнтованою, а не нормо-орієнтованою. Тобто результат іспиту має відбивати знання та компетентність екзаменованого. Як екзаменований виглядає порівняно з іншими екзаменованими, не має практичного значення для ліцензійного чи сертифікаційного процесу.

Матриця змісту тесту

Для визначення змісту та поведінкових аспектів, які будуть оцінюватися тестом, зазвичай розробляється структура тесту, або матриця. Така матриця допомагає визначити, який об'єм матеріалу включається в тест та на чому саме наголошується.

Матриця тесту – це структура організації а також певний обсяг освітнього домену (знань, умінь, навичок, цінностей), що виноситься на оцінювання.

На додаток до двох вимірів (змісту й аспектів діяльності) можна включати інші виміри, що відбивають такі параметри, як рівні засвоєння, освітній стандарт, компетентність тощо. Ці деталі спрямовують роботу зі створення тестових завдань і дають змогу уникати звуження домену оцінювання. Крім цього, така

структура допомагає робити обґрунтовані висновки у контексті різних компонентів виміру змісту.

Основні етапи, які необхідно здійснити для проектування матриці тесту, що є базисом створення якісного тесту, такі:

- формулювання мети іспиту, визначення домену;
- вибір методу вимірювання;
- опис домену, що діагностується;
- структурування домену, що діагностується (визначення важливих вимірів), наприклад:
 - теми, навчальних елементів;
 - рівнів засвоєння знань;
 - аспектів діяльності;
 - дисциплін, навчальних модулів;
- “наповнення” N-вимірної матриці детальним змістом;
- визначення співвідношення окремих елементів (“клітинок”) матриці;
- встановлення специфікації тестових завдань (відповідність “клітинці” матриці).

Домен і матриця мають бути достатньо загальними і компактними, щоб ними могли користуватися автори тестових завдань та викладачі під час аналізу результатів. Утім, домен і матриця мають достатньо конкретно окреслювати межі елемента домену, що діагностується. Для цього необхідне структурування домену. Чим краще структуровані домен і матриця, тим інформативнішим буде аналіз результатів оцінювання та ефективнішим – використання результатів.

Наведемо приклад шаблону матриці тесту.

Структура змісту тестового іспиту

Зміст	Рівні засвоєння			Кількість, %
	1	2	3	
Тема 1	5	3	7	15
Тема 2	9	–	6	15
Тема 3	10	5	5	20
Тема 4	10	10	–	20
Тема 5	5	10	15	30
Кількість, %	30	50	20	

Вибір форматів тестових завдань

Для вимірювання знань, навичок і професійного ставлення нині використовується значна кількість різноманітних форматів тестових завдань. Виходячи з того, що результати тестування мають відбивати структурування й організацію отриманих знань, навичок і ставлень екзаменованого, та враховуючи, що їх

застосування відображує міркування і мислення екзаменованого, адекватні рівню освіти та його досвіду, необхідно забезпечити якомога більшу відповідність між тим, “що вимірюється”, і тим, “як вимірюється”. Необхідним є також усвідомлення того, що оцінювання знань, навичок і ставлень є складним і багатогранним процесом, і для забезпечення відповідного оцінювання необхідно використовувати різноманітні тестові формати.

Формування тесту та його специфікація

Для забезпечення бажаних параметрів тесту (валідності, надійності, точності та передбачуваності) необхідно розробити специфікацію, яка міститиме таку інформацію:

- мета проведення іспиту;
- зміст, аспекти діяльності (елементи компетентності) у сфері оцінювання;
- формат іспиту (бланковий, комп’ютерний, лабораторний, практично-орієнтований);
- формат тестового завдання;
- формат відповідей;
- критерії та процедури оцінювання;
- психометричні властивості (наприклад, середнє значення за тест, індекс складності та дискримінації);
- інструктаж для екзаменованих.

Більшість зазначених параметрів є досить однозначними. Крім того, для визначення змісту та поведінкових аспектів, які будуть оцінюватися тестом, зазвичай розробляється структура тесту, або матриця. Така матриця допомагає визначити, який об’єм матеріалу включається в тест і на чому робиться акцент.

Докази валідності та надійності результатів

Як уже зазначалося, одним із завдань ліцензування і сертифікації є розробка тестових інструментів, які ефективно визначають елементи компетентності, необхідні для обізнаної практичної діяльності. Розглянемо дві основні характеристики вдало розробленого тестового інструменту.

Результати тесту можуть бути корисними лише тоді, коли вони *валідні* та *надійні* [39].

Валідність – це доцільність, значимість і корисність певних висновків, зроблених за результатами тестування. Валідизація тесту – це процес збору доказів для обґрунтування зроблених висновків. Для ліцензійних і сертифікаційних іспитів основний акцент робиться на доказах валідності вимірюваних результатів.

Надійність – це стійкість результатів. Якщо результати тестування нестійкі, то висновки, зроблені на підставі цих результатів, є ненадійними. Надійність також включає точність (безпомилковість) результатів. Успішність екзаменованого може коливатися під час різних тестувань з причин, що не пов'язані з метою вимірювання. Екзаменований може бути краще підготовленим, більше нервуватися, бути більше або менше втомленим, через що його успішність буде варіюватися. Такі коливання називаються похибками, що знижують надійність результатів тестування. Тому необхідно виявити основні причини похибок вимірювання, вирахувати їх величину та вплив на надійність. Оскільки більшість ліцензійних і сертифікаційних іспитів включає прохідний бал для визначення тих, хто склав чи не склав такий іспит, слід повідомляти стандартну похибку вимірювання для результатів на межі прохідного бала чи близько до неї.

Вибір раціонального та аргументованого прохідного бала

При нормативному підході прохідний бал встановлюється в одиницях стандартного відхилення на певній відстані від середнього значення розподілу екзаменаційних балів. Наприклад, прохідний бал може встановлюватися нижчим за середній на 1,5 стандартних відхилення. Варіантом такого підходу є завчасне встановлення кількості тих, хто не складе тест (у відсотковому відношенні). Наприклад, 4% найгірших результатів вважатимуться незадовільними. У будь-якій з цих ситуацій певна частина екзаменованих отримає незадовільну оцінку незалежно від складності іспиту або їхньої компетентності.

Такий підхід має ряд недоліків: по-перше, завчасно визначено відсоток тих, хто не складе іспит, незалежно від компетентності екзаменованих; по-друге, є недоречним порівняння відносної успішності екзаменованих для визначення їхньої компетентності; по-третє, такий підхід є значною мірою упередженим.

В останні роки значного поширення набувають методи визначення прохідного бала, що базуються на рішеннях експертів щодо очікуваної або необхідної успішності екзаменованих. Процедури визначення прохідного бала поділяються за типом судження, якого вимагає їх застосування: судження щодо змісту тесту чи тестових завдань та судження щодо груп екзаменованих.

Урівнювання тесту

У практиці тестування існує поняття форми тесту. В ідеалі альтернативні форми тесту мають бути взаємозамінними. Для екзаменованого не повинно мати значення, яка саме форма тесту використовується, оскільки складність тесту з року в рік не має змінюватися.

Випадкові коливання загальної складності тесту при проведенні тестувань у різний час нададуть перевагу тим екзаменованим, які складатимуть більш легку форму тесту. Оскільки екзаменовані, складаючи два тести з одним змістом, здебільшого не отримає однакових результатів, виникає необхідність у розробці процедури конвертації балів, отриманих після проведення одного тесту, в одиниці іншого тесту. Процес визначення взаємозв'язків результатів тестування називається *урівнюванням тесту* і включає трансформацію балів одного тесту в еквівалентні бали іншого. Першочергове призначення урівнювання полягає у тому, щоб максимально забезпечити ідентичність прохідного бала різним формам або методам проведення тесту. Якщо неможливо провести урівнювання для іспитів, які складала невелика кількість екзаменованих, то принаймні необхідно провести дослідження стабільності результатів протягом певного періоду часу.

Фальсифікація

Фальсифікація на екзамені переважно асоціюється зі списуванням, проте в контексті ліцензійних і сертифікаційних іспитів вона може набути таких форм:

- подання підроблених посвідчень чи неправдивої інформації;
- використання підставної особи для складання іспиту;
- користування шпаргалками під час складання іспиту;
- списування;
- несанкціонований доступ до конфіденційних екзаменаційних матеріалів.

З цих п'яти форм дві останні зустрічаються найчастіше. Залежно від результатів аналізу екзаменаційний бал може бути анульовано. *Тобто ліцензійна чи сертифікаційна організація повинна мати право анулювати результати тестування.* Оскільки екзаменовані можуть оскаржити таке рішення, необхідно описати заборонені дії, зазначити санкції, які можуть бути застосовані, повідомити про наявність юридичних або нормативних повноважень на вживання заходів у разі виявлення фальсифікації. *Ліцензійний або сертифікаційний екзамен не може бути зарахований на підставі результатів, отриманих особою за рахунок завчасного доступу до екзаменаційних матеріалів.*

2.2. Приклади стандартизованих тестів

На прикладі поширених відомих стандартизованих тестів проаналізуємо етапи побудови якісного інструменту педагогічного оцінювання, про які йшлося раніше.

По-перше, розглянемо приклади відомих, а отже, вже визнаних стандартизованих тестів: вступного тесту до вищих навчальних закладів США (SAT) і сертифікаційних тестів для студентів медичних спеціальностей України (ліцензійні інтегровані іспити).

По-друге, розглянемо приклади предметних тестів, розроблених слухачами Міжнародної літньої школи 2005-го, яку проводив Центр тестування при МОЗ України в межах україно-європейського проекту “Справедливе оцінювання” програми TEMPUS/TACIS.

Наведені приклади містять фрагменти тестів, розроблених із дотриманням основних етапів проектування тесту та таких, що відповідають певним процесуальним, змістовим і технологічним вимогам та оцінюють кінцевий рівень екзаменованого відповідно до визначених вимог.

При підготовці інструменту педагогічного оцінювання (тесту) слід дотримуватися основних етапів його створення. У наших прикладах відображено такі:

- встановлення мети оцінювання;
- опис домену змісту, що оцінюється, та навчальних цілей;
- вибір методу оцінювання;
- проектування структури тесту (матриці) ;
- вибір форматів тестових завдань.

Нагадаємо, що мета оцінювання визначає зміст оцінювання або домен змісту, який своєю чергою структурується у формі багатовимірної матриці змісту, що допомагає вибрати відповідні формати тестових завдань, з яких створюється тест.

Тест перевірки загальних академічних здібностей

*Мета та метод оцінювання
Домен змісту та навчальні цілі
Матриця тесту
Формати тестових завдань*

Розгляд стандартизованих тестів почнемо з найпоширенішого у світі, створеного ще у 1926 р. американського тесту перевірки загальних академічних здібностей, більш відомого як SAT.

При вступі до вузів США або Канади на будь-яку бакалаврську програму підготовки не потрібно здавати вступні іспити. Необхідна лише здача загальнонаціонального тесту SAT¹ (Scholastic Aptitude Test) [5] з результатом вище 700 балів, що здають усі випускники шкіл США, або АСТ (American College Test), а також обов'язкового тесту для всіх іноземних студентів TOEFL (Test of English As a Foreign Language) з результатом вище 500 балів [40]. Кожен вуз самостійно встановлює прохідний бал. Атестація проводиться в спеціальних центрах здачі національних тестів, а результати реєструються у загальнонаціональному комп'ютері. Складений один раз тест можна використовувати для вступу до будь-якого вузу США.

Процедура зарахування потребує заповнення великої кількості анкет, форм, надання свідоцтва про закінчення школи, рекомендацій, написання есе, а також виконання ряду вимог, пов'язаних з одержанням спеціальної студентської візи F. Зазвичай, термін зарахування до вузів США становить не менше чотирьох місяців, а у деяких найпрестижніших – більше ніж півроку. В університетах США запроваджено систему deadline (останній день подачі документів) – переважно за шість місяців до початку занять.

Первісно під SAT розуміли іспит на академічну придатність, а предметні тести називали тестами досягнень. У 1993 р. назви було змінено відповідно на “SAT I: Інтелектуальний тест” і “SAT II: Предметні тести”. У 2004 р. назви знову було змінено на “SAT Інтелектуальний тест” і “SAT Предметні тести”. Таким чином, іспит складається з двох тестів, які зазвичай називають: SAT² і SAT II³.

У передвипускному класі середньої школи (junior year) проводиться PSAT (Preliminary Scholastic Aptitude Test) [41], що є більш короткою версією SAT.

¹ Аббревіатура SAT означає - тест перевірки загальних академічних здібностей.

² Під SAT розуміють The SAT Reasoning Test - SAT Інтелектуальний тест (до 2004 р. — SAT I: Scholastic Aptitude Test I).

³ Під SAT II розуміють The SAT Subjects Tests - SAT Предметні тести (до 2004 р. — SAT II: Subject Tests).

Він має подвійну мету: орієнтує майбутніх абітурієнтів на задачу SAT і є кваліфікаційним іспитом для одержання стипендій, що присуджуються Національною Корпорацією Стипендій за Заслуги. Результати SAT учнів випускних класів (senior year) направляються до університетів, в які вони збираються вступати.

Кожен тест складається з кількох частин. Так, SAT оцінює загальні навички з читання, письма та математики. Частини тесту оцінюються незалежно: за кожну частину може присуджуватися оцінка від 200 до 800, а загальна оцінка за тест – від 600 до 2400. Таким чином, середня оцінка кожної частини – 500, а середня оцінка за тест – 1500 [42]. Середній бал прийнятих у 2003 р. абітурієнтів по SAT і SAT II складав 1400–1580 (за кожен тест), для ACT – 29,8 з 36 можливих.

Додатково до оцінки у балах визначається ранг у процентилях, що визначає відсоток студентів з оцінками нижче даної. Наприклад, якщо ваш перцентиль 80, то це означає, що ваш результат кращий, ніж у 80-ти з кожних 100 абітурієнтів, які писали тест.

Проведенням SAT керує Рада Коледжів¹, а розробкою, публікацією буклетів і підрахунком балів займається Служба освітнього тестування (ETS)², утворена ще в 1947 р. Фондом Карнегі з розвитку освіти та Екзаменаційною Радою зі вступу до коледжів. Рада Коледжів – це не комерційна асоціація, заснована 1900 р., метою якої є забезпечення стандартизованої форми проведення SAT для всіх коледжів і студентів. Коледжі використовують оцінки за тест як один з критеріїв при вирішенні питань щодо прийому.

Кількість тих, хто здавав SAT серед випускників середньої школи в 2005 р. досягла дуже високого рівня – 1.475.623. Випускники середньої школи з родин з низькими доходами грошовий збір не сплачують.

Мета та метод оцінювання

SAT є загальнонаціональним критеріально-орієнтованим стандартизованим іспитом, що складається з двох психометричних тестів – SAT і SAT II. Його завдання полягає у виявленні та перевірці саме тих здібностей і особливостей, які необхідні для успішного навчання в університеті або коледжі.

SAT проводиться до навчання з метою прийняття адміністративних рішень, таких як відбір і розподіл кандидатів на групи (рівні) на основі попередньо визначеного критерію – придатності до подальшого навчання. Процедури вимірювання, шкалювання та організації процесу тестування чітко регламентовані.

¹ College Board - <http://www.collegeboard.com>

² Educational Testing Service (ETS) - <http://www.ets.org>

Для апробації нових ТЗ із майбутніх форм тесту і гарантії того, що оцінки майбутніх варіантів SAT можна порівнювати з оцінками більш ранніх форм, призначено а п р о б а ц і й ну частину тесту. Ця частина не входить до оцінки. Вона може містити завдання з читання, письма або з математики.

SAT оцінює критичне мислення і здатність вирішувати проблеми, в той час як SAT II призначений для перевірки глибини знань студента в конкретній сфері, а також того, наскільки добре студент здатний застосувати ці знання. Деякі з найпрестижніших американських університетів і коледжів вимагають від абітурієнтів здавати ще й одну з двадцяти двох версій спеціального тесту SAT II, що перевіряє їхні знання в одній з п'яти галузей (англійська мова, історія, математика, природничі науки, іноземні мови), в яких вони бажають спеціалізуватися.

Багато коледжів використовують результати SAT II не тільки для вирішення питання про вступ, а й для видачі рекомендацій щодо вибору курсу (профорієнтації). При використанні в комбінації з іншою інформацією ці результати дають можливість також зробити позитивний прогноз щодо майбутніх досягнень.

Домен змісту, що оцінюється, та навчальні цілі

SAT

Для випускників 2006 р. зміст і структуру SAT було змінено з 12 березня 2005 р. Здачу тестів з навичок письма і ELPT (English Language Proficiency Test) припинено після січня 2005-го. *Нова версія* SAT не складніша за попередню і краще погоджена з існуючою програмою навчання. З неї виключено деякі розділи, зокрема аналогії та кількісні порівняння, але включено завдання з читання уривків, написання есе, багатовибіркові питання на правопис і стиль. Через включення третьої кваліфікаційної вимоги (написання есе), надзвичайно важливим для абітурієнтів є уміння грамотно писати. Розділ, відомий нині як Читання, раніше називався вербальним (усним). Розділ Математика змінено, щоб включити матеріал, який вивчався протягом трьох років на курсі математичної підготовки до коледжу в середній школі.

Отже, SAT призначений для перевірки:

- уміння читати і розбирати зміст прочитаного (critical reading);
- математичного мислення;
- уміння грамотно писати.

Він визначає загальний освітній рівень, навчальні/академічні навички й уміння, отримані в американській середній школі (переважно вербальні і математичні знання, а також уміння абстрактно і логічно мислити).

SAT II

SAT II поділяється на п'ять основних частин:

- англійська література;
- історія і суспільні науки: історія США, всесвітня історія;
- математика: математика I рівня, математика II рівня;
- природничі науки: біологія E/M¹, хімія, фізика;
- іноземні мови: китайська зі сприйняттям на слух, французька, французька із сприйняттям на слух, німецька, німецька зі сприйняттям на слух, іспанська, іспанська зі сприйняттям на слух, сучасний іврит, італійська, латинська, японська зі сприйняттям на слух і корейська зі сприйняттям на слух.

Тест визначає рівень знань з трьох конкретних дисциплін на вибір абітурієнта й залежно від напрямку підготовки, в якому він має намір спеціалізуватися. Наприклад, тест може складатися з історії, літератури і географії. Абітурієнти, що планують спеціалізуватися, наприклад, в економіці, мають обов'язково включити в SAT II математику і письмову англійську мову.

Структура (матриця) тесту

SAT

Тестування триває 3 год 45 хв. При визначенні оцінки зараховуються тільки 3 год і 20 хв – апробаційна частина не зараховується. Усього в тесті 10 частин. Структуру тесту умовно наведено в таблиці.

Структура SAT

Частини	Кількість, тип/зміст завдань	Тривалість, хв
Читання	19 – завершення речення	
	48 – розуміння прочитаного	
	Усього багатовибіркових запитань – 67	70
Математика	44 – багатовибіркові ТЗ	
	10 – із продюкованою відповіддю	
	Усього запитань – 54	70
Письмо	49 – багатовибіркові ТЗ	
	1 – есе	
	49 запитань + есе	60
Апробаційна	Читання, письмо або математика	25

ПРИМІТКА: порядок проходження частин є довільним. Однак 25-хвилинне есе завжди на початку тесту, а 10-хвилинна секція багатовибіркових ТЗ на перевірку грамотного письма – завжди наприкінці.

¹ З акцентом або на екологію (Біологія E), або молекулярну будову (Біологія M).

Розглянемо докладніше кожну з частин тесту.

Читання

Зміст цієї частини наведено нижче:

Час	Зміст	Тип / зміст завдань	Бали
70 хв (дві 25-хвилинні секції й одна 20-хвилинна)	Читання	Багатовибіркові ТЗ на завершення речень та розуміння прочитаного	200–800
I секція	10 – завершення речень, 15 – розуміння прочитаного		
II секція	9 – завершення речень, 15 – розуміння прочитаного		
III секція	18 – розуміння прочитаного		

Ця частина, раніше відома як вербальна, перевіряє словниковий запас, уміння розуміти й аналізувати зміст прочитаного. Передбачається читання як коротких, так і довгих уривків, зміст яких стосується галузі природничих, гуманітарних і соціальних наук, художньої літератури. Аналогії виключено, однак запитання на завершення речень і запитання на розуміння прочитаного залишено. При створенні ТЗ передбачалося, що екзаменовані повинні:

- розуміти значення слів і те, як різні частини речення логічно пов'язані між собою;
- робити і формулювати логічні висновки;
- диференціювати основні й допоміжні ідеї;
- додержуватися логіки суперечки;
- розпізнавати жанри.

Математика

Зміст цієї частини наведено нижче:

Тривалість	Зміст	Типи завдань	Бали
70 хв (дві 25-хвилинні частини й одна 20-хвилинна)	Числа й операції; алгебра і функції; геометрія; аналіз даних, статистика й імовірність	Багатовибіркові ТЗ, ТЗ із продукуваною відповіддю	200–800
I секція	20 – багатовибіркові		
II секція	10 – багатовибіркові 10 – із продукуваною відповіддю		
III секція	14 – багатовибіркові		

Тест містить ТЗ на розширені теми: зростання показникової функції, абсолютне значення, запис функцій з акцентом на таких розділах, як лінійні функції, дії з показниками степеня, властивості дотичних.

Необхідні уміння, що вимірялися раніше у форматі кількісного порівняння, наприклад, оцінка і значення числа, будуть вимірюватися за допомогою багатовибіркових ТЗ і ТЗ із продукованою відповіддю (student-produced responses).

Зміст конкретних розділів тесту такий:

Числа й операції

Показниково зростаючі послідовності, що називають також геометричними прогресіями, які використовуються в реальному житті, наприклад, зростання населення.

Множини. Якщо множина є набором предметів, то “предмети” можуть назватися “елементами” множини. ТЗ можуть стосуватися об’єднання двох множин (множина, що складається з елементів, присутніх у будь-якій або в обох множинах) або перетинання двох множин (набір спільних елементів).

Алгебра і функції

Абсолютна величина. Екзаменовані мають розумітися на понятті абсолютної величини і на її визначенні, бути здатними працювати з виразами, рівняннями, і функціями, що містять абсолютні величини.

Раціональні рівняння і нерівності типу $\frac{2x+1}{x-3}$.

Ірраціональні рівняння типу $5\sqrt{x}+14=29$ з цілими і раціональними показниками степеня.

Пряма і зворотна залежності. Тест може включати запитання, що стосуються прямо і обернено пропорційних величин.

Функції. Екзаменовані мають розумітися на понятті функції та її властивостях. Наприклад, якщо $f(x) = x + 2x$, то абітурієнти мусять знати, що $f(5) = 5 + 25 = 30$. Якщо f – лінійна функція і якщо $f(6) = 7$ і $f(8) = 12$, то абітурієнти повинні вміти визначити кут нахилу графіка функції $f(x)$.

Поняття області визначення й області зміни. Тест включає запитання про значення x , в яких дана функція не визначена, або про значення $f(x)$, яких вона не може набувати.

Моделювання. Тест може включати запитання, що стосуються математичних моделей реальних ситуацій, і вимагати представлення цієї моделі у вигляді графіка або рівняння.

Лінійні функції – рівняння і графіки. Тест може включати запитання, що стосуються побудови графіків лінійних функцій.

Квадратичні функції – рівняння і графіки. Тест може включати запитання, що стосуються побудови графіків квадратичних функцій.

Геометрія

Визначення довжини відрізків і сегментів; ліній, променів і конгруентності. Тест може включати визначення, що використовуються зазвичай у середній школі.

Задачі, в яких як альтернативний метод розв'язання може бути використана тригонометрія. Тест може включати дуже різні запитання про властивості 45–45–90- або 30–60–90-градусних трикутників. Ці задачі можуть бути розв'язані як тригонометричними методами, так і з використанням інших (геометричних) методів.

Властивості дотичної.

Аналітична геометрія. Деякі ТЗ можуть вимагати знання властивостей паралельних і перпендикулярних ліній, уміння складати рівняння лінії, знаходити середні точки лінійних сегментів або відстань між двома точками на координатній площині.

Якісна поведінка графіків і функцій.

Перетворення та їх вплив на графіки функцій. Деякі запитання можуть вимагати знання впливу простих перетворень на графіки функцій. Наприклад, за графіком функції $f(x)$ побудувати графік функції $f(x + 2)$.

Аналіз даних, статистика та ймовірність

Інтерпретація даних, точкових діаграм (scatterplots) і матриць. Тест може включати запитання, що стосуються побудови лінії, яка щонайкраще описує набір розрізнених точок, наведених на рисунку. Наприклад: визначати, чи має ця лінія додатний нахил, але менший 1. Передбачається, що екзаменовані мають інтерпретувати дані, відображені в таблицях, діаграмах і графіках.

Геометрична ймовірність. Деякі запитання можуть вимагати знання геометричної ймовірності. Наприклад, якщо точка довільно вибирається з області, частина якої затінена, як знайти ймовірність того, що вибрана точка буде із затіненої частини області.

Огляд змісту математичної частини нової версії SAT

Рівень нової версії тесту підвищено для долучення більш актуальних тем. Нова версія містить:

Числа й операції

Арифметичні вирази, еквівалентні текстовому опису (включаючи відсотки, коефіцієнт і пропорцію).

Властивості цілих чисел (парні, непарні, прості числа, подільність тощо).

Раціональні числа.

Логічні вирази.

Множини (об'єднання, перетинання, елементи).

Техніка рахунку.

Послідовності та ряди (включаючи геометричні прогресії).

Елементарна теорія чисел.

Алгебра і функції

Підстановка і спрощення алгебраїчних виразів.

Властивості показників степеня.

Алгебраїчні вирази, еквівалентні текстовому опису.

Розв'язання лінійних рівнянь і нерівностей.

Системи рівнянь і нерівностей.

Квадратні рівняння.

Раціональні й ірраціональні рівняння.

Рівняння лінії.

Абсолютна величина.

Прямо і обернено пропорційна залежність.

Поняття алгебраїчної функції.

Нові символи, що застосовуються у звичайних операціях.

Геометрія

Площа і периметр багатокутника.

Площа круга і довжина кола.

Об'єм прямокутника, куба і циліндра.

Теорема Піфагора і властивості рівнобедреного, рівностороннього і прямокутного трикутників.

Властивості паралельних і перпендикулярних ліній.

Аналітична геометрія.

Наочність геометричної побудови.

Нахил.

Подібність.

Перетворення.

Аналіз даних, статистика та ймовірність

Інтерпретація даних.

Статистика (середнє значення, медіана і мода).

Ймовірність.

Письмо

Зміст цієї частини наведено нижче:

Тривалість	Зміст	Типи завдань	Бали
60 хв (25- та 10-хвилинна частини й 25-хвилинне есе)	граматика, використання і добір слів	багатовибіркові ТЗ (35 хв), есе (25 хв)	200–800

Коротке есе вимірює здатність:

- використовувати доречні слова і структуру речення;
- формулювати й обґрунтовувати основну ідею;

- організовувати, логічно, ясно і правильно висловлювати свої думки;
- формулювати своє відношення до проблеми з використанням аргументації і доказів, заснованих на власному досвіді, ерудиції або спостереженнях, що підтримують вашу точку зору;
- розвивати свою точку зору на проблему, що описується в уривку;
- додержуватися традицій стандартної письмової англійської мови.

При введенні розділу есе в письмову частину тесту передбачалося, що написаний протягом короткого проміжку часу текст есе не буде відшліфований, а являтиме собою початкову фазу процесу творення — чернетку.

Багатовибіркові ТЗ вимірюють здатність:

- покращувати речення й абзаци;
- ідентифікувати помилки стилю, граматики, побудови речень, узгодження дієслів, використання слів, багатослівності;
- додержуватися традицій стандартної письмової англійської мови.

SAT II

Тест складається з п'яти основних частин. Тестування триває одну годину для кожної з частин (дисциплін). Розглянемо докладніше деякі з них.

Природничі науки

Тест з *біології* оцінює розуміння загальної біології на рівні, достатньому для вступу до коледжу. *Біологія E* приділяє більшу увагу запитанням, що стосуються біологічних угруповань, народонаселення й енергетичного обміну. *Біологія M* розглядає запитання, що стосуються біохімії, структури клітини і процесів у ній (дихання і фотосинтез). Вибір позначається на бланку відповідей.

Очікувані здібності/уміння:

- відтворювати і розуміти основні концепції біології і застосовувати вивчені принципи для вирішення біологічних завдань;
- розуміти прості алгебраїчні поняття, включаючи відносини, прямі і зворотні пропорції; застосовувати ці поняття при розгляді виразів, еквівалентних текстовому опису;
- систематизувати й інтерпретувати результати спостережень і експериментів, робити висновки з експериментальних даних, додаючи дані, наведені в графічній і/або табличній формі;
- мати добрі знання з метричної системи одиниць.

Тест з *хімії* оцінює розуміння загальної хімії на рівні, достатньому для вступу до коледжу.

Очікувані здібності/уміння:

- розуміти основні хімічні концепції і застосовувати їх для розв’язання хімічних задач;
- систематизувати й інтерпретувати результати спостережень і експериментів, робити висновки з експериментальних даних, додаючи дані, наведені в графічній і/або табличній формі;
- мати досвід роботи в лабораторії та добрі знання метричної системи одиниць;
- оперувати простими алгебраїчними співвідношеннями і застосовувати їх при розгляді виразів, еквівалентних текстовому опису;
- демонструвати добре знання таких алгебраїчних понять, як відношення, пряма і обернена пропорція, показники степеня і логарифмічне зображення числа (у вигляді мантиси і порядку).

Тест з *фізики* оцінює розуміння фізики на рівні, достатньому для вступу до коледжу.

Очікувані здібності/уміння:

- відтворювати і розуміти основні концепції фізики і застосовувати вивчені принципи для розв’язання фізичних задач;
- розуміти прості алгебраїчні, тригонометричні і графічні співвідношення, включаючи відношення, прямі й обернені пропорції; застосовувати ці поняття при розв’язанні фізичних задач;
- мати добрі знання з метричної системи одиниць.

Математика

Широкий оглядовий тест з *математики 1 рівня* призначений для абітурієнтів, які прослухали трирічний підготовчий курс математики, що включає два роки алгебри й один рік геометрії.

Зміст тесту охоплює:

- числа й операції;
- алгебру і функції;
- геометрію і виміри (Евклідова плоска, аналітична, тривимірна, тригонометрія);
- аналіз даних, статистика, ймовірність.

Широкий оглядовий тест з *математики 2 рівня* призначений для абітурієнтів, що прослухали підготовчий курс математики понад три роки, що включає два роки алгебри, один рік геометрії плюс елементарна алгебра і/або тригонометрія.

Зміст тесту охоплює:

- числа й операції;
- алгебру і функції;
- геометрію і виміри (аналітична, тривимірна, тригонометрія);
- аналіз даних, статистика, ймовірність.

Формати тестових завдань

Формати тестових завдань незалежні від конкретного підручника або методичних вказівок. Зміст ТЗ змінюється, щоб відбивати поточні тенденції у програмах середньої школи, але типи запитань з року в рік залишаються майже незмінними.

При створенні багатовибіркових ТЗ особлива увага приділяється створенню дистракторів, оскільки значно складніше написати гарний дистрактор, ніж гарну відповідь. Звичайно пропонують тільки два привабливих дистрактори: один правильний, інший або навмисно провокує неправильну відповідь, або тільки частково правильний. Інші три дистрактори – відволікаючі (fluff). Однак це може підвищувати ймовірність вгадування.

SAT

Подібно до більшості тестів, запитання, як правило, розташовуються в порядку зростання складності, за винятком розділу читання, упорядкованого хронологічно.

Тест складається переважно з багатовибіркових ТЗ й есе. Проте існують деякі особливості в написанні ТЗ для різних частин тесту. Коротко розглянемо їх нижче. Приклади ТЗ з математики, фізики та хімії за 1995–2004 рр. можна знайти на сайті [43].

Читання

ТЗ на завершення речень частини SAT Читання можуть формулюватися таким чином. Кожне речення має один або два пропуски, що означають пропущені слова. Дистрактори містять слова або набори слів і позначені літерами від А до Е. Необхідно підібрати і вставити в речення таке слово або набір слів, які щонайкраще відповідали б змісту речення в цілому.

ТЗ на читання уривків частини SAT Читання перевіряють здатність абітурієнта розуміти і ретельно обмірковувати кілька різних уривків, що відрізняються за величиною (від 100 до 850 слів). Уривки беруться з різних галузей, вони різняться за стилем. Деякі ТЗ містять пари пов'язаних уривків, які необхідно порівняти або протиставити. До них може бути поставлено запитання, пов'язані з:

- контекстним словником: ці запитання вимагають визначення значень слів у контексті прочитаного уривка;
- точним розумінням: для оцінки ступеня розуміння важливої інформації, що безпосередньо міститься в уривку;
- загальним аналізом: ці запитання вимірюють здатність синтезувати й аналізувати інформацію, а також оцінювати зроблені припущення й використаний автором стиль. Більша частина запитань припадає саме

на цю категорію. Вас можуть попросити ідентифікувати причину й наслідок, зробити висновки, визначити головну ідею, простежити за логікою оповіді.

Усі багатовибіркові завдання оцінюються дихотомно (за двома ознаками: 0 або 1).

Математика

Застосовуються стандартні *багатовибіркові* ТЗ з п'ятьма дистракторами. Як відповідь вибирається найбільш правильна з наведених.

При використанні ТЗ із *продукованою (відкритою) відповіддю* варіанти відповідей не вказуються. Екзаменованим повинен розв'язати задачу і внести свою відповідь у спеціальний бланк. Техніка внесення подібна до внесення будь-якої іншої інформації, що автоматично зчитується. Основна перевага цього методу полягає в тому, що він дозволяє вводити отриману відповідь незалежно від її вигляду: ціле число, десятковий, або простий дріб [44]. Наприклад, якщо ви одержуєте $2/5$, можете внести $2/5$. Якщо одержуєте $0,4$ – можете внести $0,4$. Таблиця містить тільки числа, що змінюються від 0 до 9999.

Зразок бланка відповідей з таблицями наведено на рис. 11.

Якщо відповідь займає діапазон значень, то внесена відповідь має бути у межах цього діапазону. Наприклад, якщо в розглянутій задачі правильними відповідями є всі числа між 4 і 5, то правильною буде відповідь $4,0002$ ($4 < x < 5$), а округлена величина $4,00$ не може вважатися правильною відповіддю на задачу, оскільки вона не знаходиться у межах зазначеного діапазону.

Тест містить 10 запитань такого типу. Завдання обох форматів оцінюються дихотомно (за двома ознаками).

Письмо

Визначення помилок у реченнях. Використовуються багатовибіркові ТЗ. Вони перевіряють здатність знаходити граматичні помилки і помилки у використанні слів. Кожне речення містить одну помилку або не містить помилок взагалі. Можливі помилки виділяються і позначаються літерами (A, B, C, D, E). Якщо речення містить помилку, слід вибрати одну виділену частину, що має бути змінена для того, щоб зробити речення коректним. Якщо речення без помилок, потрібно вибрати E. При виборі необхідно дотримуватися традицій стандартної письмової англійської мови.

Покращення речень. Використовуються багатовибіркові ТЗ. Вони перевіряють здатність знаходити помилки в побудові речень та ідентифікувати правильні речення. Частина кожного речення або ціле речення підкреслено, а нижче наведено п'ять способів перефразування підкресленого. Кожне речення містить одну помилку або не містить помилок взагалі. Дистрактор A повторює вихідний

Directions: For Student-Produced Response questions 9-18, use the grids at the bottom of the answer sheet page on which you have answered questions 1-8.

Each of the remaining 10 questions requires you to solve the problem and enter your answer by marking the circles in the special grid, as shown in the examples below. You may use any available space for scratchwork.

Answer: $\frac{7}{12}$

Write answer in boxes. →

7	/	1	2	
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○

← Fraction line

Grid in result. →

Answer: 2.5

2	.	5		
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○

← Decimal point

Answer: 201

Either position is correct.

2	0	1		
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○

Note: You may start your answers in any column, space permitting. Columns not needed should be left blank.

- Mark no more than one circle in any column.
- Because the answer sheet will be machine-scored, **you will receive credit only if the circles are filled in correctly.**
- Although not required, it is suggested that you write your answer in the boxes at the top of the columns to help you fill in the circles accurately.
- Some problems may have more than one correct answer. In such cases, grid only one answer.
- No question has a negative answer.
- **Mixed numbers** such as $3\frac{1}{2}$ must be gridded as 3.5 or 7/2. (If $\frac{311}{2}$ is gridded, it will be interpreted as $\frac{311}{2}$, not $3\frac{1}{2}$.)

- **Decimal Answers:** If you obtain a decimal answer with more digits than the grid can accommodate, it may be either rounded or truncated, but it must fill the entire grid. For example, if you obtain an answer such as 0.6666..., you should record your result as .666 or .667. **A less accurate value such as .66 or .67 will be scored as incorrect.**
- Acceptable ways to grid $\frac{2}{3}$ are:

2	/	3		
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○

.	6	6	6	
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○

.	6	6	7	
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○
○	○	○	○	○

Рис. 11. Форма для ТЗ з продукуваною (відкритою) відповіддю

варіант, решта чотири – пропонують альтернативні. Під час відповіді необхідно враховувати граматику, підбір слів, побудову речення і пунктуацію. Вибирається найкращий варіант відповіді, що робить речення зрозумілим і точним, без неоднозначності. При виборі необхідно дотримуватися традицій стандартної письмової англійської мови.

Покращення абзаців. Використовуються багатовибіркові ТЗ. Вони перевіряють здатність:

- редагувати і виправляти речення в контексті абзацу або цілого уривка;
- організовувати і викладати речення в абзаці в правильній логічній манері.

Як завдання пропонується чернетка нарису, деякі частини якого слід змінити. Одні запитання можуть стосуватися окремих речень або їх частин і вимагати

поліпшення структури речення або підбору слів. Інші – стосуються організації речень і стилю. Під час вибору необхідно дотримуватися традицій стандартної письмової англійської мови.

SAT II

Тест складається переважно з багатовибіркових ТЗ.

Структура всіх SAT II однакова, але деякі тести мають свої особливості. SAT II з Біології містить спільне ядро з 60-ти багатовибіркових ТЗ на спільні знання, з наступними 20-ма багатовибірковими ТЗ, що роблять акцент на екологію (Біологія Е) або молекулярну будову (Біологія М). Після завершення основної частини екзаменованій вибирає ту секцію, до якої він найбільш підготовлений.

Висновки

У США існує кілька форматів психометричного іспиту. Крім SAT, існують іспити АСТ – Американський тест для вступу до коледжу. Хоча і АСТ, і SAT – це базові вступні тести, розходження між ними досить істотні. *По-перше*, АСТ включає розділ, що перевіряє знання граматики, приділяючи лише невелику увагу перевірці словникового запасу, а SAT, навпаки, вимагає від абітурієнта володіння великим лексичним запасом слів, але не містить запитань на знання граматичних структур і правил. *По-друге*, АСТ побудовано на основі багатовибіркових ТЗ (MCQ – multiple-choice question), тоді як SAT включає запитання й інших типів, зокрема есе (твір). *По-третє*, в АСТ не знімаються бали за неправильні відповіді, що іноді спонукає абітурієнтів відповідати навмання на запитання, на які вони не знають відповіді, а в SAT знімаються бали за неправильні відповіді. *По-четверте*, у SAT немає запитань, що перевіряють “наукове мислення” на прикладах мало знайомих абітурієнтам галузей (наприклад, астрономії), а в АСТ є завдання такого типу. *По-п’яте*, до розділу математики в SAT входить мало запитань з тригонометрії, тоді як у подібному розділі в АСТ їх є достатньо [40].

Той факт, що, незважаючи на істотні розбіжності між іспитами, велика частина вузів приймає студентів як на основі SAT, так і на основі АСТ (хоча деякі з них і вимагають від абітурієнтів здавати винятково один конкретний тест), свідчить про неузгодженість щодо оптимальних моделей тестування.

За дорученням Ради Коледжів експерти-предметники (включаючи фахівців зі змісту, експертів з педагогічних вимірів і когнітивних психологів) проаналізували у 2005 р. систему категорій умінь, що є основою складання кожної частини SAT. Результати аналізу було використано під час кодування ТЗ відповідно до категорій умінь. Хоча при кодуванні ТЗ може бути зазначено

кілька умінь (skills), відзначається й основна для кожного ТЗ категорія умінь. Це дає можливість забезпечити зворотний зв'язок з екзаменованими шляхом оприлюднення оцінок за вміння, з урахуванням кодів основних умінь, наведених у таблиці.

Уміння, що вимірюються різними частинами SAT

Частина	Номер уміння	Зміст уміння
Читання	1	Визначення значення слів
	2	Розуміння речень
	3	Розуміння великих уривків тексту
	4	Аналіз мети, аудиторії і стратегії
Математика	1	Застосування основних знань з математики
	2	Застосування знань більш високого рівня з математики
	3	Уміння долати труднощі
	4	Моделювання і глибоке розуміння
Письмо	1	Уміння підбирати слова і граматичні зв'язки між ними
	2	Уміння змінювати і порівнювати граматичні конструкції
	3	Уміння користуватися фразами і кластерами в реченні
	4	Уміння впорядковувати і встановлювати зв'язки між реченнями і абзацами

Таке оцінювання умінь потенційно може бути більш інформативним, ніж стандартні кластерні оцінки, засновані на змісті або типі ТЗ, оскільки уміння відбивають основну модель діяльності абітурієнтів у певному домені. Докладний опис теоретичної моделі, використаної експертами для визначення категорій умінь, і способи, що використовувалися під час кодування ТЗ, наведено Huff [45], O'Callaghan [46] і Van der Veen [47]. Оцінювання надійності під час вимірювання умінь за допомогою SAT докладно розглянув M. Ewing [48].

Медичні ліцензійні тестові іспити

Починаючи з 1996 р., згідно з наказом міністра охорони здоров'я № 396 від 31.12.96 “Про підготовку до впровадження системи ліцензійних державних інтегрованих перевідних і випускних іспитів у медичних вищих закладах освіти III–IV рівнів акредитації” в Україні триває робота щодо розробки та запровадження державних медичних ліцензійних іспитів (МЛІ). На першому етапі впровадження ліцензійний іспит складався з двох екзаменів: інтегрованого екзамену з дев'яти фундаментальних медико-біологічних дисциплін (“Крок 1”) та інтегрованого екзамену з основних медичних дисциплін (“Крок 2”) як складової частини державних іспитів. Наступним етапом стало впровадження

інтегрованого екзамену зі спеціальних клінічних дисциплін після закінчення навчання в інтернатурі (“Крок 3”).

Ліцензійні інтегровані іспити як обов’язкова складова частина державної атестації проводяться для фахівців з вищою освітою при наданні кваліфікації відповідного освітньо-кваліфікаційного рівня при підготовці лікарів, лікарів-стоматологів, медичних сестер, провізорів та фармацевтів.

Мета ліцензійних інтегрованих тестових екзаменів

Ліцензійний іспит складається з метою встановлення відповідності рівня професійних знань і вмінь студентів вищих медичних закладів освіти мінімальному необхідному рівню кваліфікації згідно з державними кваліфікаційними вимогами.

Іспитами “Крок 1. Загальна лікарська підготовка”, “Крок 1. Стоматологія” оцінюється вміння використовувати знання та розуміння ключових понять фундаментальних біомедичних наук, головну увагу приділяючи принципам та механізмам, що лежать в основі здоров’я, хвороби та моделей лікування.

Іспити “Крок 2. Загальна лікарська підготовка”, “Крок 2. Стоматологія” оцінюють вміння використовувати знання та розуміння основних найважливіших медичних і стоматологічних дисциплін для забезпечення допомоги хворому під наглядом лікаря.

Метою ліцензійного інтегрованого іспиту “Крок 3” є встановлення відповідності рівня професійної компетентності кваліфікованих фахівців з вищою освітою (на етапі випуску з інтернатури) вимогам галузевих стандартів освіти та готовності фахівця розпочати самостійну ефективну та безпечну загальну (неспеціалізовану) медичну або стоматологічну практику. Професійна компетентність у “Крок 3. Загальна лікарська підготовка” діагностується як вміння застосовувати знання і розуміння основних медичних дисциплін для забезпечення самостійної та диференційної діагностики, надання допомоги хворому при невідкладних станах, організації допомоги в екстремальних ситуаціях, здійснення протиепідемічних заходів при інфекційних захворюваннях, а також розуміння інших аспектів загальної лікарської підготовки.

Професійна компетентність у “Крок 3. Стоматологія” діагностується як уміння застосовувати знання основних стоматологічних і медичних дисциплін для самостійного надання медичної допомоги відповідно до стандартів надання медичної допомоги та прийняття клінічних, етичних і організаційно-економічних рішень в умовах загальної (неспеціалізованої) стоматологічної практики.

Форма екзамену

Екзамен проводиться у формі письмового тестування. Загальний обсяг тесту становить 200 тестових завдань, загальна тривалість тестування – 200 хвилин.

Тестові завдання, з яких складається екзаменаційний тест, є завданнями багатовибіркового типу, що вимірюють застосування знань. Тест сформовано згідно зі структурою змісту іспиту, яка відповідає навчальним програмам. Структура змісту і тестові завдання створено і прийнято колективом викладачів, делегованих від 13 медичних вузів України. У 2004 р. зміст і тестові завдання переглядалися відповідно до вимог Галузевих стандартів.

Формат тестових завдань

Тест включає тестові завдання з однією правильною відповіддю.

Тестові завдання цього формату складаються із ситуаційної задачі, запитання або твердження і чотирьох або п'яти відповідей, позначених латинськими літерами (A, B, C, D, E). Правильна відповідь на тестове завдання – це одна найкраща відповідь на запитання (твердження), яка вноситься до бланка відповідей шляхом замальовування відповідної латинської літери.

Екзаменаційний буклет

Екзаменаційний буклет має індивідуальний номер. Буклет містить 200 тестових завдань. Кожне завдання пронумеровано від 1 до 200. До запитання додаються чотири-п'ять відповідей, позначених буквами. Ви маєте вибрати одну **НАЙКРАЩУ** відповідь. Декілька тестових завдань згруповано у кластер – блок з однією загальною клінічною ситуацією та різними запитаннями, які вимагають від вас вибрати також одну **НАЙМЕНШ ІМОВІРНУ** відповідь. Будьте уважні до таких завдань.

До буклету входять також загальна інструкція і таблиця нормальних лабораторних величин.

Бланк відповідей

Бланк відповідей складається з двох частин (див. рис. 12).

Ліва частина містить інформацію про іспит та особисті дані студента (ідентифікаційний код, наданий напередодні тестування, прізвище, ім'я, по батькові, курс, дату складання іспиту та додаткову інформацію). Кожному студенту надається ідентифікаційний код, який буде занесено до реєстру осіб, які позитивно складають ліцензійний іспит. Код є постійним і буде використовуватися на всіх етапах іспиту.

ID студента										Прізвище										І П										БЛАНК ВІДПОВІДЕЙ										МЛІ Україна																																																	
0	1	2	3	4	5	6	7	8	9	А	Б	В	Г	Д	Е	Ж	З	И	Й	А	Б	В	Г	Д	Е	Ж	З	И	Й	А	Б	В	Г	Д	Е	Ж	З	И	Й	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
0	1	2	3	4	5	6	7	8	9	А	Б	В	Г	Д	Е	Ж	З	И	Й	А	Б	В	Г	Д	Е	Ж	З	И	Й	А	Б	В	Г	Д	Е	Ж	З	И	Й	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
0	1	2	3	4	5	6	7	8	9	А	Б	В	Г	Д	Е	Ж	З	И	Й	А	Б	В	Г	Д	Е	Ж	З	И	Й	А	Б	В	Г	Д	Е	Ж	З	И	Й	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
0	1	2	3	4	5	6	7	8	9	А	Б	В	Г	Д	Е	Ж	З	И	Й	А	Б	В	Г	Д	Е	Ж	З	И	Й	А	Б	В	Г	Д	Е	Ж	З	И	Й	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200

КОД (код спеціальності, код року, ВНЗ, варіант)

А	Б	С	Д	Е	Ф	Г	Н	І
0	1	2	3	4	5	6	7	8
9	0	1	2	3	4	5	6	7
8	9	0	1	2	3	4	5	6
7	8	9	0	1	2	3	4	5
6	7	8	9	0	1	2	3	4
5	6	7	8	9	0	1	2	3
4	5	6	7	8	9	0	1	2
3	4	5	6	7	8	9	0	1
2	3	4	5	6	7	8	9	0
1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8

ДОДАТКОВА ІНФОРМАЦІЯ

КУРС, ФАКУЛЬТЕТ _____

НАЗВА ЕКЗАМЕНУ _____

ДАТА _____ ВНЗ _____

ПРІЗВИЩЕ, І., П. _____

ІНСТРУКЦІЯ ЗАПОВНЕННЯ БЛАНКУ ВІДПОВІДЕЙ так ні

• Використовуйте лише олівць, 2М. Правильне замальовування

• Ретельно і повністю замалюйте коло з обраною літерою. Неправильне замальовування

• Ретельно виберіть правильну відповідь.

ФОРМА № 159-9

Рис. 12. Бланк відповідей

Права частина є безпосередньо бланком відповідей на тестові завдання. Правильна відповідь, яку вибрав студент, позначається на бланку шляхом замальовування кружка з відповідною літерою біля номера, який відповідає даному тестовому завданню. Відмітка робиться так, щоб не можна було прочитати літери всередині кружка.

На кожне тестове завдання має бути тільки одна найкраща відповідь. Правильна відповідь на тестове завдання дає 1 бал, неправильна – 0 балів. Якщо ви відмічаєте більше ніж одну літеру, відповідь вважається неправильною. Якщо ви змінюєте літеру, повністю стирайте гумкою попередню відмітку так, щоб з'явилася літера. Залишки відміток можуть призвести до того, що деякі відповіді будуть вважатися неправильними, оскільки зчитування бланків відповідей здійснюється шляхом комп'ютерного сканування.

Форми представлення результатів тестування

Перелік обов'язкових форм-звітів за результатами ліцензійних інтегрованих іспитів, який готується Центром тестування при МОЗ України для кожного екзамену, має такий вигляд:

A01	Результати складання ліцензійного інтегрованого іспиту студентами (інтернами) – громадянами України та громадянами іноземних держав
A02.1	Результати складання ліцензійного інтегрованого іспиту студентами (інтернами) за групами
A03	Результати складання ліцензійного інтегрованого іспиту студентами (інтернами) за факультетами
A04	Результати складання ліцензійного інтегрованого іспиту студентами (інтернами) за спеціальностями інтернатури
B01.1	Результати складання ліцензійного інтегрованого іспиту студентами (інтернами) – громадянами України різних ВНЗ за субтестами
B01.2	Результати складання ліцензійного інтегрованого іспиту інтернами – громадянами України різних спеціальностей інтернатури за субтестами
B02.1	Результати складання ліцензійного інтегрованого іспиту студентами (інтернами) – громадянами України для порівняльного аналізу
B02.2	Результати складання ліцензійного інтегрованого іспиту студентами (інтернами) – громадянами іноземних держав
C02.1	Кількість студентів (інтернів), які не склали ліцензійний інтегрований іспит
C02.2	Кількість студентів (інтернів), які не склали б ліцензійний інтегрований іспит за різними критеріями
R01	Результати складання ліцензійного інтегрованого іспиту для рейтингу вищих навчальних закладів МОЗ України III–IV рівнів акредитації

За кожним тестовим екзаменом, крім індивідуальних і порівняльних звітних форм, розраховуються технологічні звітні форми: психометрія стандартизованого тесту, психометрія процедури тестування за кожною аудиторією та кожним вищим навчальним закладом, психометрія тестових завдань на національній виборці та виборці вищого навчального закладу, а також готується аналітична довідка.

Критерії “склав – не склав”

Критерії «склав – не склав» затверджуються щороку Міністерством охорони здоров'я України.

Для проведення ліцензійних інтегрованих іспитів у 2005 р. Міністерство охорони здоров'я України на підставі аналізу результатів тестових екзаменів минулих років та за допомогою експертних методик для іспитів, запроваджених у статусі обов'язкових, встановило такі абсолютні величини критерію “склав” (листи МОЗ України від 07.07.04 № 08.01-22/1398 і від 24.06.05 № 08.01-22/1334):

- 40,5% правильних відповідей для екзамену “Крок 1. Стоматологія”;

- 45,5% правильних відповідей для екзамену “Крок 3. Лікувальна справа і педіатрія” (весна 2005 р.);
- 55,5% правильних відповідей для екзамену “Крок 3. Лікувальна справа і педіатрія” (осінь 2005 р.);
- 45,5% правильних відповідей для екзамену “Крок 2. Загальна лікарська підготовка”;
- 45,5% правильних відповідей для екзамену “Крок 2. Стоматологія”;
- 40,5% правильних відповідей для екзамену “Крок 1. Загальна лікарська підготовка”.

Студентам, які одержали на тестовому екзамені “Крок 1” результат “не склав”, дозволяється повторне складання екзамену “Крок 1” не більше двох разів до найближчої сесії у терміни, затверджені Міністерством охорони здоров'я України. У разі невдалого перескладання екзамену “Крок 1” студент не допускається до наступної екзаменаційної сесії та відраховується з вищого навчального закладу як такий, що не виконав навчальний план.

Студенти – громадяни України, які одержали на тестовому екзамені “Крок 2” результат “не склав”, допускаються до складання державних випускних іспитів. Ці студенти не отримують сертифіката ліцензійних іспитів і вважаються такими, що не пройшли державної атестації, вони не отримують диплома про закінчення вищого навчального закладу. Такі студенти можуть повторно скласти екзамен “Крок 2” не раніше, ніж через рік протягом трьох років.

Студенти – громадяни іноземних держав, які одержали на тестовому екзамені “Крок 2” результат “не склав”, допускаються до складання державних випускних іспитів. Ці студенти не отримують сертифіката ліцензійних іспитів, але можуть отримати диплом за умови відповідного рішення вищого навчального закладу.

Інтерни, які одержали на тестовому екзамені “Крок 3” результат “не склав”, допускаються до інших етапів атестації. Вони не отримують сертифіката ліцензійних іспитів, вважаються такими, що не пройшли атестації, і не отримують сертифіката лікаря-спеціаліста. Такі інтерни можуть повторно скласти екзамен “Крок 3” у наступний термін його проведення.

Критерій «склав – не склав» застосовується лише щодо загального тестового результату; застосування затвердженого критерію для окремих субтестів є великою помилкою, оскільки субтести мають різну складність та довжину.

Після проведення іспиту й отримання статистичних даних Центр тестування надає навчальним закладам рекомендовану статистичну шкалу для конвертації балів тестового іспиту в традиційні оцінки «2», «3», «4», «5». Для конвертації первинних тестових балів ліцензійних іспитів в оцінки застосовуються нормоорієнтовані критерії, що визначаються на основі середнього показника (M) і стандартного відхилення (σ) результатів вітчизняних студентів. Ці критерії

конвертації також слід застосовувати лише до загального тестового результату. Аналогічні статистичні шкали можуть бути розроблені для субтестів мінімально достатньої довжини не менше 30 тестових завдань.

Матеріали ліцензійних інтегрованих тестових екзаменів “Крок” (приклади)

Тести для ліцензійних інтегрованих іспитів формуються з банків тестових завдань, які пройшли первинну, тестову та фахову експертизу. Тести, як правило, містять блок “якірних” тестових завдань – тих, що використовувалися під час тестування у попередні роки та мають належні психометричні показники, блок тестових завдань закритого банку, блок тестових завдань з пов’язаних банків (наприклад, у тесті “Крок 3” можуть використовуватися тестові завдання з банку “Крок 2”).

“Крок 1. Загальна лікарська підготовка”

Структура змісту екзамену

Екзамен “Крок 1” проводиться з використанням єдиного екзаменаційного тесту, до складу якого входять тестові завдання з нормальної та патологічної анатомії, гістології, ембріології, біології, біологічної хімії, нормальної та патологічної фізіології, мікробіології, фармакології.

Наведемо структуру змісту іспиту, сформовану на основі освітньо-кваліфікаційної характеристики лікаря.

Структура змісту іспиту “Крок 1. Загальна лікарська підготовка”

1	Загальні принципи
	1.1 Біохімія та молекулярна біологія
	1.2 Біологія клітини
	1.3 Розвиток людини та генетика
	1.4 Біологія тканин та їхня реакція на хворобу
	1.5 Психосоціальні, культурні впливи і впливи навколишнього середовища на поведінку, здоров’я та процеси захворювання (патоморфоз)
	1.6 Мультисистемні процеси
	1.7 Фармакодинаміка та фармакокінетика
	1.8 Мікробіологія, вірусологія, паразитологія
	1.9 Імунні реакції
	1.10 Кількісні методи
2	Гемопоетична та лімфоретикулярна системи
3	Центральна та периферична нервові системи. Сенсорні системи

4	Шкіра та сполучна тканина
5	Опорно-рухова система
6	Дихальна система
7	Серцево-судинна система
8	Система травлення
9	Система сечовиведення
10	Репродуктивна система
11	Ендокринна система

Приклад тестового завдання

Дитина зі спадковим імунodefіцитом часто хворіє на пневмонії у зв'язку з порушенням гуморального імунітету.

З порушенням функції яких імунoкомпетентних клітин це пов'язано?

- A. Т-кілерів;
- B. Т-супресорів;
- C. Т-хелперів;
- D. В-лімфоцитів;
- E. Макрофагів.

Відповідь: B.

“Крок 2. Загальна лікарська підготовка”

Структура змісту екзамену

Екзамен “Крок 2” проводиться з використанням єдиного екзаменаційного тесту, до складу якого входять тестові завдання з терапевтичного, хірургічного, педіатричного, гігієнічного та акушерсько-гінекологічного профілів. Екзамен перевіряє знання з внутрішніх, професійних та інфекційних хвороб, ендокринології, фтизіатрії, невідкладних станів, хірургічних хвороб, дитячої хірургії, онкології, урології, анестезіології, дитячих хвороб, дитячих інфекційних хвороб, акушерства, гінекології, гігієни, соціальної медицини, організації та економіки охорони здоров'я.

Екзаменаційний тест побудовано за інтегрованою структурою змісту, яка складається з двох типів запитань. Тестові завдання класифікуються за різними ознаками:

1. Патологія органів та систем.
2. Медичні дисципліни.
3. Аспекти діяльності лікаря тощо.

Структуру змісту іспиту сформовано на основі кваліфікаційної характеристики лікаря і наведено нижче.

Структура змісту іспиту “Крок 2. Загальна лікарська підготовка”

1	Загальні запитання: нормальний ріст і розвиток людини; загальні принципи медичної допомоги	
	1.1 Дитинство	
	1.2 Юність	
	1.3 Дорослий вік	(*)
	1.4 Старість	(*)
	1.5 Медична етика та правові аспекти медицини	(*)
	1.6 Прикладна біостатистика; клінічна епідеміологія	
	1.7 Інфекції та людина	
	Патологія органів і систем	
2	Захворювання крові та кровотворних органів	
3	Психічні розлади	(*)
4	Захворювання нервової та сенсорної систем	(*)
5	Захворювання серцево-судинної системи	
6	Захворювання дихальної системи	
7	Розлади системи травлення	
8	Гінекологічні розлади	
9	Сечовивідна та чоловіча репродуктивна система	
10	Патологія вагітності, пологів та післяпологового періоду	
11	Патологія шкіри та підшкірних тканин	(*)
12	Патологія опорно-рухової системи та сполучної тканини	
13	Ендокринні та метаболічні розлади	

Аспекти діяльності лікаря

Тестові завдання охоплюють такі аспекти діяльності лікаря:

1. Первинна та вторинна профілактика захворювань, підтримка здоров'я людини.
2. Розуміння механізмів захворювань (етіологія, патогенез).
3. Діагностика захворювань та ускладнень.
4. Ведення хворих (лікування, подальше обстеження, тактика, реабілітація, рекомендації).

Приклад тестового завдання

Чоловік 20-ти років перебуває на лікарняному з приводу ГРВІ протягом 14 днів. Останні 3 дні скаржиться на появу ниючого болю в ділянці серця, задишку, відчуття перебоїв серцевої діяльності. Об'єктивно: пульс 100 уд/хв, АТ 110/90 мм рт. ст. Гомілки пастозні. Межі серця незначно розширені в поперечнику, тони приглушені, поодинокі екстрасистолічні скорочення, систолічний шум на верхівці. Печінка +1,5 см. На ЕКГ: PQ — 0,18; зниження вольтажу, негативний Т в V₂ – V₄. Який діагноз найбільш імовірний:

- A. Ревматичний міокардит;
- B. Неревматичний міокардит;
- C. Бактеріальний септичний ендокардит;
- D. Дилатаційна кардіоміопатія;
- E. Міокардіодистрофія.

Відповідь: B.

Крок 2. Стоматологія

Структура змісту екзамену

Екзамен “Крок 2” проводиться з використанням єдиного екзаменаційного тесту, до складу якого входять тестові завдання з терапевтичної стоматології, стоматології дитячого віку, хірургічної стоматології, ортопедичної стоматології, профілактики стоматологічних захворювань, ортодонтії.

Екзаменаційний тест побудовано за інтегрованою структурою змісту. Тестові завдання класифікуються за різними ознаками:

1. Загальні запитання і патологія органів та систем.
2. Клінічні дисципліни.
3. Аспекти діяльності лікаря тощо.

Загальні запитання і патологія органів та систем.

Структуру змісту іспиту сформовано на основі освітньо-кваліфікаційної характеристики лікаря-стоматолога і наведено нижче.

Структура змісту іспиту “Крок 2. Стоматологія”

1.	Загальні принципи
	* анатомо-фізіологічні особливості органів та тканин щелепно-лицьової ділянки;
	* матеріалознавство
	* знеболювання
	* організація стоматологічної допомоги, медична етика та деонтологія
	* військова стоматологія
	* асептика та антисептика
	Окрема патологія
2	Некаріозні ураження твердих тканин зубів
3	Карієс
4	Ускладнення карієсу
5	Захворювання пародонта
6	Захворювання слизової оболонки порожнини рота
7	Запальні захворювання щелепно-лицьової ділянки
8	Травматичні пошкодження щелепно-лицьової ділянки
9	Онкостоматологія

10	Реконструктивно-відновлювальна та пластична хірургія
11	Операція видалення зуба
12	Ортопедичне лікування захворювань твердих тканин зубів та часткової втрати зубів
13	Ортопедичне лікування при повній відсутності зубів
14	Ортопедичне лікування захворювань пародонта
15	Атипове зубне протезування і ортопедичне лікування захворювань скронево-нижньощелепного суглоба
16	Щелепно-лицьове протезування та імплантологія
17	Загальна ортодонтія
18	Зубощелепні аномалії та деформації

Аспекти діяльності лікаря

Тестові завдання охоплюють такі аспекти діяльності лікаря:

1. Первинна і вторинна профілактика захворювань, підтримка здоров'я людини.
2. Розуміння механізмів захворювань (етіологія, патогенез).
3. Діагностика захворювань та їх ускладнень.
4. Ведення хворих (лікування, подальше обстеження, тактика, реабілітація, рекомендації).

Приклад тестового завдання

У дитини 3,5 років під час планової санації в зубі виявлено каріозну порожнину, наповнену розм'якшеним пігментованим дентином. Під час некротомії каріозної порожнини оголився ріг пульпи, який кровоточить і болить. Перкусія зуба безболісна. Який метод лікування найбільш доцільний у цьому випадку?

- A. Вітальна екстирпація.
- B. Девітальна екстирпація.
- C. Біологічний метод.
- D. Вітальна ампутація.
- E. Девітальна ампутація.

Відповідь: E.

2.3. Приклади предметних тестів

Суспільствознавство

Кінцеві вимоги до змісту оцінювання – орієнтація на Державний стандарт.

Мета оцінювання

Визначити рівень знань учнів із суспільствознавства, що включають сформованість умінь аналізувати суспільні явища і процеси, загальні змістові наскрізні лінії: людина – суспільство – природа.

З'ясувати рівень підготовленості учнів до взаємодії з соціальним середовищем, до самореалізації їх як особистостей.

Матриця змісту, що оцінюється

1. Історія:

- Всесвітня історія та Історія України.
- Світ та Україна у роки Першої та Другої світових війн.
- Протистояння демократії і тоталітаризму.

2. Правознавство:

- Держава як засіб упорядкування суспільних відносин.
- Поняття права та правової системи.

3. Філософія:

- Основи філософії.
- Філософські погляди на сутність людини, її буття і духовність, суспільство, природу і світ.

4. Людина і світ:

- Соціальна сфера суспільства.
- Етнічна структура суспільства.

Матрицю змісту наведено нижче.

Модулі \ Рівні	Знання	Розуміння	Застосування	Аналіз	Усього	Межі змін
Історія					40%	35–45%
Правознавство					20%	15–25%
Основи філософії					30%	25–35%
Людина і світ					10%	5–15%
Разом	50%	10%	20%	20%	100%	

Навчальні цілі

1. Визначення хронології основних подій та наслідків світових війн для держав та їх громадян.
2. Аналіз історичних процесів державотворення в Україні та світі, основні досягнення науки і мистецтва.
3. Ідентифікація явищ і процесів суспільного життя, встановлення зв'язку між подіями та явищами.
4. Інтерпретація основних визначень історичних і правових понять.
5. Вирізнення найпоширеніших філософських поглядів на сутність людини.
6. Поєднання в логічний ланцюг хронології, подій, персоналій.
7. Визначення причинно-наслідкових зв'язків подій і явищ.

Обґрунтування вибору

методу оцінювання досягнення навчальних цілей із суспільствознавства

Основними факторами, що впливають на вибір методу оцінювання для діагностики рівня навчальних досягнень із суспільствознавства, є рівні засвоєння навчальних цілей, що діагностуються. Навчальні цілі, що описують знання та інтелектуальні здібності учнів, стосуються переважно когнітивного домену й охоплюють чотири рівні: знання – Рз.1, розуміння – Рз.2, застосування – Рз.3, аналіз – Рз.4.

Методом вимірювання для оцінювання когнітивного домену обрано тестування.

Матриця тесту

Матриця тесту містить три осі, які відповідають:

- 1 – вимогам Стандарту до рівня підготовки;
- 2 – змісту, або предметному модулю;
- 3 – рівням засвоєння знань.

	Предметні модулі																Усього
	Історія (М.1)				Правознавство (М.2)				Основи філософії (М.3)				Людина і світ (М.4)				
Рівні	Рз. 1	Рз. 2	Рз. 3	Рз. 4	Рз. 1	Рз. 2	Рз. 3	Рз. 4	Рз. 1	Рз. 2	Рз. 3	Рз. 4	Рз. 1	Рз. 2	Рз. 3	Рз. 4	
Вимоги Стандарту																	
Сформованість вмінь аналізувати сусп. явища і процеси (В.-1)	5	1	2	2	3	–	1	1	4	1	1	2	1	–	1	–	25
	10 %	2%	4%	4%	6%	–	2%	2%	8%	2%	2%	4%	2%	–	2%	–	50 %
Рівень підготовки до взаємодії із соціальним середовищем (В.-2)	5	1	2	2	2	1	1	1	3	1	2	1	1	1	–	1	25
	10 %	2%	4%	4%	4%	2%	2%	2%	6%	2%	4%	2%	2%	2%	–	2%	50 %

Специфікація тестового завдання — це код, який у цьому випадку складається з трьох частин. На *першому* місці розташовано код, що відповідає вимогам Стандарту. Він позначається В.1–В.2. На *другому* місці розташовано код, що відповідає предметному модулю. Він позначається М.1–М.4. На *третьому* місці розташовано код, що відповідає рівню засвоєння знань. Він позначається Рз.1–Рз.4.

Зразки тестових завдань

Нижче наведено зразки тестових завдань різних форматів зі *специфікацією*, що відповідає матриці тесту. Правильні відповіді позначено “◆◆”.

Формат А

Примітка: необхідно вибрати найбільш правильну відповідь.

Формат А	Специфікація: В.1, М.1, Рз.1
Вибрати серед зазначених дат дату початку Першої світової війни: 25 жовтня 1917 14 серпня 1914 16 травня 1915 ◆◆ 1 серпня 1914 11 листопада 1917	

Формат А	Специфікація: В.1, М.1, Рз.1
<p>Визначити місце першого застосування отруйних газів у роки Першої світової війни:</p> <p>На р. Марні</p> <p>◆◆ на р. Іпрі</p> <p>на р. Соммі</p> <p>на р. Віслі</p> <p>на р. Десна</p>	

Формат А	Специфікація: В.2, М.1, Рз.1
<p>Згадати, яка з наведених подій відбулася раніше:</p> <p>◆◆ Проголошення Декларації про державний суверенітет</p> <p>Всеукраїнський референдум</p> <p>Нарада в Біловезькій пущі</p> <p>Утворення СНД</p> <p>Антидержавний „путч” у Москві.</p>	

Формат А	Специфікація: В.2, М.1, Рз.2
<p>Визначити, до якої партії належав політичний лідер, який був суперником Л. Кучми у другому турі президентських виборів 1999 року.</p> <p>СДПУ(о)</p> <p>НДП</p> <p>◆◆ КПРС</p> <p>Соцпартії</p> <p>Селянської партії</p>	

Формат Х

Примітка: правильною може бути будь-яка кількість відповідей.

Формат Х	Специфікація: В.2, М.2, Рз.1
<p>Визначити джерела (форми) права:</p> <p>◆◆ нормативний договір</p> <p>мораль</p> <p>◆◆ правовий звичай</p> <p>◆◆ правовий прецедент</p> <p>правовий акт</p>	

Формат Х	Специфікація: В.1, М.2, Рз.1
<p>Визначити основні галузі права України:</p> <p>◆◆ адміністративне</p> <p>◆◆ фінансове</p> <p>◆◆ конституційне</p> <p>громадянське</p> <p>◆◆ міжнародне</p>	

Формат В

Примітка: складання логічних пар; для кожного пронумерованого завдання необхідно вибрати один варіант відповіді, позначений буквою.

Формат В	Специфікація: В.1, М.2, Рз.1
Поєднати кожне рішення відповідної події міжнародного життя другої половини ХХ ст. з назвою документа	
<ol style="list-style-type: none"> 1. Встановлення суворого контролю за виробництвом ядерної зброї 2. Заборона ядерних випробувань у повітрі, під водою, в космосі 3. Остаточне виведення радянських військ з Афганістану 4. Становлення вільного від ядерної зброї та ненасильницького миру 	<ol style="list-style-type: none"> А. Делійська декларація В. Женевські угоди С. Стокгольмське звернення Д. Московський договір
<i>Відповіді: 1 – D; 2 – C; 3 – A; 4 – B</i>	

Формат В	Специфікація: В.1, М.1, Рз.2
Віднайти логічні пари, які характеризують вказані поняття	
<ol style="list-style-type: none"> 1. Людина, яка формується у певному суспільстві, перебуває у певному зв'язку з ним 2. Розумна фізична істота, яка сформувалася внаслідок еволюції природи 3. Особа, що має політико-правовий зв'язок із державою 4. Особа, яка володіє всіма правами та свободами, користується захистом держави 	<ol style="list-style-type: none"> А. Людина В. Особа С. Громадянин
<i>Відповіді: 1 – B; 2 – A; 3 – C</i>	

Формат В	Специфікація: В.1, М.1, Рз.4
Встановити концепції трактування філософської сутності суспільства відповідно до етапів історичного розвитку	
<ol style="list-style-type: none"> 1. Суспільство – це "град земний", який прагне наблизитися до досконалості "граду небесного" 2. Суспільство – це сукупність людей, які більш-менш задовольняються законами справедливості 3. Суспільство – це система соціальних дій людей, сенс життя яких визначається формуванням цінностей 4. Суспільство – це продукт взаємодії людей на основі їхньої спільної трудової діяльності 	<ol style="list-style-type: none"> А. Античність В. Середньовіччя С. Нові часи Д. Новітні часи
<i>Відповіді: 1 – A; 2 – B; 3 – D; 4 – C</i>	

Формат В	Специфікація: В.1, М.1, Рз.3
Віддиференціювати зміст класифікації громадських організацій України	
<ol style="list-style-type: none"> 1. "Просвіта" 2. Спілка українських студентів 3. Мафіозні організації 4. Спілка офіцерів України 5. Українська ліга християнської молоді 6. Асоціація вчителів-фізиків 	<ol style="list-style-type: none"> А. Соціально-усвідомлені громадські організації В. Ціннісно-орієнтовані громадські організації С. Соціально-ціннісні громадські організації Д. Асоціальні громадські організації Е. Політизовані громадські організації
<i>Відповіді: 1 – B; 2 – A; 3 – D; 4; 5 – E</i>	

Формат Д

Інструкція. Треба дати дві відповіді. У списку ліворуч наведено три категорії, позначені буквами. Чотири з п'яти пронумерованих тестових завдань зі списку праворуч тим чи іншим чином стосуються ОДНІЄЇ з цих категорій. Виконайте такі завдання. (1) Замалюйте букву зі списку ліворуч, якій відповідають чотири тестових завдання. (2) Замалюйте цифру завдання зі списку праворуч, яке НЕ стосується тієї ж категорії, що й інші чотири.

Формат Д	Специфікація: В.1, М.1, Рз.3
Вирізнити політику радянської влади, з якою пов'язано чотири з п'яти ознак	
А. "Військовий комунізм" Б. Нова економічна політика В. Продовольча програма	1. Продподаток 2. Госпрозрахунок 3. Відміна карткової системи 4. Товарно-грошові відносини 5. Приватна власність
Відповіді: Б. 4.	

Формат Д	Специфікація: В.1, М.1, Рз.4
Уточнити назву політичної партії, яка належала до пропрезидентської сили в Україні в 1991 — 1994 рр.	
А. Національно-демократичні сили Б. Центристські сили В. Ліві сили	1. КПУ 2. СПУ 3. НДП 4. СелПУ 5. ПсПУ
Відповіді:	

Формат Д	Специфікація: В.2, М.1, Рз.4
Поєднати представників реалізму і модернізму в літературі другої половини ХХ ст. відповідно до основних ознак цих течій, вирізнивши ознаку, яка не належить до тієї категорії, що решта чотири	
А. Джон Бойтон Прістлі Б. Ернест Хемінгуей В. Альбер Камю	1. Психологізм 2. Морально-етична проблематика 3. Усвідомлення антагонізмів у сучасній цивілізації 4. "Життя круговерть, де його Рух: погане > гірше > погане" 5. Правдиве зображення дійсності
Відповіді:	

Формат Д	Специфікація: В.1, М.3, Рз.2
Визначити категорію філософського напрямку, з яким пов'язані чотири з п'яти концепцій	
А. Фрейдизм Б. Марксизм В. Позитивізм	1. Недооцінка філософського осягнення людини 2. Визначення первинності буття стосовно свідомості у природі та людському суспільстві 3. Постійний розвиток суспільного і людського пізнання 4. Революційне перетворення – головний шлях покращення суспільства 5. Цілковита залежність людини від суспільства
Відповіді:	

Формат Д	Специфікація: В.1, М.3, Рз.4
Модифікувати теорії стародавніх філософів стосовно ідеальної держави, визначивши одну, яка не належить до тієї категорії, що решта чотири	
A. Платон B. Арістотель B. Конфуцій	1. Сім'ї в ідеальній державі не повинно бути 2. В ідеальній державі діти мусять виховуватися в державних установах 3. В ідеальній державі філософи можуть відмовитися від усіх благ 4. В ідеальній державі має існувати приватна власність 5. В ідеальній державі не має існувати приватної власності
<i>Відповіді:</i>	

Формат К

Примітка: вибрати правильну комбінацію варіантів відповідей (наприклад, тільки 1, 2; тільки 1, 3...)

Формат К	Специфікація: В.2, М.1, Рз.1
Ідентифікувати поняття "операція Вісла".	
1. Військова операція Червоної армії наприкінці 1944 р. 2. Операція КДБ зі знищення загонів УПА в 1944 р. 3. Переселення польським урядом етнічних українців у західні райони Польщі 4. Примусова міграція українців на територію Польщі	
A Тільки 1, 2, 3 B Тільки 1, 3 C Тільки 2, 4 D Тільки 4 E Всі правильні	
<i>Відповіді:</i>	

Формат К	Специфікація: В.1, М.2, Рз.4
Проаналізувати події трьох етапів десталінізації в СРСР, визначивши події другого етапу.	
1. Відновлення Калмицької АРСР 2. Ліквідація військових трибуналів 3. Реабілітація О. Довженка 4. Зняття пам'ятників Сталіну	
A Тільки 1, 2, 3 B Тільки 1, 3 C Тільки 2, 4 D Тільки 4 E Всі правильні	
<i>Відповіді:</i>	

Формат К	Специфікація: В.1, М.2, Рз.2
Віднести до категорії основних принципів правової держави комбінацію правильних відповідей	
1. Наявність території і конституції 2. Наявність органів контролю і примусу 3. Верховенство закону і права 4. Гарантія свободи особи та її прав	
A Тільки 1, 2, 3 B Тільки 1, 3 C Тільки 2, 4 D Тільки 4 E Всі правильні	
<i>Відповіді:</i>	

Формат К	Специфікація: В.1, М.2, Рз.4
Уточнити спільні і єдині норми права і моралі	
<ol style="list-style-type: none"> 1. Виражають людські принципи та ідеали 2. Виступають коректорами поведінки людей 3. Виражають загальнолюдські принципи та ідеали 4. Виступають регуляторами поведінки людей 	
<p>A Тільки 1, 2, 3 B Тільки 1, 3 C Тільки 2, 4 D Тільки 4 E Усі правильні</p>	
<i>Відповіді:</i>	

Формат К	Специфікація: В.1, М.4, Рз.2
Вказати конфесії України, віруючі яких є прихильниками протестантських церков	
<ol style="list-style-type: none"> 1. Українська греко-католицька церква 2. Українська лютеранська церква 3. Закарпатська реформаторська церква 4. Римсько-католицька церква в Україні 	
<p>A Тільки 1, 2, 3 B Тільки 1, 3 C Тільки 2, 4 D Тільки 4 E Усі правильні</p>	
<i>Відповіді:</i>	

Фізика

Матриця змісту, що оцінюється

1. Механіка:
 - Основи кінематики.
 - Основи динаміки.
 - Закони збереження в механіці.
 - Елементи механіки рідин і газів.
2. Молекулярна фізика і термодинаміка:
 - Основи молекулярно-кінетичної теорії.
 - Основи термодинаміки.
 - Властивості газів, рідин і твердих тіл.
3. Електродинаміка:
 - Основи електростатики.
 - Закони постійного струму.
 - Електричний струм у різних середовищах.
 - Магнітне поле, електромагнітна індукція.
4. Коливання і хвилі:
 - Механічні коливання і хвилі.

- Електромагнітні коливання і хвилі.
- 5. Оптика.
- 6. Елементи теорії відносності.
- 7. Квантова фізика.
 - Світлові кванти.
 - Атом і атомне ядро.

Матрицю змісту наведено у таблиці.

Назва модуля \ Рівні	Знання (Рз.1)	Розуміння (Рз.2)	Застосування (Рз.3)	Аналіз (Рз.4)	Разом	Межі змін
Механіка (М.1)					20%	15–25%
Молекулярна фізика і термодинаміка (М.2)					15%	10–25%
Електродинаміка (М.3)					15%	10–25%
Колівання і хвилі (М.4)					15%	10–25%
Оптика (М.5)					15%	10–25%
Елементи теорії відносності (М.6)					5%	5–10%
Квантова фізика (М.7)					15%	10–25%
Разом	15%	25%	40%	20%	100%	

Матриця тесту

Нижче подано фрагмент матриці тесту:

	Модулі																Разом
	Механіка (М.1)				Молекулярна фізика та термодинаміка (М.2)				...				Квантова фізика (М.7)				
Рівні	Рз.1	Рз.2	Рз.3	Рз.4	Рз.1	Рз.2	Рз.3	Рз.4	Рз.1	Рз.2	Рз.3	Рз.4	Рз.1	Рз.2	Рз.3	Рз.4	
Елементи знань																	
	Явища (Е.1)	1	1		1	1	–	1		1	–	1	–
	1%	1%		1%	1%	–	1%		1%	–	1%	–	10%
Поняття (Е.2)	2		1			1		1	1	1	–	1	15
	2%		1%			1%		1%	1%	1%	–	1%	15%
...

Під елементами знань будемо розуміти такі елементи фізичних знань, що підлягають оцінюванню:

- фізичні явища (Е.1): ознаки, за якими це явище відбувається, його зв'язок з іншими та пояснення на основі наукової теорії;
- фізичні поняття, зокрема фізичні величини (Е.2);
- формули, що з'єднують дану фізичну величину з іншими (Е.3);
- одиниці фізичних величин і способи їх вимірювання (Е.4);
- фізичні теорії (Е.5): основні положення, принципи, дослідне обґрунтування та межі застосування;
- фундаментальні фізичні досліди (Е.6): мета, схема, умови, хід і результати;
- фізичні закони (Е.7): формулювання, математичний вираз та умови застосування;
- прилади, механізми і машини (Е.8): призначення, принцип дії та схема будови.

Приклад матриці тесту містить три осі, які відповідають:

- вісь 1 – елементу фізичних знань;
- вісь 2 – структурі змісту, або модулю;
- вісь 3 – рівню засвоєння знань.

Специфікація тестового завдання являє собою код, який у даному випадку складається з трьох частин. На *першому* місці розташовано код, що відповідає елементу фізичних знань. Він позначається Е.1–Е.8. На *другому* місці – код, що відповідає модулю. Він позначається М.1–М.7. На *третьому* місці – код, що відповідає рівню засвоєння знань. Він позначається Рз.1–Рз.4.

Зразки тестових завдань

Нижче наведено зразки тестових завдань **формату А** зі *специфікацією*, яка відповідає *матриці тесту*. Правильні відповіді позначено “◆◆”.

Формат А	Специфікація: Е.1, М.2, Рз.2
Визначити, як змінюється порядковий номер хімічного елемента в таблиці Менделєєва, якщо його ядро випромінює α -частинку:	
◆◆ зменшується на 4	
збільшується на 4	
зменшується на 2	
збільшується на 2	
не змінюється	

Формат А	Специфікація: Е.2, М.2, Рз.3
В ядрі атома аргону 18 протонів і 22 нейтрони. Визначити кількість електронів у цьому атомі:	
4 електрони	
◆◆ 18 електронів	
22 електрони	
396 електронів	
40 електронів	

Формат А	Специфікація: Е.2, М.2, Рз.1
<p>Визначити, що є однаковим у різних ізотопів одного й того ж хімічного елемента:</p> <ul style="list-style-type: none"> ◆◆ кількість нейтронів кількість протонів кількість нуклонів кількість електронів і протонів кількість електронів 	
Формат А	Специфікація: Е.1, М.2, Рз.1
<p>Визначити, які елементарні частинки обов'язково вилітають разом з ядрами-уламками під час ланцюгової реакції поділу ядер урану:</p> <ul style="list-style-type: none"> α-частинки β-частинки ◆◆ нейтрони електрони протони 	
Формат А	Специфікація: Е.3, М.2, Рз.3
<p>Визначити, як зміниться тиск ідеального газу, якщо при сталій концентрації середня кінетична енергія поступального руху молекул газу збільшиться в 3 рази:</p> <ul style="list-style-type: none"> збільшиться в 9 разів ◆◆ збільшиться в 3 рази зменшиться в 3 рази зменшиться в 9 разів залишиться незмінним 	
Формат А	Специфікація: Е.1, М.2, Рз.2
<p>Визначити, як змінюється вага тіла під час переміщення з полюса Землі на екватор:</p> <ul style="list-style-type: none"> не змінюється зростає ◆◆ зменшується стає рівною нулю 	
Формат А	Специфікація: Е.3, М.2, Рз.1
<p>Визначити, яка з наведених формул не визначає енергію електричного поля:</p> $W = \frac{cu^2}{2} \qquad W = \frac{\epsilon_0 \epsilon E^2}{2} V$ $W = \frac{q^2}{2c} \qquad W = \frac{q \cdot u}{2}$ <ul style="list-style-type: none"> ◆◆ $W = \frac{LI^2}{2}$ 	
Формат А	Специфікація: Е.3, М.2, Рз.2
<p>Визначити, як зміниться по модулю напруженість електричного поля точкового заряду при збільшенні відстані до заряду вдвічі:</p> <ul style="list-style-type: none"> зменшиться в 2 рази збільшиться в 2 рази ◆◆ зменшиться в 4 рази не зміниться збільшиться в 4 рази 	
Формат А	Специфікація: Е.3, М.2, Рз.3
<p>Розрахувати магнітний потік через контур площею 100 см² в однорідному магнітному полі з індукцією 2Тл, якщо площина контуру паралельна вектору індукції:</p> <ul style="list-style-type: none"> 200 Вб 2 Вб 0,02 Вб 20 Вб ◆◆ 0 Вб 	

Формат А	Специфікація: Е.7, М.2, Рз.1
<p>Визначити, які перетворення енергії відбуваються при киданні тіла з поверхні Землі вертикально вгору:</p> <ul style="list-style-type: none"> ◆ механічна енергія переходить у внутрішню енергію ◆ кінетична енергія переходить у потенціальну енергію ◆ внутрішня енергія переходить у механічну енергію ◆ перетворення енергії не відбувається ◆ потенціальна енергія переходить у кінетичну енергію 	

Формат А	Специфікація: Е.3, М.2, Рз.3
<p>Вирахувати енергію зв'язку ядра ${}^8\text{O}^{17}$ в МеВ. Маса нейтрона 1,0087 а. о. м., маса протона 1,0073 а. о. м., маса ядра ${}^8\text{O}^{17}$ 16,9991 а. о. м.</p> <ul style="list-style-type: none"> ◆ 128 МеВ ◆ 256 МеВ ◆ 12,8 МеВ ◆ 1,28 МеВ ◆ 25,6 МеВ 	

Формат А	Специфікація: Е.1, М.2, Рз.3
<p>Доповнити рівняння радіоактивного перетворення ${}^1_0\text{n} \rightarrow {}^1_1\text{p} + ? + \nu$</p> <ul style="list-style-type: none"> ◆ e; ◆ \bar{e}; ◆ α-частинка; ◆ ${}^1_1\text{H}$ ◆ $\bar{\nu}$ 	

Формат А	Специфікація: Е.1, М.2, Рз.1
<p>Визначити вид/види радіоактивного випромінювання, яке/які не відхиляється у магнітному полі</p> <ul style="list-style-type: none"> ◆ α-випромінювання ◆ β-випромінювання ◆ α- і γ-випромінювання ◆ β- і γ-випромінювання ◆ γ-випромінювання 	

Формат А	Специфікація: Е.1, М.2, Рз.1
<p>Визначити, за якої умови рідина змочує поверхню твердого тіла:</p> <ul style="list-style-type: none"> ◆ сила взаємодії між молекулами твердого тіла більша за силу взаємодії між молекулами рідини і твердого тіла ◆ сила взаємодії між молекулами твердого тіла менша за силу взаємодії між молекулами рідини і твердого тіла ◆ сила взаємодії між молекулами твердого тіла дорівнює силі взаємодії між молекулами рідини і твердого тіла ◆ сила взаємодії між молекулами рідини і твердого тіла більша за силу взаємодії між молекулами рідини ◆ відсутня взаємодія між молекулами твердого тіла і молекулами рідини 	

Формат А	Специфікація: Е.5, М.2, Рз.3
<p>Визначити, яких значень можуть набувати n і m у формулі $\frac{1}{\lambda} = R \left(\frac{1}{n^2} - \frac{1}{m^2} \right)$ для видимої області спектра випромінювання атома водню:</p> <ul style="list-style-type: none"> ◆ $n = 2, m = 3, 4, 5, \dots$ ◆ $n = 1, m = 2, 3, 4, \dots$ ◆ $n = 4, m = 5, 6, 7, \dots$ ◆ $n = 5, m = 6, 7, 8, \dots$ ◆ $n = 3, m = 4, 5, 6$ 	

Приклад тестових завдань різної складності

Нижче наведено приклад тестових завдань різного рівня складності, що відповідають одному й тому ж елементу знань (закон радіоактивного розпаду).

Формат X	Специфікація: Е.7, М.2, Рз.1
Ідентифікувати математичний запис закону радіоактивного розпаду: ♦♦ $N = N_0 \cdot 2^{-t/T}$ $N = m/M \cdot N_A$ $N = n \cdot V$ ♦♦ $N = N_0 \cdot e^{-\lambda t}$ $N = \nu \cdot N_A$ 1.1	

Формат А	Специфікація: Е.7, М.2, Рз.2
Визначити, який із графіків залежності кількості ядер, що не розпалися (N), від часу (t) відповідає закону радіоактивного розпаду:	
a b c ♦♦ d	

Формат А	Специфікація: Е.7, М.2, Рз.3
Дано графік залежності кількості ядер ізотопу ${}_{87}\text{Fr}^{207}$, що не розпалися від часу. Визначити проміжок часу в секундах, через який залишиться половина початкової кількості ядер.	
10 с 30 с ♦♦ 20 с 40 с 60 с	

3.0

Формат А	Специфікація: Е.7, М.2, Рз.4
Визначити період напіврозпаду ізотопу ${}_{87}\text{Fr}^{207}$.	
10 с 30 с ♦♦ 20 с 40 с 60 с	

Частина 3. Додаткові матеріали

*Розширення знань про педагогічне
тестування*

*Визначення основних інформаційних ресурсів
тестування*

Створення якісних тестів

3.1. Інформаційні ресурси

TeAch-nology – англomовний веб-портал універсального характеру для освітян (<http://www.teach-nology.com>).

ОСВІТА – український освітній портал (<http://www.osvita.org.ua>), має зокрема ресурс, що стосується зовнішнього тестування (<http://www.osvita.org.ua/abitur/?page=testing>).

ALTE – англomовний сайт Європейської асоціації мовних тестових установ (<http://www.alte.org/index.cfm>). Її членами є організації з Європи, що проводять сертифікаційні тестові іспити на знання відповідних мов і видають сертифікати. Проводить міжнародні конференції.

УАО – англomовний сайт офісу оцінювання успішності університету штату Іллінойс (США) (<http://www.assessment.ilstu.edu/mission/>).

Assessment Institute in Indianapolis – (<http://www.planning.iupui.edu/>).

NCTLA Assessment Institutes – (<http://www.ed.psu.edu/cshe/nctla.asp>).

ACER – англomовний сайт Ради освітніх досліджень Австралії (<http://www.acer.edu.au>). Рада зокрема займається розробкою національної освітньої політики в галузі оцінювання.

EURYDICE – інформаційна мережа з питань освіти в Європі (<http://www.eurydice.org>). Головне призначення ресурсу – розвиток європейської співпраці в галузі освіти. Мови: англійська, німецька, французька.

Danish Evaluation Institute – сайт Данського Інституту Оцінювання (<http://www.eva.dk>). Інститут є національним центром, що акумулює національний і міжнародний досвід з оцінювання якості навчання та засвоєння знань. Мова данська.

AEA-Europe – *Association for Educational Assessment*. Англomовний сайт Асоціації педагогічного оцінювання в Європі (<http://www.aea-europe.net>). Сприяє розвитку та впровадженню педагогічного оцінювання в європейських країнах, проводить щорічну європейську конференцію з педагогічного оцінювання.

IEA – *International Association for the Evaluation of Educational Achievement* (<http://www.iea.nl>). Англomовний сайт Міжнародної асоціації оцінювання освітніх досягнень. Асоціація дослідницьких організацій, що займаються порівняльним вивченням освітніх програм та реформ, їх впливу на якість освіти.

CSTEEP – Center for the Study of Testing, Evaluation and Educational Policy (<http://www.csteep.bc.edu>). Англomовний сайт Центру вивчення тестування, оцінювання та освітньої політики. Центр сприяє впровадженню сучасних підходів і технологій педагогічного тестування, зокрема щодо великомасштабного оцінювання знань школярів.

CRESST – National Center for Research on Evaluation, Standards and Student Testing (<http://cresst96.cse.ucla.edu/index3.htm>). Англomовний сайт національного Центру досліджень оцінювання, стандартів і вузівського тестування (США). Проводить дослідження з технології оцінювання та тестування, якості навчання та засвоєння знань.

ETS – Educational Testing Service (<http://www.ets.org>). Установа з педагогічного тестування (США). Англomовний сайт провідної організації з тестування в освітній системі США.

ЕГЭ – Портал информационной поддержки единого государственного экзамена (<http://ege.edu.ru/PortalWeb/index.jsp>). Російськомовний портал інформаційної підтримки єдиного державного екзамену РФ.

PES – Public Examination Systems (<http://www1.worldbank.org/education/exams/index.htm>). Англomовний сайт «Екзаменаційні системи в освіті», заснований Світовим банком і присвячений питанням розробки, адміністрування і розвитку екзаменів та оцінювання в освітніх системах, переважно у країнах, що розвиваються.

IAEA – International Association for Educational Assessment (<http://www.unesco.org/education/partners/iaea/>). Англomовний сайт Всесвітньої асоціації педагогічного оцінювання. Асоціація існує за підтримки ЮНЕСКО і сприяє спілкуванню профільних організацій, поширює професійні знання і досвід завдяки публікаціям, конференціям, експертній активності та тренінгам.

3.2. Історія розвитку тестування

За матеріалами [7] наведемо коротко історичну довідку розвитку тестування.

Сучасну педагогіку неможливо уявити без активного використання інструментарію об'єктивних методів вимірювання та оцінювання якісних характеристик, притаманних людині, до яких належить і рівень знань. Такий інструментарій створювався протягом останніх ста років. Вдосконалюється він і тепер. Йдеться про тестологію як галузь наукових досліджень, що займається вимірюванням і оцінюванням здібностей людини.

Тестування (від англійського testing – випробування) вперше використав Дж. Фішер для перевірки рівня знань учнів за допомогою оригінальних спеціальних книг (scale books), які з'явилися ще в 1864 р. у Великобританії. Ці книги можна вважати першими зразками шкільних тестів успішності. Адже теоретичні основи тестування було розроблено пізніше, у 1883 р., англійським психологом Ф. Гальтоном у його роботі «Дослідження людських здібностей та їх розвиток». Ф. Гальтон визначив тестування як метод, в якому застосовуються однакові досліди щодо великої кількості людей зі статистичною обробкою результатів і визначенням еталонів оцінки.

У 1890 р. в роботі американських психологів Дж. Кеттела і В. Маккеона «Розумові тести та виміри» вперше було введено термін «тест». Але у науковій літературі засновником тестової діагностики одностайно вважається Дж. Кеттел. Він започаткував традицію досліджень інтелекту вступників до вищих навчальних закладів, яка зберігається в американських університетах і дотепер. Для вивчення ефективності дидактичних прийомів у 1894 р. американець Дж. Райс уперше застосував свої таблиці з перевірки знань з орфографії, а М.Стоун у 1908 р. надрукував перший тест з арифметики.

На початку ХХ ст. у розробці тестів спостерігається розмежування психологічного та педагогічного напрямків. Розробка першого педагогічного тесту належить американському психологу Е. Торндайку. Саме з розвитком тестування у психології та педагогіці починають застосовуватися математичні методи, які й вплинули на розвиток тестології. З цього моменту тестування розвивається у двох основних напрямках:

- створення та використання тестів інтелектуального розвитку;

- створення та використання педагогічних тестів, призначених для оцінки академічних знань і здібностей учнів.

Підвищення уваги до тестування було зумовлено також активізацією діяльності Асоціації американських психологів і, особливо, утворенням нею в 1895 р. Комітету з розробки і використання так званих розумових тестів. Комітет, до складу якого входили найвідоміші психологи США, створив ряд тестів для визначення рівня інтелектуального розвитку й індивідуальних особливостей учнів.

Незважаючи на розпочаті спроби щодо впровадження тестів у навчальні заклади США, підхід до тестування наприкінці XIX ст. мав теоретичний характер, а тестування все ще залишалося переважно сферою діяльності психологів.

Розробкою проблеми тестування займалися Дж. Орлеанс і Г. Сілі, які у роботі «Об’єктивні тести» (1905 р.) підкреслили об’єктивний характер одержаних результатів та визначили такі головні цілі тестування:

- забезпечення викладачів інформацією як про ступінь засвоєння вивченого матеріалу, так і про обсяг матеріалу, призначеного для подальшого вивчення;
- надання допомоги викладачам у виборі методів навчання.

Проаналізувавши різні засоби оцінки рівня знань учнів, автори роботи здійснили спробу класифікувати тести за груповою ознакою. Саме Дж. Орлеанс і Г. Сілі запровадили класифікацію тестів на індивідуальні та групові.

Автор одного з перших групових тестів, створеного у 1913 р., — американець В. Пайл. Розроблений ним тест призначався для одночасної перевірки найпростіших процесів пам’яті учнів усього класу. Мета вимірів за цим тестом — встановлення норм пам’яті учнів різного віку, виявлення розумового і фізичного зростання.

Груповий тест В. Пайла удосконалив Р. Пінтер, який розробив арифметичний тест, а також запропонував оцінювати результати, одержані при тестуванні, у відсотках.

Поштовхом для подальшого розвитку тестології, як і інших галузей науки, стала Перша світова війна. Після вступу США у війну перед американськими психологами постало завдання створити тести, які могли б використовуватися для тестування великої кількості військовослужбовців із застосуванням мінімальної кількості іспитів. У вересні 1917 р. було створено та випробувано на практиці тести з визначення рівня інтелектуального розвитку службовців американської армії — так звані тести «Альфа» і «Бета». Тест «Альфа» призначався для тих, хто вмів читати і писати, а тест «Бета» — для тих, хто такими навичками володів менше. З вересня 1917 р. по січень 1919 р. таке тестування пройшли 1 726 966 осіб.

Після того як Спілка американських шкільних рад у 1914 р. затвердила тестування як об'єктивний метод оцінки знань, у США почався справжній бум застосування тестів успішності.

В. Макколл, визначаючи місце вимірів у навчальному процесі, розділив тести на педагогічні (Educational Tests) і тести визначення рівня розумового розвитку (Intelligence Tests), а також сформулював головну мету використання цих тестів. Це об'єднання у групи учнів, що засвоїли однаковий за обсягом матеріал, а також учнів, які засвоюють матеріал з однаковою швидкістю. Для визначення рівня розвитку учня В. Макколл ввів поняття педагогічного та розумового віку.

У 1925 р. Р. Пінтер у роботі «Тестування за певним рівнем розумового розвитку. Методи та результати» аналізує широкий історичний матеріал із проблеми тестування і пропонує класифікацію тестів, поділяючи їх на три основні групи:

- мовні тести (Test Involving Language);
- немовні тести (Test not Involving Language);
- комбіновані розумово-педагогічні тести (Combined Mental-Educational Tests).

У 20–30-х рр. минулого століття перед американськими психологами постало питання про взаємозв'язок тестів з визначення рівня розумового розвитку з експериментальними дослідженнями в галузі психології. Вчені дійшли висновку, що загальним для тестів і психолого-педагогічних дослідів є можливість їх проведення у певних умовах з можливим повторенням. Різниця полягає у тому, з якою метою проводяться тестування та дослідження. На думку американського психолога Г. Віппла, тести мають скоріше діагностичну, ніж теоретичну мету. Під час тестування здійснюється лише визначення рівня певних здібностей учнів. Мета психолого-педагогічних дослідів – відкриття нових закономірностей.

Чільне місце серед американських тестологів посідає Ф. Фрімен, який у роботі «Розумові тести. Їх історія, принципи та застосування» (1926 р.) дав широкий історичний огляд розвитку тестування у США і визначив основні напрямки цього розвитку, а саме: створення педагогічних тестів, метою яких було оцінювання результатів на різних рівнях навчання, а також розробка так званих розумових тестів, мета яких – визначення рівня загального розумового розвитку і спеціальних здібностей для подальшого навчання.

У монографії «Тести нового типу» (1926 р.) Ч. Грін проаналізував переваги і недоліки цих тестів, запропонувавши свою класифікацію, а також дав рекомендації щодо створення тестів:

- точно визначити обсяг матеріалу, призначеного для тестування;
- дослідним шляхом визначити найкращу форму тесту для даного матеріалу;
- точно визначити тривалість тестування, маючи на увазі дослідницькі дані про середню швидкість відповідей учнів протягом тестування;
- перевірити правильність мови і логічність включених до тестів тверджень;

- розташувати запитання відповідно до зростання складності, уникаючи регулярності у чергуванні правильних та неправильних відповідей.

У ті ж роки з'являється праця Дж. Міллера «Об'єктивні тести у вищій школі». Протягом 10 років автор використовував у процесі викладання педагогіки та психології в Оклахомському університеті так звані об'єктивні тести (Objective Tests). Аналізуючи основні моменти процедури тестування, Дж. Міллер відзначав, що найкращою є ситуація, коли кожний учень має копію тесту, але можливі й усні запитання та запис на дошці.

Д. Снедден розробив головні вимоги щодо інтерв'ю (усної форми тестування), яке використовувалося при прийомі на роботу. Він відзначив, що складність інтерв'ю полягає в тому, що воно є інструментом для виміру властивостей характеру, які не можуть бути визначені об'єктивно.

Прогностичне тестування завдячує своєю появою вченим Т. Келлею і П. Сімондзу. Перший у праці «Педагогічне керівництво» запропонував чотири тести для прогнозування можливих успіхів першокурсників, а другий увів поняття «тест поведінки» (Test of Conduct).

У 30–50-ті рр. ХХ ст. фахівці приділяли менше уваги проблемі тестування, водночас посилилася критика тестології як науки. Проте дослідження в цій галузі тривали, вводилися нові тести. Так, у 1932 р. Г. Луфброу розробив і проаналізував такі тести:

- тест оцінки провини;
- тест на розсіювання уваги;
- тест на виявлення схильності до обману;
- тест на виявлення загальної позиції;
- словниковий тест;
- тест на виявлення гідності;
- тест-лабіринт;
- тест на виявлення моральних поглядів.

Багато з названих тестів стали складовими тесту на виявлення морально-політичних якостей під час перевірки лояльності кандидатів на державну службу. У цьому випадку для тестування почали використовуватися тести із застосуванням технічних засобів, які одержали назву «детекторів брехні».

Розширення сфер впровадження тестів вимагало розробки загальних вимог до них. Саме С. Пріссей і Ф. Робінсон у праці «Психологія і нова освіта» (1933 р.) виступили з підтримкою стандартизованих тестів, визначивши такі їх особливості:

- ретельний відбір матеріалу для тестування;
- чітка ясність і недвозначність вказівок;
- правильне, з точки зору мовних норм, формулювання запитань;
- об'єктивне і по можливості просте оцінювання одержаних результатів.

Поступово тестування із психолого-педагогічної проблеми перейшло в соціальну та ідеологічну сферу. Тести почали піддавати гострій критиці, оскільки за результатами тестування особи, які відносилися до чорної раси, вихідці з менш забезпечених сімей, мали гірші показники, на відміну від представників більш забезпечених прошарків населення. Не утруднюючи себе глибоким аналізом, прибічники ідеологічних наук поспішно наклеїли тестам ярлик «буржуазні», як дискримінаційним за класовою та расовою ознакою. Детально ці проблеми аналізує Ю. Гільбух у працях [49, 50], Н. Нікандров у праці «Сучасна вища школа капіталістичних країн» (1978 р.), Б. Саймон [51].

У колишньому СРСР розвиток і використання діагностичних методів має свою історію, проаналізовану в праці І. Цатурової [52]. Загалом ця історія не збігається із всесвітньою. Умовно можна виділити три періоди.

Перший – з початку 20-х до середини 30-х рр.

Цей період визначається поширенням різних тестових методик і наукових пошуків у даній галузі. Але відома постанова ЦК ВКП(б), видана у 1936 р. під назвою «Про педагогічні викривлення у системі наркомпросів», у якій засуджувалася практика використання тестів, зробила свою справу. Тестування було пов'язано з так званою педологічною наукою і практикою, що припустилася абсолютизації одержаних за допомогою тестів результатів дослідження школярів. Категоричне та загальномасштабне засудження педології як псевдонауки супроводжувалося відкиданням позитивних досягнень радянських педагогів і психологів у галузі оцінки знань і вмінь учнів, які в цілому творчо розвивали педагогіку та психологію. Тому проблема тестування в педагогічній і психологічній літературі протягом тривалого часу ігнорувалася, а припинення всіх досліджень, пов'язаних з розробкою та використанням тестів, стало суттєвою перешкодою на шляху подальшого розвитку психодіагностики і дидактодіагностики. Отже, протягом приблизно двадцяти років в СРСР проблема діагностичних методик навіть не поставала.

Початок другого етапу розвитку методів тестування можна віднести до 60-х рр., коли тестування як метод вимірювання знань почали використовувати у військових училищах Міністерства оборони, Міністерства внутрішніх справ та інших спеціалізованих закладах. Цей період відзначається також активізацією досліджень у галузі програмованого навчання. Виходять збірники наукових праць: «Питання алгоритмізації і програмованого навчання» (1973 р.), «Питання теорії і методики програмованого навчання», «Програмоване навчання за кордоном» (1968 р.), роботи Т. Ільїної «Програмоване навчання і школа» (1968 р.), М. Нікандрова «Програмоване навчання та ідеї кібернетики» (1970 р.), Н. Талізїної «Теорія програмованого навчання» (1975 р.) та інших.

Нова хвиля повторного використання тестів, починаючи з 60-х рр., мала загальноосвітній характер і пов'язана з іменем відомого американського педагога Р. Тайлера. За його ініціативою було розроблено програму NAEP – National Assessment of Educational Progress, яка суттєво вплинула на впровадження

тестування у систему освіти США. Головний принцип Р. Тайлера полягав у тому, що його програма мала на меті не здійснення контролю за місцевими системами шкільної освіти «зверху», вона повинна була викликати зацікавленість у вдосконаленні цього процесу.

У 1983 р. проведенням програми NAEP починає займатися ETS, яка поставила за мету систематизацію робіт з тестування, зокрема стандартизацію тестів, встановлення єдиних правил процедури тестування, а також розробку критеріїв для визначення якості освіти, набутої американськими школярами та студентами.

ETS розроблено понад 50 програм тестування, до яких належать як загальнонаціональні програми визначення рівня знань школярів протягом навчання у середніх школах, студентів коледжів та університетів, так і програми визначення професійної придатності – MMPI. Розроблено також :

- методики складання тестів;
- програми збирання даних про абітурієнтів;
- методи соціологічних обстежень абітурієнтів і студентів;
- методи адаптації випускників вищих навчальних закладів до умов праці.

Головний центр ETS розташований у місті Принстон, штат Нью-Джерсі. Під впливом NAEP і завдяки діяльності ETS більшість штатів встановили нижню межу мінімально необхідної шкільної успішності і дали їй визначення. Ці дані є точкою відліку для шкіл щодо призначення повторних курсів навчання.

Важливим кроком було введення до системи освіти США стандартизованих загальнонаціональних тестів, які набули характеру обов'язкових.

Усі абітурієнти для вступу до вищих навчальних закладів США мають пройти тестування для визначення їхньої академічної успішності. Тести, які використовуються, є стандартизованими, що дає можливість проводити порівняльний аналіз у масштабах країни.

Найпоширенішими для вступу в американські коледжі та університети є два тести:

- тест перевірки загальних академічних здібностей (SAT – Scholastic Aptitude Test);
- тест на визначення глибини знань з окремих дисциплін (ACT – American College Test).

Тест SAT як тест допуску до вищої освіти, запроваджений у США в 1926 р., до 2005 р. складався з двох розділів:

- вербального (SAT-Verbal), що вимірює розуміння прочитаного матеріалу, обсяг словникового запасу, здатність логічно мислити. З 2005 р. доповнено розділ з перевірки вміння грамотно і чітко висловлювати власні думки у письмовій формі – есе (essay);

- математичного (SAT-Math), що вимірює рівень знань з елементарної математики, знання математичних понять, а також використання математичних методів при розв'язанні задач.

Тривалість цього тестування три години. При цьому загальний тест складається з 78 запитань для SAT-Verbal та 60 запитань для SAT-Math [53].

Частини	Тип/зміст завдань	Тривалість, хв
Вербальна	10 – завершення речення 13 – аналогії 12 – розуміння прочитаного	30
Вербальна	9 – завершення речення 6 – аналогії 15 – розуміння прочитаного	30
Мінівербальна	13 – розуміння прочитаного	15
Математика	25 – багатовибіркові ТЗ	30
Математика	15 – кількісні порівняння 10 – із продукуваною відповіддю	30
Мініматематика	10 – багатовибіркові ТЗ	15
Апробаційна	Різні, що відповідають вербальним, або математичним частинам	30

Сьогодні у США щорічно понад 1 млн школярів, випускників середніх шкіл за особистим бажанням виконують завдання тесту SAT. Цей тест допомагає кожному оцінити рівень своєї підготовки за загальноприйнятими критеріями, порівняти свої результати з середніми по країні і відповідно до цього вибрати навчальний заклад для продовження освіти.

Тест АСТ, як тест із визначення рівня знань із загальноосвітніх дисциплін, введено як обов'язковий у більшості престижних коледжів та університетів США з 1959 р.

Цей тест складається з трьох одноденних тестувань з таких дисциплін: суспільні науки та американська історія, біологія, хімія, твір англійською мовою, історія Європи та всесвітня культура, математика, фізика, література. Додатково за вибором: французька, німецька, італійська, японська, латинська, іспанська мови [54].

Тестування проходить у спеціальних центрах, створених по всій країні та за її межами.

Усі тестові завдання та ключі до них знаходяться у Головному центрі служби тестування в м. Принстон. Поширення цієї інформації є злочином, що карається законом. З метою попередження можливості втрати інформації база даних з тестових завдань поповнюється, а ті завдання, які вже були використані при тестуванні, вилучаються з неї і потім друкуються.

Тестування вийшло далеко за межі системи освіти і використовується під час масових іспитів на професійну придатність, масових наборах до армії й комплектації спеціальних підрозділів поліції та інших служб [55].

Але вже наприкінці 70-х рр. деякі американські вчені дійшли висновку, що процес оцінки знань не повинен бути обмежений використанням лише стандартизованих тестів. Дослідження, проведені Х. Уеллсом, показали: щоб допомогти учням у засвоєнні навчального матеріалу, необхідні такі тести, які визначали б щоденні успіхи учня. Такими, на його думку, є тести з відповідних галузей знань (Domain Referenced Tests – DRT).

Дж. Мак і В. Хант [56] підтримують ідею використання тестів типу (DRT) і дають їм назву критеріально-орієнтованих (Criterion-Referenced Tests).

Більшість спеціалістів із проблем тестування вважають, що для учнів тести є безпосередньо тим критерієм, який дає їм змогу краще оцінити себе, з'ясувати мету та методи навчання. При правильному використанні тести допомагають як викладачам, так і учням. А при неправильному – можуть загальмувати процес навчання. Висловлюється також думка про неприпустимість і шкідливість використання з метою навчання одного тесту для цілого класу учнів, тому що для деяких учнів він може виявитися дуже легким, а для інших – важким.

Отже, тести мають бути настільки індивідуалізовані, щоб кожен учень одержував тест, який відповідає його рівню засвоєння навчального матеріалу. Викладач готує усі тести на початку курсу, а учні проходять тестування в міру підготовки. Але якщо не буде враховано навчальну функцію тестів, то вони можуть завдати шкоди як учням, так і навчальній програмі. Частина вчених має іншу думку. До них належить Р. Ебел [57], який вважає, що саме нестандартизовані тести (складені окремими викладачами) частіше містять помилки. Вони обмежені за обсягом матеріалу, а також не включають розділи, важливі для розвитку учня.

В останні роки тестування широко використовується і в європейських країнах, зокрема у Німеччині, Швеції, Норвегії, Великобританії та інших. Так, у Німеччині, на відміну, наприклад, від Англії, впровадження тестування стримувалося, по-перше, політичними обставинами, а по-друге, тим, що там мала перевагу гуманітарна педагогіка. І тільки з середини 50-х рр. у деяких галузях системи освіти почали застосовувати тести. Поштовх поширенню тестування дала Міжнародна конференція в 1967 р. у Берліні, яка стимулювала розвиток тестів, орієнтованих на критерії. Але масове застосування цього методу в системі освіти Німеччини відбулося у середині 80-х рр. і було наслідком поширення американської системи. На цей період припадає введення при вступі до найпрестижніших університетів обов'язкових тестувань [3].

Отже, аналіз історичного розвитку тестології засвідчує, що ця галузь набула широкого розвитку у США та розвинених західних країнах, де накопичено великий досвід щодо розробки і практичного застосування тестів у різних сферах діяльності.

Наприкінці 70-х рр. група соціологів університету Карнегі-Мелона в м. Пітсбурзі провела дослідження з метою з'ясування причин, які привели до широкого використання тестів у навчальних закладах США. «Тестування глибоко пустило корені в американське життя. Ніхто не може продовжити навчання в американських школах, не проходячи тестування за стандартизованими тестами різного роду, а тестування, пов'язане з влаштуванням на роботу та підвищенням кваліфікації, настільки поширене, що безпосередньо стосується кожного» [58]. У цих дослідженнях, виданих у 1981 р., вказано на такі причини поширення тестування у США:

- бізнесовий інтерес до максимально ефективного використання людських ресурсів,
- прагнення винагороджувати талант незалежно від його соціального походження,
- загальновизнана прихильність американців до національних стандартів.

Розглядаючи історію та тенденції розвитку педагогічних вимірювань як за кордоном, так і в Україні, висвітлимо чинники, які обумовили нинішній стан.

Надзвичайно важливим кроком є розпочате у 1993–1994 рр. впровадження тестування на випускних іспитах у середніх школах, результати яких зараховувалися як складова частина вступних іспитів до вищих навчальних закладів України. Незважаючи на численні недоліки першої спроби, варто підкреслити дві надзвичайно важливі особливості цього процесу:

- по-перше, процес розпочався;
- по-друге, він мав державний і загальнонаціональний характер.

Приємно констатувати, що розпочався новий етап впровадження тестування у нашій державі. Це означає, що вже можна аналізувати його перші результати, робити висновки та прогнози, критикувати з метою вдосконалення. Ця спроба стандартизувати в масштабах країни вимірювання й оцінювання рівня знань усіх випускників середніх шкіл за єдиними об'єктивними критеріями є надзвичайно важливою.

3.3. Критерії валідності результатів вимірювання

Наведені вище (розділ 1.1.) критерії аналізу об'єктивності результатів дають змогу на практиці провести визначення валідності інструментарію та процедури вимірювання, процедури оцінювання; візуалізувати наявність порушення валідності, якщо це має місце; встановити чинники, що викликали порушення. Кількісний аналіз параметрів інструментарію може бути проведений із використанням основних положень і висновків теорії тестування лише після експерименту, тобто на апостеріорному рівні. Зупинимось лише на аналізі характеристик тесту.

Аналіз характеристик тесту, як і аналіз тестових завдань, проводиться за єдиною методикою, яка ґрунтується на системному аналізі функції розподілу результатів тестування та характеристичної кривої тесту.

Залежно від значення і знаків асиметрії та ексцесу закону розподілу можна виділити дев'ять таких критеріїв:

Критерій I

Тест є валідним, якщо асиметрія та ексцес закону розподілу дорівнюють нулю ($A_s = 0$, $E_x = 0$).

Критерій II

Якщо асиметрія є від'ємною, а ексцес додатним ($A_s = \text{“-“}$, $E_x = \text{“+“}$), то тест є невалідним (легкий і з малою розподільною здатністю).

Це результат дії кількох факторів:

- тест сконструйовано з невалідних тестових завдань (він мусить валідизуватися шляхом їх ускладнення та збільшення їх розподільної здатності);
- процедура тестування є невалідною.

Критерій III

Якщо асиметрія та ексцес є від'ємними ($A_s = \text{“-“}$, $E_x = \text{“-“}$), то тест є легким, але з великою розподільною здатністю.

Його валідизація полягає в ускладненні тестових завдань.

Критерій IV

Якщо асиметрія є від'ємною, а ексцес відсутній ($A_s = \text{“-“}$, $E_x = 0$), то тест є легким.

Його валідація полягає в ускладненні тестових завдань.

Критерій V

Якщо асиметрія є додатною, а ексцес від'ємним ($A_s = \text{“+“}$, $E_x = \text{“-“}$), то тест є складним з великою розподільною здатністю.

Його валідація полягає у валідації змісту шляхом вилучення тестових завдань, які не відповідають цій вимозі.

Критерій VI

При додатних асиметрії та ексцесі ($A_s = \text{“+“}$, $E_x = \text{“+“}$) тест є складним з малою розподільною здатністю.

Це є наслідком дії двох факторів, що впливають один на одного: порушення валідності змісту, невалідність тестових завдань.

Валідація тесту полягає у валідації тестових завдань.

Критерій VII

При додатній асиметрії та відсутності ексцесу ($A_s = \text{“+“}$, $E_x = 0$) тест є складним.

Це є наслідком дії двох факторів, що впливають один на одного: порушення валідності змісту, невалідність тестових завдань.

Валідація тесту полягає у валідації тестових завдань.

Особливої уваги потребують два останні критерії (VIII і IX), які дають можливість визначити конструктивну валідність тесту, а саме перевірити тест на однорідність.

Нагадаємо, що тест вважається однорідним, якщо його сконструйовано з тестових завдань з близькими характеристиками. Неоднорідним вважається такий тест, який складається з окремих субтестів, кожний з яких можна розглядати як тест з певними характеристиками.

Критерій VIII

Якщо крива розподілу симетрична, асиметрія дорівнює нулю, а ексцес додатний ($A_s = 0$, $E_x = \text{“+“}$), то тест є неоднорідним і сконструйованим із двох субтестів: складного і легкого.

Валідація тесту полягає в його поділі на два окремих тести.

Критерій IX

Якщо крива розподілу симетрична, асиметрія дорівнює нулю, а ексцес від'ємний ($A_s = 0$, $E_x = \text{“-“}$), то тест є неоднорідним, сконструйованим із трьох субтестів: легкого, середнього і складного.

Валідизація тесту полягає в його поділі на окремі тести.

Узагальнення. Визначаючи основні категорії вимірювання рівня знань, можна зробити висновок про те, що валідність методу вимірювання залежить від валідності кожної з означених категорій. Таким чином, поряд з валідністю інструментарію важливе місце посідає валідність процедури вимірювання, через те, що невідповідність характеристик цієї категорії критеріям валідності робить невалідним метод вимірювання, навіть якщо інструментарій тестування атестовано як валідний.

Наведені загальні критерії об'єктивності результатів тестування та критерії кількісного аналізу тестів дають можливість ідентифікувати порушення валідності процедури вимірювання.

Найчастіше наслідком порушення валідності процедури вимірювання є випадки, класифіковані як *II* та *V критерії* загальних критеріїв об'єктивності результатів. Тоді порушення нормального закону розподілу результатів тестування пов'язані з наявністю додатного ексцесу, що засвідчує зменшення розподільної здатності тесту.

3.4. Основні закони та закономірності

Вимірюванням й оцінюванням якісних характеристик займається тестологія. Тестологія – це напрям наукових досліджень у галузі теорії та практики вимірювання й оцінювання якісних характеристик, притаманних людині, в якому методом вимірювання є тестування.

Метод тестування передбачає, що інструментом вимірювання є тест, складений із тестових завдань; процедурою вимірювання є тестування, методом оцінювання – шкалювання.

Тестологія спирається на сукупність законів, які складають основу її аксіоматики. Як галузь науки вона знаходиться у стані активного розвитку, тому на сучасному етапі єдиного стійкого погляду на тестологію немає. Найчастіше в науковій літературі застосовується *система з п'яти законів*, сформульованих на основі узагальнення загальносистемних законів природи, які мають характер гіпотез [4].

Ізакон – закон збереження. Результати тестування X_i для i -го опитуваного завжди складаються з дійсної компоненти T_{ij} , характерної для кожного j -го тесту та похибки E_j , таким чином:

$$X_i = T_{ij} + E_j.$$

Наявність похибки зумовлено дією об'єктивних і суб'єктивних чинників, що певною мірою залежить від методу вимірювання. Зупинимося на природі, видах та особливостях похибок при педагогічних вимірюваннях. Наявність похибок при вимірюванні будь-якої величини є наслідком впливу:

- недосконалої інструменту вимірювання;
- недосконалої методики вимірювання;
- дії неконтрольованих факторів тощо.

Ці чинники призводять до виникнення похибок, які поділяються на три класи: промахи, або грубі похибки, систематичні похибки, випадкові похибки.

Промач є наслідком грубих порушень процесу або процедури вимірювання, а також неправильної методики вимірювання. Промачи після детального аналізу одержаних результатів можуть бути усунені.

Систематичні похибки вимірювання є стійкими або такими, що закономірно змінюються від одного вимірювання до іншого. Вони, як правило, пов'язані як

з інструментарієм, так і з методикою вимірювання. Тому їх можна виявити і мінімізувати.

Випадкові похибки є неконтрольованими, такими, що не піддаються мінімізації і мають стохастичний характер.

Похибки можуть з'явитися і через нестабільність характеристики, яку вимірюють.

Отже, коливання результатів вимірювання викликано *природною нестабільністю ознаки та помилками вимірювання*. Але порівняно з природним коливанням ознаки, випадкові помилки вимірювання, здебільшого, є незначними.

Відрізнити дійсну компоненту T_{ij} від похибки E_i з певним ступенем надійності можна під час повторних багаторазових вимірювань. Справді, середнє значення повторних вимірювань знаходимо за формулою:

$$\bar{X} = 1/n \sum_{i=1}^n X_i = T_i + 1/n \sum_{i=1}^n E_i.$$

Якщо похибка E_i не містить систематичних помилок і промахів —

$$\frac{1}{n} \sum_{i=1}^n E_i$$

є середнім значенням випадкової величини, при аналізі якої можна застосувати методи математичної статистики. Якщо припустити, що математичне очікування випадкової величини E_i тотожно дорівнює нулю, помилка при її визначенні вираховуватиметься як:

$$E = \frac{\sigma_{Ei}}{\sqrt{n}},$$

де, σ_{Ei} — середньоквадратичне відхилення. А оскільки T_{ij} — величина стала для даного j -го тесту, формулу розрахунку результатів тестування для i -ї особи, яка тестувалася, в остаточному вигляді можна записати так:

$$X_i = T_{ij} + \frac{\sigma_{Ei}}{\sqrt{n}}.$$

II закон — закон напряму та умов — системне узагальнення закону ентропії, де визначаються спрямованість процесу тестування та обмеження на діагностичні можливості тестів як інструменту вимірювання. Virізнюються два стани тесту:

- 1) еталонне рішення як найменш імовірне, яке має більш упорядковану інформацію про характеристику досліджуваної ознаки X_1 ;
- 2) довільне, тобто відмінне від еталонного рішення як найбільш імовірне, яке має найменшу впорядкованість щодо характеристики ознаки X_2 .

Якщо діагностичну функцію тесту I розглядати як міру складності визначення дійсного значення характеристики ознаки, то можна записати:

$$\Delta I = f[(X_1 - X_2)] \geq 0,$$

де ΔI – приріст діагностичної функції тесту. При виконанні тесту приріст діагностичної функції з часом поступово зменшується і при досягненні еталонної відповіді стає рівним нулю: $\Delta I/\Delta t \rightarrow 0$.

Це дає підстави сформулювати **II закон**:

Процес виконання тесту кінцевої довжини за фіксований проміжок часу надає тестові такого кінцевого стану, який визначається мінімальним приростом його діагностичної функції.

З цього закону випливає дуже корисний висновок: подальше збільшення часу тестування за умови відсутності впливу зовнішніх факторів не дає нової інформації щодо досліджуваної характеристики.

III закон – закон індукованих дій – ґрунтується на гіпотезі загальносистемного закону індукованих дій, що викривають механізм консервативності, який тією чи іншою мірою притаманний будь-якій системі.

Діагностична функція відкритої тест-системи у процесі її прогресивного розвитку завжди зменшується за рахунок дій інших діагностичних систем, що досліджують однаково характеристику особистості. З цього закону випливають важливі наслідки:

- для здійснення діагностичних функцій тест-система мусить мати інформаційний потенціал, який перевищує потенціал інших джерел, що протидіють діагностуванню;
- інформаційний потенціал необхідний для здійснення діагностичних функцій тест-системи тим менший, чим більший інформаційний потенціал інших тест-систем, які реалізують ту ж саму діагностичну функцію.

IV закон – закон граничного розвитку тест-систем. Тест-системи, прогресивно розвиваючись (вдосконалюючись та еволюціонуючи), досягають характерної для певної сукупності зовнішніх і внутрішніх умов межі, яка відповідає мініимальному приросту діагностичної функції тесту і вказує на неефективність їх подальшого функціонування.

Цей закон визначає межу між двома сторонами діагностичного процесу, які визначаються у III законі, та вказує момент, коли тест-система повинна еволюціонувати, змінюючи не лише свою структуру, а й сукупність зовнішніх умов її існування, щоб задовольнити вимоги діагностики.

На основі дії цього закону виникає поняття **валідності** тесту, при невиконанні якого має бути застосовано інший тест, що відповідає діагностичним вимогам і вимогам дослідження.

V закон – закон конкуренції тест-систем. Серед тест-систем розвиваються ті, що за певної сукупності зовнішніх і внутрішніх умов досягають максимального значення критерію оптимальності.

У соціально-психологічній і педагогічній сферах критерії оптимальності не мають абсолютного характеру. Таким критерієм для тестів обирається максимально досяжне значення коефіцієнта надійності тесту: $R \rightarrow \max$.

Закон V сприяє розвитку й поширенню найбільш надійних тестів як інструментів вимірювання, що мають найвищу точність.

У сучасній тестології існує два напрями вимірювання, які полягають у диференціації не лише тестів, а й тактики тестування на два типи: вимірювання, орієнтоване на норми, і вимірювання, орієнтоване на критерії. Крім відмінностей у стратегії та тактиці процесу вимірювання, ці напрями суттєво різняться у закономірностях, яким мають підпорядковуватися результати вимірювань.

Так, вимірювання, орієнтовані на *норми*, ґрунтуються на класичній теорії тестів, основою якої є:

- гіпотеза нормального розподілу, яка припускає, що якісні характеристики, притаманні людині, розподіляються відповідно до кривої Гауса (за нормальним законом);
- ранговий принцип, який базується на уяві, що якісний параметр визначається відносно середньостатистичного значення досліджуваної групи.

Водночас при вимірюваннях, орієнтованих на *критерії*, спостерігається відхід від нормального закону розподілу результатів тестування, оскільки ймовірності дати правильну p або неправильну q відповідь не дорівнюють одна одній. Це призводить до відхилення від нормального закону розподілу та накопичення результатів тестування на правому боці кривої розподілу. При вимірюваннях, орієнтованих на критерії, необхідно визначити:

- коли можна прийняти рішення щодо досягнення навчальної мети;
- яку кількість завдань для цього потрібно виконати;
- до якої групи за показниками навчальної успішності слід віднести особу, яка тестувалася.

У теорії вимірювання, орієнтованій на критерії, використовуються *дихотомні*, *квантитативні* та *політомні* моделі обробки результатів. Усі ці моделі виходять з того, що результати тестування мають розподілятися за *біноміальним* законом розподілу. Але варто враховувати те, що з математичної точки зору при однакових ймовірностях дати правильну p або неправильну q відповідь, біноміальна крива є симетричною і при збільшенні кількості випробувань наближається до своєї межі – *нормальної* кривої. Біноміальна крива має суттєву асиметрію лише при дуже великій різниці ймовірностей дати правильну p або неправильну q відповідь. Ці міркування, що ґрунтуються на властивостях законів розподілу, дають змогу визначити умови, за яких навіть при критеріальних вимірюваннях одержані

результати розподіляються за нормальним законом. Такою умовою є рівність ймовірностей дати правильну та неправильну відповідь $p = q$. Крім того, усі моделі обробки даних передбачають наявність нормативної бази прийняття рішень, яка створюється лише при вимірюваннях, орієнтованих на норми. Все це підтверджує висновок Р. Фріке про інтегральну узгодженість обох методів вимірювання.

Література

1. Lienert G. A. Testaufbau und Testanalyse//Wienheim. – Beltz 3, Aufl. – 1969. – P. 7–14.
2. Wiczerkowski W., Schumann M. Klassische Testtheorie//In: Klauer K.J. Handbuch der Paragogischer Diagnostik. – Band I. – Dusseldorf: Schwann. – 1978.
3. Ингенкамп К. Педагогическая диагностика/ Зарубежная школа и педагогика. – М.: Педагогика, 1991. – 239 с.
4. Булах І. Є. Комп'ютерна діагностика навчальної успішності. – К.: ЦМК МОЗ України, УДМУ, 1995. – 221 с.
5. SAT Program – <http://www.collegeboard.com/student/testing/sat>
6. Анастаси А. Психологическое тестирование: Пер. с англ. в 2 кн. – Кн. 1. – М.: Педагогика, 1982. – 316 с.
7. Булах І. Є. Історія розвитку та сучасний стан педагогічної тестології. – К.: ЦМК МОЗ України, 1994. – 21 с.
8. Аванесов В. С. Форма тестовых заданий: Учебное пособие. – М.: Иссл. Центр Гособразования СССР по проблемам качества подготовки специалистов, 1991.
9. Аванесов В. С. Основы научной организации педагогического контроля в высшей школе. – М.: Моск. ин-т стали и сплавов, 1989.
10. Беспалько В. П. Основы теории педагогических систем. – Воронеж: Воронежский университет, 1977. – 240 с.
11. Кыверялг А. А. Методы исследования в профессиональной педагогике. – Таллинн: Залгус, 1980. – 334 с.
12. Розенберг Н. М. Проблемы измерений в дидактике. – К.: Вища школа, 1979.
13. Агрусці Г., Артемчук Л., Булах І., Вілмут Дж., Лукіна Т., Мруга М. Основи педагогічного оцінювання. Частина І. Теорія/Під заг. ред. І. Булах. – К.: “Майстер клас”, 2006. – 113 с.

14. Бурлачук Л. Ф., Морозов С. М. Словарь-справочник по психологической диагностике. — К.: Наук. думка, 1989. — 197 с.
15. Булах І. Є., Волосовец В. Ф., Вороненко Ю. В. Система управління якістю медичної освіти в Україні. Монографія. — Д., “АРТ-ПРЕС”, 2003. — 212 с.
16. Howard Wainer etc. Computerized Adaptive Testing. A primer. Lawrence Erlbaum Associated, Publishers, 1990. — 300 с.
17. Булах І. Є. Теорія комп’ютерного тестування. — К.: ЦМК МОЗ України, 1994. — 59 с.
18. Булах І. Є. Кількісний аналіз результатів тестування. — К.: ЦМК МОЗ України, 1994, — 54 с.
19. Bloom B. Taxonomy of educational objectives, handbook I: The cognitive domain. — New York: Me Kay, 1956. — 128 p.
20. Krathwohl D., Bloom B., Masia B. Taxonomy of educational objectives, handbook II: The affective domain. — New York: Me Kay, 1964. — 210 p.
21. Simpson E. The classification objectives in the psychomotor domain — Washington:, Gryphon House, 1972. — 213 p.
22. Rychen D. S. & Salganik L. H. (Eds.) Defining and selecting key competencies. — Güttingen, Germany: Hogrefe & Huber, 2001. — 256 с.
23. Key Competencies in a knowledge-based economy: a first STEP towards selection, definition and description. Concept document of the Commission expert group on “Key competencies”. 27 March 2002.
24. Case S. M., Swanson D. B. Constructing written test questions for the basic and clinical sciences/National Board of Medical Examiners. — Philadelphia, 1996.
25. Ripkey R. R., Case S. M., Swanson D. B. A “new” item format for assessing aspects of clinical competence. *Academic Medicine*, 71 (10), October Supplement, — P. 34–36.
26. Nedelsky, L. Absolute grading standards for objective tests//*Educational and Psychological Measurement*, № 1. — P. 3–19.
27. Angoff W. H. Scales, norms, and equivalent scores. In R. L.Thorndike (Ed.), *Educational Measurement* (2nd edition). Washington, DC: American Council on Education. — P. 508–600.
28. Ebel R. L. *Essentials of educational measurement*, Englewood Cliffs, NJ: Prentice-Hall.

29. Livingston S. A. Zieky M. J. Passing scores: A manual for setting standards of performance on educational and occupational tests. Princeton, NJ: Educational Testing Service. – 1982.
30. Berk R. A. Determination of optimal cutting scores in criterion-referenced measurement//Journal of Experimental Education. – 1976. – 15. – P. 4–9.
31. Hofstee W. K. The case for compromise in educational selection and grading/In S. B. Anderson & J. S. Helmick On educational testing. San Francisco, CA: Jossey-Bass. – 1983.
32. Beuk C. H. A method for reaching a compromise between absolute and relative standards in examinations//Journal of Educational Measurement. – 1984. – 21. – P. 147–152.
33. De Gruijter D. N. Compromise methods for establishing examination standards//Journal of Educational Measurement. – 1985. – 22. – P. 263–269.
34. Hambleton R. K., Zaal J. N., Pieters H. J. Computerized adaptive testing: theory, applications and standards//Kluwer Academic Publishers. – Boston, MA, US. – 1991. – P. 458.
35. Impara J. C. Licensure testing: purposes, procedures and practices. Buros Institute of mental measurements. University of Nebraska. – Lincoln, 1995. – 362 p.
36. Gronlund N. E. Measurement and evaluation in teaching (5th edition). – New York: Macmillan.
37. Mehrens W. A., Lehman I. J. Measurement and evaluation in education and psychology (3rd edition). – New York: Holt; Rinehart & Winston, 1991. – 592 p.
38. Standards for educational and psychological testing/American Educational Research Association. American Psychological Association. National Council on Measurement in Education. – Washington, DC: American Psychological Association, 1999.
39. Образование в США. – <http://www.russianamerica.com/common/cmn/usa-info.php/school2.htm>
40. The PSAT (Preliminary Scholastic Aptitude Test) – http://www.eabjm.com/counseling/standardized_tests/psat.htm
41. SAT*: About the Scholastic Aptitude Test. – <http://www.studentmarket.com/studentmarket/aboutsat.html>
42. IIT Study Circle. Scholastic Aptitude Test. Previous Question Papers. – <http://surrendranath.tripod.com/sat.html>

43. Student-Produced Responses: Directions – <http://www.collegeboard.com/student/testing/sat>
44. Huff. K. A practical application of evidence centered design principles: Coding items for skills. In K. Huff (Organizer), Connecting curriculum and assessment through meaningful score reports. Symposium conducted at the meeting of the National Council on Measurement in Education, San Diego, 2004.
45. O’Callaghan R., Morley M., & Schwartz A. Developing skill categories for the SAT math section. In K. Huff (Organizer), Connecting curriculum and assessment through meaningful score reports. Symposium conducted at the meeting of the National Council on Measurement in Education, San Diego, 2004.
46. Van der Veen. A. Toward a construct of critical reading for the new SAT. In K. Huff (Organizer), Connecting curriculum and assessment through meaningful score reports. Symposium conducted at the meeting of the National Council on Measurement in Education, San Diego, 2004.
47. Ewing M., Huff K., Andrews M. and King K. Assessing the Reliability of Skills Measured by the SAT. Research Notes. Office of Research and Analysis. RN-24, December, 2005.
48. Гильбух Ю. З. Интеллектуальные тесты в школьной политике капиталистических государств.//Критика основных направлений буржуазной политики в области образования. – М.:1976; Интеллектуальное тестирование на Западе: итоги и проблемы//Советская педагогика, 1980, № 2, С. 128–136.
49. Гильбух Ю. З. Умственно одаренный ребенок. Психология, диагностика, педагогика. – К.: РОВО “Укрвузполиграф”, 1992.
50. Саймон Б. Английская школа и интеллектуальные тесты/Пер. с англ. – М.: АПН РСФСР, 1958.
51. Цатурова И. Из истории развития тестов в СССР и за рубежом. – Таганрог, 1969.
52. All about the SAT. – <http://www.collegedegreeguide.com/articles-fr/sat.htm>
53. Bulletin of the ACT Program. – Netherlands: International Edition, 1994–95 registration.
54. Мышко С. А. Тестирование как психолого-педагогическое средство оценивания академических способностей в системе образования США// Обзор. информ. Сер.: Высшая и средняя спец. школа за рубежом. – М.: НИИВШ, 1980, Вып. 6.
55. Hunt V., Muc J. Psychological development and the educational enterprise//Educational theory, 1975, vol. 25., N 4.

56. Ebel R. L. Educational testing: Valid? Biased? Useful? // Phi. Delta Kappun., 1975, October, vol. 57, N 2.
57. Resnick L. B., Melnick De. A new perspective on the use of standardized tests//Phi. Delta Kappun, 1981, May, N 5, P. 623–6.



Навчальне видання

БУЛАХ Ірина Євгенівна
МРУГА Марина Рашидівна

СТВОРЮЄМО ЯКІСНИЙ ТЕСТ

Навчальний посібник для викладачів,
учителів та методистів

Відповідальний редактор *В. І. Отрішко*
Літературний редактор *Л. Г. Дончевська*
Коректор *Л. Г. Дончевська*
Комп'ютерне макетування *Т. Я. Кушка*

Підписано до друку 15.12.2006. Формат 70 x 100/16.
Папір офсетний. Гарнітура Newton. Друк офсетний.
Обл.-вид. арк. 8,0. Наклад 1000 прим.

ВИДАВНИЦТВО МАЙСТЕР-КЛАС

Головний офіс
Редакція "МК-Харків"
Редакція "МК ІАВО"
(Інформаційна агенція
"Вища освіта")
"МК-НМВ"
(Навчально-методичний відділ)

01042, м. Київ, вул. Івана Кудрі, 5;
e-mail: kiev@mastecklassbook.com
61166, м. Харків, вул. Бакуліна, 11;
e-mail: kharkov@mastecklassbook.com
61103, м. Харків, а/с 492;
e-mail: iavo@mastecklassbook.com
61166, м. Харків, вул. Бакуліна, 11;
e-mail: nmo@mastecklassbook.com

Свідоцтво про внесення суб'єкта видавничої справи до державного реєстру видавців, виготівників та розповсюджувачів видавничої продукції ДК № 2178

Ексклюзивний дистриб'ютор
ЗАТ "Національний книжковий
проект"

036880, м. Київ, вул. Сім'ї Сосніних, 3
Тел.: 8 (044) 273-57-16, 273-57-18

Надруковано у друкарні ПП "Олена-Фарма"
61070, м. Харків, вул. Старошишківська, 12.

ОГОЛОШУЄМО БОРОТЬБУ ЗА ЯКІСТЬ ПІДГОТОВКИ ДО ТЕСТУВАННЯ!

під час зовнішнього оцінювання та вступних іспитів



**Центр тестування при МОЗ України
Видавництво МАЙСТЕР-КЛАС**

НМВ МК (Навчально-методичний відділ видавництва «Майстер-клас») проведе у вас майстер-клас з основ тестування.
Замовлення приймаються за телефоном 8 057 7518937
м. Харків, 61166, вул. Бакуліна, 11.

ТО-ТЕСТ

тематичне оцінювання

Готові матеріали для проведення тематичного, поточного, семестрового та річного оцінювання, що забезпечать проведення справедливого оцінювання знань учнів екзаменатором та значно зменшать трудозатрати на проведення оцінювання.

**Центр тестування при МОЗ України
Видавництво МАЙСТЕР-КЛАС**

