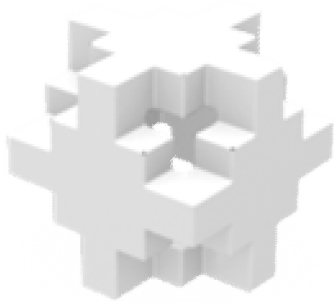


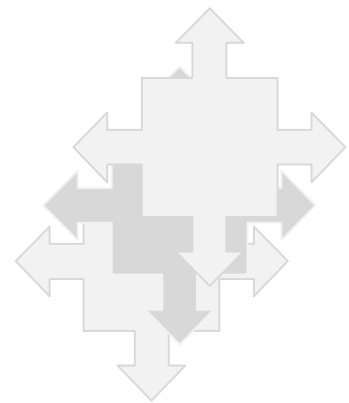
Кухар Л.О., Сергієнко В.П.



Конструювання тестів

Курс лекцій

Навчальний посібник



Луцьк – 2010

УДК 37.091.26(175.8)
ББК 74.580.28 я73
К 95

Рецензенти:

Лапінський В.В. – кандидат фізико-математичних наук, доцент, завідувач лабораторією інформатики Інституту педагогіки НАПН України;

Малежик М.П. – доктор фізико-математичних наук, професор кафедри комп'ютерної інженерії Національного педагогічного університету імені М.П. Драгоманова

Упорядники – автори:

Кухар Людмила Олександрівна, Сергієнко Володимир Петрович

К 95 **Конструювання тестів. Курс лекцій:** навч. посіб. /
Л.О. Кухар, В.П. Сергієнко. – Луцьк, 2010. – 182 с.

Посібник містить курс лекцій нової навчальної дисципліни «Конструювання тестів» для засвоєння майбутніми педагогами знань щодо теорії та практики конструювання тестів. Значну увагу приділено класифікації педагогічних тестів, опрацюванню результатів тестування, особливостям застосування комп'ютерних технологій у тестуванні, правилам складання тестових завдань.

Книга адресована науковцям, педагогам-практикам, працівникам системи підвищення кваліфікації вчителів, студентам педагогічних навчальних закладів, усім тим, хто цікавиться тестуванням та оцінюванням.

Рекомендовано до друку рішенням кафедри комп'ютерної інженерії Інституту інформатики Національного педагогічного університету імені М.П. Драгоманова (протокол № 1 від 27.08.2010 р.)

УДК 37.091.26 (175.8)
ББК 74.580.28 я 73

© Кухар Л.О., Сергієнко В.П.

Зміст

Передмова	5
§ 1. Введення в проблему тестового контролю. Роль тестів у сучасному навчальному процесі	6
1.1. Тест як засіб педагогічного вимірювання.....	6
1.2. Поняття «якості освіти».....	7
1.3. Моніторинг якості освіти	9
1.4. Теоретичні аспекти освітнього моніторингу.....	11
1.5. Педагогічне оцінювання	20
1.6. Система педагогічного оцінювання.....	25
1.7. Класифікація цілей навчання. Таксономія Блума	27
1.8. Педагогічне вимірювання та тестування	37
1.9. Переваги та недоліки тестового контролю.....	38
1.10. Політика з оцінювання та тестування.....	41
1.11. Учасники процесу оцінювання (тестування).....	43
§ 2. Педагогічні тести. Терміни і визначення. Педагогічний контроль, предмет і об'єкт контролю. Принципи педагогічного контролю	45
2.1. Історія становлення та розвитку тестування	45
2.2. Роль тестування у становленні інших країн	49
2.3. Понятійний апарат тестології. Поняття тесту	51
2.4. Класифікація педагогічних тестів	55
2.5. Адаптивне комп'ютерне тестування.....	69
2.6. Якість тесту	72
§ 3. Зміст тесту. Принципи добору змісту	73
3.1. Проблеми конструювання тестових завдань	73
3.2. Зміст тестів для поточного, проміжного та результуючого контролю.....	74
§ 4. Етапи конструювання тесту. Специфікація тесту	76
4.1. Етапи створення тесту	76
4.2. Стандартизація тесту	82

4.3. Варіативність тестів.....	83
4.4. Створення паралельних варіантів.....	84
4.5. Фасет.....	85
§ 5. Тестові завдання: правила конструювання, формати, вимоги.....	87
5.1. Правила конструювання тестових завдань.....	87
5.2. Форми (формати) тестових завдань.....	90
5.3. Створення тестових завдань різних когнітивних рівнів.....	93
5.4. Типи тестових завдань.....	95
§ 6. Опрацювання та представлення результатів тестування. Статистичне обґрунтування якості тесту.....	109
6.1. Критерії якості методу вимірювання. Поняття валідності і надійності тестів.....	109
6.2. Статистичні характеристики тесту.....	118
6.3. Класичні статистичні методи аналізу результатів тестування..	120
6.4. Зразки задач на визначення статистичних характеристик результатів тестування.....	126
6.5. Визначення складності тесту засобами електронних таблиць Microsoft Excel.....	134
6.6. Комплексне оцінювання якості тестових завдань і тестів.....	142
§ 7. Розрахунок результатів. Методи шкалювання та представлення результатів тестування.....	153
7.1. Одновимірне шкалювання. Види одновимірного шкалювання	153
7.2. Стандартизація та норми.....	158
7.3. Основні типи шкал в педагогічних вимірюваннях.....	163
7.4. Шкалювання результатів тестових вимірювань.....	169
§ 8. Методика і технологія тестування.....	175
8.1. Розробка інструкцій для проведення тестування.....	175
8.2. Стандартизація процедури та умов тестування.....	176
8.3. Зразки інструкцій для проведення тестування.....	178
СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ.....	180

Передмова

У запропонованому навчальному посібнику подано орієнтовний зміст лекцій навчального курсу «Конструювання тестів», який за навчальним планом підготовки бакалавра (напряму 0403 Системні науки та кібернетика, напрям підготовки 6.040302 *Інформатика**) структурно входить до циклу «Освітні вимірювання» за вибором навчального закладу.

Дисципліна «Конструювання тестів» реалізує додаткову підготовку вчителя інформатики в галузі теорії та практики педагогічних вимірювань з використанням тестів.

Одним з найважливіших завдань дисципліни є підготовка майбутніх учителів, а також підвищення кваліфікації викладачів у галузі теорії і практики педагогічного тестування.

Сучасний стан системи освіти зобов'язує педагогів шукати нові підходи до оцінювання навчальних досягнень учнів. Це означає, що необхідно ознайомитися з підходами до оцінювання результатів навчання в інших країнах, результатами міжнародних досліджень, в яких бере участь Україна, визначати сучасні пріоритети в освіті.

Дисципліна передбачає вивчення питань прикладного спрямування: конструювання тестів різних рівнів застосування, контрольного матеріалу за цілями використання, інтерпретація результатів тестування, виконання статистичного опрацювання даних, тлумачення його результатів та формування висновків щодо його відповідності меті, завданням, розуміння основних характеристик якості контрольного матеріалу і можливості використання комп'ютерних технологій при створенні, проведенні і опрацюванні результатів тестування.

Цей навчальний посібник стане в нагоді майбутнім педагогам, учителям, викладачам, та усім, хто цікавиться тестуванням, технологіями конструювання тестів та їх якістю.

§ 1. Введення в проблему тестового контролю. Роль тестів у сучасному навчальному процесі

1.1. Тест як засіб педагогічного вимірювання

Процес європейської інтеграції все помітніше впливає на усі сфери життя держави, не обійшов він і вищої освіти. Україна чітко визначила орієнтири на входження в освітній і науковий простір Європи, удосконалює освітню діяльність в контексті європейських вимог, зробила конкретні кроки для практичного приєднання до Болонського процесу.

Основним завданням на перехідний період є реалізація передбаченої Болонською декларацією системи академічних кредитів ECTS (European Credit Transfer System – Європейської кредитно-трансферної системи). Саме її розглядають як засіб підвищення мобільності студентів під час переходу із однієї навчальної програми на іншу. ECTS повинна стати багатоцільовим інструментом визнання і мобільності, засобом реформування навчальних програм. Надзвичайно важливий акумулюючий момент впровадження кредитної системи, що полягає в можливості враховувати усі досягнення студента, а не лише навчальне навантаження, наприклад, участь в наукових дослідженнях, конференціях, олімпіадах і тому подібне.

Проте одним із найважливіших показників є рівень знань студентів з навчальних дисциплін, що вивчаються. Найоб'єктивнішим засобом оцінювання рівня знань в даний час вважають тести, які дозволяють неупереджено оцінити навчальні досягнення студентів. Тому необхідно розглянути основні вимоги до тестів, їх переваги і недоліки, а також рекомендації щодо їх конструювання і використання.

Проблемі організації навчання з використанням кредитно-модульної технології присвячений ряд робіт, опублікованих ще в середині 90-х років, проте в них лише була поставлена проблема застосування кредитно-модульної технології.

У деяких роботах дано рекомендації з організації тестування знань студентів в умовах кредитно-модульної

технології навчання, проте наразі відсутні роботи, присвячені системному аналізу впливу тестування на якість знань, класифікації тестів і особливостям конструювання тестів з метою підвищення ефективності організації навчання студентів.

1.2. **Поняття «якості освіти»**

Поняття «якість освіти» науковці тлумачать по-різному. Для практичних цілей під якістю освіти почали розуміти «якість підготовки».

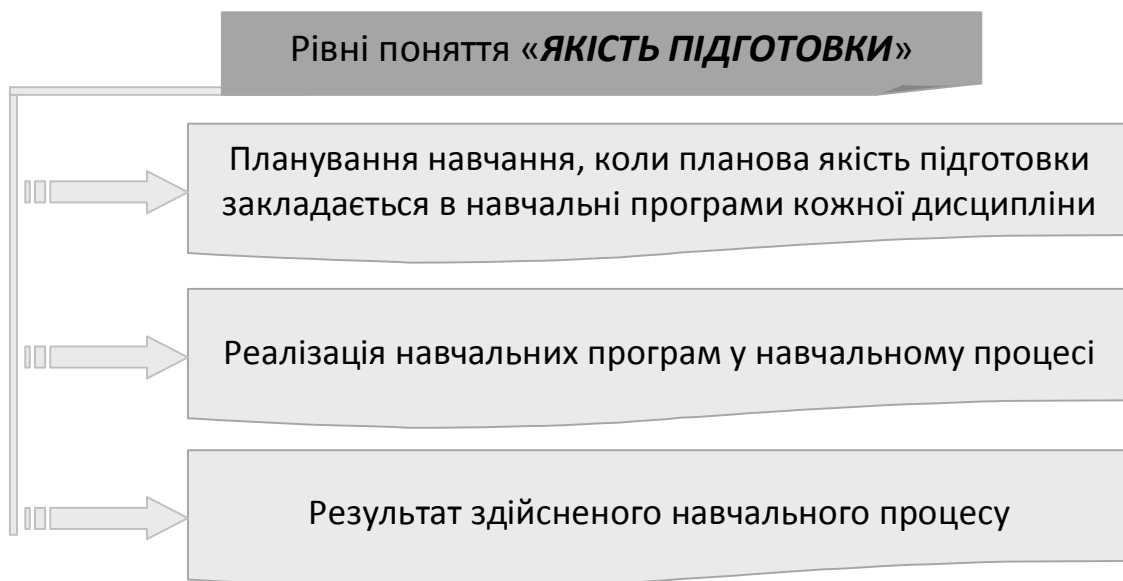


Рис. 1.1. Рівні поняття «якість підготовки»

Часто використовують термін «навчальні досягнення» як синонім поняття «якість підготовки». Закони України «Про освіту», «Про вищу освіту» не дають визначення поняттю «якість освіти», проте опосередковано це поняття використовується. У законах інших держав СНД дане поняття застосовується ширше.

Під якістю освіти можна розуміти міру відповідності фактично досягнутих результатів заздалегідь запланованим. Гарантії якості освіти без вимірювань його результатів носять багато у чому декларативний характер. Як інструменти вимірювання знань студентів виступає шкала оцінок, яка рекомендована МОН України.

Якість освіти є категорією, що за своєю сутністю відображає різні аспекти освітнього процесу – філософські,

соціальні, педагогічні, політико-правові, демографічні, економічні тощо. Це поняття поєднує властивості й характеристики освітнього процесу та його результату, які спроможні задовольняти освітні потреби всіх суб'єктів навчально-виховного процесу, – учнів і студентів, їхніх батьків, викладачів, роботодавців, управлінців, тобто суспільства в цілому.

Тому соціально-психологічний контекст проблеми якості освіти, тобто групі уявлення педагогів, учнівської та студентської молоді, батьківського загалу, слід враховувати в розбудові системи управління освітою. Опитування різних категорій учасників навчально-виховного процесу, проведене Інститутом соціальної і політичної психології АПН України, показало, що більшість з них має чіткі уявлення про якість освіти і визначає її таким чином:

Розвиток учня (студента) як самодостатньої особистості, сформованість його світогляду, моральних переконань	40,3 %
Зв'язок знань із життям, здатність застосовувати їх на практиці	37,9 %
Високий рівень одержаних знань	31,9 %
Позитивна мотивація до навчання, готовність учня (студента) навчатися протягом усього життя	28,8 %
Реалізація оздоровчої функції освіти	25,7 %
Оснащеність навчальних закладів сучасними, передусім комп'ютерними, засобами навчання	23,9 %
Якість управління освітніми процесами з боку державних органів освіти	16,1 %

Очікувана якість освіти формується під впливом сукупності умов і чинників, які охоплюють методологічні, теоретичні й практичні аспекти управління освітою.

Якість – це основа європейських домовленостей у рамках Болонського процесу, до якого приєдналася Україна. Відповідно до його принципів, відповідальність за якість вищої освіти лежить насамперед на кожному окремому навчальному закладі, й у такий спосіб забезпечується можливість перевірки

якості системи навчання на національному і міжнародному рівнях. Мова йде вже не стільки про якість освіти як таку, скільки про систему гарантій такої якості, котра б забезпечувала привабливість вищої освіти.

Поліпшення якості освіти та рівний доступ до неї має стати одним із головних завдань сучасної державної політики в галузі освіти, необхідною умовою реалізації права громадян на освіту. Адже світове співтовариство визнало, що освіта, добробут і здоров'я людини є головними чинниками якості її життя, а якість освіти – головною метою, пріоритетом розвитку громадянського суспільства. Не випадково, серед характеристик так званого індексу розвитку людського потенціалу, за яким ООН порівнює рівень соціального та економічного розвитку різних країн, показник освітньої діяльності є одним із трьох основних індикаторів в інтегрованій оцінці людського розвитку. Більше того, сьогодні у Європі вона сприймається як об'єкт суспільного єднання і консолідації різних народів, національних освітніх систем.

1.3. Моніторинг якості освіти

Найважливішою цінністю сучасного світу є особистість, перед якою відкрита можливість самореалізації та саморозвитку, входження до сучасного суспільства. Побудова освіти в новітньому контексті, модернізація усіх її складових перетворюється на першочергову потребу. Відповідь на поставлені запитання може дати тільки висока якість освіти, що стає визначальним принципом інноваційної освітньої системи, в якій основою усіх зрушень є створення платформи для розкриття потенційних можливостей дітей, прогнозування потреб і моделей розвитку особистості.

Інструментом здійснення високоякісної освітньої моделі є моніторинг якості освіти, головною складовою якого є моніторинг навчальних досягнень учнів. Сучасна практика та проведені дослідження дозволяють стверджувати про відсутність на сьогодні комплексної моделі моніторингу якості освіти учня, яка б дозволяла систематично, прозоро та

ефективно відслідковувати динаміку розвитку особистості, встановлювати причини проблем і розробляти прогноз.

Відповідь на поставлені запитання може дати тільки високоякісна освіта. Якість у цьому контексті набуває особливої актуальності на сучасному етапі.

Інструментом реалізації високоякісної освітньої моделі є моніторинг якості освіти, головною складовою якого є моніторинг навчальних досягнень учнів, що характеризується систематичністю, тривалістю в часі, прозорістю та ефективною системою відслідковування та ставить завдання встановлення причин і невідповідностей результату поставленим цілям.

Важливо зазначити, що моніторинг сприяє удосконаленню інформаційного забезпечення освітнього процесу як важливої умови ефективного управління якістю освіти, дозволяє спрямувати інформаційне забезпечення на досягнення пріоритетів щодо домінування чи недостатніх суб'єкт-суб'єктних відносин і розвитку учнів, гуманізації та гармонізації освіти.

Найвагомішим орієнтиром щодо переваг моніторингу якості навчання є те, що таке відслідковування надає змогу зорієнтуватися на результат, який співвідноситься не лише із державним замовленням, але і з особистісними досягненнями учнів, під якими розуміється

- ступінь прогресу особистості по відношенню до її попередніх досягнень;
- просування учня в процесі засвоєння знань, умінь, розвитку психічних характеристик і моральних якостей, а саме:
 - фонд комунікативних умінь;
 - творча пізнавальна діяльність;
 - особистісні соціальні досягнення в межах школи та в соціумі (особистісний компонент, пов'язаний із розвитком здібностей до самопізнання, самореалізації, самоконтролю, самоорганізації);
- стан здоров'я, фізичного розвитку (природо відповідний компонент) тощо.

1.4. Теоретичні аспекти освітнього моніторингу

Система управління освітою потребує об'єктивної картини її реального стану для забезпечення обґрунтованості рішень які приймаються.

Сучасні соціально-економічні та політичні зміни в державі стали передумовою реформування національної системи освіти. Основною соціальною вимогою суспільства є якісна освіта. Усі функції, які виконує педагогічний моніторинг як частина педагогічної діяльності, підпорядковані підвищенню якості навчання і виховання, забезпеченню наукового підходу до управління навчальною та виховною діяльністю студентів, підвищенню якості освіти в цілому.

Визнання цінності людської індивідуальності, процеси гуманізації та стандартизації в галузі освіти потребують якісних змін в галузі контролю та моніторингу процесу навчання та виховання.

Головним завданням освітньої політики на сучасному етапі її розвитку та становлення є досягнення високого рівня якості освіти, що відповідає запитам особистості, суспільства та держави.

Діагностика якості є обов'язковим компонентом будь-якої освітньої системи. В ході численних досліджень було доведено, що найефективнішим у цьому відношенні є моніторинг.

Управління якістю професійної освіти здійснюється за допомогою використання різноманітних методів, засобів у рамках систем якості різного рівня. Їх вибір, композиція визначаються рівневістю системи якості, ступенем її розвитку та підходами до забезпечення якості освіти.

Для підвищення результативності освіти важливим є створення та використання систем моніторингу, без яких неможливе якісне управління її розвитком. Проблема здійснення моніторингу в освіті, на рівні держави та окремого навчального закладу, на сьогодні є надзвичайно актуальною, оскільки саме моніторингові дослідження дають можливість цілісно побачити реальний стан освіти. Але важливим є не лише моніторинг освіти, а й прогнозування та відстеження її неспинного розвитку.

Моніторинг є ланкою управління якістю освіти. Останнім часом моніторинг якості освіти став водночас об'єктом теоретичних досліджень і сферою педагогічної діяльності.

При аналізі проблем, що стосуються моніторингу як механізму управління якістю освіти, виникає ряд невирішених завдань:

- немає чіткого тлумачення поняття «якість освіти»;
- немає спеціальних методик оцінювання якості освіти та її результатів.

Розробленням теорії та деяких аспектів практичного застосування освітнього моніторингу займалися В.І. Байденко, В.Г. Вікторов, В.В. Вербець, П.В. Дмитренко, Т.Ф. Єсенкова, В.О. Кальней, О.М. Касьянова, Г.А. Лісьєв, О.І. Локшина, Т.О. Лукіна, О.І. Ляшенко, А.Н. Майоров, Т.Невілл Послтвейт, О.О. Островерх, О.О. Патрикєєва, А.І.Субетто, А.І. Севрук, А. Тайджнман, Д. Уїлмс, С.Є. Шишов, Р.С. Яков.

Моніторинг як засіб удосконалення системи інформаційного забезпечення управління освітою розглядають О. Абдуліна, В. Горб, О. Локшина, О. Майоров, Л. Мишанська, С. Сіліна та ін.

Дослідження наукової та педагогічної літератури з проблем моніторингу свідчить про неоднозначне сприйняття цього поняття.

Аналіз літератури показав, що не існує єдиної загальнодержавної системи моніторингу, хоча є набір інструментів та технологій, що використовуються освітніми закладами для цього. Звідси доцільно та необхідно продовжувати роботу зі створення альтернативних систем моніторингу.

Моніторинг як ефективний засіб людської діяльності здавна застосовує спостереження, як особливий спосіб пізнання, що ґрунтується на відносно тривалому, цілеспрямованому і планомірному сприйнятті предметів і явищ освітньої діяльності.

Прикладами організації спостережень за навколишнім середовищем були тридцять сім томів "Природної історії" Гая Секунда Плінія (старшого) написані ще в першому столітті нашої ери. У книгах містяться відомості з астрономії, фізики, географії, зоології, ботаніки, медицини, історії. Вони стали

найповнішою енциклопедією аж до епохи середньовіччя. Набагато пізніше, в ХХ сторіччі, в наукових колах виник термін моніторинг для позначення повторних цілеспрямованих спостережень за одним або кількома елементами довкілля у просторі та часі.

Це поняття не має точного, однозначного тлумачення, оскільки застосовується в різних сферах науково-практичної діяльності людини.

«*Моніторинг*» (від англ. monitoring – контроль, відстеження) – це і форма дослідження і спосіб забезпечення сфери управління своєчасною та якісною інформацією.

У педагогічній літературі термін «моніторинг» вживають у значенні відстеження результативності навчально-виховного процесу (моніторинг навчальних досягнень учнів); інколи під ним розуміють звичайний педагогічний контроль (моніторинг успішності учнів з предмету); частіше його використовують для вивчення певних параметрів функціонування системи освіти чи окремих її елементів або суб'єктів освітнього процесу (моніторинг якості підготовки фахівців з певної спеціальності, моніторинг матеріально-технічного забезпечення навчальних закладів, стану здоров'я учнів тощо). Тому моніторинг має статус дослідження, а не емпіричного збирання інформації про певні характеристики та властивості освітньої системи.

Поняття "моніторинг" вивчається і використовується в різних сферах науково-практичної діяльності, у тому числі педагогічній. Водночас окремі аспекти цієї важливої проблеми ще потребують вивчення. Існування різноманітних понять "початковий моніторинг", "педагогічний моніторинг", "моніторинг якості освіти", "моніторинг якості навчання", категорії яких неоднозначно розкриваються сучасними ученими – педагогами, свідчать про різні підходи до їх використання.

Деякі вчені (Д. Матрос, Д. Полев, Н. Мельников) розглядають моніторинг як механізм контролю та відстеження якості освіти, постійне спостереження за навчально-виховним процесом з метою виявлення його відповідності бажаному результату або першочерговим завданням. На думку учених, поняття "освітній моніторинг" як педагогічна та управлінська

категорія перекладає загальні положення теорії інформації мовою педагогіки, психології та управління. Дослідники вважають, що завдання учених полягає в конкретизації положень даної теорії.

Моніторинг став самостійним напрямом управлінської діяльності в освіті, у якому проходить інтеграція вимірювань, досліджень, експерименту, інформатики та управління. Моніторингова система набула комплексного, цілісного характеру.

В загальному розумінні моніторинг можна визначити як постійне спостереження за яким-небудь процесом з метою виявлення його відповідності бажаному результату або вихідному стану. Результати моніторингу мають цінність досить обмежений проміжок часу, і чим більше динаміка розвитку системи, тим менший період цінності даних моніторингу.

Моніторинг є досить складним та неоднозначним явищем. В різноманітних сферах він використовується з різною метою, але володіє спільними властивостями та характеристиками.

Основні сфери застосування моніторингу як наукового дослідження – це екологія, біологія, педагогіка, економіка, психологія, теорія управління тощо. Головна сфера практичного застосування моніторингу це управління, а саме інформаційне обслуговування управління в різноманітних галузях діяльності.

В деяких сферах науково-практичної діяльності моніторинг знаходиться лише на початковій стадії дослідження, в інших його дослідження та вивчення знаходиться на завершальному етапі.

Проаналізувавши принципи моніторингу в таких сферах як екологія, соціологія та інші можна виділити декілька загальних принципів проведення моніторингу:

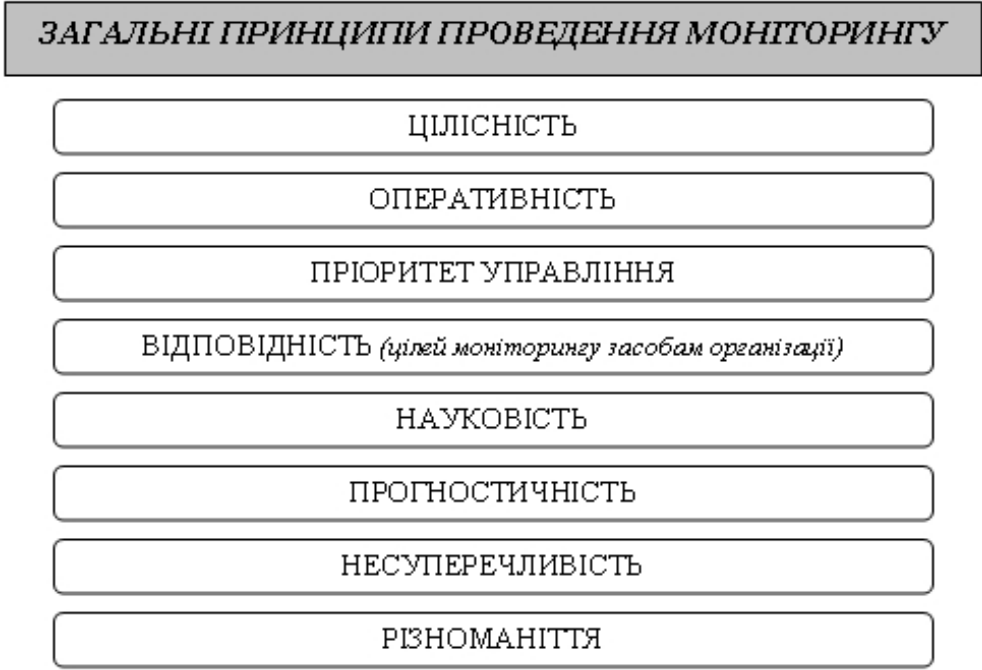


Рис. 1.2. Загальні принципи проведення моніторингу

Функції освітнього моніторингу

- ✓ *інформаційна* – створення вірогідного масиву інформації щодо якості освіти у навчальному закладі;
- ✓ *діагностична* – фіксування реального стану якості освіти у навчальному закладі;
- ✓ *прогнозуюча* – виявлення стратегії і тактики розвитку освіти у навчальному закладі;
- ✓ *управлінська* – вплив на мету, зміст і методи управлінської діяльності;
- ✓ *педагогічна* – створення цілісності процесу навчання, виховання та розвитку учнів.

Завдяки своїй структурі та функціям, моніторингова система набуває комплексного та цілісного характеру, забезпечує універсальні можливості в освітній сфері.

Системою моніторингу – це сукупність елементів, взаємодія яких забезпечує здійснення моніторингових процедур.

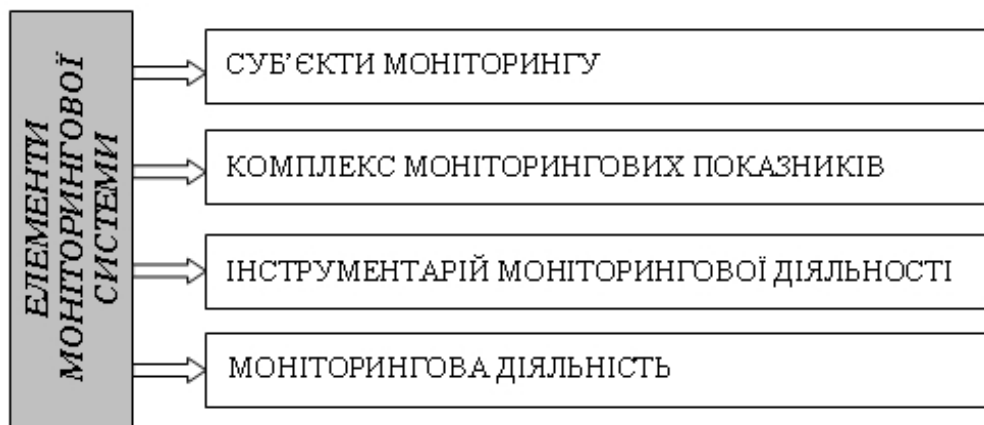


Рис. 1.3. Елементи моніторингової системи

Суб'єктами моніторингу є носії моніторингових функцій, що їх здійснюють. Вони умовно поділяються на дві великі групи: тих хто добирає і тих, хто опрацьовує. Таким чином, суб'єктами в системі освітнього моніторингу виступають як школи, вищі навчальні заклади, районні управління освітою, МОН, так і окремі люди (учні, студенти, викладачі).

Об'єктом моніторингу є система освітніх установ: дошкільна освіта; початкова загальна освіта; основна загальна освіта; загальна середня (повна) освіта; додаткова освіта; вища освіта, підвищення кваліфікації.

Предметом моніторингу є якість освіти як системоутворювальний чинник освітніх установ.

Метою моніторингу є відстеження динаміки якості освітніх послуг, освітніх установ, ефективність управління якістю освіти.

Завданнями моніторингу є:

- ✓ безперервне спостереження за станом системи освіти в межах своєї компетенції і отримання оперативної інформації про неї;
- ✓ своєчасне виявлення змін, що відбуваються в системі освіти, і чинників, що спричиняють їх;
- ✓ попередження негативних тенденцій в системі утворення міста; здійснення короткострокового прогнозування розвитку найважливіших процесів в системі освіти;
- ✓ оцінювання ефективності і повноти реалізації методичного забезпечення освіти.

Інструментарієм моніторингової діяльності виступає сукупність таких елементів, а саме: форм статистичної

звітності, інформаційних стандартів, анкет, опитних листів, які використовують суб'єкти моніторингу у своїй діяльності.

Види моніторингу.

Залежно від поставленої мети, розрізняють такі види освітнього моніторингу:

- ✓ *Інформаційний;*
- ✓ *Базовий;*
- ✓ *Проблемний:*
 - функціонування,
 - розвитку;
- ✓ *Управлінський.*

Інформаційний моніторинг передбачає структурування, накопичення, поширення інформації та не передбачає спеціально організованого дослідження на етапі збирання інформації.

Базовий моніторинг виявляє нові проблеми та небезпеки до того, як вони будуть апробовані в сфері управління.

Проблемний моніторинг дозволяє досліджувати закономірності процесів, ступені небезпеки, типологію проблем. Цей вид моніторингу можна розділити на дві складові залежно від поставлених завдань:

1. *моніторинг проблем функціонування*, який застосовується до одного завдання чи проблеми та не має обмежень у часі;
2. *моніторинг проблем розвитку* (досить динамічний), розглядає питання щодо якості, інструментарію, забезпечення, які вирішуються в короткі терміни. Управлінський моніторинг відстежує, оцінює ефективність, наслідки та результати прийнятих рішень.

Залежно від результатів експертизи, моніторинг розподіляють на:

- ✓ *динамічний* – досліджуються дані щодо динаміки розвитку об'єкта, явища, показника;
- ✓ *конкурентний* – обираються результати аналогічного дослідження освітніх систем;

- ✓ *порівняльний* – в якості бази використовуються результати ідентичного дослідження однієї чи кількох систем вищого рівня;
- ✓ *комплексний* – використовується кілька підстав для експертизи;
- ✓ *безосновний* для аналізу використовуються результати одного дослідження.

Залежно від *ієрархії систем управління*, моніторинг якості освіти можна поділити на такі види:

- ✓ *шкільний*;
- ✓ *районний*;
- ✓ *регіональний*;
- ✓ *всеукраїнський*.

Моніторинг, як і будь-яке наукове дослідження проводиться за певним алгоритмом.

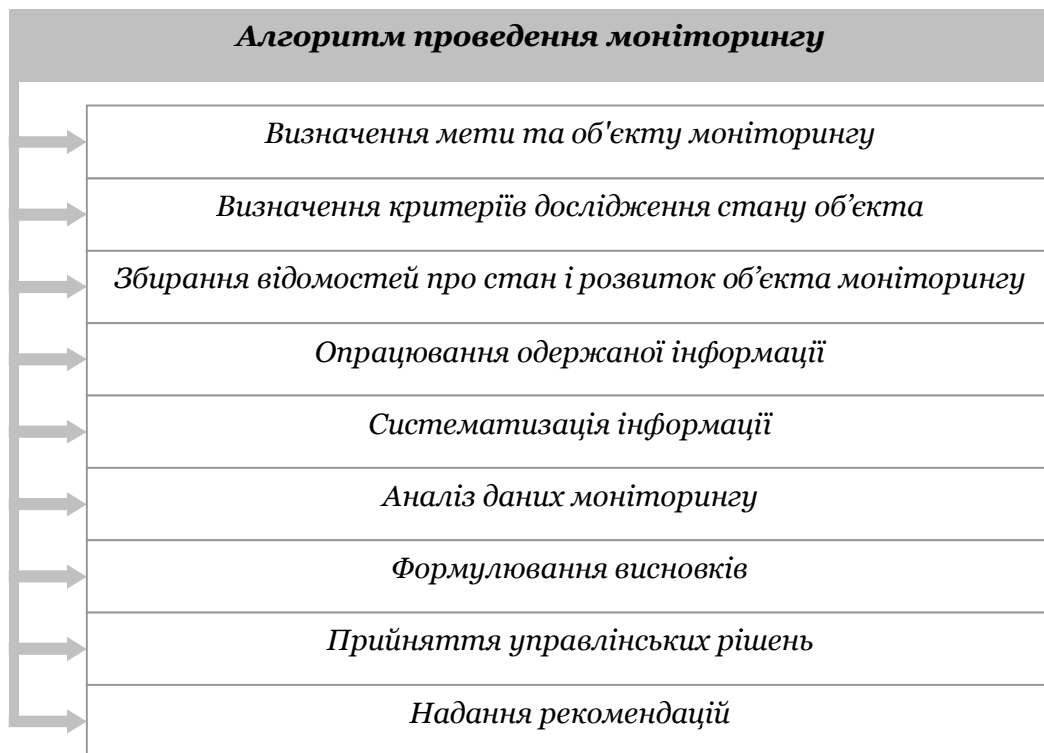


Рис. 1.4. Приклад схеми проведення моніторингу

Створення інформаційних передумов для формування цілісного уявлення про стан системи освіти, про якісні та кількісні зміни в ній визначає основну мету моніторингу розвитку освітньої системи.

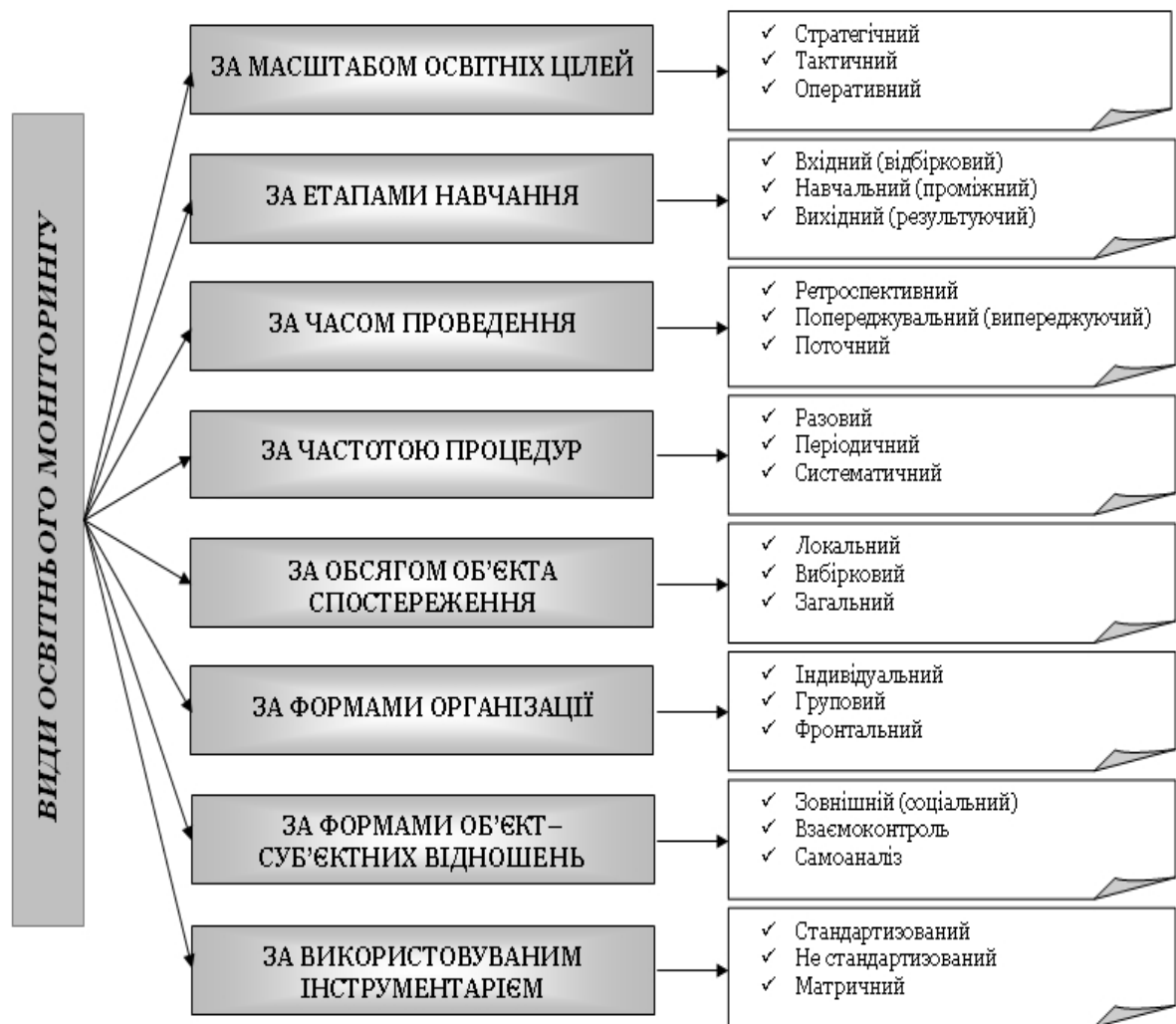


Рис. 1.5. Види освітнього моніторингу

Тому, основні завдання освітнього моніторингу зводяться до наступного:

- ✓ узгодження та виявлення комплексу показників, що забезпечують цілісне уявлення про стан системи освіти, про якісні та кількісні зміни в ній;
- ✓ систематизація відомостей щодо стану і розвитку системи освіти;
- ✓ забезпечення регулярного і наочного представлення інформації щодо процесів, які відбуваються в системі освіти;
- ✓ інформаційне забезпечення аналізу і прогнозування стану і розвитку системи освіти, вироблення управлінських рішень.

Проаналізувавши сучасну наукову літературу, напрями роботи Міністерства освіти та науки України, можна дійти висновку, що нині досить перспективним та затребуваним є

розгляд та вирішення питання моніторингу як технології управління якістю освіти.

1.5. Педагогічне оцінювання

Моніторинг та оцінювання – невіддільні. Вони є необхідними інструментами управління якістю освіти. Призначення як моніторингу, так і оцінювання полягає в тому, щоб впливати на прийняття рішень, щодо вдосконалення, переорієнтації, зміни стратегії навчання тощо.

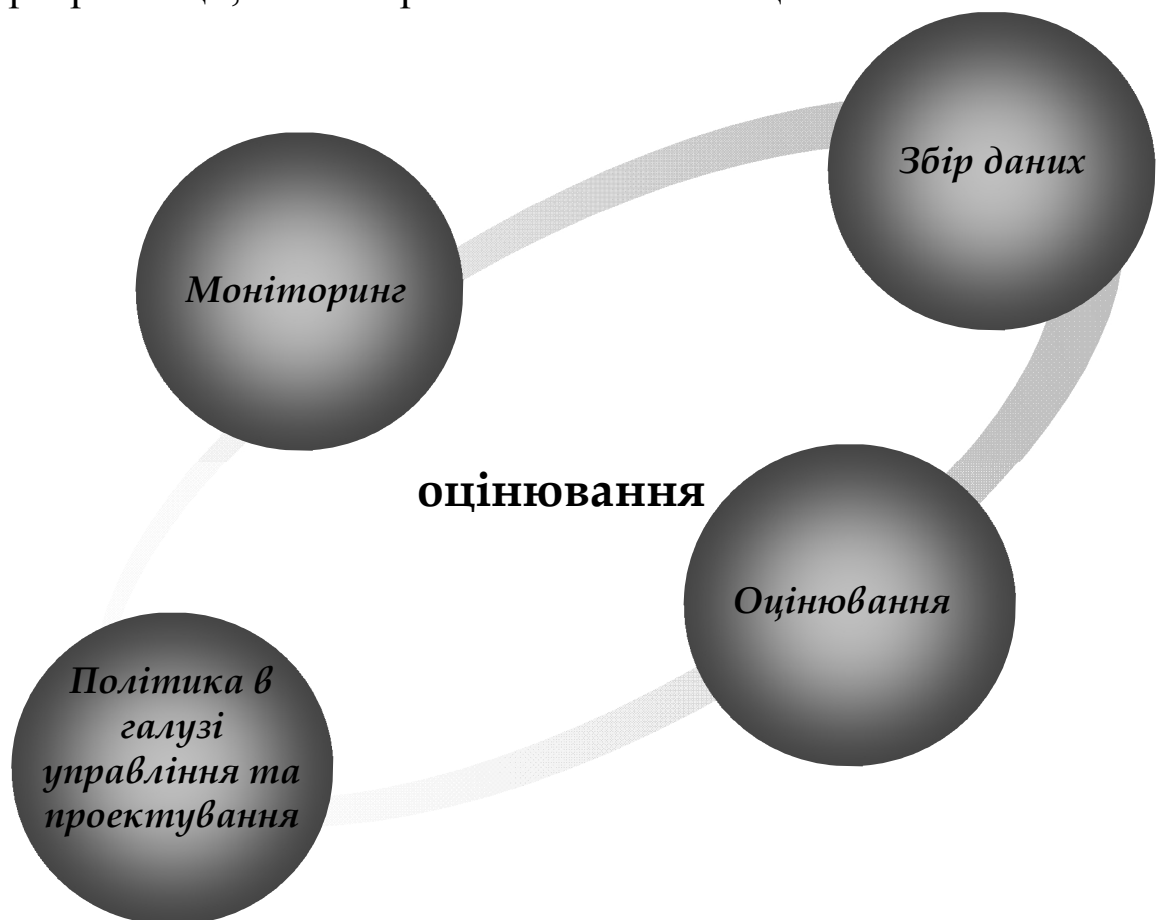


Рис. 1.6. Цикл оцінювання

Одне з перших визначень поняття «оцінювання» належить Ралфу Тайлеру (1949 р.): «оцінювання це процес виявлення того, наскільки виконано освітні завдання».

Загальноприйнятим вважається визначення, яке належить Кларенсу Едварду Бібі (1977 р.): «оцінювання освіти – це систематичне збирання і тлумачення фактів, за якими йде наступний етап – судження про їх цінність і відповідне планування подальших дій».

Керол Г. Вайс, розглядаючи в якості об'єкта оцінювання програми розвитку системи, формулює таке визначення: «**оцінюванням** називається систематичне оцінювання операцій та результатів програм чи політики порівняно з комплексом явних та неявних стандартів, з метою вдосконалення програми чи політики».

Розрізняють власне оцінювання (*evaluation*) та педагогічне оцінювання (*assessment*). У розумінні «*evaluation*» передбачено винесення ціннісного судження про процес, предмет або явище, які піддаються вимірюванню. Функціями такого оцінювання є:

- ✓ оцінювання для контролювання і звітування,
- ✓ оцінювання для розвитку,
- ✓ оцінювання для набуття знань і розуміння.

Поняття «педагогічне оцінювання» (*assessment*) є частиною оцінювання *evaluation* і означає низку процесів для визначення того, якою мірою окремі учні (студенти) досягли запланованих результатів навчання. Ці процеси тісно пов'язані з процесами навчання.



У світі педагогічне оцінювання як науково-практичний напрям існує понад 100 років. Західні освітяни давно і назавжди засвоїли тезу, підтверджену досвідом різних країн та науковими дослідженнями, що «**оцінювання керує навчанням**», і на власному досвіді переконалися, що модернізація навчальних планів та програм є неефективною, поки вона не підтримана відповідними змінами змісту, а інколи і форм оцінювання. Безумовно, цей 100-річний шлях не був легким; і одночасно з досягненнями розвиток оцінювання супроводжувався певними «хворобами розвитку» з вигляді надмірного «захоплення» деякими методами оцінювання, некоректними способами інтерпретації результатів, розквітом та спадом у використанні певних методів та підходів до оцінювання. Питання забезпечення якості педагогічних вимірювань гостро постало приблизно 50 років тому, коли масові тестові системи набули широкого розповсюдження та

стали основою для прийняття важливих освітніх рішень, які могли вплинути на долю учня або дорослого. Тоді ж були створені перші документи, спрямовані на забезпечення якості у тестуванні та оцінюванні: у 1954 році Американська психологічна асоціація розробила «Технічні рекомендації для психологічних тестів та діагностичних методик», а Американська асоціація педагогічних досліджень та Національна рада вимірювань в освіті спільно розробили «Технічні рекомендації для тестів досягнень». Ці два документи стали основою для найбільш давніх з існуючих на теперішній день «Стандартів педагогічного та психологічного тестування».

В Україні стандартизоване оцінювання, зокрема тестування, знаходиться на етапі розвитку. В країні існує дві системи національного оцінювання – система медичних ліцензійних іспитів (функціонує з 1999 року), в межах якої реалізується стандартизоване тестування всіх студентів та випускників галузі охорони здоров'я, та стандартизоване тестування школярів у вигляді зовнішнього незалежного оцінювання (функціонує як обов'язкова система з 2007-2008 року). Водночас країну заповнила низка тестових заходів (системи у школах, коледжах, університетах) та тестової продукції (друкованої, комп'ютерної), значна частина якої не відповідає найнижчим стандартам якості. Здебільшого, у країні відсутні знання та відомості, якими є ці стандарти.

Процедура оцінювання в освіті завжди займала пріоритетне місце. Проте, вона довгий час була недосконалою і пов'язувалась лише з оцінюванням знань, умінь і навичок. Значення оцінювання освітньої діяльності передусім у тому, що воно може допомогти простежити рівень розвитку освітнього процесу в Україні та в регіоні, зокрема.

Педагогічне оцінювання означає низку процесів, які використовують для визначення того, що учень знає, розуміє та вміє.

Контроль та оцінювання виступають обов'язковими складовими, та посідають важливе місце в структурі педагогічної діяльності. Щодня учитель оцінює навчальні досягнення школярів, витрачаючи на це близько третини свого

робочого часу. Протягом останніх років процес оцінювання результатів навчальної діяльності учнів зазнав істотних змін, пов'язаних із введенням у середні навчальні заклади 12-бальної шкали. У сучасній школі здійснювати оцінювальну діяльність спроможний лише педагог, який має відповідний рівень підготовки в цій галузі.

Питанню оцінювання знань присвячено чимало досліджень, наукових праць у вітчизняній та зарубіжній педагогіці.

Детально проаналізувавши, їх можна згрупувати за напрямками наукового висвітлення:

- ✓ моніторинг якості освіти (О. Байназарова, О. Локшина, О. Ляшенко, А. Сбруєва);
- ✓ розроблення сучасних інноваційних технологій, які включені в систему оцінювання знань (В. Мадзігон, В. Бочарнікова, Т. Ільїна, М. Лещенко, С. Різниченко, О. Романовський, Л. Романишина, Р. Барро (Barro R.), К. Бібі (Beeby C.), Б. Блум (Bloom B.), С. Докінз (Dawkins S.), У. Дреус (Drews U.), Д. Гехт (Heght D.), Е. Льюїс (Lewis A.), Т. Маррей (Murrey T.));
- ✓ використання багатобальної шкали оцінювання знань, умінь, навичок (В. Беспалько, М. Єрецький, В. Козаков, М. Ефрік (Africk M.), Р. Молл (Moll R.), Дж. Ноах (Noah J.), П. Діедеріч (Diederich P.));
- ✓ теоретичні підходи до оцінювання знань студентів, їх розроблення і вдосконалення (Л. Артемчук, І. Булах, Г. Костюк, В. Кремень, І. Кулібаба, Л. Одерій, Є. Перовський),
- ✓ психологічні механізми цього процесу досліджували Б. Ананьєв, І. Галян, М. Наумов, І. Якиманська;
- ✓ педагогічні аспекти цієї проблеми вивчали та аналізували Ш. Амонашвілі, С. Близнюк, К. Делікатний, В. Паламарчук, В. Полонський, С. Подмазін, Л. Романишина, П. Сікорський.

Незважаючи на те, що в педагогіці і методиці проведено багато досліджень з проблеми педагогічного оцінювання навчальних досягнень учнів, проблема системного підходу до

процесу педагогічного оцінювання в школі залишається актуальною.

Наукові праці, в яких досліджується питання якості знань студентів, досить часто містять синонімічне вживання категорій **«освітній моніторинг»** та **«педагогічне оцінювання»**. Ці поняття мають бути розмежовані на основі підходів до визначення категорії *«якість освіти»*, яка тлумачиться у політичному, управлінському і педагогічному контекстах.

Педагогічне оцінювання відображає процес взаємодії викладача і студента, в ході якого визначається якість засвоєних знань, що істотно впливає на подальший процес навчання. Педагогічне оцінювання поєднує чуттєво-емоційні та раціональні чинники, що посилюють його значення в контексті забезпечення успішного індивідуального професійного розвитку студентів.

Сутністю *педагогічних засад оцінювання* є теоретичні положення, що визначають зміст педагогічного процесу оцінювання і конкретизуються у відповідних педагогічних принципах, підходах, методах, прийомах, засобах, способах, які використовуються педагогом у визначенні якісного рівня знань студентів з навчального предмета.

Відповідно, ***педагогічні принципи*** відображають закономірності організації процесу оцінювання засвоєних студентами знань, які бере за основу педагог, визначаючи якість засвоєння навчального матеріалу та прогнозуючи перспективний професійний розвиток студентів.

У переважній більшості наукових праць, присвячених оцінюванню знань, висвітлено лише дії викладача, без аналізу взаємодії викладач – студент, реакції студентів щодо вимог педагога і т. ін. Недостатньо проаналізованою залишилася також і чуттєво-емоційна атмосфера оцінювального процесу. Педагогічні дослідження авторів, які певною мірою торкнулися питань оцінювання знань, поверхово висвітлюють педагогічні засади оцінювання знань студентів, якими керується викладач у процесі педагогічної діяльності.

Аналіз наукових досліджень, які пов'язані з тематикою педагогічного оцінювання знань, переконує у наявності деяких

суперечностей, притаманних процесові підготовки фахівців, а саме: між метою, завданнями вищої освіти і реальними результатами їх досягнення в опануванні програмового матеріалу з навчального предмета; раціонально-прагматичним підходом до забезпечення якості професійної освіти та доцільністю реалізації гуманістично спрямованого й особистісно-орієнтованого професійного навчання.

1.6. Система педагогічного оцінювання.

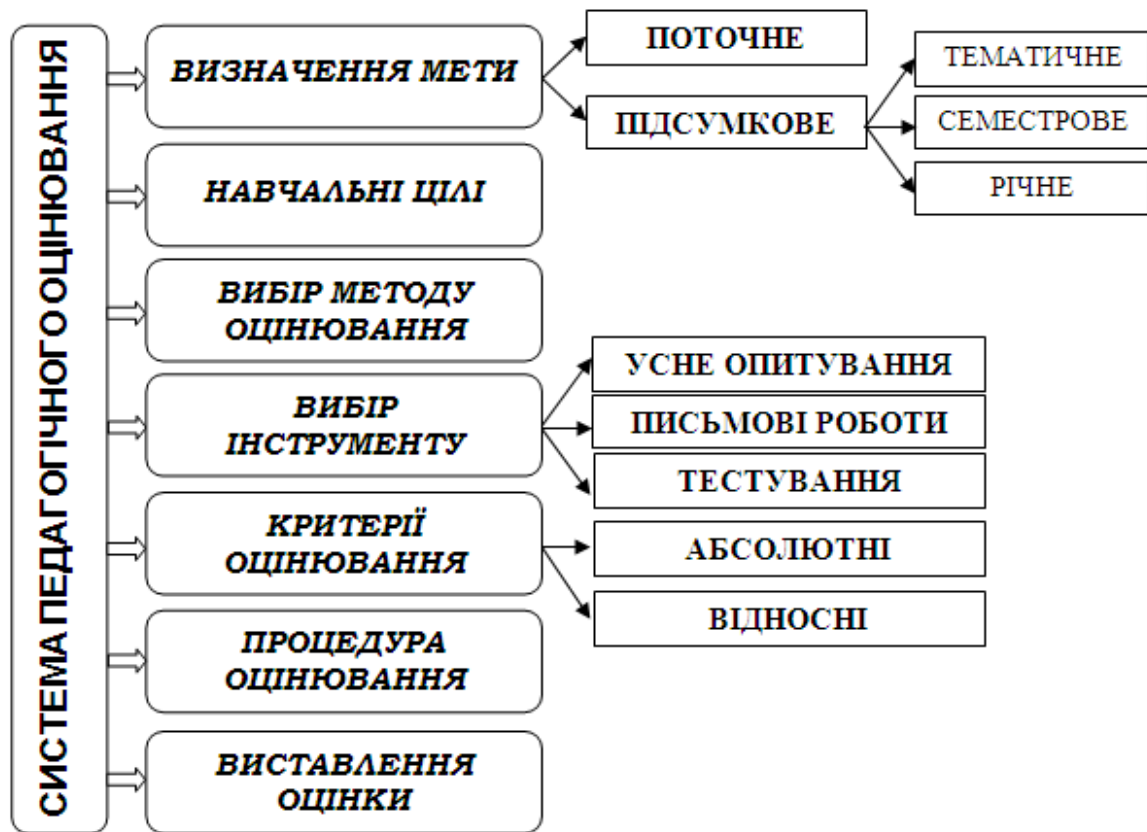


Рис. 1.7. Система педагогічного оцінювання навчальних досягнень учнів

Відповідно до мети, педагогічне оцінювання навчальних досягнень учнів може бути поточним, підсумковим (тематичне, семестрове, річне оцінювання) та державна підсумкова атестація.

Поточне оцінювання – це процес встановлення рівня навчальних досягнень учня (учениці) в оволодінні змістом предмета, уміннями та навичками відповідно до вимог навчальних програм.

Поточне оцінювання здійснюється у процесі поурочного вивчення теми. Його основними завданнями є:

- ✓ встановлення й оцінювання рівнів розуміння і первинного засвоєння окремих елементів змісту теми;
- ✓ встановлення зв'язків між ними та засвоєним змістом попередніх тем;
- ✓ закріплення знань, умінь і навичок.

Підсумковому оцінюванню навчальних досягнень підлягають основні результати вивчення теми (розділу).

Підсумкова оцінка виставляється на підставі результатів опанування учнями матеріалу теми впродовж її вивчення з урахуванням поточних оцінок, різних видів навчальних робіт (практичних, лабораторних, самостійних, творчих, контрольних) та навчальної активності школярів.

Державна підсумкова атестація – це форма контролю відповідно до освітнього рівня випускників навчальних закладів системи загальної середньої освіти.

За змістом навчального матеріалу контроль поділяється на три види:

- контроль теоретичних знань;
- контроль виконання вправ та розв'язування задач;
- контроль практичних робіт.

Усі ці види контролю можуть здійснюватись як в усній, так і в письмовій формі з обов'язковою рівневою диференціацією складності завдань.

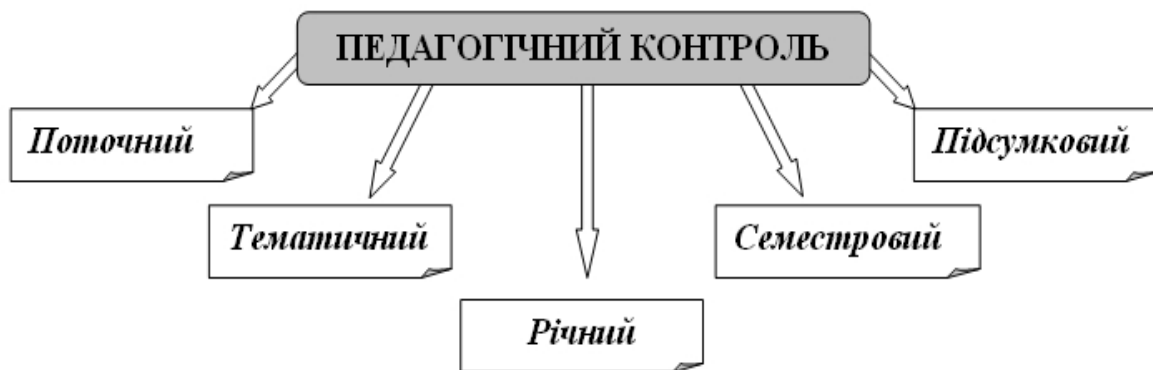


Рис. 1.8. Види педагогічного контролю

1.7. Класифікація цілей навчання. Таксономія Блума

Цілі навчання можуть бути виражені в таких поняттях і категоріях, як знання, уміння, навички, способи творчої діяльності й емоційно оцінні норми. Кожна з них, має не просте значення, однак у своїй сукупності вони дають повне уявлення про те, чому маємо вчити. Головна мета навчання і виховання полягає в передачі молодому поколінню нагромадженої попередніми поколіннями людської культури.



Навчальні цілі – це результат, який прогнозується і має бути досягнутий у процесі всього навчання і на його обмежених етапах.

Вимога чіткого та однозначного опису цілей є базовою для більш узагальненої вимоги діагностичності досягнення цілей навчання, яка, крім певного однозначного опису цілей навчання, означає також визначення методу вимірювання результатів навчання для оцінювання ступеня реалізації навчальних цілей.

Чіткому визначенню цілей навчання сприяє їх класифікація (таксономія).

Таксономія цілей – класифікація цілей, групування за певними ознаками.

Процес розробки таксономії цілей навчання може здійснюватися у декілька етапів. На першому етапі визначаються загальні цілі та можливий вплив засвоєного змісту дисципліни на спрямованість особистості студентів. На другому етапі цілі навчання уточнюються в процесі розроблення моделі засвоєння бази знань.

Модель засвоєння бази знань навчальної теми – це перелік елементів бази знань з визначенням рівня засвоєння кожного елемента.

До елементів бази знань відносяться:

- поняття, терміни, факти, символи, судження.
- тенденції, властивості, теорії, критерії, закони
- правила, принципи, норми, методи, процеси, алгоритми.



Рис. 1.9. Елементи бази знань

Таксономія цілей навчання є їх описом у термінах професійної діяльності (поведінки):

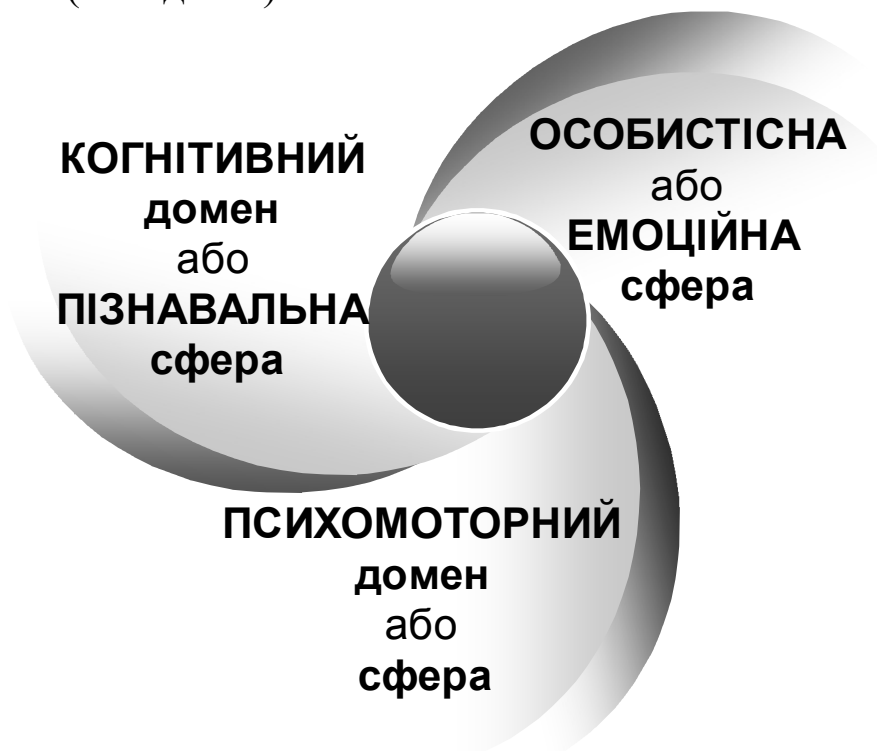


Рис. 1.10. Складові елементи таксономії цілей навчання

Організація процесу навчання перш за все пов'язана з чітким визначенням його цілей, а також усвідомленням і прийняттям їх учнями. Цільові установки навчання сприяють розумінню студентами суті і способів організації навчально-пізнавальної діяльності, істотно впливаючи на її активізацію.

Як у процесі навчання, так і під час проведення кожного навчального заняття реалізуються *три основні групи взаємопов'язаних цілей*.

До першої з них належать *навчальні цілі*: оволодіння знаннями, уміннями, навичками; до другої – *розвивальні цілі*: розвиток інтелектуальної, емоційно-вольової, діяльнісно-поведінкової сфери особистості; до третьої – *виховні цілі*: формування наукового світогляду, моральної, художньо-естетичної, правової, трудової, екологічної культури тощо.

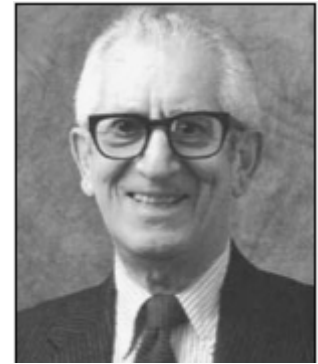
Перехід від загального уявлення про результати навчання до конкретнішого передбачає чіткий опис того, що учень може досягти в результаті навчання.

Загальний прийом конкретизації цілей – використання дієслів, які вказують на конкретну дію: аналізувати, синтезувати, інтерпретувати, оцінювати, розуміти, перетворювати, застосовувати, висловлювати, створювати тощо.

Для вчителя-практика важливо використовувати впорядковану ієрархічну класифікацію цілей тому, що вона:

- 1) забезпечує концентрацію зусиль на головному в діяльності, визначенні першочергових завдань і перспектив подальшої роботи;
- 2) створює можливості для роз'яснення учням орієнтирів навчальної роботи, спільної зацікавленої роботи педагога й учнів;
- 3) створює еталони оцінки результатів навчання, які можна розробляти й уточнювати разом з учнями (чіткі формулювання цілей, які відображені в результатах діяльності, забезпечують надійність і об'єктивність оцінки).

Автором однієї з перших схем педагогічних цілей був американський вчений Б. Блум. Після закінчення Другої світової війни група американських педагогів і психологів під керівництвом відомого вченого Б. Блума розробила загальні способи і правила чіткого формулювання та впорядкування педагогічних цілей. Ним випущена в світ перша частина «Таксономії» (1956 р.), де описуються цілі пізнавальної (когнітивної) області. Поняття «таксономія» взято з біології (у перекладі з грецької означає «розміщення в порядку»; «закон»).



Б. Блум

До цілей першої, когнітивної групи, входять такі, які передбачають запам'ятовування і відтворення вивченого матеріалу, а також розв'язання проблем, у ході яких необхідно переосмислити наявні знання, будувати їх нові об'єднання, структури, створювати нові знання. Цілі цієї групи в основному представлені в навчальних програмах, підручниках та посібниках, у повсякденній шкільній практиці.

Другу групу цілей (афективна, емоційно-ціннісна сфера) становлять цілі формування емоційно-особистісного ставлення до навколишнього світу. Вони виражаються через сприймання, інтерес, нахили, здібності, переживання почуттів, формування ставлення, його осмислення і вияв у діяльності.

Цілі навчання психомоторної сфери становлять третю групу. Вони включають ті чи інші види моторної (рухливої) маніпулятивної діяльності нервово-м'язової координації. Це навички письма, мовні, фізичні та трудові навички. Використання чіткої, впорядкованої системи цілей навчання, на думку Б. Блума, дуже важливе для побудови навчального процесу у зв'язку з тим, що, по-перше, знаючи цілі навчання, учитель впорядковує їх, визначає першочергові, базові, перспективу подальшої роботи; по-друге, знання вчителем конкретних цілей дає можливість пояснити учням орієнтири в їх спільній роботі; по-третє, чітке формулювання цілей, які виражені через результати діяльності, піддається надійному і об'єктивному оцінюванню.

Елементи таксономії Блума

Цілі навчання у когнітивній сфері можуть бути виражені через такі елементи засвоєння (їх ще називають елементами таксономії Блума): знання (відтворення), розуміння, застосування, аналіз, синтез і оцінка (оцінювання). Розглянемо їх зміст, а також ту діяльність, виконання якої передбачається поданим рівнем засвоєння змісту.

Таблиця 1

Зміст діяльності учня відповідно до цілей навчання

1. Знання (відтворення)	Ця категорія означає запам'ятовування та відтворення матеріалу, який вивчається. Йдеться про запам'ятовування і відтворення термінів, конкретних фактів, методів і процедур, основних понять, правил, принципів, цілісних теорій.
2. Розуміння	Показником розуміння вивченого може бути здатність людини встановлювати зв'язок одного матеріалу з іншим, перетворювати його із однієї форми вираження в іншу, переводити його з однієї «мови» на іншу (наприклад, зі словесної у графічну, математичну і навпаки). Як показник розуміння може також бути інтерпретація матеріалу учнем (пояснення, короткий виклад), прогнозування майбутніх наслідків, що впливають із наявних даних.
3. Застосування	Цей елемент засвоєння означає вміння використовувати вивчений матеріал у конкретних умовах і нових ситуаціях. Сюди входить застосування правил, методів, вміння поділяти матеріал на складові поняття, законів, принципів, теорій.
4. Аналіз	До цієї категорії належить виділення частин цілого, виявлення взаємозв'язку між ними, осмислення принципів організації цілого. Навчальні результати характеризуються осмисленням не тільки змісту навчального матеріалу, а і його внутрішньої структури. Учень, який добре оволодів цією категорією навчальних цілей, бачить помилки й огріхи в логіці міркувань, бачить різницю між фактами і наслідками, оцінює значущість даних.
5. Синтез	Ця категорія означає вміння комбінувати елементи, щоб одержати ціле з новою системною властивістю. Таким новим продуктом може бути повідомлення, план дій, нова схема тощо.

6. Оцінювання	Як категорія навчальних цілей, вона означає вміння оцінювати значення того чи іншого матеріалу для конкретної мети. Судження і результати дії учня мають засновуватися на чітких критеріях. Учень оцінює логіку побудови матеріалу у вигляді письмового тексту, оцінює відповідність висновків уже даним і т.д.
----------------------	---

Таблиця 2

Категорії навчальних цілей в когнітивній сфері

<i>Категорії навчальних цілей</i>	<i>Приклади узагальнених типів навчальних цілей</i>
1. Знання	Учень: <ul style="list-style-type: none"> ✓ знає терміни, що використовуються, ✓ знає конкретні факти, ✓ знає методи і процедури, ✓ знає основні поняття, ✓ знає правила і принципи.
2. Розуміння	Учень: <ul style="list-style-type: none"> ✓ розуміє факти, правила, принципи, ✓ інтерпретує словесний матеріал, схеми, графіки, діаграми, ✓ перетворює словесний матеріал в математичні вирази; ✓ приблизно описує можливі наслідки.
3. Застосування	Учень: <ul style="list-style-type: none"> ✓ використовує поняття і принципи в нових ситуаціях; ✓ застосовує закони і теорії в конкретних практичних ситуаціях; ✓ демонструє правильне застосування методу чи процедури.
4. Аналіз	Учень: <ul style="list-style-type: none"> ✓ виділяє приховані припущення; ✓ бачить помилки і недоліки в логіці міркувань; ✓ розмежовує факти і наслідки; ✓ оцінює значущість даних.

Категорії навчальних цілей	Приклади узагальнених типів навчальних цілей
5. Синтез	Учень: ✓ пише невеликий твір; ✓ пропонує план проведення експерименту; ✓ використовує знання з різних галузей, щоб скласти план розв'язання тої чи іншої проблеми.
6. Оцінювання	Учень: ✓ письмово оцінює логіку побудови матеріалу; ✓ оцінює відповідність висновків наявним даним, значущість того чи іншого продукту діяльності, використовуючи внутрішні критерії; ✓ оцінює значущість того чи іншого продукту діяльності, використовуючи зовнішні критерії.

Рівні оцінювання знань. Когнітивний домен. Bloom, 1956



Науковці подають різні рівні класифікації когнітивної сфери, наведемо порівняння різних таксономічних систем для когнітивної сфери Б. Блума, В. Беспалька, М. Скаткіна, Г. Клауса і Ж. Піаже (за Л. Артемчук):

Таблиця 3

Порівняльний аналіз таксономічних систем для когнітивної сфери

Б. Блум	В. Беспалько	М. Скаткін	Г. Клаус і Ж. Піаже
Знання (запам'ятовування)	Знайомство (ідентифікація)	Сприймання, осмислення, запам'ятовування	Примітивна форма сприймання
Розуміння (змісту вивченого)	Відтворення (типів завдань)	Застосування (у типових ситуаціях)	Початкове осмислення
Застосування (у нових ситуаціях)	Вміння (нетипові завдання)		Змістовий (формування понять)
Аналіз (розділити ціле на складові)	Трансформація, творчість (побудова методу розв'язування проблем і пошукових задач)	Застосування (у нових ситуаціях)	
Синтез (створювання цілого з частин)			
Оцінювання (визначення цінності засобів для досягнення мети)			

Існують системи описання навчальних результатів через характеристику рівнів засвоєння знань.

Так, **І. Лернер** пропонує розрізняти три рівні засвоєння знань:

- 1) первинне засвоєння, впізнавання-відтворення;
- 2) застосування в знайомій ситуації (за зразком);
- 3) застосування в новій ситуації (творче).

Загально визнаною є класифікація рівнів засвоєння, яку запропонував **В. Беспалько**:

- 1) впізнавання об'єктів, властивостей, процесів під час повторного сприймання відомостей про них чи дій з ними (знання-знайомство);
- 2) відтворення, репродуктивна дія – самостійне відтворення і застосування відомостей для виконання даної дії (знання-копії);

- 3) застосування, продуктивна дія – пошук і використання суб'єктивно нових відомостей для самостійного виконання нової дії (знання, уміння і навички);
- 4) творчість, творча дія – самостійне конструювання способу діяльності, пошук нових відомостей (знання-трансформації).

Спосіб постановки цілей, який пропонує педагогічна технологія, полягає в тому, що цілі навчання формулюються через результати навчання, які виражені в діях учнів. Вимоги до цілей навчання: цілі навчання повинні бути обґрунтовані, практично досяжними, діагностичними та відповідати змісту Державного стандарту.

14 січня 2004 року постановою Кабінету Міністрів України затверджено Державний стандарт базової і повної середньої освіти, яким було унормовано систему показників освіченості особи. Державний стандарт базової і повної середньої освіти – це зведення норм і положень, що визначають державні вимоги до освіченості учнів і випускників шкіл на рівні початкової, базової та повної загальної середньої освіти, а також гарантії держави щодо її здобуття.

Реалізується він у нормативних документах, які визначають суспільно зумовлений зміст загальної середньої освіти, вимоги та гарантії держави щодо її одержання громадянами. Обов'язковим компонентом процедури педагогічного оцінювання є вибір методу оцінювання.



Метод оцінювання – це категорія, що з'ясовує, яким чином визначаються показники, які вимірюють.



Вимірювання – це процес надання чисельного значення певному показнику відповідно до його кількісного прояву із застосуванням чітко визначених правил вимірювання.


У нашому випадку, це відповідність засвоєного матеріалу відносно державного предметного стандарту.



Метод вимірювання – це спосіб, за допомогою якого здійснюється надання кількісного значення показнику,

який вимірюється, тобто виставлення відповідної оцінки.

Для реалізації процедури вимірювання обов'язково має бути обрано інструмент вимірювання.

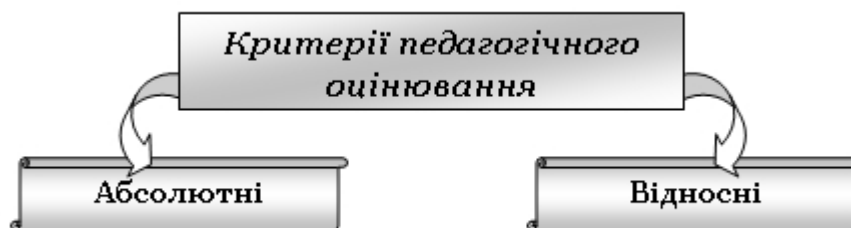
 **Інструмент вимірювання** – це категорія, що визначає засіб, за допомогою якого здійснюється вимірювання.

Наприклад, під час усного опитування і письмової роботи інструментом вимірювання є, відповідно, запитання поставлене в усній формі та білет або структуроване завдання, а під час тестування – тест, складений із тестових завдань.

Важливим і обов'язковим компонентом педагогічного оцінювання є добір критеріїв.

Критерії оцінювання навчальних досягнень реалізуються в нормах оцінок, які встановлюють чітке співвідношення між вимогами до знань, умінь і навичок, які оцінюються, та показником оцінки в балах.

Критерії педагогічного оцінювання



Відносний критерій залежить від порівняння між собою результатів, здобутих учнями.

Абсолютний критерій, на відміну від відносного, не залежить від результатів усіх учнів, яких оцінюють за даним критерієм.

Для учителя, який складає завдання для перевірки рівня успішності за абсолютним критерієм, не має значення, які результати здобувають інші, оскільки здобута ними оцінка не впливатиме на критерій.

Визначення критеріїв для педагогічного оцінювання має обов'язково відбуватися з урахуванням мети оцінювання.

Наступна складова системи педагогічного оцінювання – це *проведення процедури оцінювання* із урахуванням мети,

навчальних цілей, методу, інструменту та критеріїв педагогічного оцінювання.

Остання складова системи педагогічного оцінювання – виставлення оцінки з урахуванням усіх складових системи педагогічного оцінювання.

1.8. Педагогічне вимірювання та тестування

Під педагогічним вимірюванням можна розуміти процес встановлення відповідності між оцінюваними характеристиками учнів (студентів), і точками емпіричної шкали, в якій відношення між різними оцінками характеристик виражені властивостями числового ряду.

Як характеристики учнів, виступають їх знання і уміння на момент виконання завдань (тесту). Роль одиниці вимірювання відіграють тестові завдання. Об'єктом оцінювання стають особи, що навчаються. Результатом вимірювання – шкала балів тестованих.

Найефективнішим засобом вимірювання, що забезпечує об'єктивність і зіставність оцінок в процесі контролю, є педагогічні тести.

Негативне відношення до тестування є наслідком недостатньо повного та чіткого уявлення про нього як про інструментальний засіб навчального процесу. Якщо розглядати тестування як діагностичну процедуру, то зникає багато побоювань, пов'язаних з його використанням в навчальному процесі.

Широко методи діагностування застосовуються при визначенні рівня володіння навчальним матеріалом (педагогічна діагностика).

Основна мета педагогічного діагностування – отримання діагностичних відомостей. Для ухвалення вірних рішень відомості про реальне положення справ повинні бути об'єктивними, інформативними і придатними для контролю.

Роль діагностичних властивостей при педагогічному діагностуванні відіграють тестові завдання, які повинні відповідати певним вимогам (інформативність, придатність для контролю і т.д.). Тільки для тестів існують науково-

обґрунтовані критерії якості. Тест утворює не будь-яка система завдань, а лише та, яка забезпечує інформативні оцінки рівня і якості підготовки осіб, що навчаються.

Критерій інформативності є ключовим для характеристики тесту як засобу вимірювання.

Інформативність тесту залежить як від характеристики тестових завдань (для оцінки інформативності завдань необхідне емпіричне дослідження тесту на репрезентативній вибірці осіб, що навчаються), так і від рівня підготовки учнів (студентів).

Рівні сучасного розуміння тесту і тестування

«**Побутовий**» рівень. На цьому рівні тест розуміють як набір питань з варіантами відповідей (на одному рівні з кросвордами, головоломками)

«**Словниковий**» рівень. Для нього характерні різні погляди і суперечності в поняттях і визначеннях. Саме на цьому етапі знаходиться сучасна тестологія на пострадянському просторі.

«**Науковий**» рівень. Є найточнішим, враховує особливості тестів і відображає вимоги до них.



Тест – це інструмент, що складається з вивіреної системи тестових завдань, стандартизованої процедури проведення тестування, заздалегідь спроектованої технології опрацювання і аналізу результатів, призначений для вимірювання якостей і властивостей особи, зміна яких можлива в процесі систематичного навчання.

В основі тесту лежить спеціально підготовлений і випробуваний набір завдань, що дозволяє об'єктивно і надійно оцінити досліджувані якості і властивості на основі використання статистичних методів.

1.9. Переваги та недоліки тестового контролю

Відмінна риса тесту – наявність вимірювання, функція якого полягає в тому, щоб надавати кількісну і якісну інформацію щодо прогресу навчання, діагностики недоліків, прогнозування успішності.

Тестовий контроль відрізняється від інших методів контролю (усні і письмові іспити, заліки, контрольні роботи і тому подібне) тим, що він є спеціально підготовленим контрольним набором завдань, що дозволяє надійно та адекватно кількісно оцінити знання учнів (студентів) за допомогою статистичних методів.

Тести – якісний і об'єктивний спосіб оцінювання

Об'єктивність досягається шляхом стандартизації процедури проведення тестування, стандартизації і перевірки показників якості завдань і тесту в цілому. Тести мають широкі можливості для ухвалення управлінських рішень.

Порівняно з іншими формами контролю знань, тестування має свої переваги і недоліки.

Переваги

- ✓ Тестування є якіснішим і об'єктивнішим способом оцінювання, його об'єктивність досягається шляхом стандартизації процедури проведення, перевірки показників якості тестових завдань і тестів взагалі.
- ✓ Тестування – справедливіший, порівняно з іншими, метод, воно ставить всіх учнів в рівні умови, як у процесі контролю, так і в процесі оцінювання, практично виключаючи суб'єктивізм викладача. За даними англійської асоціації NEAB, що займається підсумковою атестацією учнів Великобританії, тестування дозволяє знизити кількість апеляцій більш ніж у три рази, зробити процедуру оцінювання однаковою для всіх учнів незалежно від місця проживання, типу і виду освітньої установи, в якій навчаються учні.
- ✓ Тести це об'ємніший інструмент, оскільки він може включати в себе завдання з усіх тем курсу, у той час як на усний іспит, зазвичай, виноситься 2 – 4 теми, а на письмовий – 3 – 5. Це дозволяє виявити знання учня по всьому курсу, виключивши елемент випадковості при «витягуванні» екзаменаційного білета. За допомогою тестування можна встановити рівень знань учня з предмету в цілому і з окремих його розділів.

- ✓ Тест це точніший інструмент, так, наприклад, шкала оцінювання тесту із 20 питань, складається з 20 поділок, в той час, як звичайна шкала оцінки знань – лише з чотирьох.
- ✓ Тестування ефективніше з економічної точки зору. Основні витрати при тестуванні припадають на розроблення якісного інструментарію, тобто мають разовий характер. Витрати ж на проведення тестування значно нижчі, ніж при письмовому або усному контролі. Проведення тестування і контроль результатів у групі з 30 чоловік займає півтори–дві години, усний або письмовий іспит – не менше чотирьох годин.
- ✓ Тестування – більш «м'який» інструмент, воно ставить усіх учнів в рівні умови, використовуючи єдину процедуру і єдині критерії оцінювання, що призводить до зниження передекзаменаційних нервових напружень.

Недоліки

- ✓ Дані, отримані викладачем в результаті тестування, хоча й містять у собі відомості про прогалини в знаннях з конкретних розділів, але не завжди дозволяють судити про причини цих прогалин.
- ✓ Тест не дозволяє перевіряти й оцінювати високі, продуктивні рівні знань, пов'язані із творчістю, тобто ймовірнісні, абстрактні та методологічні знання.
- ✓ Широта охоплення тем у тестуванні має і зворотній бік. Учень під час тестування, на відміну від усного або письмового іспиту, не має достатньо часу для глибокого аналізу теми.
- ✓ Забезпечення об'єктивності і справедливості тесту вимагає спеціальних заходів щодо забезпечення конфіденційності тестових завдань. При повторному застосуванні тесту бажаним є внесення змін до завдань.
- ✓ У тестуванні присутній елемент випадковості. Наприклад, учень, не відповівши на просте запитання, може дати правильну відповідь на складніше. Причиною цього може бути, як випадкова помилка в першому запитанні, так і вгадування відповіді у

другому. Це спотворює результати тестування і призводить до необхідності обліку ймовірнісної складової при їх аналізі.

1.10. Політика з оцінювання та тестування

В Національній доктрині розвитку освіти підкреслюється, що “модернізація системи освіти спрямована на забезпечення її якості відповідно до найновіших досягнень вітчизняної і світової науки, культури та соціальної політики”.

До концептуальних характеристик якості національної освіти, окреслених у нормативно-правових документах віднесено такі положення:

- якість освіти є національним пріоритетом;
- висока якість освіти передбачає органічний зв'язок освіти і науки, педагогічної теорії і практики;
- модернізація системи освіти спрямована на забезпечення її якості відповідно до найновіших досягнень вітчизняної і світової науки, культури та соціальної політики.

Концепція загальної середньої освіти (12-річної школи) також проголошує якісну освіту стратегічним напрямом розвитку української освіти: “Загальноосвітня школа України має здійснити прорив до якісно нової освіти всіх дітей шкільного віку”.

Отже, якість освіти стає визначальним принципом інноваційної освітньої системи, в якій основою всіх зрушень є створення платформи для розкриття потенційних можливостей дітей, прогнозування потреб і моделей розвитку особистості.

Стандарти якості – документи

Аналіз опублікованих документів, що стосуються якості оцінювання і тестування, дозволяє згрупувати їх таким чином:

1^{ша} група – технічні стандарти

- Стандарти освітнього та психологічного тестування (США, 1999), надалі – АРА Стандарти;
- ETS Стандарти для забезпечення якості та справедливості (США, 2002);

- Рекомендації з освітнього та психологічного тестування (CPA, 1996);
- EFPA Модель Аналізу та Оцінювання Психологічних Тестів).

2^{га} група – концептуальні стандарти

- JСТР Кодекс справедливого тестування в освіті,
- JСТР Права та відповідальність користувачів тестів: рекомендації та очікування,
- Міжнародні рекомендації ІТС з використання тестів (2000),
- Міжнародні Рекомендації ІТС з Комп'ютерного та Інтернет тестування (2005),
- NFA принципи та індикатори для систем оцінювання студентів.

3^я група – вимоги до професійної відповідальності різних категорій учасників процесу оцінювання

- NCME Кодекс професійної відповідальності при педагогічних вимірюваннях,
- NCME Стандарти компетентності вчителів з оцінювання учнів,
- NCME Стандарти компетентності освітніх адміністраторів з оцінювання учнів,
- ІТС Міжнародні рекомендації з використання тестів).

Технічні стандарти спрямовані на визначення характеристик якості продукту (тесту) з технічними деталями та інструкціями. Вони використовуються для регулювання використання тестів залежно від контексту, для формулювання технічних вимог до високоякісної професійної практики, визначення валідності, надійності та інших властивостей тестового інструменту та результатів.

Концептуальні стандарти стосуються принципів, на яких базується система оцінювання та діяльність різних учасників процесу. Концептуальні стандарти менше торкаються технічних аспектів тестування, а приділяють більше уваги етичному використанню оцінювання та взаємовідносинам між різними учасниками в системі оцінювання. В цілому вони намагаються змістити акценти з надмірного розповсюдження

традиційного стандартизованого тестування і «поставити навчання у центр процесу оцінювання».

Стандарти професійної відповідальності ставлять чіткі вимоги до осіб, задіяних в оцінюванні.

Деякі стандарти, наприклад, міжнародні рекомендації комісії ІТС, відносяться до двох категорій (технічні стандарти та концептуальні стандарти). На відміну від інших документів їхні розробники намагались створити універсальний шаблон для розробки національних документів якості, тому в них немає національної специфіки.

Незважаючи на структурну та змістовну різницю всі документи базують на однакових принципах тестування і оцінювання, які походять з сучасних теорій оцінювання.

Стандарти якості в галузі педагогічного оцінювання

Найвпливовішим та відомим документом першої групи є АРА Стандарти освітнього та психологічного тестування (АРА Standards for Educational and Psychological Testing, 1999) спільно розроблені Американською асоціацією освітніх досліджень (АЕРА), Американською психологічною асоціацією (АРА) та Національною радою вимірювання в освіті (NCME).

«Представником» другої групи є ІТС Міжнародні рекомендації з використання тестів, які не мають «наміру “винайти” нові рекомендації, а радше об'єднати спільні принципи, що відбиті у існуючих рекомендаціях, кодексах належної практики, стандартах та інших відповідних документах, а також створити цілісну структуру, в якій їх можна зрозуміти та використовувати». На відміну від інших документів Рекомендації ІТС містять опис не лише того, що мають робити особи, задіяні у оцінюванні, але й те, які знання, вміння та навички повинні мати ці особи.

1.11. Учасники процесу оцінювання (тестування)

Оцінювання впливає на осіб, організації та суспільство у цілому. Прикладами індивідів, які мають відношення до оцінювання, є учні, студенти, батьки, вчителі, керівники шкіл,

освітні адміністратори, претенденти на робоче місце, працівники та роботодавці, інспектори, державні чиновники тощо. Інтереси різних сторін та учасників оцінювання можуть бути гармонійними або суперечливими.

Можна виділити три основні **категорії учасників тестування**:

- *розробники тесту* (системи оцінювання), які розробляють, видають, а також адмініструють та обробляють тестові (оціночні) матеріали;
- *користувачі тестів* (оцінювання), яким потрібні результати тестування (оцінювання) для прийняття певних рішень;
- *особи, які оцінюються*.


Додатково можна виділити таких учасників як:

- *замовники* (спонсори тесту, оцінювання),
- *адміністратори* тесту (оцінювання),
- *експерти-рецензенти* тесту та інші.

Функції та ролі цих учасників можуть поєднуватися, а інколи можуть поділятися ще з виділенням більш вузьких або спеціалізованих функцій залежно від завдань, які виконуються у процесі оцінювання. Слід зауважити, що структура праці в Україні в галузі оцінювання лише формується і потреба в деяких суб'єктах оцінювання, що існують в західних країнах, в наших умовах поки що не виникла.

§ 2. Педагогічні тести. Терміни і визначення. Педагогічний контроль, предмет і об'єкт контролю. Принципи педагогічного контролю

2.1. Історія становлення та розвитку тестування

 Слово «тест» походить від англійського «test» та означає «випробування, дослідження», а корені тестування (тобто перевірки на придатність до певних функцій) губляться в далекій давнині.

Тестологія – (з англ. *test* – проба, випробування) – наука про вимірювання психофізіологічних та особистісних характеристик.

Тест – сукупність завдань, які переважно вимагають однозначної відповіді, укладений за певними правилами та процедурами, передбачає попередню експериментальну перевірку й відповідає таким характеристикам ефективності, як валідність і надійність.

Тестування – це одночасно і метод, і результат вимірювання.

Мабуть, неможливо відновити повну і точну історію тестів і тестування. Проте спробуємо вказати деякі ключові віхи в цій історії.

1. У древньому Вавилоні проводили випробування випускників у школах професійних писарів.
2. У давньому Єгипті використовувалася система випробувань для співбесід і подальшої перевірки умінь.
3. У древній Греції проводили випробування (наприклад, в школі Піфагора) за допомогою спеціальних математичних завдань підвищеної складності.
4. У древньому Китаї використовували перші професійно орієнтовані випробування для претендентів на посаду урядовця (тих, хто пройшов таке випробування з кращими результатами, приймали на державну службу).
5. У древній Індії застосовували випробування на мудрість за допомогою різних ігор, наприклад, з використанням шахів, нард.

6. У древньому В'єтнамі проводили регулярну атестацію і переатестацію всіх цивільних чиновників і військовослужбовців.
7. В Англії вперше використовувалися письмові роботи для підвищення мотивації навчальної роботи. Родоначальником тестування часто називають англійського ученого Френсиса Гальтона, який в 1884 – 1885 рр. проводив серію випробувань для відвідувачів своєї лабораторії.
8. У США з'явилася перша книга з тестовими завданнями з математики, історії, граматики, навігації, текстами для творів з використанням простих статистичних розрахунків у педагогічній роботі.
9. У Німеччині проводили усні (на відміну від традиційних англійських письмових) іспити з комісією не менше, ніж у кількості з двох екзаменаторів.
10. У Франції з'явилися перші спроби тестування інтелектуальних здібностей дітей за допомогою спеціально створених тестів для перевірки інтелектуальності.
11. У Росії (дорадянський період) до кінця ХІХ століття будь-яка серйозна робота з тестування не описана. Перші тести з'явилися на початку ХХ століття і використовувалися як допоміжний інструментарій перевірки знань.
12. В СРСР тести використовували як допоміжний варіант. Були навіть спроби заборони тестування. Наприклад, в 1936 році було прийнято Постанову Ради Народних Комісарів "Про педологічні викривлення в системі Наркомпросу". Все це призвело до застою в розвитку тестування на період з середини 30-х до кінця 70-х років.
13. Родоначальниками сучасної тестології вчені вважають представників американських і європейських шкіл психології ХІХ і ХХ століть: Біне, Векслера, Гілфорда. Відлік сучасного тестування ведуть, зазвичай, від тестів, розроблених А. Біне і Т. Симоном за замовленням Міністерства освіти Франції для добору в спецшколи

дітей із затримкою інтелектуального розвитку. Останнім часом наголошується підвищене зростання інтересу до тестів і тестування в багатьох країнах з метою підтримати державний рівень загальної освіченості (наприклад, у США, Німеччині, Франції, Австралії та інших країнах).

Ідея проведення спеціальних обстежень, вимірювань з метою оцінювання досягнутих результатів і визначення подальших шляхів розвитку системи освіти, формування моделі її запланованого стану для нашої країни не нова. Ще наприкінці XIX ст. земським діячем бароном М. Корфом були зроблені спроби виміряти у кількох повітах міцність здобутих селянами знань.

Що стосується розвитку тестології в Україні, то це тісно пов'язано з Росією, адже Україна у той час була у складі Російської імперії, а потім СРСР. На початку XX ст. в Росії та Україні було здійснено перші спроби запровадити тести в практику шкільного контролю. Певні дослідження з тестології пізніше проводили й у Радянському Союзі.

Розвиток тестування в СРСР відбувався в три етапи:

Перший етап – 20-ті – середина 30-х років XX ст. Відбувається процес накопичення відомостей з тестових технологій. У 1925 р. при педологічному відділі Інституту методів шкільної роботи створено спеціальну тестову комісію, яка навесні 1926 р. підготувала стандартизовані тести з природознавства, суспільствознавства, арифметики (обчислювальні операції та розв'язування задач), географії (знання карти), на правопис і розуміння прочитаного. В основу цих матеріалів були покладені американські зразки тестів. До них додавали інструкції й особову картку для обліку прогресу учня в навчанні. У СРСР у цей час, як і на Заході, визначальним став прагматичний підхід, що ґрунтувався не на психолого-педагогічній теорії, а на математичній статистиці. У період сталінізму тестологію почали розглядати як ворожу й буржуазну. Після постанови ЦК ВКП (б) 1936 р. «Про педологічні викривлення в системі Наркомпросу» були ліквідовані не тільки інтелектуальні тести, а й тести

успішності. Як наслідок – застій в експериментальній педагогіці та психології.

Другий етап – 30 – 50 роки ХХ ст., коли тести використовували переважно у сфері психологічних досліджень на обмеженому контингенті учнів зазвичай за зарубіжними зразками.

Третій етап – 60 – 90 роки ХХ ст. Тестування як метод вимірювання знань почали використовувати у військових училищах Міністерства оборони, Міністерства внутрішніх справ та інших спеціальних навчальних закладах. Однак спроби відродити тестування в 70-х роках ХХ ст. були невдалими, так само як невдалою була організація тестування успішності випускників загальноосвітніх навчальних закладів, здійснена Міністерством освіти 1993 – 1994 рр. Після здобуття Україною незалежності розпочинається новий етап у розвитку тестології в нашій країні. Хоч наслідки політики колишнього СРСР відчутні й донині, що позначається на відсутності національної школи тестування, Україна робить кроки щодо запровадження тестування в освіті. Починаючи з 90-х років ХХ ст, багато вищих навчальних закладів почали переходити на нові умови прийому до ВНЗ із використанням тестових технологій. У 1993 р. Міністерство охорони здоров'я України за підтримки міжнародних фондів і безпосередньої участі західних спеціалістів запровадило тестування для ліцензування лікарів. Система перевірки й оцінювання професійних знань і майстерності студентів та випускників медичних навчальних закладів «Кроки» є першою в Україні, хоча і вузькопрофесійною, системою тестування загальнодержавного рівня. У рамках цього масштабного проекту створено банк завдань, у який збирають емпіричні дані, проводять їх психометричний аналіз.

У 2002 – 2005 рр. Центр тестових технологій за підтримки Міністерства освіти й науки України та Міжнародного фонду «Відродження» започаткував експеримент щодо впровадження зовнішнього незалежного оцінювання в систему загальної середньої освіти. Поетапно запроваджували тестування з історії України та всесвітньої історії, математики, української мови, географії, біології, фізики, хімії, результати яких протягом 2003

– 2007 рр. випускникам загальноосвітніх навчальних закладів за бажанням могли зарахувати під час вступу до вищих навчальних закладів (2003 р. – 3121 учасник тестування, в експерименті брали участь 4 університети; 2004 р. – 4500 учасників тестування, 31 університет зараховував результати тестування за згодою абітурієнтів; 2005 р. – 8700 випускників ЗНЗ, в експерименті брали участь усі ВНЗ України). Згідно з указом Президента України від 4 липня 2005 р. (указ № 1013/2005 «Про невідкладні заходи щодо забезпечення функціонування та розвитку освіти в Україні») з 2006 р. в Україні впроваджується національна система зовнішнього незалежного оцінювання, в основу якої покладено модель зовнішнього стандартизованого тестування.

У 2006 р. створено Український центр оцінювання якості освіти, який здійснює оцінювання й сертифікацію навчальних досягнень випускників. У 2006 р. Український центр оцінювання якості освіти за підтримки МБО «Центр тестових технологій і моніторингу якості освіти» та міжнародного фонду «Відродження» вперше провів зовнішнє незалежне оцінювання навчальних досягнень на національному рівні, у якому брали участь 41 000 випускників ЗНЗ (загалом 82 000 тестувань з української мови, математики та історії). З 2008 р. вступ до ВНЗ здійснюють тільки за результатами зовнішнього незалежного оцінювання (української мови та літератури, зарубіжної літератури, історії України, всесвітньої історії, математики, біології, фізики, хімії, географії, основ економіки та основ правознавства).

2.2. Роль тестування у становленні інших країн

Тестування – метод діагностики із застосуванням стандартизованих запитань та завдань, що мають певну шкалу значень.

До нього вдаються для стандартизованого визначення індивідуальних відмінностей особистості в усьому світі. Тестові методи використовують у трьох сферах:

- 1) освіті;
- 2) при виборі професії та професійній підготовці;

3) при психологічному консультуванні й психотерапії.

Вони дають змогу з певною ймовірністю визначити рівень розвитку в індивіда психологічних властивостей (пам'яті, мислення, уяви та ін.), особистісних характеристик, ступінь готовності до певної діяльності, засвоєння знань і навичок тощо.

У різних країнах існує ціла система тестування особистості: чи то майбутнього працівника, чи студента, військового чи держслужбовця. Переважно у країнах, де тестологія розвинена як наука (у Німеччині, Англії, Франції, США, Ізраїлі, Австрії, Нідерландах, Швеції, Швейцарії, Великобританії, Австралії та ін.), при вступі до навчального закладу, при влаштуванні на роботу обов'язково потрібно пройти тести двох видів. Зазвичай це національні та особистісні психологічні тести.

Наприклад, для того щоб вступити в німецьку школу, школяр повинен скласти вступні тести та пройти співбесіду.

А іноземний громадянин, який претендує на громадянство в Німеччині, повинен пройти тест на знання німецької мови, скласти тест на знання законів, громадського порядку і життя в Німеччині.

Для тих громадян, які хочуть працювати у державних службах Німеччини, крім того, що психологи вивчають індивідуальні особливості претендента за допомогою тестів, вивчається навіть його релігійна позиція.

У збройних силах США провідним органом здійснення професійно-психологічного контролю є штаб командування з відбору. Щорічно на наукові дослідження військових органів професійно-психологічного добору американський уряд виділяє 200 млн. доларів, із них 7 млн. – на розробку нових психодіагностичних методик. Подібні системи відбору армійців існують у Німеччині, Франції, Великій Британії.

В Індонезії, окрім тестування, вимагається сертифікат поліції про бездоганну поведінку абітурієнта, який хоче вступити до вищого навчального закладу.

В Україні сучасні психологічні тести, що використовуються фахівцями, поділяються на дві основні групи: перша – тести для дослідження здібностей, знань і

умінь, наприклад, тести щодо поглядів, рахунку, пам'яті, мови, просторової уяви, уваги, практичного і технічного досвіду; друга – тести, за допомогою яких досліджують характер, темперамент, емоції, психомоторний і сенсорний типи. Згідно з результатами цих тестів судять про людину. Вони і до цього часу не втратили свого значення при визначенні профілю навчання і майбутньої професії. Так, тести широко використовуються нашими педагогічно-психологічними консультаціями. При виборі деяких професій, наприклад, льотчика, шофера або машиніста, за допомогою тестів дізнаються, чи не послугує причиною небезпеки при виконанні даної роботи недолік, який до цього часу не давав про себе знати і якого ніяким іншим шляхом проявити не можна. Тести також допомагають в медицині, психології, при дослідженні впливу середовища на людину, застосовують їх і в освіті. Але завжди користуються ними виключно фахівці.

2.3. Понятійний апарат тестології. Поняття тесту



Тестологія – наука про вимірювання психофізіологічних та особистісних характеристик, а також обсягу та якості знань, умінь, навичок.

Тестологи вивчають і створюють способи, методи, технології вимірювань психофізіологічних та особистісних характеристик, а також обсягу та якості знань, умінь, навичок. Тестологи створюють тестові комплекси, у яких реалізовано досягнення тестології у вигляді сукупності технологій, рекомендацій, тестів, автоматизованих систем, пристроїв.

Тестові комплекси застосовують для атестації ЗУН учнів та абітурієнтів, для вимірювання обсягу та якості навичок і умінь при прийомі людей на роботу та ін..

Визначення поняття тест та тестування намагалися дати багато вчених, та не всі з них були зрозумілими та відображали дійсний зміст цих понять. Наведемо деякі з визначень.

Так *К. Ингенкамп* у своїй праці “ *Педагогическая диагностика* ” дає таке визначення поняттю тестування:



Тестування – це метод педагогічної діагностики, за допомогою якого вибірка поведінки, що репрезентує

передумови чи результати навчального процесу, повинна максимально відповідати принципам співставлення, об'єктивності, надійності та валідності вимірювань, повинна пройти опрацювання та інтерпретацію та бути готовою до використання в педагогічній практиці.

Російський дослідник *А.Н. Майоров* у роботі “*Теорія і практика створення тестів для системи освіти*” дає таке визначення:



Тест – це інструмент, що складається із кваліметрично вивіреної системи тестових завдань, стандартизованої процедури проведення, завчасно спланованої технології опрацювання і аналізу результатів, призначений для вимірювання властивостей і якостей особистості.

Ключовим поняттям тестології є поняття «педагогічний тест».

Педагогічний тест –



- ✓ така система завдань, результат виконання яких групою претендентів дозволяє досить надійно ранжувати їх (надати їм порядкові номери) за якістю навчання, кількістю наявних знань;
- ✓ система стандартизованих завдань, результат виконання яких дозволяє за заданим ступенем точності виміряти знання, навички та вміння випробуваного.

Іноді під поняттям «тест» розуміють засіб педагогічного вимірювання або стандартизований метод дослідження, призначений для кількісних та якісних оцінок характеристик людини.

Історично склалися два основні теоретичні підходи (теорії) конструювання та використання педагогічних тестів.

- ✓ Перший на основі класичної теорії тестів (***Classical Test Theory – CTT***);
- ✓ другий в рамках теорії латентно-структурного аналізу, сучасна теорія конструювання тестів (***Item Response Theory – IRT***), яку правильніше було б назвати методологією, оскільки це теорія розробки і

застосування методів дослідження емпіричних даних (теорія методів).

Очевидно виникає проблема вибору, проблема припущення того чи іншого підходу до розробки і застосування тестів. Теорія LSA або правильніше основана на ній IRT (сучасна теорія тестування) більш ефективна, але складніша у використанні. Необхідно застосовувати потужніший математичний апарат, розробляти дорожче і більш складніше програмне забезпечення. Використання IRT потребує високої кваліфікації як від розробників, так і від користувачів тестів.

Таблиця 4

Аналіз основних теорій конструювання тестів

<u>Теорія СТТ</u>	<u>Теорія IRT</u>
Переваги	
<ul style="list-style-type: none"> ✓ Базується на нескладних припущеннях. 	<ul style="list-style-type: none"> ✓ Інтегрованість, міждисциплінарність. ✓ Інваріантність завдань. ✓ Передбачає конкретизацію функції та моделі завдання. ✓ Характеристики завдань не залежать від досліджуваної групи, за допомогою якої вони були отримані. ✓ Враховується ступінь значущості відповіді на кожне запитання для отримання кінцевого балу.
Недоліки	
<ul style="list-style-type: none"> ✓ Показники завдання не є постійними серед учасників тесту. ✓ Оцінювання здібностей (бали тесту) залежать від специфіки конкретного тесту. ✓ Надійність тесту визначається паралельними варіантами. ✓ Чіткість оцінювання має бути однаковою для всіх рівнів здібностей. ✓ Теорія орієнтована на тест, а не на завдання. 	<ul style="list-style-type: none"> ✓ Використання складного математично-статистичного апарату для опрацювання результатів. ✓ Розробка спеціальних програмних продуктів та застосування комп'ютерної техніки.

У ІРТ не ставляться і не вирішуються фундаментальні проблеми валідності і надійності тесту: тест там заздалегідь вважається надійним і валідним. Обчислення ІРТ зводяться до отримання оцінок параметрів складності завдання і до вимірювання рівня знань тих, що беруть участь у випробуванні.

До досягнень ІРТ відносять використання однієї шкали в вимірах значень параметрів випробуваних та завдань тесту. Це дозволяє співвідносити рівні ЗУН випробуваних з мірою складності тестових завдань. На цій можливості спільних оцінок параметрів випробуваних та завдань заснований застосовуваний більшістю тестологів математичний апарат.

В Україні поки що не склалося уявлення про педагогічну тестологію як виділену науку або наукову область:

- Немає посади тестолога в реєстрах відділів кадрів;
- Немає вакансій для академіків-тестологів в Академії Наук та ін.;
- Педагоги не сприймають тестології, так як у ній занадто багато математики;
- Програмісти вважають, що тестологія – пройдений етап, і там уже нема чого робити творчій особистості.



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

- 1) відповідність меті;
- 2) лаконічність;
- 3) логічна форма висловлювання;
- 4) наявність місця для відповіді;
- 5) однакові правила оцінювання відповідей;
- 6) правильне розміщення елементів завдання;
- 7) наявність однакової інструкції для всіх тестованих;
- 8) адекватність інструкції, формі та змісту завдання.



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

Вимоги до тестових завдань:

- а) певна складність, що відповідає меті й рівню оцінювання;

- б) диференційна здатність (достатня варіативність тестових балів);
- в) позитивна кореляція балів завдань із балами всього тесту;
- г) відповідає вимогам чистоти форми й предметної чистоти змісту.



Завдання тестові паралельні – тестові завдання однакового рівня складності та диференційної здатності, що належать до того самого розділу чи теми.

Комплект тестовий – зошит, у якому містяться всі завдання, призначені для одного учасника тестування, і бланк (бланки) відповідей. За наявності тестового комплекту тестовий зошит можна використовувати як чернетку, його не перевіряють і не оцінюють. Якщо тест складається із завдань на вибір відповіді й завдань відкритого типу (визначення відповіді самостійно), тестованому пропонують два бланки відповідей (№ 1 і № 2). У бланку № 1 тестований певними символами позначає правильні значення відповіді, його перевіряють за допомогою спеціальної комп'ютерної програми, у бланку № 2 тестований виконує завдання відкритого типу, які перевіряють експерти.

2.4. Класифікація педагогічних тестів

Застосування тестування як методу вимірювання призвело 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, В. Аванесов, А. Анастасі, І. Булах, В. Беспалько, О. Киверялг, М. Розенберг та ін.

Найпоширенішою є узагальнена класифікація, 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).



Рис. 2.1. Класифікація педагогічних тестів

Структура тесту залежить від мети тестової перевірки знань. Поки що у суспільстві домінує хибна думка, що тести дають змогу перевірити лише репродуктивні знання. Це пов'язано з поверхневим ознайомленням із тестовими технологіями. Насправді правильно розроблені тестові

завдання перевіряють всю пізнавальну сферу, яка містить (відповідно до складності): знання, розуміння, застосування, аналіз, синтез, оцінювання.

Наведемо приклад іншої класифікації педагогічних тестів, яка ґрунтується на меті тестування та залежить від етапу навчання, на якому воно проводиться.



Рис. 2.2. Класифікація педагогічних тестів залежно від мети використання

За функціональною ознакою (призначення тестів щодо предмета дослідження) розрізняють:

- ✓ тести інтелекту – методики психологічної діагностики, призначені для виявлення розумового потенціалу індивіда;
- ✓ тести креативності – сукупність методик для вивчення та оцінювання творчих здібностей особистості (креативності);

- ✓ тести досягнень – методики, за допомогою яких визначають ступінь конкретних знань, навичок, умінь особистості;
- ✓ тести особистісні – методики психодіагностики, за допомогою яких визначають різні якості особистості та її характеристики: уподобання, цінності, ставлення, емоційні, мотиваційні та міжособистісні властивості, типові форми поведінки;
- ✓ проектні тести – методики, скеровані на виявлення певних психічних властивостей людини. Вони передбачають стимули, реагуючи на які, людина виявляє найхарактерніші свої якості. Для цього досліджуваній особі пропонують витлумачити події, відновити ціле за деталями, надати сенс неоформленому матеріалу, створити оповідання за малюнком із невизначеним змістом.

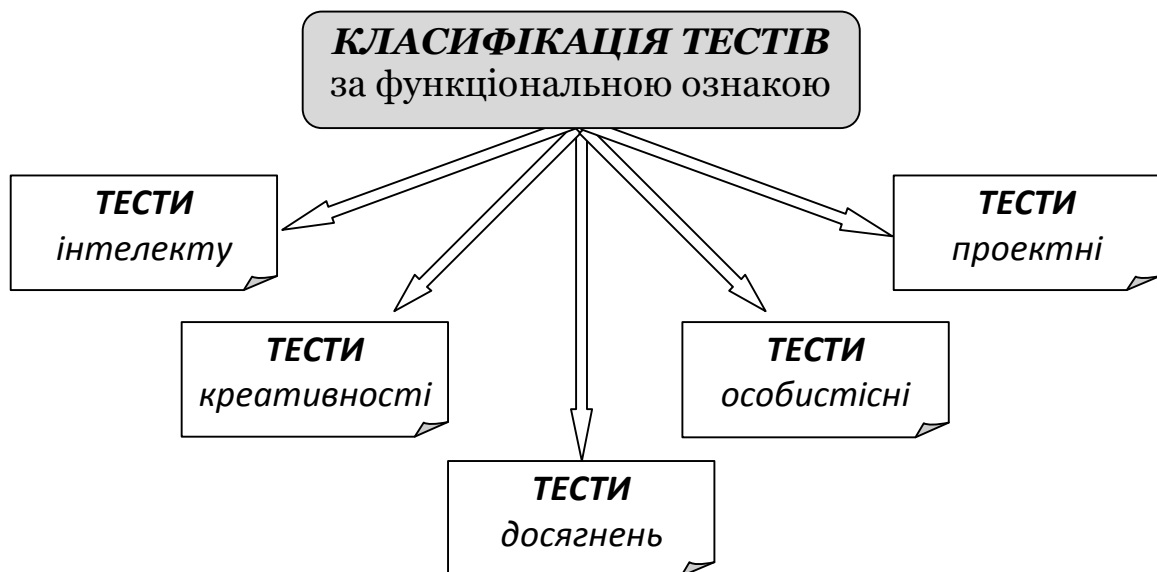


Рис. 2.3. Класифікація тестів за функціональною ознакою

У міжнародних моніторингових дослідженнях (напр., ТІМ88) найчастіше користуються трьома когнітивними рівнями: знання, застосування, обґрунтування.

Завдання для тесту добирають залежно від того, що хоче перевірити учитель, який рівень дослідити.

Рівень «**знання**» перевіряє базові знання учнів, а саме знання фактів, понять, процесів тощо. Це найпростіший рівень

засвоєння предмета. Але здійснювати складні види пізнавальної роботи учні зможуть лише оволодівши глибокими знаннями та відповідними вміннями. Під час оцінювання цього рівня від учнів очікується: уміння пригадати основні закономірності; володіння термінологією, фактами, поняттями; вміння узагальнювати, обирати відповідний науковий апарат, обладнання, пристрої для вимірювання й експериментальної діяльності тощо.

Завдання для оцінювання рівня «*застосування*» добирають таким чином, щоб учні виявили вміння застосовувати знання і розуміння понять у простих ситуаціях. Ці завдання дають змогу перевірити вміння учнів порівнювати, протиставляти й класифікувати, тлумачити наукову інформацію, а також використовувати своє розуміння наукових понять і принципів для знаходження рішення.

Рівень «*обґрунтування*» не обмежується перевіркою вмінь, потрібних для розв'язання стандартних проблем. Завдання цього рівня дають змогу перевірити: уміння, які потрібні при зіткненні з незнайомими ситуаціями, складними задачами та багаторівневими проблемами; здатність аналізувати проблему, добирати принципи, які слід використати при вирішенні завдання; уміння вибрати та застосувати відповідні рівняння, формули; спроможність зробити висновки на базі наукових даних і фактів; розуміння причини й наслідку дослідження; здатність оцінювати й приймати відповідні рішення тощо.

Завдання і час, відведений на їх розв'язання, повинні відповідати віковим особливостям учнів. При доборі варіантів завдань учителеві слід робити їх рівноцінними. Завдання повинні формулюватися чітко, без зайвої інформації та деталізації. Бажано, щоб запитання починалося зі слів: «Укажіть...», «Зазначте...», «Оберіть...» тощо.

До варіантів відповідей слід включати *дистрактори* (неправильні варіанти), які є зрозумілими для учнів, наближені до правильних і становлять собою найпоширеніші помилки при відповідях на подібні завдання. Такий підхід унеможливорює просте відгадування учнями правильної відповіді та дає змогу глибше перевірити рівень навчальних досягнень. Щодо

кількості варіантів відповідей, учням 10-11 класів варто пропонувати 4-5 варіантів відповідей, учням 7-9 класів - 4 варіанти.

При визначенні рівня складності завдань можливо використовувати поширений у світовій практиці принцип – повністю правильно розв'язати весь набір тестових завдань повинно близько 5 відсотків учнів класу.

У міжнародній практиці на завдання закритого типу з вибором однієї правильної відповіді з декількох наведених відводиться близько 1 хв. На розв'язання завдань на встановлення відповідностей і визначення послідовності можна дати близько 1,5-2 хв. Загалом, учням 7-9 класів доцільно формувати тест з 25-30 завдань на 35-40 хв. уроку, учням 10-11 класів - з 35-40 завдань.

За співвідношенням із нормами або критеріями тести поділяють на орієнтовані на норму (нормативно зорієнтовані) та орієнтовані на критерії (критеріально-зорієнтовані).



Рис. 2.4. Класифікація тестів за співвідношенням із нормами або критеріями

Історично склалися два підходи щодо добору атестаційних тестів і до інтерпретації результатів їх виконання.



Тест нормативно зорієнтований – інструмент оцінювання, що зорієнтований на статистичні норми, які встановлені для певної сукупності тестованих.

Навчальні досягнення окремого учня (студента) інтерпретуються залежно від досягнень усієї сукупності тестованих вище або нижче від середнього показника – норми. Тестову норму визначають за результатами апробації тесту на репрезентативній вибірці або після проведення тестування,

якщо репрезентативну вибірку визначити до початку тестування неможливо. Тестованих групують за рангами, тому незалежно від використаної шкали не можна зробити висновок про рівень знань і вмінь тестованих або про досягнення ними конкретних цілей навчання. Такий підхід не співвідноситься зі змістом процесу навчання, а якщо оцінювання здійснює вчитель (викладач), його оцінки зазвичай суб'єктивні, тому що він виставляє їх відносно середнього рівня підготовки класу / групи. Цей тест доцільно використовувати для зарахування результатів до вищих навчальних закладів, нагородження успішніших у навчанні тощо. У нормативно зорієнтованому тесті відсоток правильних виконаних завдань має становити не менш як 50%.

В другому підході результати виконання інтерпретуються відносно деякої визначеної області змісту, яка визначена як 100 % запропонованого для засвоєння обсягу матеріалу.



Тест критеріально-зорієнтований – інструмент оцінювання, призначений для визначення рівня індивідуальних досягнень відносно певного критерію, що ґрунтується на логіко-функціональному аналізі змісту завдань.

Він має на меті оцінити, чи досягнув тестований установленого рівня засвоєння матеріалу навчальної програми або її частини. Результати тестування в такому разі порівнюють із певним критерієм рівня підготовленості тестованого (критичним рівнем обсягу знань, умінь і навичок), який визначають до початку тестування. Результат вказуватиме, чи відповідає рівень досягнень тестованого вимогам стандарту або іншим критеріям. Існує два способи інтерпретації результатів: у першому випадку робиться висновок, засвоєний чи не засвоєний навчальний матеріал (досягнув тестований стандарту чи ні), у другому – подається процент засвоєння матеріалу, що перевіряється (на якому рівні засвоєно стандарт або який відсоток із усіх вимог стандарту засвоєно). Отже, критеріально-зорієнтований тест, на відміну від нормативно зорієнтованого, оцінює, що вміє робити тестований відносно інших. У критеріально-зорієнтованому тесті відсоток виконання завдань

має коливатися в межах 80 – 90 %. В останнє десятиріччя надають перевагу тестам, зорієнтованим на певні критерії. Перша інформація про критеріально-зорієнтовані тести з'явилася на Заході в 60-х роках ХХ століття.

Здавалось б, бажано поєднати дві інтерпретації, оскільки перша та друга представляє інтерес в рамках атестаційного тестування учнів і абітурієнтів. Однак, ідея поєднання породжує додаткові труднощі, причому труднощі не пов'язані з опрацюванням, а мають методичний характер.

Нормативно орієнтовані і критеріально–зорієнтовані тести повинні відповідати різним, часто протилежним один одному критеріям якості. Також існують різні підходи в доборі змісту і до формування тесту.

Може бути інший, продуктивніший шлях, який припускає інше рішення, коли частина тесту, орієнтована на вимоги стандартів будується в логіці критеріально-зорієнтованого підходу і дозволяє диференціювати за знаннями найбільш підготовлених учнів і абітурієнтів.

Досить привабливо приймати в престижні ВНЗ за результатами критеріально-зорієнтованих тестів. У цьому випадку перевага при вступі надається не тому, що один кращий за іншого, а тому, що випробуваний володіє достатнім обсягом матеріалу, який необхідний для навчання в даному виші.

Методично – обґрунтований, системний підхід до розробки атестаційних тестів припускає також вирішення питання щодо вибору деякого критерію володіння стандартом, вибору визначених пропозицій змісту стандартів і варіативної частини в тесті. Включення в атестаційний тест виключно змісту стандартів, що означає орієнтацію на необхідний мінімум, призведе до зниження загального рівня підготовленості.

До числа концептуальних проблем відноситься вибір стратегії атестаційного тестування. Можливо, що атестаційне тестування слід зробити дворівневим.

Перший рівень – це критеріально-зорієнтоване тестування, метою якого є добір тих, хто задовольняє мінімальним вимогам, а другий рівень – нормативно зорієнтований, який дає можливість здійснити диференціацію випробуваних.

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: "зараховано – не зараховано", "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 із попередньо встановленими критеріями.

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

За рівнем впровадження і статусом використання тести поділяються на загальнодержавні (національні), відомчі, або міністерські, тести навчального закладу, кафедральні, особисті або неформальні.

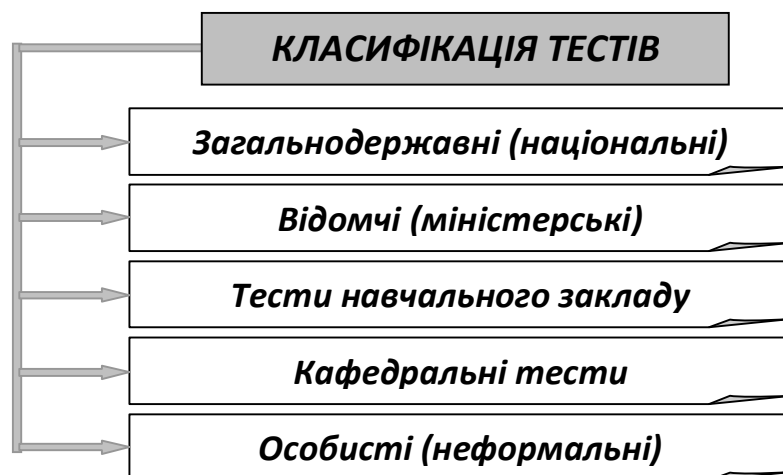


Рис. 2.5. Класифікація тестів за рівнем впровадження та статусом використання

При цьому залежно від статусу використання тести можуть бути обов'язкові, пілотні й дослідницькі. Ця класифікація тестів визначається різними вимогами до цих типів тестування, 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.

Пілотне тестування 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.

Стандартизовані тести

Стандартизація (з *англ.* standart – 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.

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 опрацювання 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 в шкалу оцінок.

Застосуємо термін "стандартизація" у першому розумінні до тесту.



Рис. 2.6. Класифікація тестів за рівнем уніфікації

!
Стандартизованим називається тест, що має комплексну характеристику, яка визначається його властивостями, процедурами вимірювання і 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 відображає. 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.

!
Тест стандартизований – інструмент оцінювання, що вимірює стандартний набір широко визначених результатів навчання, використовує стандартні інструкції та стандартну методику оцінювання, а також дає змогу порівнювати результат окремого тестованого з результатами інших, які виконували той самий тест за аналогічних умов.

Ці тести прив'язані переважно до норм, що призначені для порівняння рівня успішності тестованих певного навчального закладу з рівнем успішності репрезентативної вибірки групи

тестованих (наприклад, на національному чи регіональному рівнях). Раніше у світовій практиці такі тести ґрунтувалися на завданнях на вибір відповіді, передусім на завданнях з вибором однієї відповіді. Останнім часом тести модифікують, включаючи завдання на надання відповіді, а також групують у кластери для подальшої інтерпретації.

Якісно розроблені стандартизовані тести для перевірки успішності, як правило, мають такі характеристики:

1. Зміст тесту ґрунтується на матеріалах широко загальновизнаних підручників і методичних посібників.
2. Тестові завдання ґрунтуються на чіткому наборі специфікацій.
3. Тестові завдання апробовано, переглянуто, проаналізовано на предмет складності й дискримінаційної здатності, завдання з незадовільними статистичними характеристиками перероблено або вилучено з тесту.
4. Остаточний набір завдань дібрано на основі специфікації тесту.
5. Інструкції щодо проведення тесту й виставлення оцінок (балів навчальних досягнень) чітко описані.
6. Тест проводять із вибраною групою тестованих, щоб установити національні та / або регіональні норми для інтерпретації результатів тестів.
7. Остаточний варіант тесту публікують разом із методичним посібником, який описує статистичні характеристики тесту, методику його проведення й оцінювання, а також інтерпретацію та використання результатів. Найвідомішими у світі є стандартизовані тести SAT, ORE, GMAT, LSAT, MCAT, TOEFL тощо.

Гомогенні та гетерогенні тести



Рис. 2.7. Класифікація тестів за тематикою завдань

! *Тест гомогенний* – сукупність стандартизованих завдань з однієї навчальної дисципліни (розділу дисципліни).

Система завдань зростаючої складності для об'єктивного та ефективного оцінювання рівня підготовленості учнів до однієї із навчальних дисциплін. Гомогенні тести поширені більше інших. Вони створюються для контролю знань з однієї навчальної дисципліни або з одного розділу об'ємної навчальної дисципліни.

! *Тест гетерогенний* – сукупність стандартизованих завдань, що створені з метою вимірювання знань з кількох навчальних дисциплін (розділів дисциплін), але не вимагає одночасного залучення знань з різних дисциплін (розділів дисциплін).

Гетерогенні тести використовуються для комплексного оцінювання випускника школи, для добору абітурієнтів при прийомі у ВНЗ. Гетерогенний тест складається з гомогенних. Опрацювання результату ведеться для кожного гомогенного тесту окремо з наступною інтеграцією в єдиний рейтинг.

! *Тест інтегративний* – система завдань для загальної діагностики готовності випускника освітнього закладу. Завдання тесту такі, що для їх виконання потрібні синтетичні знання з кількох навчальних дисциплін.

Класифікація тестів за процедурою формування черги подання



Рис. 2.8. Класифікація тестів за процедурою формування черги подання

❗ **Тест адаптивний** – система завдань з відомими властивостями і параметрами складності.

Завдання подаються за допомогою комп'ютера, в порядку, що залежить від відповіді випробуваного на попереднє завдання. При правильному виконанні попереднього завдання – ступінь складності наступного зростає і навпаки. За допомогою різних стратегій подання завдань вдається значно скоротити час комп'ютерних тестувань без втрати точності вимірювань.

❗ **Тест змішаний** – сукупність стандартизованих завдань різної складності, що подаються у випадковому порядку.

❗ **Тест зростаючої складності** – сукупність стандартизованих завдань, що подаються в порядку зростаючої складності.

Залежно від кількості осіб, які виконують тест, тести поділяють на індивідуальні та групові.

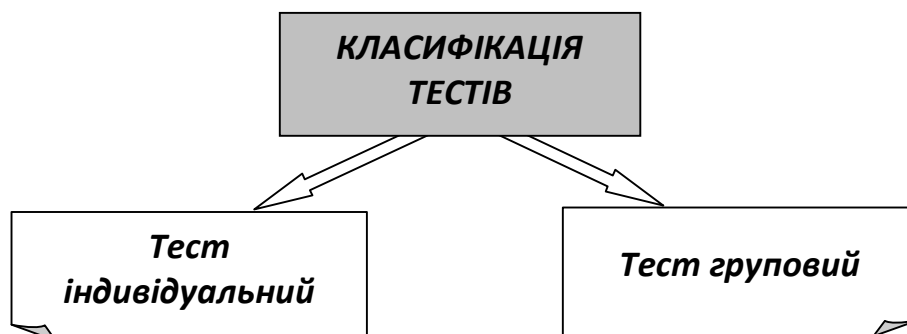


Рис. 2.9. Класифікація тестів залежно від кількості тестованих



Тест індивідуальний – тест, створений для окремого випробуваного або окремого типу випробуваних.



Тест груповий – тест, який група випробуваних виконує разом (колективно, спільно).

2.5. Адаптивне комп'ютерне тестування

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



Рис. 2.10. Класифікація тестів за методом тестування



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. Часто бланки перевіряють за допомогою спеціального графарету. Особливість бланкового тестування – його можна проводити тільки один раз: як тільки одна група пройшла тестування, цей варіант тесту не можна використати вдруге, тому що отримані результати не даватимуть достовірних відомостей про знання тестованого.

Комп'ютерне тестування, яке в зарубіжній літературі отримало назву computer-based testing (CBT), 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 порівняно з бланковим. Основна перевага – отримання результату тестування одразу після його закінчення. Основним недоліком є необхідність мати мінімальні навички роботи з комп'ютером та в інтерфейсному середовищі тестової програми.

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** (CAT), є тестуванням 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. Тому різні екзаменовані, що тестуються за системою CAT, найімовірніше, відповідатимуть на 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 ж завдання, які 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 запитання зі складного набору. Менш компетентні дадуть правильні відповіді на половину запитань зі свого «легкого» набору. Об'єднуюча основа, яка утримує всі ці тести разом, – особливий тип психометричних теорій, так звана Теорія Відповідей на Тестові Завдання (ІКТ). 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 полягає у доборі завдань, які є найбільш інформативними і водночас відповідають специфіці змісту. 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.

Переваги:

1. Дає можливість більш гнучко виміряти знання тестованих.
2. Дає можливість виміряти знання меншою кількістю завдань, ніж у класичній моделі.
3. Виявляє теми, які тестований знає погано, і дає можливість перевірити знання з цієї теми додатково.

Недоліки:

1. Заздалегідь невідомо, скільки запитань треба поставити тестованому, щоб визначити рівень його знань.

2. Можна застосувати модель тільки за наявності комп'ютера. Надійність результатів такого тестування найвища, тому що програма тестування підлаштовується під рівень знань тестованого.

2.6. Якість тесту

Якість освіти – ряд системно-соціальних властивостей і характеристик, які визначають відповідність (адекватність) системи освіти прийнятим вимогам, соціальним нормам, державним освітнім стандартам.

Якість тесту це окрема категорія, що базується на певних теоріях і закономірностях. Вона, безумовно залежить від того, на який тип вимірювання ми орієнтувались 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).

§ 3. Зміст тесту. Принципи добору змісту

3.1. Проблеми конструювання тестових завдань

Багато хто вважає, що тестові завдання повинні створюватися у формі тверджень, ввівши у вимоги до тестових завдань логічну форму висловлювання. Це пов'язано, по-перше, зі словесним складом завдання у тестовій формі: зміст тестового твердження засвоюється завжди краще, ніж зміст запитання, оскільки, в тестових твердженнях немає жодного зайвого слова і навіть знаку, в той час як запитання потребує ряду додаткових слів і знаків для вираження потрібного змісту, значення та інтонації.

Проте добре сформульоване завдання в запитальній формі нічим не поступається добре сформульованому запитанню у формі твердження. Але якщо у формі запитання спробувати поставити два заперечення, то таке завдання стає абсолютно незрозумілим. Мало того, в рекомендаціях по складанню тестових завдань голландського інституту СІТО дається така рекомендація: «Використовуйте прямі запитання. Переважно застосовуйте прямі запитання, що утворюють повне речення зі знаком запитання в кінці». Крім того, аналіз завдань, що використовуються при атестації учнів, показує, що переважною формою завдань (понад 90 %) є пряме запитання. Вибір способу формулювання тестового завдання у формі твердження або у формі запитання, повинен визначатися двома обставинами: по-перше, необхідно вибирати ту форму, яка буде максимально зрозумілою для випробуваних, і, по-друге, одна з форм завдань вимагає неодмінно формулювання запитання у формі твердження – це завдання альтернативних відповідей, оскільки вони вимагають відповіді так чи ні, які краще підходять до тверджень.

Один із приводів для критики тестування полягає у тому, що для більшості завдань існує ймовірність випадкового вибору правильних відповідей. У загальному випадку це проблема лише завдань закритого типу, і чим менше можливих альтернатив пропонується випробуваному в рамках одного завдання, тим більша ймовірність добору. Ця проблема гостро

стоїть для завдань альтернативних відповідей, де ймовірність добору максимальна і становить 50 %, тому не рекомендується використовувати такі завдання для тестів навчальних досягнень. Однак серія з п'яти завдань альтернативних відповідей дає ймовірність добору 0,0313, а з десяти завдань всього лише 0,00098.

Проблема добору пов'язана ще й з питанням часу, який виділяється на тестування і, деякою мірою, з питанням мотивації. Чим вище мотивація, тим більша ймовірність того, що випробуваний спробує дати відповідь випадково на ті запитання, на які він не встигає відповісти, тобто намагатиметься вгадати. У свою чергу, кількість питань, з якими можливим є добір, залежить від часу, який виділяється на тестування. Чим менше часу відводиться на тестування, тим більше запитань, на які поширюється випадковий добір. Таким чином, для часу, який виділяється на тестування, повинен бути знайдений баланс: при недостатній тривалості тестування зростає ймовірність добору, при великій – зростає ймовірність підказок та інших порушень дисципліни.

3.2. Зміст тестів для поточного, проміжного та результуючого контролю

Вид тестового контролю визначається цілями тестування – отриманням відомостей про рівень знань студентів на певному етапі навчання. Відповідно з цілями визначаються єдині хронотопні (змістовно-часові) точки тестування в навчальному процесі.

Тестування можна поділити на:

- поточний та оперативний тест-контроль знань студентів;
- проміжний тест-контроль знань студентів;
- рубіжний тест-контроль знань студентів ;
- підсумковий (результуючий) тест-контроль знань студентів;
- тренінгові тестування.

Поточне та оперативне тестування (діагностичне) проводиться викладачем на семінарах для виявлення вихідного

рівня готовності студента до подальшого навчання, перевірки якості засвоєння знань за певними темами семінарів, розділів програми дисципліни.

Проміжний контроль проводиться на підсумкових заняттях шляхом перевірки письмових відповідей на теоретичні запитання та розв'язання ситуаційних задач з інтерпретацією результатів дослідження.

Рубіжне тестування проводиться наприкінці модуля, семестру, навчального року, для рубіжного контролю: з дисципліни, за модулями дисциплін, з дисциплін спеціалізації або сукупності дисциплін. Результати тестування використовуються викладачем для формування оцінки з дисципліни. Метою рубіжного тестування є визначення ступеня засвоєння студентами знань і вмінь (рівня компетентності) за професійною освітньою програмою.

Підсумкове тестування проводиться з метою контролю залишкових знань студентів; перевірки рівня готовності студента до атестаційних випробувань. Тестування проводиться згідно із затвердженим графіком, але не пізніше 6 місяців з моменту здачі заліку / іспиту з тестованої дисципліни. Результати підсумкового тестування документуються і є необхідною умовою допуску до державних атестаційних випробувань.

Тренінгові тестування є різновидом самостійної роботи студентів, використовуються викладачем як навчальна технологія для самостійного відпрацювання студентами теми, типу завдань і т.д. Методичне забезпечення самостійної роботи студентів передбачає три блоки матеріалів:

- власне методичні матеріали
- рекомендації, вказівки, інструкції до методичних матеріалів;
- оціночні засоби контролю навчання (контролю залишкових знань).

§ 4. Етапи конструювання тесту. Специфікація тесту

4.1. Етапи створення тесту

Останнім часом в Україні поширюється розуміння того, що для розроблення науково-обґрунтованих педагогічних тестів блискучого знання змісту предмета недостатньо. Необхідна методика, яка включає в себе сукупність методів аналізу змісту і форм завдань, а також науковий підхід для процесу накопичення і опрацювання відомостей, потрібні методи розрахунку параметрів завдань, методи експертизи якості тестів і тестових завдань. Тобто потрібна теорія тестових методів, що у філософській літературі називається методологією.

Взагалі кажучи, питання доводиться ставити ширше, оскільки, якщо для конструювання тестів достатньо знання теорії тестів, то для наукової організації усього процесу в цілому доводиться дбати про методологію, яка охоплює окрім розробки тестових методів і застосування тестів для різних цілей, інтерпретацію тестових результатів, розробку принципів наукової організації діяльності тестових центрів і ціле коло інших питань, які потребують негайного вирішення.

Цих питань досить багато, тому мова повинна йти не про методологічні основи, а про методологічні проблеми, що потребують свого невідкладного розв'язання вже сьогодні. Хоча ці проблеми носять виключно науковий характер, але вони не надумані, а продиктовані самою практикою. В основному вони викликані намаганням створити в Україні професійно організовані тестові структури в сфері практики без випереджувального розвитку методології та теорії, без створення певного кадрового потенціалу, без належного розвитку культурного і наукового рівня в такій досить непростій галузі освітнього вимірювання.

Перед тим як перейти до самих проблем хотілося б нагадати про необхідність їх розв'язання. Оскільки часто вважають, що достатньо назвати набір завдань тестами і вони ними стануть. При цьому якимось забувається про те, що без

належної наукової основи тест може бути причиною ілюзійного враження про знання чи незнання учня або абітурієнта. В цьому тест може стати значно гіршим від екзамену, де є можливість скоригувати своє враження від відповіді шляхом додаткових запитань.

Зовнішнє Незалежне Оцінювання стає найважливішою складовою контрольно-оцінювальної системи, що виконує роль головної функції моніторингу якості освіти. При цьому можна виділити системоутворюючі чинники, які активно впливають на освітній процес:

- *проектувальний*, що орієнтує на науково-обґрунтоване формулювання цілей діяльності системи, її функцій, конкретизацію очікуваних результатів і термінів досягнення мети, досяжності поставлених цілей, визначення величини витратних матеріалів і часу;
- *комунікативний*, який включає усі ієрархічні рівні управління освітньою системою, стиль суб'єктно-об'єктних і суб'єктно-суб'єктних взаємин;
- *змістовно-організаційний*, який стандартизує зміст, організацію, технології та процедури тестування, що задає єдність вимог щодо якості підготовки;
- *аналітико-результативний*, що включає способи отримання, накопичення, узагальнення та аналізу результатів, методи педагогічної інтерпретації, техніку і технологію оформлення та представлення результатів, підведення підсумків тестування та навчання.

Як і будь-яка система, незалежне тестування має свою організаційну структуру, діяльністю якої керує і координує центральний орган. Останній відіграє домінуючу роль у функціонуванні усієї системи і має свої підсистеми, в тому числі на етапі проведення ЗНО. Взаємозв'язку цілей, контрольно-оцінювальних засобів і суб'єктів тестування схематично наведено на рисунку 4.1.

До перелічених вище чинників необхідно додати теоретико-концептуальний. Діяльність системи масового тестування багато в чому забезпечується рівнем розвитку усіх його компонентів, точністю та науковою обґрунтованістю концептуального апарату. У процесі розвитку системи відбувається безперервне зміцнення науково-теоретичної та матеріально-технічної бази, розробляється теорія педагогічних вимірювань та моніторингу, вдосконалюються педагогічні контрольні-вимірювальні матеріали, техніка та технологія тестування, підготовка учнів, методи опрацювання статистичних даних, форми і способи їх подання для користувачів, в контрольній-оцінювальній процес включається все більша кількість загальноосвітніх установ країни.



Рис. 4.1. Елементи стандартизованого тестування

Для тестів навчальних досягнень існують загальні правила та етапи розробки. Залежно від тієї мети, яка стоїть перед розробниками, процедура розробки тестів має деякі

особливості, однак перелік етапів створення тестового інструментарію однаковий для усіх видів тестів.

Етапи конструювання тестів

1. Визначення мети тестування.
2. Визначення ресурсних можливостей розробників.
3. Добір змісту навчального матеріалу.
4. Конструювання технологічної матриці та її експертиза.
5. Вибір форматів та створення тестових завдань, їх експертиза.
6. Побудова вибірки для апробації завдань і тестів.
7. Компонування завдань для апробації.
8. Апробація тестових завдань.
9. Визначення та розрахунок показників якості тестових завдань.
10. Вибраковка завдань і формування тесту.
11. Апробація тесту.
12. Визначення та розрахунок показників якості тесту.
13. Побудова кінцевого варіанту тесту.
14. Стандартизація тесту.
15. Нормування тесту.
16. Забезпечення тесту.

Цей перелік повний, оскільки в ньому представлені усі етапи конструювання тестів. Створення тестів починається із визначення мети тестування. Звичайно, для тестів, які ви маєте намір використовувати для порівняння результатів між кількома класами в одному освітньому закладі та тестами, призначеними для підсумкової атестації учнів, існує значна відмінність.

Етапи розробки тестів для різноманітних цілей використання мають бути різними.

З точки зору вимірювання тестові завдання можна поділити на дві загальні категорії: ті, в яких екзаменовані *вибирають* правильні відповіді, базуючись на відомостях, поданих у тестовому завданні, та ті, в яких вони мають *надати* (створити, згенерувати) відповіді самостійно. Когнітивні здібності, що потрібні для відповіді на завдання з двох зазначених категорій, є різними незалежно від змісту завдання.

Завдання першої категорії називаються запитаннями багатовибіркового типу (MCQ – multiple-choice question). До другої категорії належать завдання з короткою або довгою відповіддю, есе.

Розглянемо формати тестових завдань багатовибіркового типу, які протягом більш ніж сторічної практики застосування методу тестування набули найширшого 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 правильним, а решта дистракторів є неправильними.

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

Таблиця 5

Етапи робіт по створенню критеріально та нормативно орієнтованих тестів

Назва етапів / робіт	Нормативно орієнтовані	Критеріально - зорієнтовані
<i>1. Визначення мети тестування</i>	Оцінювання учня відносно інших досліджуваних.	Описання засвоєного чи незасвоєного навчального матеріалу. Оцінювання учня відносно набору характеристик достатніх для досягнення результату.
<i>2. Визначення ресурсних можливостей розробників (різниця в необхідності залучення різних фахівців)</i>	Орієнтація на статистичні методи оцінювання. Фахівці необхідні для статистичного опрацювання результатів.	Орієнтація на експертне оцінювання, проведення робіт з визначення переліку та обсягу навчальних завдань, критерію досягнення необхідного рівня підготовки. Спеціалісти для експертизи навчального матеріалу.

<i>Назва етапів / робіт</i>	<i>Нормативно орієнтовані</i>	<i>Критеріально - орієнтовані</i>
<i>3. Добір змісту навчального матеріалу</i>	Може бути дібраний широкий спектр матеріалу.	Добирається матеріал, який стосується обмеженої тематики.
<i>4. Конструювання технологічної матриці та її експертиза</i>	Можуть бути використані складні матриці.	Складність матриці обмежується вибраним для оцінювання критерієм.
<i>5. Створення тестових завдань та їх експертиза</i>	Вимоги до змісту та якості завдань істотно не відрізняються. Визначаються загальними вимогами до тестових завдань.	
<i>6. Побудова вибірки для апробації завдань і тестів</i>	Більш жорсткі вимоги до обсягу та балансу вибірки, що визначаються вимогами розрахунку статистичних норм.	Більш м'які вимоги до обсягу та балансу вибірки, що визначаються вимогами відпрацювання якості тестових завдань.
<i>7. Компонування завдань для апробації</i>	Вимоги до конструювання завдань для апробації істотно не відрізняються. Визначаються способом проведення тестування.	
<i>8. Апробація тестових завдань</i>	В том числі для визначення складності та дискримінантності.	Для відпрацювання змісту завдань.
<i>9. Визначення та розрахунок показників якості тестових завдань</i>	Істотно не відрізняються	
<i>10. Вибраковка завдань та формування тесту</i>	Відбираються завдання з максимальною дискримінантністю. Занадто прості і занадто складні завдання відкидаються	Відбираються завдання, які максимально покривають досліджувану область критерію. Складні та прості завдання включаються в тест

<i>Назва етапів / робіт</i>	<i>Нормативно орієнтовані</i>	<i>Критеріально - орієнтовані</i>
<i>11. Апробація тесту</i>	Істотно не відрізняються	
<i>12. Визначення та розрахунок показників якості тесту</i>	Істотно не відрізняються	
<i>13. Побудова кінцевого варіанту тесту</i>	Відмінності визначаються особливостями розроблених матриць	цілями та технологічних
<i>14. Стандартизація тесту</i>	Визначається рівнем використання тесту	
<i>15. Нормування тесту</i>	Статистичні норми, що розраховуються для нормативної вибірки, чи визначення місця досліджуваного відносно інших, що приймали участь в тестуванні	Описання критерію досягнення результату чи описання тих навчальних задач, з якими учень повинен справитися
<i>16. Забезпечення тесту</i>	Визначається рівнем використання тесту	

Тест, створений для перевірки знань учнів після вивченої теми, і тест для підсумкової перевірки повинні відрізнятися.

4.2. Стандартизація тесту

Стандартизація тесту – сукупність експериментальних, методичних і статистичних процедур, що забезпечують створення суворо фіксованих компонентів тесту (інструкції, набору завдань, методу опрацювання протоколів і підрахунку балів, способу інтерпретації). В окремому випадку під стандартизацією розуміється збирання репрезентативних тестових норм і побудову стандартної шкали тестових балів.

Що, безсумнівно, повинен знати і вміти робити кожен освічений користувач тесту – це розуміти, що таке тестові норми і як ними користуватися.

Початковий сумарний бал, підрахований з допомогою ключа, не є показником, який можна діагностично

інтерпретувати. Його називають в тестології «*сирим*» *тестовим балом*. Застосування тестових норм у професійно організованій системі тестування ґрунтується на переведенні тестових балів із «сирої» шкали в стандартну. Ця процедура називається *стандартизацією тестового балу*.

Вибірка, на якій визначаються статистичні тестові норми, називається *вибіркою стандартизації*. Кількість її елементів, як правило, не менше 200. Стільки тестованих мають взяти участь в експерименті з визначення тестових норм – в експерименті зі стандартизації тесту.

4.3. Варіативність тестів

Варіативність змісту означає можливість заміни елементів завдання іншими схожими елементами, так, щоб відповіді у більшості, якщо не усіх випробуваних, були різні. Наприклад, у тесті з математики усі випробувані можуть розв'язувати квадратне рівняння, але у кожного випробуваного параметри рівняння можуть відрізнитися. Отже, різними будуть і відповіді. Отже, завдання та відповіді різні, а перевіряється одна й та ж тема. Не випадково, в центрах тестування різних країн вимагають, щоб кожен розробник готував не менше п'ятнадцяти варіантів кожного завдання свого тесту.

Таким чином, варіативність змісту кожного завдання тесту виключає можливості списування. Списування – традиційна вада середньої і вищої школи. Воно є результатом втрат не тільки етичних, але і професійно-педагогічних. Професорсько-викладацький склад, в більшості своїй, не навчений створювати тестові форми з варіативним змістом.

Деякі автори вважають завдання відкритої форми кращими і важливішими за завдання з вибором однієї або кількох правильних відповідей. Але ця думка, в деяких випадках, є помилковою. Не усі завдання відкритої форми насправді виявляються важчими і кращими ніж завдання з вибором однієї правильної відповіді. Відкрита форма, в бланковому варіанті програє за критерієм технологічності. Тому справжні тестові системи спираються головним чином на завдання з вибором однієї або декількох правильних

відповідей. Завдання ж відкритої форми застосовуються переважно тоді, коли треба перевірити правопис слова, терміна, формули. За кордоном їх застосовують, головним чином, у процесі емпіричної апробації завдань, для продукування безлічі неправильних, але правдоподібних відповідей, які називаються дистракторами. З тим, щоб потім створити якісні завдання з вибором однієї або декількох правильних відповідей.

4.4. Створення паралельних варіантів

Паралельними називаються завдання, які ґрунтуються на принципі варіативності змісту. Крім того, вони мають приблизно однакову емпіричну міру складності усіх варіантів завдань та подібні кореляції з вектором сум вихідних балів випробуваних. З цього визначення видно, що паралельність завдань забезпечується не тільки варіативністю змісту, але й статистичною перевіркою міри складності і кореляції варіантів завдання. За змістом принцип паралельності завдань ширший, ніж принцип варіативності змісту. Паралельність включає в себе ще й формальні критерії, спирається на емпіричні дані результатів випробуваних.

Різний і статус цих принципів. Варіативність є змістовно-методичним принципом конструювання тестових завдань, у той час як принцип паралельності завдань є загальнотеоретичним, що базується на статистиці, а тому використовується у всіх формальних теоріях педагогічних вимірювань. Треба підкреслити, що паралельність за змістом є лише однією із необхідних умов паралельності тестових завдань. Не усі варіативні завдання можуть відповідати принципу паралельності змісту. Інакше кажучи, варіативність є істотною, але не достатньою умовою для виникнення феномена паралельності варіантів тестового завдання.

Ідея розроблення та існування паралельних варіантів тестових завдань є важливою для розробників тестів. Це шлях до сучасного педагогічного тесту. У західній літературі часто помилково пишуть про «паралельні тести»; так само помилково це перекладається і в текстах російською та українською мовою. Насправді, тест створюється один, а до нього робляться

паралельні варіанти кожного завдання тесту. Створення паралельних варіантів створює передумови для застосування ефективних методів перевірки надійності результатів тесту.

4.5. Фасет

Варіативність змісту тестового завдання досягається застосуванням фасетів.

Фасетом називається набір змінних елементів завдання, що представляються в фігурних дужках для випадкового вибору в автоматизованому тестуванні

Застосування фасета дозволяє в процесі контролю виключити списування та інші спотворення, відомі з досвіду проведення іспитів, заліків та інших традиційних форм перевірки знань. Ідея фасета асоціюється з відомою педагогічною практикою створення кількох варіантів однаково складних завдань для проведення контрольних робіт. Принцип фасетності змісту завдання є головним при композиції професійно розроблених тестових завдань; він дозволяє створювати відразу декілька варіантів одного й того ж завдання, а, отже, і варіантів тесту.

Приклади фасета:

1. {Шевченко} ПИСАВ:
 - 1) драми та вірші
 - 2) вірші та поеми
 - 3) поеми та драми

Змінювати можна не тільки прізвища авторів, а також функції, назви і багато іншого:

2. ФУНКЦІЯ {.....} є
 - 1) парною
 - 2) непарною
 - 3) ні парною, ні непарною

Хотілося б підкреслити, що цей принцип відноситься не до добору відповідей, а до добору змісту завдань. Під час тестування, проведеного за допомогою ЕОМ і записаної в ній інструментальної програми, випробуваний отримує з фасета тільки один варіант змісту завдання.

Наприклад, якщо потрібно перевірити знання фонетики російської мови в підсумковому контролі, то випробуваному, швидше за все, дістанеться тільки одна літера з фасета:

3. БУКВА {б, в, г, д, ж, к, л, м, н, п, р, с, т, ф, ц, ч, ш, щ} позначає звук
- 1) м'який
 - 2) твердий
 - 3) м'який і твердий

У поточному ж контролі, випробуваний може потренувати себе на всіх буквах представленого фасета. Ця форма є особливо ефективною для організації самоконтролю – найгуманнішої форми контролю знань.

4. {Сила, прискорення}, – ВЕКТОР, що характеризуються
- 1) величиною
 - 2) величиною і напрямком
 - 3) величиною, напрямом і точкою програми.

Завдання, утворені заміною елементів з фасета в багатьох випадках можна називати паралельними за змістом. Паралельність завдань за змістом є дуже важливим, але не єдиним критерієм для розробки так званих паралельних тестів. Ідея розробки і існування паралельних тестів є основоположною для розробників тестів.

§ 5. Тестові завдання: правила конструювання, формати, вимоги

5.1. Правила конструювання тестових завдань

1. Кожне тестове завдання має оцінювати досягнення важливої та істотно освітньої цілі. Слід уникати перевірки тривіальних або надмірно вузькоспеціальних 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. Відомості, що містяться в одному тестовому завданні, не повинна давати відповідь на інше тестове завдання.
7. Не рекомендується використовувати як правильну відповідь чи дистрактор фраз "все з вищевказаного", "нічого з вищевказаного".
8. Не рекомендується використовувати як правильну відповідь чи дистрактор фрази "жоден варіант відповіді неправильний", "немає правильної відповіді", "усі відповіді правильні", "інколи", "ймовірно" тощо.
9. Умова має бути сформульована позитивно.
10. Необхідно уникати при формулюванні умови 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 ним проблему. Умова може містити лише завдання або складатися із вступних відомостей та запитання. 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 чіткіше завдання. **Формулювання** запитання доцільно починати з **дієслова**. 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, незважаючи на список варіантів відповідей.

Формулюйте умови **позитивно**. Уникайте негативних формулювань, які вимагають протилежної, порівняно з більшістю тестових завдань дії (вибір неправильного, гіршого), та є складними для розуміння екзаменованими.

Правила написання варіантів відповідей

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

Під час добору дистракторів доцільно використовувати поширені помилки, хибні уявлення, об'єкти, що відповідають лише частині характеристик, наведених в умові, тощо. Водночас у дистракторах не повинно бути каверзних, фальшивих та хибних відомостей. Використовуйте правильні

твердження, але такі, що не належать до поданого 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.

Варіанти відповідей треба розміщувати системно (в логічному, в алфавітному порядку, у порядку збільшення або зменшення чисел тощо).

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

Відповіді мають бути **однорідними за змістом** і належати до однієї навчальної мети.

Відповіді мають бути короткими та простими за структурою. Слова, що повторюються, 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, просте речення. Складні речення у відповідях використовувати не рекомендується.

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.

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

Не використовуйте фраз на зразок «все з вищевказаного», «нічого з вищевказаного», «немає правильної відповіді», «усі відповіді правильні», «інколи», «ймовірно» тощо.

5.2. *Форми (формати) тестових завдань*

За період від 50-х років минулого сторіччя, коли вперше було використано тестові завдання багатовибіркового типу, розроблялося й апробувалось більше 40 різних форматів тестових завдань. Одні з них підтвердили свої діагностичні властивості, інші – ні.

Перший апробований формат, – це формат А. Він і дотепер є найуживанішим, оскільки, по-перше, дає можливість перевіряти знання на різних рівнях засвоєння. По-друге, при створенні тестового завдання формату А автори роблять менше помилок.

ФОРМАТ А

тестові завдання з **однією** найкращою відповіддю.

Діагностує знання, розуміння, застосування.

Структура цього тестового завдання така: умова (завдання), запитання за цією умовою та чотири-п'ять варіантів відповідей, один з яких є найправильнішим. Неправильні відповіді можуть бути не абсолютно неправильними, але вони все ж таки менш правильні, ніж еталонна відповідь.

Принципи добору відповідей:

1. Суперечність
2. Протилежність
3. Однорідність
4. Кумуляція
5. Поєднання
6. Градування
7. Подвоєне протиставлення
8. Поєднання принципів

ФОРМАТ X

тестові завдання *множинного вибору* «правильно – неправильно».

Діагностує знання.

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

До основних елементів завдань з вибором кількох правильних відповідей відносяться:

Інструкція для тих хто тестується. Вона пишеться так: «Обведіть номери усіх правильних відповідей». При автоматизованому контролі інструкція змінюється: «Відповідаючи на завдання тесту, натисніть на клавіші з номерами усіх правильних відповідей».

Змістовна основа в основному співпадає з іншими формами і видами завдань. Ефективний спосіб збагачення змісту завдань – використання фасету.

Не рекомендується заперечне формулювання завдань, коли запитують про те, що не використовується, не відноситься, не робиться та ін.

Відповіді на завдання обов'язково відносяться до одного роду чи виду термінів. Загальну кількість відповідей бажано мати в діапазоні від 5 до 14. Особливістю завдань з вибором кількох правильних відповідей є те, що в них потрібно визначати не лише правильні відповіді, але й оцінювати повноту своєї відповіді.

В завданнях цього формату найчастіше перевіряються лише засвоєння першого рівня когнітивного домену – знання

фактів та концепцій, тобто декларативних, а не процедурних знань.

ФОРМАТ N

тестові завдання з *кількома* найкращими відповідями.

Діагностує знання, розуміння, застосування.

У тестових завданнях такого формату екзаменований має вибрати певну кількість (наприклад 2, 3 чи 4) 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.

ФОРМАТ R

тестові завдання *розширеного вибору*.

Діагностує знання, розуміння, застосування.

Кожен блок запитання такого формату складається із трьох-чотирьох умов і спільного для них переліку варіантів відповідей, а інколи із вставного запитання. Кількість варіантів відповідей може коливатися від 4 до 20. Як правило, варіанти відповідей розміщено на початку блоку, перед умовами.

Структура запитань такого формату є такою: маємо, наприклад, п'ять варіантів відповідей, які можуть використовуватися лише раз, або не використовуватися жодного разу, одне запитання до умов і три-чотири умови.

ФОРМАТ B

тестові завдання *логічних пар* або *визначення відповідності*.

Тестові завдання формату B належать до категорії логічних пар, що складається 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 дібрати до кожного пронумерованого завдання один найкращий варіант відповіді, позначений літерою.

Оскільки кожен варіант відповіді може використовуватись більше одного разу або не використовуватись взагалі, тестові завдання формату В неможливо виконати методом виключення.

5.3. Створення тестових завдань різних когнітивних рівнів

Створення тестових завдань на вимірювання вищих когнітивних рівнів є складним процесом, що вимагає від того хто створює тест багато часу. Автор має проявити творчий підхід для створення привабливих і якісних завдань.

Поради для створення якісних завдань:

- ✓ Використання в умові практичних або реальних ситуацій.
- ✓ Надавайте схему обладнання апарату та поставте завдання проаналізувати або оцінити його. Наприклад, «Що відбудеться з графіком якщо його перенести», «як слід змінити координати центра кола, щоб воно проходило через точку.....», «Якого елемента не вистачає щоб....»
- ✓ Використовуйте реальні цитати з книг, газет 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.

Особливості завдань різних когнітивних рівнів

Рівень знань

Цей рівень 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.

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.

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.

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 відомостей. Аналіз вважається вищим когнітивним рівнем.

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.

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 і 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.

5.4. Типи тестових завдань

Нагадаємо, що вперше термін «test» (анг. випробування, проба, перевірка) був введений в наукову літературу американським психологом Дж. Кеттелом.

За метою використання та місцем у навчальному процесі розмежовують тести *навчальні* й *контрольні*. Навчальні тести використовуються на всіх етапах роботи над матеріалом і покликані відстежити рівень оволодіння матеріалом, закріпити або повторити його. Їх головна мета – виявити прогалини в знаннях, уміннях, навичках школярів / студентів, спрямувати їх на усунення недоліків у підготовці. З огляду на це перевіряти такі тести можуть самі учні / студенти. Контрольні тести проводяться як певний підсумок роботи над вивченням теми, вони мають комплексний характер, тобто перевіряють знання й уміння, здобуті й вироблені школярами / студентами в межах одного або кількох тематичних блоків.

Тестове завдання – складова одиниця тесту (кожне завдання є невід'ємною його частиною, тому видалення хоча б одного завдання неминуче призведе до виникнення прогалин в оцінюванні), яка відповідає не тільки формальним, але й певним статистичним вимогам:

- ✓ частка неправильних відповідей в кожному завданні;
- ✓ співвідношення балів за завданнями та балів за весь тест.

Характеристики завдань в тестовій формі:

- ✓ стислість, тобто ретельний добір слів, символів, графіків, які дозволяють використовуючи мінімум засобів досягнути максимальної зрозумілості змісту завдання;
- ✓ технологічність, тобто компонуванням завдань, яке дозволяє комп'ютеризувати процес тестування;
- ✓ визначеність місця для відповідей;

- ✓ єдиність правил оцінювання відповідей;
- ✓ адекватність інструкції формі та змісту завдання.

Існує багато думок щодо класифікації тестових завдань. Наприклад, І. Є. Булах класифікує їх *за видом тестового завдання*. Багато вчених поділяють таку думку. Отож, тести бувають з відкритими та закритими тестовими завданнями.



Рис. 5.1. Класифікація тестів за видом тестового завдання

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.

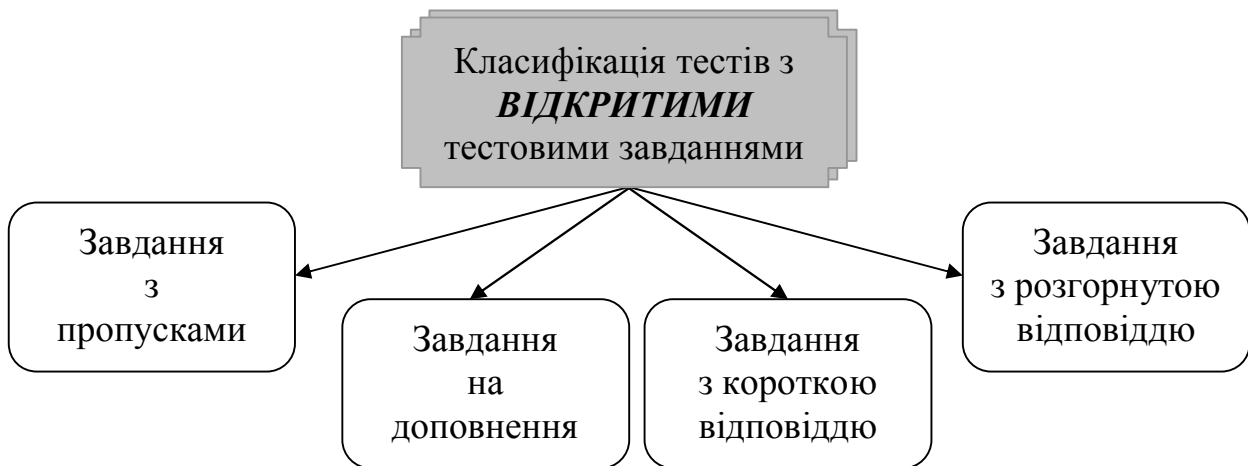


Рис. 5.2. Класифікація тестів з відкритими тестовими завданнями

Роз'яснень потребують завдання з короткою (продукованою) відповіддю. Вони можуть бути сформульовані в різній формі подання знань, але так, що запитання потребують короткої відповіді, яка є, переважно, результатом математичних розрахунків.

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

Види відкритих тестових завдань

Завдання відкритого типу (для самостійної відповіді) – тестові завдання без указаних можливих варіантів відповідей. Вимагають від тестованого самостійно сформулювати правильну відповідь на поставлене запитання та відповідно її оформити (записати одним словом, цифрою, буквою, словосполученням або підготувати розгорнуту відповідь на кілька сторінок). До завдань відкритого типу (для самостійної відповіді) належать:

- 1) завдання з пропусками.
- 2) завдання на доповнення.
- 3) завдання з короткою відповіддю.
- 4) завдання з розгорнутою відповіддю. (Е. Гронлунд до цієї групи завдань зараховує есе зі стислою та розширеною відповідями).

Завдання з короткою відповіддю відзначаються високим ступенем регулювання відповіді. Оцінювальні результати навчання – знання та навички обчислення, знання фактів, правил, термінів тощо. Завдання з розгорнутою відповіддю характеризують високий ступінь свободи у відповідях, оцінювальні результати навчання – розуміння, застосування, аналіз, синтез, оцінювання.

Правила конструювання завдань

1. Завдання має відповідати навчальним цілям.
2. Завдання має оцінювати різні рівні знань і вмінь.
3. Завдання має передбачати інтегрування знань і вмінь.
4. Завдання має бути сформульоване чітко, його умова повинна бути повністю зрозумілою для тестованих.
5. Завдання не повинне містити чинників, що перешкоджають правильному виконанню.
6. Обсяг завдання має бути розрахований так, щоб тестований міг його виконати за відведений проміжок часу.
7. Складаючи завдання, необхідно врахувати наявне обладнання й ресурси.
8. Завдання має відповідати віковим особливостям тестованих.

9. Завдання має виявляти, чи вміє тестований виконувати аналогічні завдання.
10. Завдання має передбачати відповідний метод оцінювання результатів виконання.

Завдання відкритої форми, у яких немає варіантів відповіді

До завдань такого типу належать:

- завдання з короткою відповіддю;
- завдання з розгорнутою відповіддю.

Завдання з розгорнутою відповіддю

Таке завдання перевіряє, як учні засвоїли комунікативну лінію програми, сформованість їхньої мовленнєвої компетенції. Учень мусить побудувати письмове висловлення, що передбачає обґрунтоване доведення (через ствердження чи спростування) запропонованої тези.

Схожу форму тестових завдань, а саме есе, широко використовують у Європі. Під час написання есе потрібно не просто пригадати знання, щоб вибрати правильну відповідь, а також зв'язати між собою факти та принципи, забезпечити логічність викладу і обґрунтувати добір цих фактів. Таке завдання містить відомості про процеси мислення й такі якості випускника, як критичне мислення, здатність інтерпретувати ідеї, відрізнити факти від суджень, робити обґрунтовані висновки тощо.

Такими завданнями найліпше перевіряти, як тестовані вміють створювати власні висловлювання, формулювати думки, ідеї, дотримуватися певної структури тексту. Відповіді в таких завданнях істотно варіюються від стислої й точної, сформульованої кількома реченнями, до розгорнутої довільної форми й обсягу. В останньому разі тестованого можуть досить жорстко обмежувати щодо часу виконання роботи чи її обсягу, матеріалу, на який можна спиратися, або структури тексту.

Зразок інструкції до завдання з розгорнутою відповіддю

Сформулюйте тезу, наведіть два-три переконливі аргументи, які найкраще підтвердять Ваші міркування. Проілюструйте Ваші думки посиланнями на приклади з

художньої літератури (зазначте назву твору, укажіть проблему, порушену письменником, назву твору, художній образ, через який проблему розкрито, наведіть цитату з твору тощо), історичними фактами або випадками з життя. Не переказуйте змісту, не давайте повної характеристики образів. Сформулюйте висновки.

Роботу запишіть на бланку Б. У разі потреби використовуйте чернетку.

Орієнтовний обсяг роботи – 1 сторінка (200 – 250 слів).

Таблиця 6

Аналіз переваг та недоліків завдань з розгорнутою відповіддю

Переваги та недоліки завдань з розгорнутою відповіддю	
Переваги	Недоліки
<ul style="list-style-type: none"> ➤ Оцінюють результати навчання найвищого рівня (уміння аналізувати, синтезувати, оцінювати). ➤ Підготовка завдань потребує менше часу порівняно із завданнями закритої форми. ➤ Перевіряють уміння тестованих інтегрувати й застосовувати ідеї, висловлювати власне ставлення щодо запропонованої теми. 	<ul style="list-style-type: none"> ➤ Оцінювання результатів є трудомістким і потребує багато часу. ➤ Оцінювання таких завдань є досить суб'єктивним порівняно з іншими формами завдань і часто знижує надійність тесту. ➤ Складно розрахувати час, необхідний для виконання завдання. ➤ Складно передбачити результати тестованих через високий ступінь свободи щодо структури твору (есе), стилю викладу та глибини розкриття теми. ➤ Каліграфічний почерк, добре естетичне оформлення роботи й фальсифікації під час проведення тестування (списування, порушення процедур тощо) підвищують оцінки, а поганий почерк, орфографічні й пунктуаційні помилки їх знижують.

Тести з закритими тестовими завданнями

Тести із закритими тестовими завданнями передбачають, що варіанти відповідей до тестового завдання надаються також особі, яка тестується.

Закриті тестові завдання перевіряють повноту засвоєння лінгвістичної змістової лінії навчальної програми та рівень сформованості в учнів загальноосвітніх навчальних закладів мовної компетенції.

Завдання закритої форми з варіантами відповідей на вибір (завдання на вибір відповіді) – тестові завдання, що пропонують тестованому перелік можливих варіантів відповідей, з яких вони мають вибрати одну або кілька правильних.

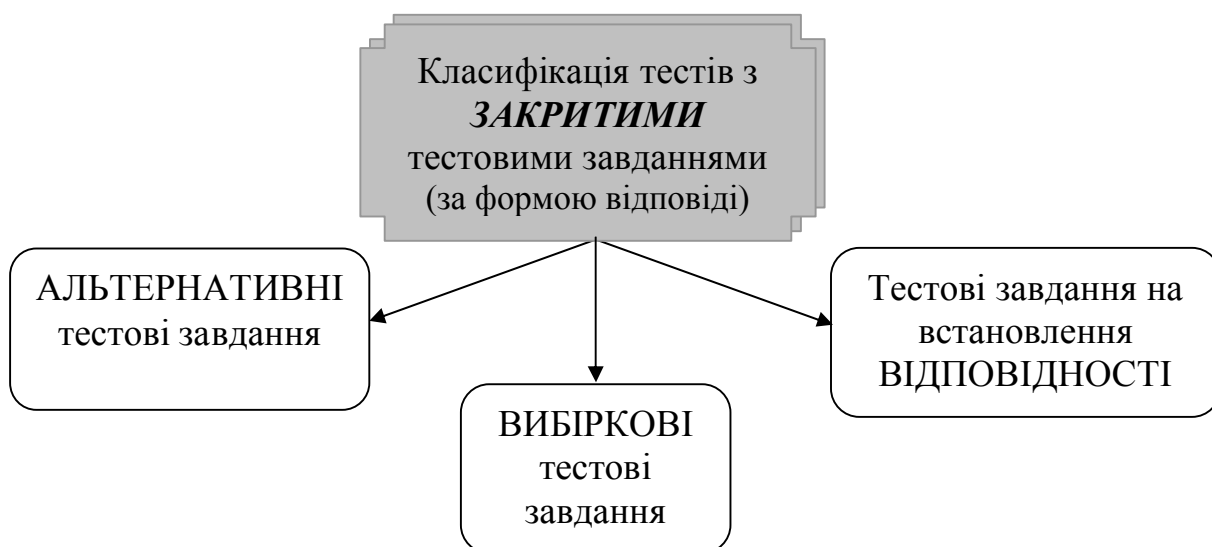


Рис. 5.3. Класифікація тестів з закритими тестовими завданнями (за формою відповіді)

До завдань на вибір відповіді належать:

- ✓ завдання з вибором однієї або кількох правильних відповідей;
- ✓ завдання на встановлення відповідності (утворення логічних пар);
- ✓ завдання на встановлення правильної послідовності.

Вимоги до завдань на вибір відповіді

1. У тексті завдання не повинно бути жодної двозначності чи неточності у формулюваннях.
2. Основу завдання доцільно формулювати в стверджувальній формі, яка після підстановки відповіді перетворюється на істинне або хибне твердження.
3. Основу завдання треба формулювати дуже коротко, одним реченням (до 8 слів) простої синтаксичної конструкції.

4. Якщо варіанти відповідей починаються з одного слова (кількох однакових слів), їх потрібно перенести в основу завдання, формулюючи варіанти відповідей 2-3 словами.
5. В основі завдання не повинно бути жодного іррелевантного для цієї проблеми матеріалу.
6. Усі варіанти відповідей мають бути приблизно однакові завдовжки або в деяких завданнях правильний варіант відповіді може бути трохи коротший за дистрактори.
7. Із завдання необхідно вилучити всі вербальні асоціації, що підказуватимуть вибір правильної відповіді.
8. Частота вибору варіанта правильної відповіді в усіх завданнях тесту має бути приблизно однаковою.
9. У відповідях не рекомендується використовувати слова «все», «жоден», «ніколи», «завжди», «всі названі», «з названих жоден».
10. Не можна використовувати завдання, у яких одна відповідь впливає з іншою.
11. Не можна включати в тест завдання, що містять оцінні судження або передбачають формулювання висловлювань з власною думкою щодо певних запитань.
12. Усі дистрактори мають бути однаково привабливі для тестованих, які не знають правильної відповіді.
13. Жоден із дистракторів не повинен бути частково правильною відповіддю, яка за певних умов може розглядатися як правильна відповідь.
14. Усі варіанти відповідей мають бути граматично узгодженими з основою завдання.

Види закритих тестових завдань

Тестові завдання з вибором однієї правильної відповіді

Як стверджують науковці, цей формат апробовано ще в 50-х роках минулого століття. Він і дотепер є найуживанішим. Цей формат тестових завдань дає змогу оцінити навчальні досягнення випускників на різних рівнях: знання, розуміння і застосування знань.

Аналіз переваг та недоліків завдань з однією правильною відповіддю

Переваги та недоліки завдань з однією правильною відповіддю	
Переваги:	Недоліки:
<ul style="list-style-type: none"> ➤ Оцінювання результатів швидке, легке, об'єктивне, надійне. ➤ Завдання добре структуровані й чіткі. ➤ Дають можливість вимірювати результати навчання від простих до складних. ➤ Неправильні варіанти відповідей дають діагностичні відомості про рівень сформованості знань і вмінь. ➤ Результати менше залежать від угадування порівняно із завданнями з альтернативними відповідями. 	<ul style="list-style-type: none"> ➤ Складання запитань потребує багато часу. ➤ Часто складно знайти правдоподібні дистрактори. ➤ Завдання неефективні для оцінювання вмінь розв'язувати проблеми. ➤ Рівень сформованості навичок читання може впливати на результати оцінювання.

Тестові завдання з вибором кількох правильних відповідей

У практиці масового педагогічного тестування застосовують дві модифікації: завдання із заздалегідь указаною кількістю правильних відповідей і завдання з не вказаною кількістю правильних відповідей. В інструкції треба обов'язково написати, скільки правильних відповідей має завдання, або зазначити, що відповідей може бути кілька. Відповіді до завдання обов'язково повинні належати до одного роду або виду термінів. Кількість відповідей бажано мати від 5 до 12. У тестових завданнях такого формату екзаменованій мусить вибрати усі правильні відповіді із запропонованого списку варіантів.

У цьому типі тестів кожен варіант відповіді повинен бути або абсолютно правильним, або абсолютно неправильним, для того щоб уникнути неоднозначності. Тестове завдання цього формату так само, як і попереднього, діагностує знання, розуміння і застосування.

Тестові завдання на встановлення відповідності

Цей формат належить до категорії логічних пар і називається ще форматом розширеного вибору. Завдання цього типу складається з інструкції-завдання та двох колонок. Одна колонка (ліворуч) – це перелік вихідних умов (слів, словосполучень, речень, дат, формул, термінів тощо), до яких тестований має відшукати відповідь у другій колонці (праворуч), яку називають списком відповідей. Тестований має порівняти матеріал лівої й правої колонок і утворити правильні логічні пари. Перед основою записують інструкцію. Визначають дві модифікації завдання: завдання, у яких відповідь можна вибирати тільки один раз, і завдання, у яких відповідь можна вибирати кілька разів. Завдання можуть бути складені у формі вибору відповідей і у формі вписування відповідей (до умови в лівій колонці самотійно вписати відповідь, утворивши логічну пару).

Таблиця 8

Аналіз переваг та недоліків завдань на встановлення відповідності

Переваги та недоліки завдань на встановлення відповідності	
Переваги	Недоліки
<ul style="list-style-type: none">➤ Оцінювання результатів легке, об'єктивне й надійне.➤ Запитання мають компакту форму, тому що набір відповідей відповідає тій самій основі, завдання перевіряє кілька подібних фактів.➤ Щоб прочитати завдання та вибрати варіанти відповідей, тестований витрачає небагато часу.➤ Завдання легко складати, якщо об'єднати кілька завдань, що мають однаковий набір варіантів відповідей.	<ul style="list-style-type: none">➤ Переважно оцінюють прості результати навчання на рівні знань, що базуються на асоціаціях.➤ Важко складати завдання, які містять достатню кількість однорідних відповідей.➤ Значна чутливість до невідповідних ключів.

Завдання на встановлення правильної послідовності – тестове завдання, у якому тестований має встановити правильну послідовність елементів, дій, подій, операцій тощо. У зарубіжній класифікації цю форму не виокремлюють, а розглядають як варіант завдання з вибором правильної відповіді або як завдання відкритого типу, коли тестований сам має вписати правильну, на його думку, послідовність. Ця форма завдань призначена насамперед для перевірки сформованості алгоритмічного мислення та алгоритмічних знань і умінь, навичок пізнавальної діяльності. Цей формат завдань природніший для гуманітарних дисциплін, наприклад, історії. У математиці йдеться про впорядкованість елементів деякої множини за лінійним порядком, послідовність дій та операцій, ланцюжок розумових дій, який утворює систему знань, умінь, уявлень, перевірку знань певних понять і термінів.

Таблиця 9

*Аналіз переваг та недоліків завдань на встановлення
правильної послідовності*

Переваги та недоліки завдань на встановлення правильної послідовності	
Переваги	Недоліки
<ul style="list-style-type: none"> ➤ Перевіряють сформованість алгоритмічного мислення. ➤ Читання варіантів відповідей потребує мало часу. ➤ Завдання добре структуроване й чітке. ➤ Результати менше залежать від угадування, ніж у завданнях з однією правильною відповіддю. 	<ul style="list-style-type: none"> ➤ Обмежуються переважно оцінюванням простих результатів навчання. ➤ Такі завдання обмежують творчу діяльність тестованих.

Вимоги до завдань в тестовій формі

Перевірка рівня навчальних досягнень учнів є одним із важливих етапів навчального процесу. Застосування комп'ютерів дає змогу забезпечити об'єктивність оцінювання знань учнів. Перевірку засвоєння теоретичних знань найпростіше реалізувати на комп'ютері у формі тестів.

Більшість комп'ютерних програм дозволяють розробляти тести, в яких учням пропонується вводити правильну відповідь

одним з трьох способів: обирати одну правильну відповідь з декількох запропонованих (одноваріантний вибір), обирати кілька можливих значень з числа запропонованих (поліваріантний вибір), вводити відповідь власноруч (тест з відкритою відповіддю).

Кожна з форм введення відповіді має власні позитивні та негативні риси. Так, найпростіший в реалізації тест з одноваріантним вибором відповіді надає високу ймовірність випадкового вибору правильного варіанту, що знижує об'єктивність оцінювання. При поліваріантному виборі потребує чіткого трактування питання оцінювання часткового вибору правильних варіантів. Важко провести межу при виставленні балів за вибір одного або двох варіантів з трьох потрібних або вибір зайвих варіантів поруч з правильними. Введення відповіді з клавіатури ускладнено можливостями неістотних для предмету оцінювання граматичних або синтаксичних помилок та накладає додаткові вимоги на довжину рядка, який потрібно вводити з клавіатури.

Для об'єктивності оцінювання знань учнів при тестовій системі перевірки знань, крім вибору форми введення відповідей, важливо забезпечити зміст запитань, який дозволить вважати тест валідним.

Валідність – це міра відповідності тесту вимірюваним знанням, умінням і навичкам, для перевірки яких був розроблений тест, міра відповідності стандартам і програмам навчання, а також результатам тестування. Це найважливіша комплексна характеристика, що відображає точність тестування.

Рекомендації щодо добору завдань для тестової перевірки знань були запропоновані у статті В.М.Казієва "Готовимся к ЕГЭ по информатике" в журналі "Информатика в школе" № 2' 2006. Наведемо основні правила розробки якісних тестів, запропоновані в названій публікації.

Тест повинен мати необхідний і достатній рівень складності.

Складність – міра розумових зусиль, потрібних для вибору відповіді. Часто цей параметр називається вагою і

визначається як частка правильних або неправильних відповідей при відповіді на завдання в групі тестованих.

Тест повинен бути об'єктивним і надійним.

Надійність – міра правильності і адекватності відображення тестом рівня знань учнів. Надійний тест забезпечує рівні права кожній групі учнів і кожному учню в групі.

Тест повинен бути стійким і таким, що може бути співставлений певній шкалі.

Стійкість тесту – міра збереження надійності і валідності при перенесенні тесту в інше, аналогічне середовище, міра рівнозначності і однорідності тестів для різних груп учнів.

Шкалювання тесту – здатність тесту відображати результати навчальних досягнень в деякій системі (шкалі) оцінок або балів.

Тест повинен бути репрезентативним.

Репрезентативність тесту (або бази тестів) – міра повноти охоплення завданнями навчального матеріалу, програми, відображення тестами різних рівнів навчання.

Тест повинен бути значущим і дискримінантним.

Значущість тесту – міра необхідності, актуальності включення в тест ключових знань.

Дискримінантність тесту – міра диференціації тестованих щодо максимального або мінімального рівня навчання.

Тест повинен бути достовірним, науковим, несуперечливим.

Достовірність, науковість, несуперечність тесту – міра істинності тесту, узгодженості, відповідності сучасному стану науки і технології, методиці навчання.

Необхідно забезпечити узгодженість завдань тесту як між собою (внутрішня узгодженість), так і з іншими тестами (зовнішня узгодженість).

Тестове завдання повинно бути сформульовано ясно і чітко, не допускати двозначного тлумачення і сприяти формулюванню правильної відповіді.

Час, виділений на тестове завдання, повинен бути витрачений на пошук відповіді, а не на розуміння умови запитання. Потрібно намагатися формулювати завдання у

вигляді одного речення. Бажано використовувати просту, граматично правильну стверджувальну форму завдання у вигляді одного речення з 5-20 слів. Слова, що повторюються, і словосполучення у відповідях повинні бути виключені і перенесені в основну частину умови. У учня не повинне виникнути питання по уточненню умови або дистрактора у процесі пошуку правильної відповіді.

Кожне завдання тесту повинне бути функціонально завершеним, тобто перевіряти конкретне знання, уміння або навичку.

Бажано якомога рідше використовувати «нестрогі» слова типу «іноді», «часто», «завжди», «все», «ніколи», «великий», «невеликий», «малий», «багато», «менше», «більше» і граматичні обороти типу «Чому не може не...», «Чи правда, що ...», «Чи можливо...», подвійні заперечення і т.д.

Відповіді повинні містити не більше 2—3 ключових слів за умови завдання. Бажано будувати відповіді однакової форми і довжини. Кількісні відповіді бажано упорядковувати за зростанням або спаданням, а якщо відповіді легко обчислити, то першою не повинна бути вказана правильна відповідь.

Необхідно виключити можливість вибору правильної (або неправильної) відповіді інтуїтивно і асоціативно, здогадкою, вербально, а місце правильної відповіді вибрати випадково.

Відповіді повинні бути незалежними, однаково привабливими для вибору, а вплив вірогідності вгадування правильної відповіді на загальний результат тестування повинен бути якомога меншим.

У кожному завданні кількість дистракторів (правдоподібних варіантів відповіді) повинна бути від 3 до 5. Всі дистрактори повинні бути дібрані грамотно, вміло, без каверз, але не довільно, а відповідно до типових помилок, що можуть бути припущені (прогнозованими) за даним завданням.

З відповіді на одне тестове завдання не можна одержувати яким-небудь чином (наприклад, евристично) відповідь на інше завдання.

Якщо можлива побудова якісного тестового завдання закритої форми (структури), то бажано використання цієї форми. Недоліком тестових завдань в закритій формі

вважається можливість дізнатися відповідь, не знаючи її. До переваг можна віднести просту структуру, доступність сприйняття, що сприяють отриманню оптимального результату.

Тест повинен складатися з мінімальної кількості завдань, достатніх для повного визначення рівня навчальних досягнень. Рекомендується уникати непотрібного дублювання знань, що перевіряються, умінь і навичок, а також передбачити всі необхідні.

Необхідно вживати заходи з мінімізації «привабливості» використання шпаргалок, списування і добору відповідей «навмання», оскільки вірогідність вгадування звичайно висока і складає 0,2—0,25. Тільки грамотні, цікаві запитання і ситуації можуть викликати у учня прагнення відповісти, а не дібрати відповідь.

Списування можна виключити такими методами: підготовкою достатньої кількості рівноцінних тестів, «перемішуванням» одного і того ж набору завдань усередині групи, чітким вимірюванням витрат часу і обмеженням в часі.

Можна порекомендувати прийом: включати до кількості дистракторів варіанти, за якими можна судити про явну невідповідність в цій підтемі або випадковий добір відповіді. Місце таких відповідей в кожному завданні фіксується, і за ці відповіді нараховуються як не позитивні, а негативні, «штрафні» бали. Про наявність таких відповідей учнів наперед потрібно попередити перед початком тестування.

Якщо тестування проводиться в системі оцінок з двома варіантами відповідей («так», «ні»), то і результат тестування повинен бути оцінений в полярній шкалі: «залік — не залік», «зданий — не зданий».

Таким чином, тест повинен відповідати необхідним і вимірюваним знанням, умінням і навичкам, стандартам і програмам навчання, повинен забезпечувати однакові права і близькі показники для кожної групи тестованих, відображати результати навчальних досягнень в системі оцінок, що задається, бути актуальним.

§ 6. Опрацювання та представлення результатів тестування. Статистичне обґрунтування якості тесту

6.1. Критерії якості методу вимірювання. Поняття валідності і надійності тестів

Вибір того чи іншого методу вимірювання рівня знань ґрунтується на критеріях, які визначають якість засобу вимірювання. Найважливішими критеріями якості є валідність, об'єктивність, надійність і точність. Розглянемо їх визначення та визначимо зміст кожного з них.

Фундаментальним є критерій *валідності*. Цей термін походить від англійського "valid" і означає обґрунтований, дійсний, придатний, той, що має силу. Вдале визначення валідності педагогічного тесту дала Анна Анастасі, американський психолог (1908 – 2001).

Валідність тесту – це «поняття, яке визначає, що вимірює тест і наскільки якісно це здійснюється» (Анна Анастасі).

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

Невалідний тест – такий, у якому 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 базового матеріалу. 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 комплексом відомостей про характеристики тесту, процедури тестування та методики оцінювання.

Багато дослідників (І. Є. Булах, М. Р. Мруга) пропонують 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:

- ✓ 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 тому, 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 (змістова);
- ✓ валідність відповідності;
- ✓ валідність прогнозу (прогностична).

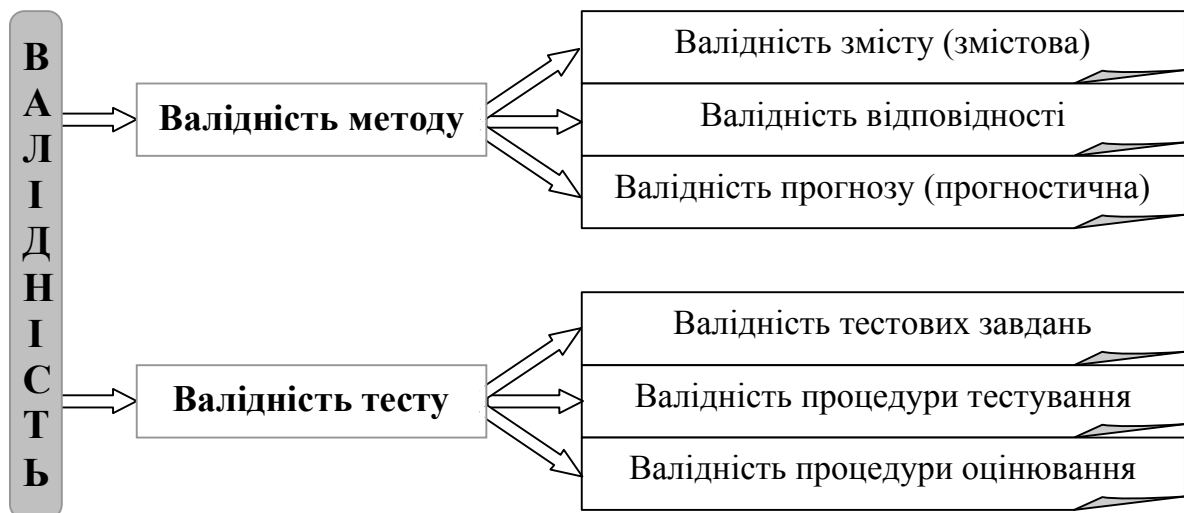


Рис. 6.1. Класифікація валідності за функціональною ознакою

Валідність змісту (змістова) – характеристика тесту, що відображає ступінь упевненості, завдання тесту досить повно охоплює зміст певної сфери знань, тест точно визначає істотні

для цієї діяльності навички, але водночас не перевіряє другорядні в даному разі знання й навички, залежить від глибини попередньо проведеного авторами тесту аналізу навчального плану й програми, змісту підручників, від точності й повноти формулювання мети навчання

Валідність відповідності – це відповідність результатів вимірювання та оцінювання, здобутих 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, І. Є. Булах та М. Р. Мруга пропонують таке визначення коефіцієнта валідності:

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

Валідність прогнозу (прогностична) – характеристика тесту, яка відображає ступінь упевненості, що отримані за тест оцінки добре прогнозують майбутні досягнення тестованого (наприклад, у США результати за складання тесту SAT передбачають певний рівень успішності студента у ВНЗ). Для виявлення прогностичної валідності результати одного тесту порівнюють із результатами іншого тесту, який проводять, наприклад, через півроку або рік навчання. Учитель / викладач може використати індивідуальні бали студента як такі, що передбачають їхній успіх у майбутньому навчанні чи професійній діяльності тощо. Прогностичну валідність оцінюють за допомогою коефіцієнта кореляції між результатами тестування й певним зовнішнім критерієм (зовнішньою змінною), що характеризуватиме вимірювану властивість у майбутньому. Основна складність визначення такої валідності – вибір вагомого зовнішнього критерію (зовнішньої змінної) в майбутньому.

Валідність методу – це комплексна характеристика, яка визначається як параметрами засобів і процедури вимірювання, так і властивостями ознаки, яка досліджується. Отже, валідність методу – це відповідність тому, що вимірюється даним методом, тому, що він повинен вимірювати. Цей критерій установлює сферу дійсності, для якої метод дає статистично ймовірні результати. Якщо мова йде про тестування діяльності певного рівня, то кваліфікаційні завдання, що пропонуються в тесті, повинні відповідати саме такому рівню складності, тобто їх не можна виконати засобами діяльності більш низького рівня. При цьому говорять про **функціональну валідність тесту**.

Валідність методу при вимірюванні успішності можна диференціювати за такими критеріями: валідність змісту; валідність відповідності; валідність прогнозу.

Вчені пропонують такі класифікації валідності.

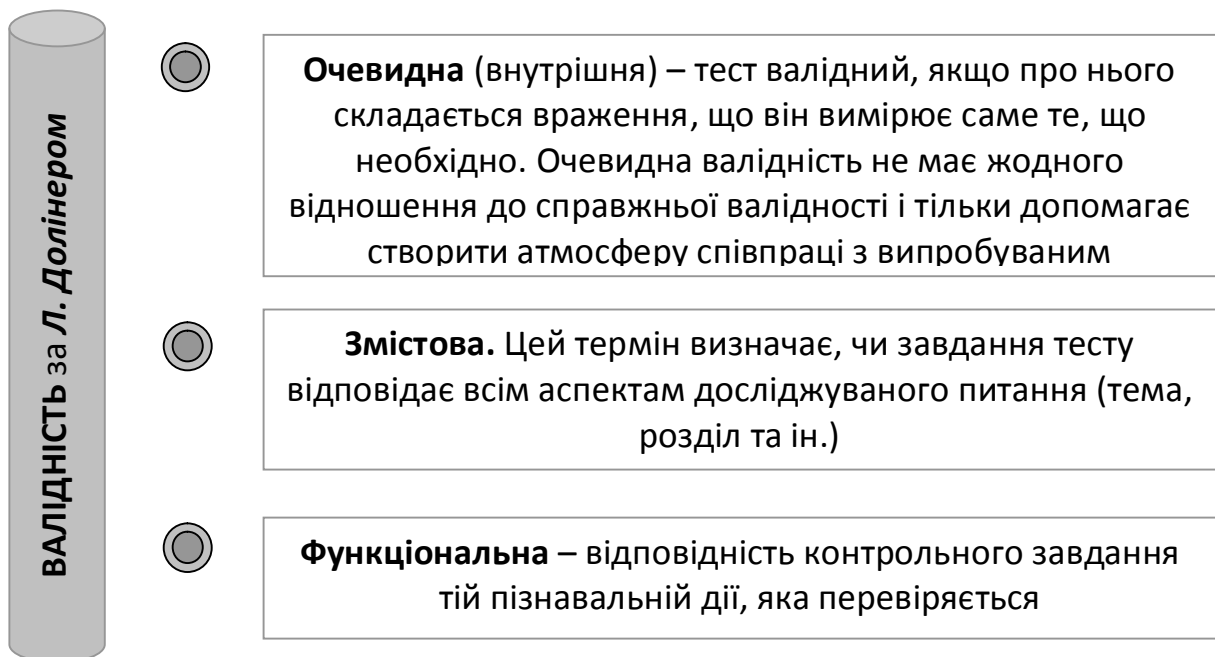


Рис. 6.2. Валідність за Л. Долінером

- **Очевидна** – тест валідний, якщо про нього в опитуваних складається враження, що він вимірює саме те, що необхідно
- **Конкурентна** – оцінюється за кореляцією результатів даного тесту з результатами інших тестів
- **Прогностична** – вивчаються кореляції між показниками тесту та деяким критерієм, що характеризує вимірювану властивість, але через певний час
- **Змістова** – якщо можна показати, що завдання відображають усі аспекти досліджуваної сфери за умови, коли інструкція до нього чітко сформульована
- **Конструктивна** – включає в себе всі підходи до визначення валідності, перераховані вище

Рис. 6.3. Валідність за П. Клайном

Надійність методу вимірювання – це міра стійкості результатів, що впливає на точність, з якою можна виміряти ту чи іншу конкретну ознаку. Перевірка надійності методу стосується насамперед відновлення результатів при повторних вимірюваннях.



Рис. 6.4. Критерії надійності методу тестування

Під **надійністю** контрольного завдання розуміють ступінь точності, з якою може бути визначена та чи інша ознака. Тобто визначено, наскільки можна довіряти результатам даного тесту. Про надійність тестів іноді роблять висновки за такою ознакою: якщо в усіх випадках перевірки тесту чи його варіантів

виявиться, що учні в розподілі за показниками успішності займуть ті самі місця, то такий тест можна вважати надійним.

Надійність тесту залежить від кількості тестових завдань. Тому для достатньої надійності підсумкового контролю тест має містити не менше ніж 40 завдань.

П. Клайн вважає, що у психометрії термін «надійність» має два значення. Тест називається надійним, якщо він є внутрішньо узгодженим. Тест також називається надійним, якщо він дає одні й ті ж показники для кожного тестованого (за умови, що тестований не змінився) при повторному тестуванні. Надійність при повторному тестуванні через певний час називають **ретестовою надійністю**.

В. Аванесов вважає, що в міру становлення нової теорії тестів починає відчуватись необхідність розширення кількості критеріїв хоча б до трьох. В якості третього педагог пропонує критерій ефективності.

Ефективність – порівняльний критерій, який дозволяє порівняти тести. Ефективним можна назвати тест, який краще, ніж інші тести, вимірює знання учнів потрібного рівня підготовки з меншою кількістю завдань, якісніше, швидше, дешевше, і все це, – якщо можливо, одночасно.

Якщо порівняти поняття "ефективність" із поняттями "надійність" і "валідність", то найістотніша відмінність нового поняття від двох традиційних полягає в переході від середнього показника до диференційованого. Достатньо пригадати, що надійність стосується тесту, який складається із зафіксованої кількості завдань, що пред'являються всім випробовуваним; тільки тоді можна знайти коефіцієнт надійності тесту як середню міру точності вимірювання. Аналогічно це стосується й валідності тесту. Ефективний же тест, навпаки, припускає відхід від фіксованої для всіх випробовуваних кількості завдань.

Психологи Л. Бурлачук і С. Морозов до ефективних тестів висувають іще одну істотну вимогу – складність завдання тесту.

Складність завдання тесту – характеристика завдання тесту, що відображає статистичний рівень розв'язання в даній вибірці проблеми стандартизації. Показником складності

тестового завдання є частка вибірки випробуваних, які виконали чи не виконали дане завдання. Наприклад, якщо лише 20 % випробуваних виконали завдання, його можна вважати складним для даної вибірки, якщо 80 % – легким.

Добір завдань за показниками складності важливий для успішного використання тестів. Так, при доборі надто складних тестових завдань валідність і надійність тесту різко зменшуються. Надто прості тестові завдання призведуть до одноманітності тесту та його неефективності.

Ще однією особливістю ефективних тестів, на думку П. Клайна, є **дискримінативність**. Досягнення задовільного розподілу показників є однією з цілей автора тестів. Нема необхідності підкреслювати те, що є очевидним: яка цінність тесту, за яким усі випробувані показали однаковий результат? За допомогою ретельного конструювання тесту можна забезпечити відповідний рівень дискримінативності, а це саме те, у чому тести виграють порівняно з іншими формами випробувань.

Як видно, вимоги, які ставлять тестологи до ефективних тестів, між собою дещо відрізняються. Мета вимірювань завжди полягає в отриманні відомостей про ознаки об'єктів чи подій. Вимірюється не сам об'єкт, а тільки властивості або відмінні ознаки об'єкта.

У педагогічній діагностиці намагаються визначити якість результатів вимірювання (тут – вимірювання успішності). Варто зазначити, що всі вищезазвані критерії спрямовані на єдину загальну вимогу: тест повинен давати якнайточнішу відомості про досліджуване явище. Вони лише конкретизують цю вимогу й характеризують її з різних боків.

Чинники, що знижують валідність результатів оцінювання:

1. Завдання не відповідають цілям тестування.
2. Мала кількість завдань, яка не охоплює усього змісту навчальної дисципліни або теми, що перевіряється.
3. Завдання є неоднозначними за змістом, некоректно сформульовані, у тесті містяться ключі до відповідей тощо.
4. У тесті вміщено завдання без урахування їх складності.

5. Завдання в тесті неправильно упорядковані (не враховано розташування завдань за складністю, завдання не згруповано за формою тощо).
6. Наявність нечітких або неповних інструкцій, що перешкоджає тестованим правильно зрозуміти, як треба працювати з тестом.
7. Суб'єктивність під час оцінювання завдань при відповіді або помилки при обчисленні балів у завданнях на вибір відповіді.
8. Не враховано час, необхідний для виконання тесту.
9. Порушення процедур проведення тестування, незадовільні умови контролю тощо.

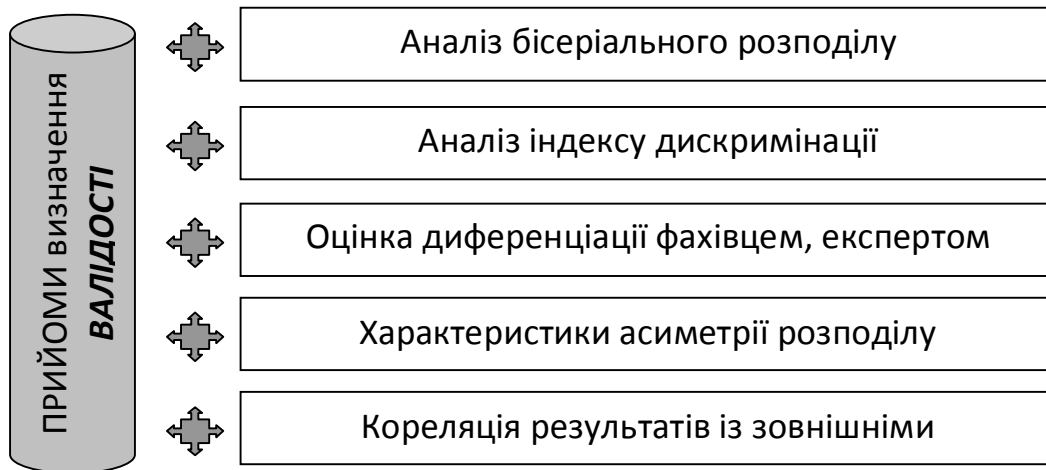


Рис. 6.5. Прийоми визначення валідності методу тестування

Ще один критерій методу вимірювання – **об'єктивність**. Він показує, наскільки мінімізовано вплив суб'єктивних чинників. Об'єктивність досягається шляхом стандартизації умов проведення вимірювання, умов оцінювання, умов аналізу результатів і забезпечується за допомогою об'єктивності проведення вимірювання, опрацювання даних та 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.

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 зіставлення 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 (*формула для коригування надійності Спірмена-Брауна* – формула, яку використовують для визначення коефіцієнта надійності тесту після застосування методу внутрішньої послідовності. Тест поділяють навпіл, розраховують надійність, наприклад, парних завдань. Знаючи надійність половини тесту, за формулою Спірмена-Брауна можна визначити надійність усього тесту. Формула має вигляд: коефіцієнт показує, наскільки підвищується надійність тесту, якщо збільшити його обсяг) 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. 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 в 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.

6.2. Статистичні характеристики тесту

Одним із останніх етапів конструювання тестів є математико-статистичне опрацювання результатів тестування.

Показники, які найчастіше використовують під час аналізу це рівень складності завдань, дисперсія, показники варіації тестових балів і рівня складності завдань (середнє квадратичне відхилення, коефіцієнт варіації), показники диференційної здатності завдань (коефіцієнт дискриміна- тивності, точково-бісеріальний коефіцієнт кореляції, бісері- альний коефіцієнт кореляції).

Коефіцієнт кореляції – характеристика ступеня взаємозв'язку між двома співзалежними ознаками (наборами тестових балів чи іншими видами оцінювання). Набуває значень від + 1 до – 1.

Додатне значення коефіцієнта кореляції свідчить про прямий зв'язок між ознаками, а від'ємне про зворотний. Чим ближче коефіцієнт кореляції до одиниці, тим щільніший зв'язок.

Дисперсія – показник варіації кількісних ознак і середній квадрат відхилень окремих значень ознаки (бала кожного тестованого) від її середнього значення (середнього арифметичного бала всіх тестованих).

Розглянемо ще деякі необхідні поняття математичної статистики і теорії ймовірностей.

Нехай дано деякий статистичний ряд з елементів x_1, x_2, \dots, x_n . Якщо ці елементи можуть мати всі допустимі значення, а об'єкт з цими характеристиками розглядається як система, то така сукупність називається *генеральною сукупністю*. Часто при цьому припускаємо, що вона скінченна і впорядкована за зростанням: $x_1 < x_2 < \dots < x_n$.

Будь-яка непорожня підмножина генеральної сукупності називається *вибіркою*. Якщо елементи вибірки вибрані випадково, то вона називається *випадковою вибіркою*.

Середня величина генеральної сукупності в цілому називається *загальною середньою*.

Вона відображає спільні риси усієї сукупності. Середня величина для окремої вибірки називається *вибірковою середньою*. Вона відображає спільні риси групи.

Існують різні міри середніх величин. Найчастіше використовується середня арифметична характеристика:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Вона називається також *вибірковою середньою* або *емпіричною середньою*.

Середня гармонійна величина, як і середня арифметична, може бути простою і ваговою. Якщо всі вагові коефіцієнти рівні між собою, то можна використати середню гармонійну у вигляді

$$\bar{x}_{\text{гарм}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Середню квадратичну вагову величину вираховуємо за формулою:

$$x_{кв} = \sqrt{\frac{\sum_{i=1}^n x_i^2 \omega_i}{\sum_{i=1}^n \omega_i}}$$

Якщо ваговий коефіцієнт $\omega_i = 1$, для всіх $i = 1, 2, \dots, n$, то отримуємо просто *середнє квадратичне*. Ці величини характеризують «концентрацію» даних вибірки навколо середнього.

До середніх величин, які характеризують структурні зміни, належать мода і медіана.

Мода – значення елемента даної сукупності, яке найчастіше повторюється. Вона відповідає певному значенню ознаки.

Медіана – значення ознаки, яке поділяє елементи рангової вибірки на дві рівні частини. Це середина рангового ряду.

Подія – будь-яка підмножина вибіркового простору. Порожню подію позначають, як і в теорії множин, символом \emptyset . Подією можна вважати і весь вибіркового простір (універсальна подія).

Випробування – перевірка усіх можливих результатів події. Два випробування незалежні, якщо будь-яка подія, визначена на основі лише одного з них, не залежить від події, визначеної на основі іншої. Оскільки подія – це множина, то для неї повинні виконуватись основні операції з множинами: об'єднання, перетин і доповнення.

6.3. Класичні статистичні методи аналізу результатів тестування

Історично виділяють два основні підходи до створення тестів. Перший з них набув широкого розвитку в рамках класичної теорії тестів. Згідно з ними, рівень знань учасників тестування оцінюється за допомогою їх індивідуальних балів. Бал обчислюють як алгебраїчну суму оцінок виконання кожного завдання тесту.

Класична теорія тестів ґрунтується на статистичних методах аналізу результатів тестування. Розглянемо найпростіші і необхідні процедури статистичного опрацювання результатів тестування знань і методи оцінки якості тесту.

В усіх відомих теоріях тестування розглядається як процес протистояння учасника із запропонованими йому завданнями. Позначимо через x_{ij} кількісну оцінку успішності виконання j -ого завдання i -им студентом. Результати тестування звичайно представляються у вигляді матриці $\{x_{ij}\}$ з n рядками та m стовпцями ($i = \overline{1, n}, j = \overline{1, m}$). Матриця тестових результатів показує результат виконання всіх завдань учасниками тестування. На практиці прийнято, як правило, використовувати дихотомічну шкалу оцінок результатів. У результаті правильного виконання завдання студент отримує один бал, $x_{ij} = 1$, в протилежному випадку – нуль балів, $x_{ij} = 0$.

Якщо за правильне виконання завдання студент отримує одиницю, а за неправильне – нуль, то бал виражає кількість правильно виконаних завдань. Результат можна оцінювати не лише нулем чи одиницею, але й присвоювати певний ваговий коефіцієнт, що відповідає складності завдання.

Процес статистичного опрацювання матриці результатів тестування будемо розглядати послідовно. На першому кроці обчислюємо індивідуальні початкові бали всіх студентів

$y_i, i = \overline{1, n}$, $y_i = \sum_{j=1}^m x_{ij}$ результат (індивідуальний бал) i -го студента після проходження тесту (кількість усіх правильних відповідей).

Обчислюємо середній результат \bar{y} сумарних балів учасників тестування та середній результат \bar{x}_j студентів за кожним завданням:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}, \quad \bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}.$$

Важливою вимогою до тестових завдань є їх об'єктивний рівень складності. В тесті немає місця завданням з невідомою

мірою складності. Завдання стають тестовими лише після емпіричної перевірки міри їх складності.

Складність завдань можна визначати двома способами :

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

– на основі емпіричної перевірки завдань, з підрахунком частки неправильних відповідей. У класичній теорії тестів багато років розглядалися тільки емпіричні показники складності. У сучасних теоріях навчальних тестів, які використовуються в дистанційному навчанні, більше уваги приділяється характеру розумової діяльності у процесі виконання тестових завдань різних форм.

Емпірично складність завдання визначається додаванням елементів матриці по рядках і дорівнює числу правильних відповідей, отриманих за кожним завданням (R_j). Чим більше правильних відповідей на завдання, тим воно легше для даної групи студентів.

У міру простоти показник R_j зручний, але доти, поки не з'являться інші групи з іншою кількістю студентів. Тому для одержання об'єктивних характеристик R_j ділять на кількість студентів у кожній групі (обсяг вибірки):

$$p_j = \frac{R_j}{n}.$$

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

$$q_j = \frac{W_j}{n}, p_j + q_j = 1$$

Наступною вимогою до тестових завдань є варіація балів.

Якщо на деяке завдання правильно відповідають всі студенти, то таке завдання стає не тестовим. Учасники тестування відповідають на нього однаково: між ними немає варіації. Відповідно з даного завдання в матриці будуть стояти лише одиниці. Нетестовим вважається завдання, на яке немає жодної правильної відповіді. Варіація по ньому теж дорівнює нулеві. Нульова варіація означає практичну потребу вилучення завдання із тесту.

Зручною мірою варіації є значення дисперсії s_y^2 і стандартне відхилення s_y сумарних балів учасників тестування:

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

та величина s_j^2 дисперсія результатів студентів по j -ому завданню:

$$s_j^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}, \quad j = \overline{1, m}$$

Якщо успішність виконання j -ого завдання оцінюється балами 0 чи 1, то міра варіації визначається формулою $s_j^2 = p_j(1-p_j)$ або $s_j^2 = p_j \cdot q_j$

Обчисливши дисперсію, можна знайти і стандартне відхилення $s_j = \sqrt{s_j^2}$

Завдання в тестовій формі не можна назвати тестовим, якщо воно не корелює із сумою балів по всьому тесту. Для цього можна використовувати коефіцієнт кореляції Пірсона:

$$R_j = \frac{\frac{\sum_{i=1}^n (x_{ij} \times y_i)}{n} - \bar{x}_j \times \bar{y}}{s_j \times s_y} \times \frac{n}{n-1}$$

або бісеріальний коефіцієнт кореляції:

$$B_j = \frac{M_{j1} - M_{j0}}{s_y} \cdot \sqrt{\frac{n_{j0} \cdot n_{j1}}{n(n-1)}}$$

де $n_{j1} = \sum_{i=1}^n x_{ij}$ – кількість тих студентів, що одержали за даним завданням 1 бал;

$n_{j0} = n - n_{j1}$ – кількість тих студентів, що відповіли неправильно на j -е завдання;

M_{j1} – середнє арифметичне сум балів по всьому тесту для тих студентів, які одержали за даним завданням 1 бал, M_{j0} – нуль балів:

$$M_{j0} = \frac{\sum_{i=1}^n (1 - x_{ij}) y_i}{n_{j0}}, \quad M_{j1} = \frac{\sum_{i=1}^n x_{ij} y_i}{n_{j1}}$$

Попарний кореляційний зв'язок завдань між собою можна обчислити за формулою

$$\varphi_{jk} = \frac{AD - BC}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

$$A = \sum_{i=1}^n x_{ij} x_{ik}$$

де A – кількість учасників тестування, які вірно виконали завдання j та k ;

$$B = \sum_{i=1}^n x_{ij} (1 - x_{ik})$$

вірно виконали завдання j та невірно k .

$$C = \sum_{i=1}^n (1 - x_{ij}) x_{ik}, \quad D = \sum_{i=1}^n (1 - x_{ij})(1 - x_{ik})$$

Із збірника завдань викидаємо завдання, що не володіють дискримінативністю $p_j > 0,9$ (надто легкі), $p_j < 0,2$ (надто важкі). Виключають завдання, що недостатньо корелюють із сумою балів ($B_j < 0,15$) і негативні коефіцієнти кореляції. Для зменшеного списку завдань складається нова впорядкована таблиця, для якої перераховуються вищезгадані показники.

Крім того, отримані тестові завдання повинні задовольняти критерій надійності та валідності.

Надійність тесту тим вища, чим більше погоджені результати учасника тестування при повторній перевірці знань за допомогою того самого тесту. Погодженість можна вимірювати коефіцієнтом надійності Кьюдера-Річардсона:

$$\rho = \frac{s_y^2 - \sum_{j=1}^m s_j^2}{2s_y^2} + \sqrt{\left(\frac{s_y^2 - \sum_{j=1}^m s_j^2}{2s_y^2} \right)^2 + \frac{\sum_{j=1}^m B_j^2 s_j^2}{2s_y^2}}$$

Якщо $0,8 \leq \rho \leq 0,89$ тест має високу надійність, якщо $\rho \geq 0,9$ надійність дуже висока. Чим вищий показник надійності, тим менша помилка виміру індивідуального результату.

Валідність тесту показує, наскільки якісно робить тест те, для чого він був створений.

Визначити коефіцієнт валідності тесту – означає визначити, як результати тестування співвідносяться з результатами інших незалежних методів оцінювання. Для визначення валідності необхідним є незалежний зовнішній критерій, тобто оцінка експерта (викладача). За коефіцієнт валідності приймають коефіцієнт кореляції результатів тестових вимірів і критерію. Якщо експертна оцінка знань студентів, отримана незалежно від процедури тестування, представлена числовою послідовністю Y_1, Y_2, \dots, Y_n , то коефіцієнт валідності тесту може бути обчислений так:

$$V = \frac{\frac{\sum_{i=1}^n (Y_i \cdot y_i)}{n} - \bar{Y} \cdot \bar{y}}{s_Y \cdot s_y} \cdot \frac{n}{n-1},$$

де $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$, $s_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}}$ – стандартне відхилення експертних оцінок.

6.4. Зразки задач на визначення статистичних характеристик результатів тестування

Наведемо ряд задач, які найчастіше розв'язуються і зустрічаються при підготовці до тестування і математичному опрацюванні результатів тестування, а також алгоритми їх розв'язання. При цьому ми не вдаватимемося до обґрунтування використаних математичних фактів як інструментарію.

Задача 1. Визначення основних статистичних показників тестування.

Нехай тепер дано результати тестування групи, яка складається з n тестованих для заданого тесту з m різних запитань. Зазвичай ці дані представлено у вигляді деякої матриці A балів (на зразок "тестований – завдання") розмірності n на m :

$$A(n \times m) = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{vmatrix} = \left\| a_{ij} \right\|_{\substack{j=\overline{1,m} \\ i=\overline{1,n}}}$$

Елемент a_{ij} матриці A – результат виконання j -го завдання для i -го тестованого.

Необхідно на основі результатів x_1, x_2, \dots, x_n тестування для кожного з n тестованих вирахувати основні статистичні показники тестування (оцінити «сирі» результати) для випадково обраної групи тестованих.

Алгоритм розв'язання:

1. Впорядковуємо ряд за зростанням (знаходимо генеральну сукупність): $x_1 < x_2 < \dots < x_n$.
2. Вибираємо підмножину тестованих, яка нас цікавить (вибірку).
3. Знаходимо середнє арифметичне за вибіркою:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

4. Знаходимо середню гармонійну величину вибірки:

$$\bar{x}_{\text{гарм}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

5. Знаходимо величини, що характеризують структурні зміни, наприклад, моду і медіану. Для даних, що мають "хорошу поведінку", медіана завжди лежить в проміжку між середнім арифметичним і модою. Ці величини розташовуються за зростанням таким чином: середнє, медіана, мода, або ж у зворотному порядку. Прямий або зворотний порядок їх розташування можна визначити, обчисливши так званий *коефіцієнт асиметрії*:

$$K = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right)^3}$$

Цей коефіцієнт відображає відносну зміну даних.

6. Знаходимо міри розсіювання, розкидання або варіації, що показують, як останні елементи сукупності (вибірки) групуються біля середніх величин.

Наприклад,

1) розмах:

$$r_x = x_{\max} - x_{\min} = x_n - x_1$$

1) середнє абсолютне відхилення:

$$x_{сер} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

2) середньоквадратичне відхилення:

$$\delta_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} (\sum_{i=1}^n x_i^2 - nx^{-2})$$

3) дисперсія:

$$D = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} (x_i^2 - nx^{-2})$$

4) стандартне відхилення:

$$\sigma = \sqrt{D}$$

6) коефіцієнт варіації:

$$v = \frac{\sigma}{\bar{x}} ;$$

7) кінець алгоритму.

Задача 2. Визначення «ваги» (міри складності) тестового завдання.

Дано результати тестування для кожного з n тестованих і тесту довжини m у вигляді матриці, а також вектор еталонних відповідей $B = (b_1, b_2, \dots, b_m)$, де b_j – еталонна відповідь на завдання номер j . Необхідно визначити "вагу" (міру складності) конкретного завдання тесту.

Алгоритм розв'язання:

1. Для чергового завдання тесту за матрицею A визначаємо кількість тестованих, які правильно відповіли на поставлене завдання.
2. За «вагу» завдання беремо дріб c_j :

$$c_j = \frac{\text{кількість тестованих}}{\text{кількість тестованих, які правильно відповіли на всі завдання}}$$

3. Обчислюємо суміжні ваги d_i :

$\frac{\text{кількість тестованих, які неправильно відповіли на завдання } j}{\text{кількість тестованих, які неправильно відповіли на всі завдання}}$

Інколи в знаменник підставляємо кількість всіх тестованих.

4. Знаходимо вектор ваги виконання $c = (c_1, c_2, \dots, c_m)$ для заданого вектора b еталонних відповідей.
5. Знаходимо вектор ваги невиконання $d = (d_1, d_2, \dots, d_m)$ для заданого вектора b еталонних відповідей.
6. Оцінюємо дисперсію кожного j -го завдання $D_j = c_i d_i$ і стандартне відхилення $\sigma_j = \sqrt{D_j}$.
7. Кінець алгоритму.

Задача 3. Оцінювання коефіцієнта кореляції (валідності) кожного тестового завдання.

Дано результати тестування для кожного з n тестованих і тесту довжини m у вигляді матриці A , а також вектор еталонних відповідей $B = (b_1, b_2, \dots, b_m)$, де b_j – еталонна відповідь на завдання номер j . Необхідно оцінити валідність кожного тестового завдання.

Алгоритм розв'язання:

1. Для чергового завдання тесту за матрицею A визначаємо кількість тестованих, які правильно відповіли на завдання j і знаходимо їх середній бал x_j .
2. Аналогічно знаходимо кількість тестованих, які неправильно відповіли на j завдання і їх середній бал y_j .
3. Знаходимо дріб c_j :
$$\frac{\text{кількість тестованих}}{\text{кількість тестованих, які правильно відповіли на завдання } j}$$
4. Знаходимо дріб d_i :
$$\frac{\text{кількість тестованих}}{\text{кількість тестованих, які неправильно відповіли на завдання } j}$$

5. Оцінюємо дисперсію кожного j завдання $D_j = c_i d_i$ і стандартне відхилення $\sigma_j = \sqrt{D_j}$
6. Знаходимо стандартне відхилення $\sigma = \sqrt{D}$ у всьому тесті.
7. Знаходимо коефіцієнт кореляції (міру валідності завдання):

$$r_j = \frac{(x_j - y_j)\sigma_j}{\sigma}$$
8. Якщо $r_j > 0,3$, то завдання можемо вважати валідним, інакше – не валідним (відзначимо, що з точки зору критеріальної валідності завдання, виконані всіма або невиконані ніким, – невалідні).
9. Кінець алгоритму.

Задача 4. Оцінювання надійності тестів.

Дано результати нормативно орієнтованого тесту для кожного з n тестованих і тесту довжини m у вигляді матриці A , а також вектор еталонних відповідей $B = (b_1, b_2, \dots, b_m)$, де b_j – еталонна відповідь на завдання номер j . Необхідно оцінити надійність тесту (ступінь стійкості результатів тестування кожного тестованого за умови, що тестування було проведено в однакових умовах).

Для того щоб вирахувати надійність нормативно орієнтованого тесту, використовуємо коефіцієнт кореляції між результатами двох паралельних тестів. Порівнюючи коефіцієнти кореляції, робимо висновок про надійність (внутрішню надійність) тесту. Якщо дві половини тесту корельовані, то і тест надійний; якщо ж навпаки – не надійний.

Алгоритм розв'язання:

1. Ділимо тест на дві рівні частини X і Y , наприклад, за парними і непарними номерами завдань. Цей метод називається методом розщеплення тесту. Таким чином, ми отримуємо дані за двома паралельними тестами X і Y –

індивідуальні бали $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n$, де n – кількість тестованих.

2. Для кожного завдання групи X виконуємо попередній алгоритм.
3. Для кожного завдання групи Y виконуємо попередній алгоритм.
4. Знаходимо коефіцієнт кореляції X і Y за формулою

$$r_{XY} = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i\right)^2} \cdot \sqrt{\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i\right)^2}}$$

5. Знаходимо надійність r усього тесту за формулою (Спірмена-Браумена):

$$r = \frac{2r_{XY}}{1 + r_{XY}}$$

6. Кінець алгоритму.

Задача 5. Задача класифікації

Нам необхідно на основі отриманих результатів (матриця A) для кожного із n тестованих отримати інтегральний (узагальнений) показник виконання тесту довжини m , а тоді за отриманими значеннями цього інтегрального показника розбити усіх тестованих на k груп (завдання класифікації).

Алгоритм розв'язання:

1. Якщо для j -го завдання збільшення значень результатів вимірювань свідчить про поліпшення відповідної властивості, то з ним пов'яжемо ознаку $z_j = 1$, а якщо свідчить про погіршення – ознаку $z_j = -1$
2. Виконуємо нормування елементів вихідної матриці так, щоб у кожному стовпці вони змінювалися в "одному напрямі": для кожного завдання (при фіксованому

$j = 1, 2, \dots, m$) і для кожного тестованого $i = 1, 2, \dots, n$ обчислимо нове значення:

$$a_{ij} := \frac{a_{ij} - m_j}{M_j - m_j}$$

де M_j, m_j , – найбільше і найменше значення елементів j -го стовпця і застосовуємо перетворення типу

$$a_{ij} := \begin{cases} a_{ij}, z_j = 1 \\ 1 - a_{ij}, z_{ij} = -1 \end{cases}$$

3. Для кожного стовпця отриманої нової матриці A (нормованої) вираховуємо середнє квадратичне відхилення за формулою

$$c_i = \sqrt{\frac{\sum_{i=1}^n (a_{ij} - \bar{a}_j)^2}{n-1}}$$

де \bar{a}_j – середнє арифметичне стовпця j .

4. Вираховуємо класифікаційний інтегральний показник

$$y_i = \sum_{j=1}^m a_{ij} c_j$$

де y_i – значення інтегрального показника для i -го учня $i = 1, 2, \dots, n$, c_j – ваговий коефіцієнт j -го завдання у тесті чи в

банку всіх завдань, a_{ij} – елемент матриці A чи його перетворення (нормоване, наприклад, стосовно максимального елемента або до норми матриці).

5. Знаходимо найменше Y_{\min} і найбільше Y_{\max} значення інтегрального показника (за кількістю усіх тестованих).

Відрізок $[Y_{\min}; Y_{\max}]$ поділяємо на задане число k інтервалів.

Часто беруть (при побудові, наприклад, гістограми) $k = 1 + 3,2 \cdot \lg n$. Усіх тестованих, для яких обчислені

значення інтегрального показника потрапляють в один і той самий інтервал, ототожнюємо і відносимо до одного класу.

2. Представляємо результати: значення інтегрального показника для кожного тестованого, а також його клас (або класифікацію тестованих за інтегральним показником).
3. Кінець алгоритму.

Задача 6. Формування груп за інтегральними показниками.

Дано інтегральну норму Y тестових результатів. Необхідно сформулювати з однієї кілька груп тестованих за їх інтегральними показниками (щодо норми).

Алгоритм розв'язання:

1. Введемо вхідні дані: n, m, A .
2. Для кожного тестованого визначаємо сумарний бал:

$$b_i = \sum_{j=1}^m a_{ij}$$

3. Поділяємо усю вибірку тестованих на три групи.
Перша група – тестовані, які набрали високі бали. Найнижчий сумарний бал того, хто потрапляє в цю групу, дорівнює

$$\min \{b_i\} + b(\max \{b_i\} - b_i).$$

- Друга група* – тестовані, які набрали середні бали. Найбільший сумарний бал того, хто потрапляє в цю групу, дорівнює

$$\min |b_i| + (1 - b)(\max |b_i| - b_i),$$

де b – коефіцієнт, $b > 0,5$.

4. Кінець алгоритму.

6.5. Визначення складності тесту засобами електронних таблиць Microsoft Excel

Крок 1. Формування матриці результатів тесту

Результати відповідей тестованих на завдання тестів оцінюються в дихотомічній шкалі: за кожну правильну відповідь тестований отримує 1 бал, а за неправильну або відсутню відповідь – 0 балів.

	A	B	C	D	E	F	G	H	I	J	K
1	№	Номер завдання									
2	тестованого	1	2	3	4	5	6	7	8	9	10
3	1	1	1	1	1	1	1	0	0	0	0
4	2	1	1	0	0	0	0	0	0	0	0
5	3	0	0	0	0	0	0	0	1	0	0
6	4	1	1	0	1	1	1	1	1	1	1
7	5	1	0	1	0	1	1	0	0	0	0
8	6	1	1	1	0	0	0	0	1	0	0
9	7	1	1	1	1	0	1	0	0	0	0
10	8	1	1	1	1	0	0	0	0	0	0
11	9	1	1	1	1	1	1	1	1	1	0
12	10	1	1	1	1	1	0	1	0	0	0
13	11	0	0	0	0	0	0	0	0	0	0
14	12	1	1	1	1	1	1	1	1	1	1
15											

Рис. 6.6. Матриця результатів тестування

Крок 2. Перетворення матриці результатів тесту

На другому кроці з матриці тестових результатів вилучаємо рядки і стовпці, які містять лише нулі або одиниці. У наведеному прикладі таких стовпців немає, а рядків лише два. Один із них, нульовий, відповідає відповідям одинадцятого тестованого, який не зміг правильно виконати жодного завдання тесту.

У цьому випадку висновок такий: даний тест непридатний для оцінювання знань тестованого. Для виявлення його рівня знань тест необхідно полегшити кількома завданнями, які, скоріш за все, виконає правильно більшість інших студентів / учнів.

Також непридатним, але з іншої причини є тест для оцінювання знань 12 тестованого, який виконав правильно усі без виключення завдання тесту. Причина непридатності полягає в його надмірній легкості (простоті), яка не дозволяє виявити реальний рівень знань дванадцятого учня / студента.

Таким чином, на даному етапі необхідно видалити із матриці даних 11 та 12 рядки.

	A	B	C	D	E	F	G	H	I	J	K
1	№	Номер завдання									
2	тестованого	1	2	3	4	5	6	7	8	9	10
3	1	1	1	1	1	1	1	0	0	0	0
4	2	1	1	0	0	0	0	0	0	0	0
5	3	0	0	0	0	0	0	0	1	0	0
6	4	1	1	0	1	1	1	1	1	1	1
7	5	1	0	1	0	1	1	0	0	0	0
8	6	1	1	1	0	0	0	0	1	0	0
9	7	1	1	1	1	0	1	0	0	0	0
10	8	1	1	1	1	0	0	0	0	0	0
11	9	1	1	1	1	1	1	1	1	1	0
12	10	1	1	1	1	1	0	1	0	0	0
13											

Рис. 6.7. Матриця результатів тестування після перетворення
Крок 3. Підрахунок індивідуальних балів тестованих та кількість відповідей на кожне завдання тесту

Індивідуальний бал тестованого отримуємо, підсумувавши всі бали, отримані за виконання завдань тесту.

Для зручності індивідуальні бали (x_i) поміщаємо в останній стовпець матриці результатів.

Кількість правильних відповідей на завдання тесту (y_i) також отримуємо обчисливши суму балів, розміщених у стовпцях.

	A	B	C	D	E	F	G	H	I	J	K	L
1	№	Номер завдання										Індивідуальні бали (X)
2	тестованого	1	2	3	4	5	6	7	8	9	10	
3	1	1	1	1	1	1	1	0	0	0	0	6
4	2	1	1	0	0	0	0	0	0	0	0	2
5	3	0	0	0	0	0	0	0	1	0	0	1
6	4	1	1	0	1	1	1	1	1	1	1	9
7	5	1	0	1	0	1	1	0	0	0	0	4
8	6	1	1	1	0	0	0	0	1	0	0	4
9	7	1	1	1	1	0	1	0	0	0	0	5
10	8	1	1	1	1	0	0	0	0	0	0	4
11	9	1	1	1	1	1	1	1	1	1	0	9
12	10	1	1	1	1	1	0	1	0	0	0	6
13	Кількість правильних відповідей (Y)	9	8	7	6	5	5	3	4	2	1	50
14												

Рис. 6.8. Матриця результатів із кінцевими сумами балів

Крок 4. Упорядкування матриці результатів

Значення індивідуальних балів необхідно відсортувати за зростанням. Для цього в MS Excel необхідно:

1. Виділити блок комірок, які містять номери тестованих, матрицю результатів та індивідуальні бали (діапазон комірок В3:Л12)
2. В меню «Данные» виберемо команду «Сортировка».

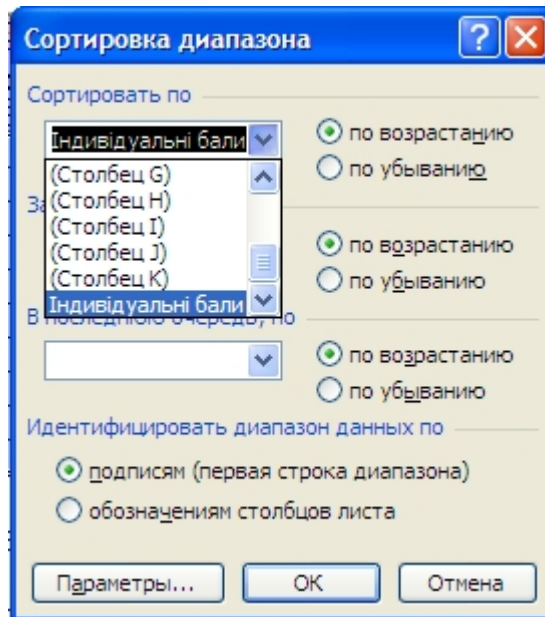


Рис. 6.9. Діалогове вікно «Сортировка»

В цьому діалоговому вікні необхідно вибрати стовпець для сортування (у нашому випадку – «Індивідуальні бали»). Матриця результатів набуде такого вигляду:

	A	B	C	D	E	F	G	H	I	J	K	L
1	№	Номер завдання										Індивідуальні бали (X)
2	тестованого	1	2	3	4	5	6	7	8	9	10	
3	3	0	0	0	0	0	0	0	1	0	0	1
4	2	1	1	0	0	0	0	0	0	0	0	2
5	5	1	0	1	0	1	1	0	0	0	0	4
6	6	1	1	1	0	0	0	0	1	0	0	4
7	8	1	1	1	1	0	0	0	0	0	0	4
8	7	1	1	1	1	0	1	0	0	0	0	5
9	1	1	1	1	1	1	1	0	0	0	0	6
10	10	1	1	1	1	1	0	1	0	0	0	6
11	4	1	1	0	1	1	1	1	1	1	1	9
12	9	1	1	1	1	1	1	1	1	1	0	9
13	Кількість правильних відповідей (Y)	9	8	7	6	5	5	3	4	2	1	50
14												

Рис. 6.10. Упорядкована матриця результатів тестування

Крок 5. Графічне представлення даних

Емпіричні результати тестування можна подати у вигляді полігону частот, гістограми або графіка.

Для побудови кривих упорядкуємо результати експерименту і підрахуємо частоту одержаних балів.

	А	В
1	№ тестованого	Індивідуальні бали (X)
2	3	1
3	2	2
4	5	4
5	6	4
6	8	4
7	7	5
8	1	6
9	10	6
10	4	9
11	9	9
12		

Рис. 6.11. Незгрупований ряд

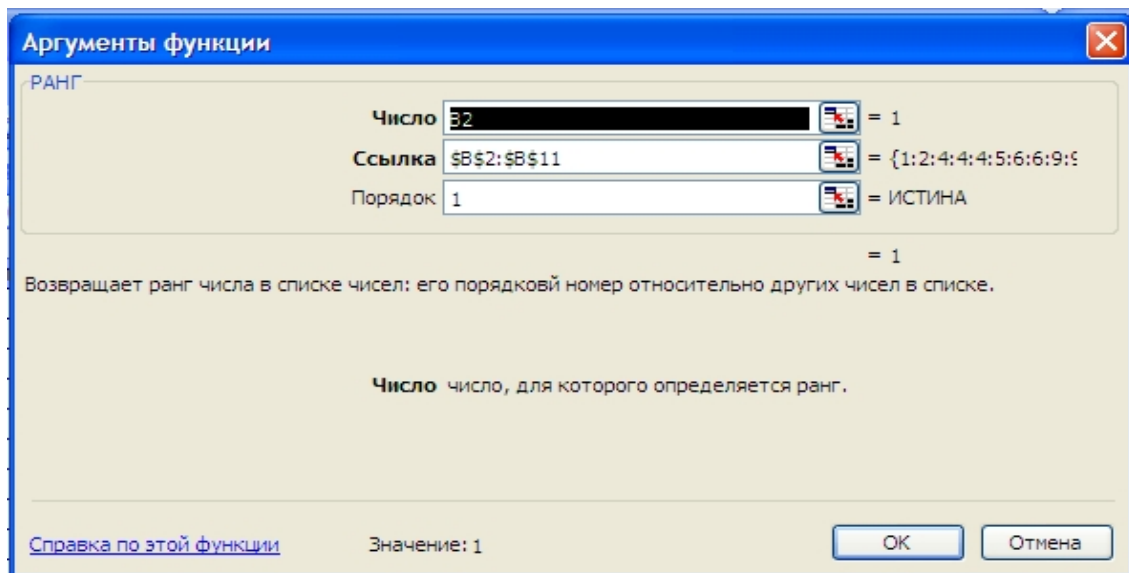


Рис. 6.12. Вікно функції РАНГ для комірки С2

Для підрахунку рейтингу (рангу) кожного тестованого за індивідуальними балами необхідно скористатися функцією РАНГ, яка визначає ранг числа зі списку. Ранг числа – це його величина відносно інших значень у списку.

Синтаксис функції РАНГ:

РАНГ (число; ссылка; порядок), де
число – посилання на комірку, для якої визначається ранг;
ссылка – посилання на діапазон індивідуальних балів (вибірка);
порядок – число, яке вказує на спосіб упорядкування. Якщо порядок дорівнює 0 (нулю), або не вказаний, то програма визначає ранг числа так, якщо б посилання

було списком, відсортованим за спаданням. Якщо порядок – будь яке ненульове значення, то ранг числа визначається так, якщо б посилення було списком, відсортованим за зростанням.

	A	B	C
1	№ тестованого	Індивідуальні бали (X)	Ранг
2	3	1	1
3	2	2	2
4	5	4	3
5	6	4	3
6	8	4	3
7	7	5	6
8	1	6	7
9	10	6	7
10	4	9	9
11	9	9	9

Рис. 6.13. Ранжований ряд

	A	B
1	Бал	Частота
2	1	1
3	2	1
4	4	3
5	5	1
6	6	2
7	9	2

Рис. 6.14. Таблиця частотного розподілу

За частотним розподілом можна побудувати гістограму.

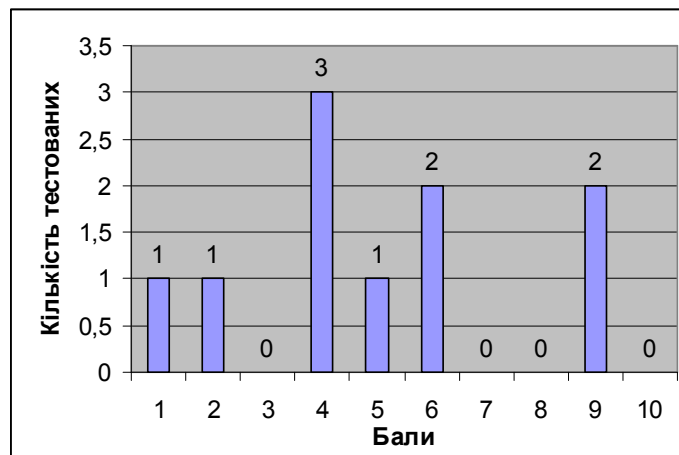


Рис. 6.15. Гістограма частот розподілу балів.

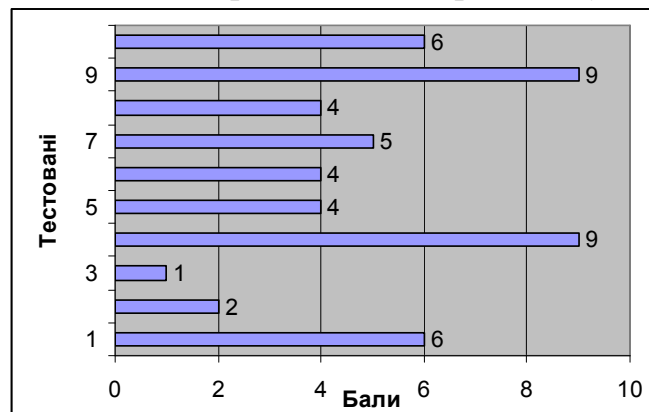
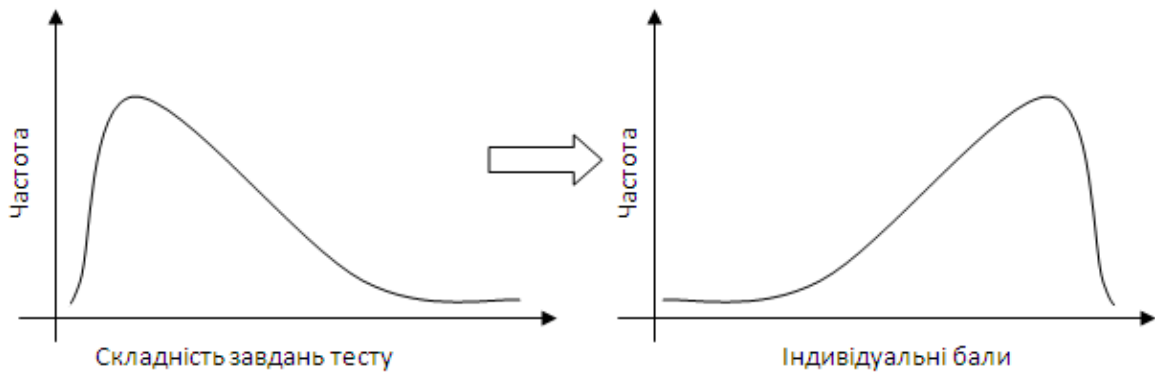


Рис. 6.16. Гістограма розподілу індивідуальних балів

При конструюванні тестів необхідно пам'ятати про те, що крива розподілу індивідуальних балів, отриманих із

репрезентативної вибірки, є результатом кривої розподілу складності завдань тесту.

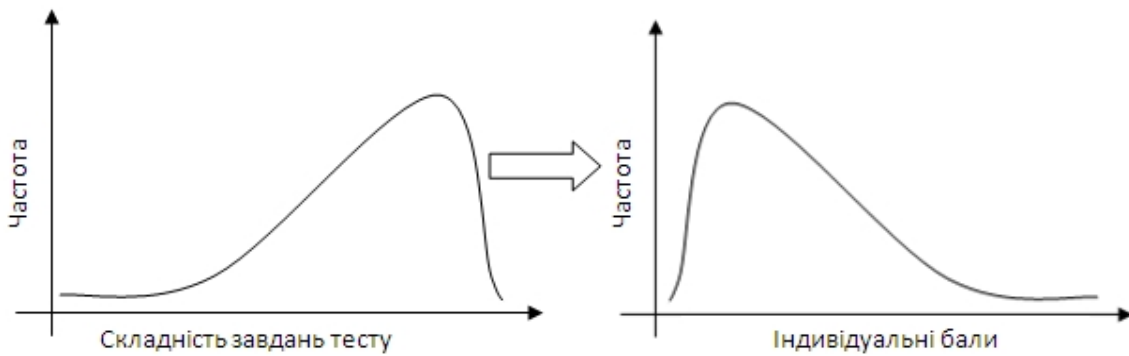
Проілюструємо це рисунками:



а) лівостороння асиметрія

б) правостороння асиметрія

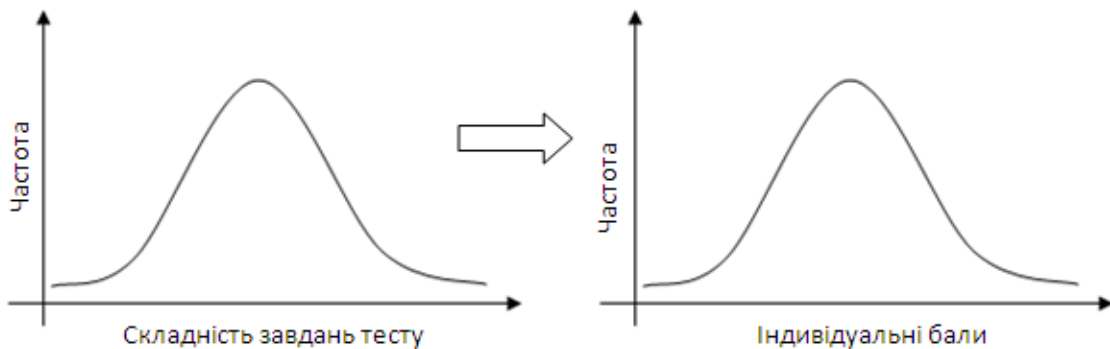
Рис. 6.17. Криві розподілу для легкого тесту



а) лівостороння асиметрія

б) правостороння асиметрія

Рис. 6.18. Крива розподілу для складного тесту



а)

б)

Рис. 6.19. Крива розподілу для тесту оптимальної складності

Для кривої на рис. 6.17 а) характерним є зміщення в тесті в бік легких завдань, що, безперечно, стане причиною

виникнення більшої кількості завищених балів у репрезентативній вибірці тестованих. Більшість тестованих виконає майже усі завдання тесту (рис. 6.17 б).

Рис. 6.18 а) відображає істотне зміщення в бік складних завдань. Це відображається на результатах тестованих, тому розподіл індивідуальних балів має яскраво виражений стрибок поблизу початку горизонтальної осі. Основна частина тестованих виконає мало найлегших завдань тесту (рис. 6.18 б).

На рис. 6.19 а) завдання тесту мають оптимальну складність, оскільки розподіл має вигляд кривої нормального розподілу. Звідси можна говорити про нормальний розподіл індивідуальних балів репрезентативної вибірки тестованих (рис. 6.19 б), що в свою чергу дозволяє вважати отриманий розподіл стійким відносно генеральної сукупності.

У професійно розроблених нормативно орієнтованих тестах типовим є результат, коли приблизно 70 % тестованих виконують правильно від 30 до 70% завдань тесту, а найчастіше зустрічається результат 50 %.

Крок 6. Визначення вибірових характеристик результатів

На даному етапі необхідно визначити середнє значення, моду, медіану, дисперсію, стандартне відхилення вибірки, асиметрію та ексцес.

Ступінь відхилення розподілу частот вибірки від симетричного розподілу, характерного для нормальної кривої, оцінюється за допомогою асиметрії. Наявність асиметрії легко встановити візуально, аналізуючи полігон частот або гістограму. Детальніше аналіз можна провести за допомогою узагальнених статистичних характеристик, призначених для оцінювання величини асиметрії в розподілі.

За допомогою функції СКОС можна обчислити асиметрію вибірки.

Синтаксис функції СКОС

СКОС (число 1; число 2),

де *число 1* – посилання на масив даних, які містять індивідуальні бали тестованих.

	А	В
1	№ тестованого	Індивідуальні бали (X)
2	1	6
3	2	2
4	3	1
5	4	9
6	5	4
7	6	4
8	7	5
9	8	4
10	9	9
11	10	6
12	Середнє значення	5
13	Мода	4
14	Медіана	4,5
15	Дисперсія	6,889
16	Стандартне відхилення	2,625
17	Асиметрія	0,277
18	Ексцес	-0,412

Рис. 6.20. Вибіркові характеристики результатів тестування

12	Середнє значення	=СРЗНАЧ(В2:В11)
13	Мода	=МОДА(В2:В11)
14	Медіана	=МЕДИАНА(В2:В11)
15	Дисперсія	=ДИСП(В2:В11)
16	Стандартне відхилення	=СТАНДОТКЛОН(В2:В11)
17	Асиметрія	=СКОС(В2:В11)
18	Ексцес	=ЭКЦЕСС(В2:В11)

Рис. 6.21. Формули для обчислення статистичних параметрів вибірки

При інтерпретації отриманого значення асиметрії 0,277 необхідно звернути увагу на те, що величина асиметрії додатна та невелика.

Асиметрія розподілу додатна, якщо основна частина значень індивідуальних балів лежить справа від середнього значення, що зазвичай характерно для надто легких завдань.

Асиметрія розподілу балів від'ємна, якщо більшість тестованих одержали оцінки нижчі середнього балу. Ефект від'ємної асиметрії діагностується у випадку занадто складних тестів, які неправильно збалансовані за складністю при доборі завдань. У добре збалансованому за складністю тесті розподіл балів має вигляд нормального розподілу. Для нормального розподілу характерною є нульова асиметрія, що зрозуміло, оскільки при повній симетрії кожне значення балу, меншого середнього значення, урівноважується іншим симетричним, більшим ніж середнє.

За допомогою ексцесу можна отримати уявлення про те, чи є функція розподілу частот гостроверхою, середньоверхою чи плоскою. (*Ексцес* – міра кривизни розподілу одиниць сукупності тестованих за значенням ознаки (бали тестів). Характеризує величину відхилення верхівки фактичного розподілу від верхівки нормального розподілу вгору (гостроверхий) чи донизу (плосковерхий)).

Для розрахунку даного параметра, використовують функцію ЕКСЦЕСС (число 1; число 2;...), де число 1 – посилання на масив даних, які містять індивідуальні бали тестованих.

У тому випадку, коли розподіл даних бімодальний (має дві моди), необхідно говорити про ексцес в околі кожної моди. Бімодальна конфігурація вказує на те, що за результатами виконання тесту вибірка тестованих розділилася на дві групи. Одна група виконала правильно легші завдання, а друга – складніші.

6.6. Комплексне оцінювання якості тестових завдань і тестів

Процес раціоналізації контролю та оцінювання якості навчальних досягнень ініціював масштабне застосування тестових технологій на всіх стадіях процесу навчання. Тестові технології, що застосовуються в системі вищої професійної освіти, покликані забезпечувати отримання оперативної і достовірної інформації про якість навчальних досягнень студентів. Але, незважаючи на вагомий перелік переваг тестових технологій, по цей день багато супротивників і прихильників впровадження тестових технологій у систему контролю та оцінювання якості навчання. При цьому в якості головної висувається думка про низьку якість тестових матеріалів. Сучасна дидактична тестологія вважає, що необ'єктивність тестів залежить в основному від низької якості їх розроблення – якщо вони зроблені нашвидку, не перевірені на надійність та валідність. Підвищення якості тестових матеріалів на стадії їх розроблення забезпечується їх експертизою і коригування після експертизи. Тестові завдання і

тести призначені для вимірювання властивостей сукупності досліджуваних об'єктів, що відносяться до багатьох потенційних випробуваних.

Тест як система складається із взаємопов'язаних тестових завдань, розташованих у певній послідовності. При експертизі якості тестових завдань і тесту в цілому необхідно оцінити кожен компонент структури тестового завдання не тільки окремо, але і в системі відносин з іншими тестовими завданнями. При цьому потрібно враховувати, що у кожного тестового завдання є ряд структурних елементів, кожен з яких характеризується низкою внутрішніх і зовнішніх властивостей. Кожен елемент тестового завдання має свою топологію, а властивості тестового завдання і тесту в цілому можуть визначатися особливостями безлічі випробуваних, показниками якості тестових завдань і тестів (далі: тестових вимірювачів – ТВ), особливостями технології опрацювання результатів тестування та показниками компетентності що беруть участь в експертизі фахівців. Таким чином, з огляду на складність структури ТВ і сукупності критеріїв, що визначають їхню якість, експертизу тесту і його елементів необхідно проводити на основі системного підходу. Метою експертизи якості ТВ є комплексне оцінювання їх якості.

При комплексному оцінюванні якості тестових завдань і тестів проводиться:

- оцінювання ступеня відповідності тестових завдань і тестів мінімуму змісту навчальної дисципліни, що визначається як державний компонент;
- оцінювання ступеня відповідності тестових завдань і тестів цілям вивчення даної навчальної дисципліни;
- оцінювання якості нормативних документів, що супроводжують розробку і застосування тестових завдань і тестів – специфікації, кодифікатор та ін.;
- оцінювання відповідності форми, змісту (значення, змісту) морфологічних (слово, символ, знак) і синтаксичних (вирази, звороти) одиниць тестового завдання формою та змістом аналогічних одиниць, використаних

(використовуються) при викладі цієї навчальної дисципліни;

- залучення викладачів до культури конструювання, експертизи та застосування тестових вимірювачів для контролю результатів навчальних досягнень студентів;
- формування основ критичного (аналітичного) підходу до оцінювання якості освітніх досягнень учнів; збирання необхідних відомостей щодо якості тестових вимірювачів для постановки і рішення цілей і завдань стандартизації баз тестових завдань і тестів.

Система комплексної експертизи якості тестових матеріалів складається з чотирьох основних етапів:



Рис. 6.22. Послідовність етапів комплексної експертизи якості ТВ

1. Попередня експертиза якості тестових матеріалів, метою якої є добір авторського матеріалу, що відповідає вимогам тестів.

2. Експертне оцінювання (внутрішня експертиза) якості тестів, що проводиться після складання першого варіанта тесту.

3. Тестологічна експертиза, що полягає у проведенні апробаційних тестувань і розрахунку на їх основі результатів статистичних характеристик якості тестів та тестових завдань.

4. Комплексне оцінювання якості тестових завдань та тестів.

На етапі попередньої експертизи якості тестових завдань та тестів проводиться первинне оцінювання якості ТВ, яке включає в себе:

✓ оцінювання якості специфікації і кодифікатор тесту, що є одним з найважливіших етапів комплексної експертизи в силу того, що на етапі експертного оцінювання внутрішній зміст якості ТВ оцінюється експертами відповідно до тих характеристик, які заявлені в цих документах;

✓ оцінювання ТВ на відповідність формальним вимогам, полягає в тому, що відповідно до вимог оформлення та змісту тестових суджень експерти–предметники проводять змістовний аналіз формулювань тестових завдань і оцінюють коректність їх формулювань.

Підсумками попередньої експертизи є перелік ТВ, що відповідають усім вимогам, і перелік ТВ, що не відповідають тим чи іншим вимогам. У першому випадку - комплекти ТВ відправляють на наступний етап комплексної експертизи – експертне оцінювання, у другому – повертають авторам з коментарями з причин відбраковування з вказівкою на корекцію ТВ.

Експертне оцінювання якості ТВ включає чотири етапи:

Перший етап - оцінювання семантики, в тому числі і коректності формулювання тестового завдання. Будь-який зміст тестового завдання може бути представлено вербально (словесно) і невербально (малюнки, графіки, таблиці). У цьому напрямку експерт проводить багатокомпонентний аналіз змісту тестового завдання, а саме:

1) *Предметно-змістовний аналіз* - експерту необхідно оцінити адекватність відображення фактичного матеріалу навчальної програми, який можна представити двома компонентами:

Фактологічний - точність відображення фактів, подій, предметів, явищ матеріалу викладається навчальної дисципліни у формі судження;

Концептуальний - коректність подання змісту у формулюванні тестового завдання логічно виділеному дидактичним модулю відповідно до структури змісту навчальної дисципліни.

2) *Композиційний аналіз*. Композиція тестового завдання відображає єдність форми, змісту, інструкції з виконання завдання та різноманітних допоміжних компонентів (такі як таблиці, малюнки, графіки). Від гармонійності оформлення всіх складових елементів структури залежить логічність, змістовність, семантична завантаженість завдання, що, в свою чергу впливає на рівень сприйняття тестового завдання екзаменованим.

3) *Функціональний аналіз*. Тести в педагогічному процесі виконують ряд функцій: діагностичну, прогностичну, коригувальну, виховну, навчальну. Тестове завдання, як компонент тестування, має функціональну навантаженість. Кожне тестове завдання поліфункціональне, але ступінь вираженості тієї чи іншої функції різний. Від точності визначення функціональної значущості тестового завдання залежить ефективність його застосування у процедурі тестування.

4) *вербальний аналіз* - побудова граматичної форми тестового судження є одним з найважливіших компонентів завдання. Від правильності граматичного оформлення всіх компонентів композиції залежить чіткість, логічність формулювання й однозначно сприйняття тестового завдання.

Другий етап - прогнозування успішності виконання даного завдання різними за рівнем підготовки навчаються. Одна з найскладніших і трудомістких робіт експерта, полягає:

По-перше, в точності виділення логічних операцій, що відповідають результатами когнітивних дій учнів (студентів), що є основою для визначення рівня складності завдання;

По-друге, експерт встановлює орієнтовний час, необхідний для виконання завдання. Особливість прогностичного напрямку оцінювальної діяльності експерта

полягає в тому, що в процесі створення конкретного тестового завдання формуються нормативи його виконання.

Третій етап - оцінювання тесту в цілому за такими критеріями:

1. *Відповідність тесту цілям навчання та тестування:* наскільки розроблений тест здатний максимально точно діагностувати рівень підготовки учнів з відповідного виду контролю (вхідний, рубіжний або підсумковий).

2. *Практичність тесту,* що полягає в доступності інструкцій і змісту завдань тесту для розуміння учнем (студентом).

3. *Валідність тесту* – визначається за наявності необхідного з кожної навчальної теми кількості тестових завдань відповідно до кодифікатора навчальної дисципліни і специфікації тесту з зазначенням у ній відносної значущості (ваги) кожної теми.

4. *Композиція тесту* - оцінюється внутрішня узгодженість завдань в тесті в залежності від його призначення і гомогенності або гетерогенності. Основна задача оцінювання внутрішньої узгодженості завдань у тесті – це перевірка поєднання дібраних завдань, яке має відображати структурну ієрархію моделі підготовки з навчальної дисципліни. При цьому експерту необхідно оцінити ефективність запропонованої розробником схеми і способу розташування завдань у тесті. Поняття збалансованості включає в себе пропорційне наповнення тесту завданнями різного рівня складності. Таким чином, аналіз композиції тесту показує ступінь гармонійного подання ключових елементів змісту навчальної дисципліни та адекватність їх відображення в тесті.

5. *Представленість дидактичних одиниць в тесті.*

6. *Ефективність форми пред'явлення тесту:* бланкового тестування, комп'ютерне тестування, адаптивне тестування.

7. *Конструктна валідність,* яка визначається на основі аналізу змісту тесту тому наскільки об'єкт тестування і характер завдань враховують психологічні особливості учнів (студентів). При цьому аналізуються: складність використовуваних ТЗ, вік учасників тестування і відповідно до їх фізіологічних можливостей оптимальний час виконання

тесту з урахуванням часу необхідного для виконання кожного ТЗ, довжини і загальної труднощі тесту.

8. *Функціональну валідність тесту*, яка визначається на основі аналізу відповідності конкретного тесту тому рівню засвоєння, виду діяльності, для якого він був створений.

9. *Змістову валідність*, яка визначається за репрезентативності результатів перевіряються тестом способів розумової або практичної діяльності, знань, умінь та навичок відповідно до вимог, програми та кодифікатора навчальної дисципліни, специфікації тесту.

10. *Критеріальну валідність тесту*, що відображає значущості результатів тестування порівняно з деякою зовнішньою змінною, зовнішнім критерієм. Основні труднощі в експертному оцінюванні носить методологічний характер, оскільки вона визначається у виборі значущого зовнішнього критерію.

Для педагогічних тестів як критерій зазвичай беруться оцінки експертів, виставлені ними за традиційної перевірки знань учнів без використання тестів. Четвертий етап - оформлення підсумків експертного оцінювання та рекомендацій з доопрацювання або коригування завдання, вважається одним з найважливіших етапів робіт. Результати експертизи повинні бути представлені в строгому формальному вигляді для того, щоб вони були доступні і зрозумілі не тільки для експерта, але й для тестології, який буде проводити тестологічну експертизу якості ТЗ і тестів після їх апробації.

Підсумки експертного оцінювання можуть бути виражені формально в протоколі, призначеному для автоматизованого опрацювання. При цьому експерт заповнює відведені для міток поля. Специфіка експертних суджень полягає в тому, що результати експертного оцінювання по суті своїй якісні (оціночно-атрибутивні), повинні бути представлені числовими величинами. Частіше за все для вираження інформації числовими значеннями використовують трьохбальну або п'ятибальну шкалу. За результатами експертного оцінювання виносяться судження про якість та придатності тесту для тестування. Одним з найважливіших етапів комплексної експертизи якості тестів та тестових завдань є проведення

апробаційного тестування з метою встановлення, перевірки та оцінювання його вимірювальних можливостей шляхом апробації на репрезентативних вибірках.

Для цього після апробаційних тестувань, провадиться:

Крок 1. Формування матриці відповідей випробуваних на завдання, в якій міститься інформація про відповіді в дихотомічній або політомічній шкалі, номер варіанту, виконаного тесту, відомості про вибір варіантів відповідей на кожне тестове завдання з запропонованих тестів. Матриці відповідей випробуваних формуються методом перетворення вихідних даних у набір ознак з двома градаціями. Ця процедура носить назву дихотомізації відповідей випробуваних.

Крок 2. Опрацювання матриць відповідей з метою отримання та інтерпретації кількісних показників ТЗ: складності/легкості, дискримінативності ТЗ, підрахунок коефіцієнтів кореляції дібраних дистракторів з правильною відповіддю (для завдань закритого типу), підрахунок значень коефіцієнтів кореляції ТЗ з підсумковим балом по тесту.

Крок 3. Розрахунок показника валідності тесту, як кореляції між показниками тесту і певним зовнішнім критерієм.

Крок 4. Розрахунок коефіцієнтів кореляції завдання із завданням з метою оцінювання внутрішньої узгодженості тесту. Внутрішня узгодженість завдань тесту - істотна характеристика тесту, яка визначає його гомогенної (однорідність, взаємодія, близькість завдань).

Крок 5. Оцінювання паралельності варіантів тестів за результатами апробаційного тестування. З метою підвищення об'єктивності та надійності педагогічних вимірювань для тестування використовуються варіанти тестів, при цьому вони повинні відповідати властивості паралельності.

Крок 6. Розрахунок і оцінювання показників надійності тесту, що характеризує точність вимірювання. Чим вища надійність, тим вища точність, з якою він вимірює рівень навчальних досягнень студентів. На заключному етапі експертизи здійснюється формування комплексної оцінки якості тестових завдань і тестів.

При комплексному оцінюванні якості ТВ знімається проблема незрівняного оцінювання внутрішніх і зовнішніх

властивостей ТВ за рахунок агрегування показників якості, заснованих на побудові оптимальної ієрархічної системи критеріїв та показників.

Цей етап включає в себе такі процедури:

1. Формування зведеної таблиці показників якості ТВ, отриманих на попередніх етапах експертизи;
2. Побудова комплексних показників якості ТВ;
3. Нормування показників якості ТВ;
4. Визначення довірчих інтервалів комплексних оцінок;
5. Формування вербально-числової шкали якості ТВ;
6. Формування підсумкового висновку у вигляді рекомендацій: – щодо застосування стандартизованих ТВ для контролю якості УД; – для авторів з корегування та доробці ТВ із зазначенням причини невідповідності вимогам до якості. Проведення комплексної експертизи якості тестів дозволить поліпшити якість ТВ на етапі їх розробки і сформувати банк стандартизованих тестових завдань та тестів, що забезпечить контрольні-оціночні процедури надійними і валідними тестовими вимірювачами.

Побудова матриці результатів тестування

Формування матриці відповідей випробуваних на завдання, в якій міститься інформація про відповіді в дихотомічній або політомічній шкалі, номер варіанту, виконаного тесту, відомості про вибір варіантів відповідей на кожне тестове завдання з запропонованих тестів. Матриці відповідей випробуваних формуються методом перетворення вихідних даних у набір ознак з двома градаціями. Ця процедура носить назву дихотомізації відповідей випробуваних.

З точки зору розробника мінімальні вимоги до складу тестового завдання полягають у наявності трьох частин:

1. Інструкції.
2. Тексту завдання (запитання).
3. Правильної відповіді.

Перераховані три складові частини тестового завдання є мінімально необхідними для конструювання тестів. Крім цього,

укладачам тестових завдань доцільно вказувати ще ряд необхідних відомостей.

Таблиця 10

Дані розробників про завдання і їх цільове призначення

<i>Дані про тестове завдання</i>	<i>Подальше використання</i>
Вік (курс, спеціальність), на який розраховано це завдання	Для експертизи є метою визначення валідності
Дані про тестове завдання	Подальше використання
Тему (предмет чи предметну область у відповідності з технологічною матрицею)	Для перевірки відповідності завдань технологічній матриці
Передбачуваний автором час виконання завдання для компонування тесту, призначеного для апробації	Для компоновки тесту, призначеного для апробації
Терміни пред'явлення (календарні терміни, оскільки одне і те ж завдання, будучи пред'явленим, наприклад, в жовтні та лютому дасть різні результати і, відповідно, повинно мати різні характеристики)	Для включення в інструкцію по проведенню тесту, а також для складання плану тестування
Передбачену статистичну складність	Для правильного розміщення завдань в тесті, призначеному для апробації
Рівень, який відповідає даному завданню, чи вміння, яке воно перевіряє	Для перевірки відповідності завдань технологічній матриці
Відповідність стандарту чи програмному матеріалу	Для експертизи з метою визначення валідності
Дані про автора	Для отримання довідок та уточнень
Можливі варіанти невербальної підтримки	Для художника и технічного редактора, для підготовки невербальних матеріалів

Основна вимога до тестових завдань – тестове завдання повинно мати однозначну правильну відповідь. Поняття однозначності відповіді трактується не стільки як вимога єдиничності відповіді, але обов'язкова наявність передбачуваного зразка.

Ясна схема оцінювання повинна забезпечити користувача тестом апаратом оцінювання саме в рамках закладеної в тест оцінки розробника. Багато питань тлумачення можуть бути зняті при розробленні чіткої і недвозначної схеми оцінювання, яка містить найбільш можливі варіанти відповідей, які можна прийняти до розгляду і оцінити як залікові. Схема оцінювання повинна повністю відповідати конкретному запитанню. Усі формулювання очікуваних відповідей повинні бути гранично зрозумілими і недвозначними.

§ 7. Розрахунок результатів. Методи шкалювання та представлення результатів тестування

7.1. Одновимірне шкалювання. Види одновимірного шкалювання

Що таке шкалювання?



Шкалювання це процес побудови шкали, проведений за певними правилами.

Для одновимірних методів шкалювання характерним є обмеження одним, частіше всього емоційним, компонентом теоретичної конструкції.

При розробленні одновимірного шкалювання автори задавалися такими цілями. Створити спосіб настільки простим, щоб при цьому отримані дані були адекватні існуючим умовам. Створити спосіб з настільки високим рівнем вимірювання, щоб мати змогу застосовувати традиційні числові методи (зокрема при отриманні статистичних даних). Створити спосіб настільки функціональний, щоб на його підставі переходити від вибірки до генеральної сукупності.

Їм це вдалося і в результаті виникла маса шкал. Кожна з них важлива і якщо не використовується на практиці, то все одно, безумовно, є науковим досягненням.

1. Психофізичний вимір.

Цей метод був запропонований Терстоуном і вважається передумовою одновимірного соціологічного шкалювання. Полягає він у такому: Терстоун, аналізуючи відповіді респондентів під час відповідей на питання, що мають об'єктивний характер, зауважив, що у кожного респондента свій поріг розрізнення. Терстоун прийшов до висновку, що шкала, побудована таким чином, вирішує і проблеми вимірювання в соціології. Ці висновки активно використовувалися й іншими авторами для подальших розробок одновимірних шкал.

2. Метод Терстоуна.

Так званий метод опитування експертів. Цей метод направлений, перш за все, на вимірювання соціально-

психологічних характеристик і перш за все соціальних установок (Атитюди), ціннісних орієнтацій, емоційних станів респондента. На першому етапі дослідження набирається група експертів. Це група, чисельністю не більше 50 чоловік, що являє собою модель досліджуваної сукупності. Експерти проводять відбір і еталонування запропонованих суджень. У результаті відбираються ті з них, які отримали найбільш узгоджені оцінки експертів. У підсумку залишається 15-30 суджень, шкально значення яких і складають шкалу дослідження. Потім ці судження, що пройшли відбір експертів, пропонуються респондентам, відповіді яких і "розкидаються" по шкалі.

3. Шкала ранжування.

Вона вирізняється тією особливістю, що результати вимірювання установок за її допомогою аналізуються відповідно до правил, що застосовуються для рангових шкал. Найбільш простий прийом вимірювання установок за правилами такої шкали – ранжування респондентами тих об'єктів, ставлення до яких з їхнього боку цікавлять дослідника. Коли хочуть, щоб мета виявлення установок для респондентів залишалася прихованою, в цю шкалу вводяться "фіктивні об'єкти". Більш складним варіантом виміру установок за допомогою рангової шкали – метод парних порівнянь.

4. Метод парних порівнянь.

Метод парних порівнянь – це метод побудови оцінної шкали, заснований на парному порівнянні усіх об'єктів вивчення. Респондент намагається вибрати, що йому більше симпатизує із запропонованої пари. Для опрацювання отриманих даних складається асиметрична матриця. При чому асиметрія вказує на вірність ходу думки респондента і реальність даних. Цей метод часто "обмежує" дослідника, тому як респондент не може вибрати, що йому подобається більше із запропонованої пари і не може відповісти, а це призводить до недостатньої кількості інформації або її вибраковуванні.

5. Факторний аналіз.

Цей метод нерозривно пов'язаний з тестовими розробками соціологічних досліджень. Він не належить до числа основних методів, а цікавить дослідника лише тому, що його розуміння

важливо для осмислення ідей, закладених у деяких соціологічних методах шкалювання. Основна ідея факторного аналізу полягає в такому: кожна спостережувана ознака може бути представлена у вигляді лінійної комбінації нормально розподілених факторів.

6. Шкала Лайкерта.

Часто цей метод зустрічається в літературі під назвою *метод сумарних оцінок*. Лайкерт запропонував вимірювати латентну змінну шляхом побудови індексу, а так само запропонував будувати фрагмент анкети, спрямований на вимірювання латентної змінної у вигляді так званого кафетерію – таблиці, рядки якої відповідають спостережуваній змінній, а стовпцями – значення цих змінних. Значення фактора в результаті визначається як сума значень спостережуваних змінних. Цей метод легкий і не вимагає використання опрацювання на комп'ютері.

7. Шкала Гуттмана.

Тут значення латентної змінної розраховується як сума позитивних відповідей, даних респондентом на питання, які розглядаються. Розглянуті дихотомічні ознаки можна впорядкувати і представити у вигляді діагональної матриці. Гуттман запропонував простий алгоритм, який приводить її до цього вигляду.

Особливість її у тому, що кожна оцінка автоматично включає в себе усі наступні та виключає попередні. Шкали подібного типу називаються кумулятивними. Вони використовувалися задовго до Гуттмана. Приміром, такою є шкала Богардуса.

8. Шкала Богардуса.

Її основне призначення – вимірювання національних і расових установок. Найчастіше її застосовують для визначення соціальної дистанції.

9. Латентно-структурний аналіз.

Цей метод був вперше запропонований Лазерсфельдом. Сам він неодноразово заявляв, що цей метод має безпосереднє відношення до теорії тестів у соціології. Полягає метод латентно-структурного аналізу у виявленні статистичного зв'язку між спостережуваною змінною, який можна пояснити

дією латентної змінної. Це виявляється в тому, що при фіксації значення латентної змінної цей зв'язок зникає. Лазерсфельд назвав це явище аксіомою локальної незалежності. Основна специфічна риса латентно-структурного аналізу: він працює з частотними таблицями, а значить, з шкалами будь-яких типів.

10. Метод семантичного диференціала.

Цей метод вперше був запропонований Осгудом. Мета його – розкриття афективних компонентів змісту, що вкладаються людьми в ті або інші об'єкти, явища, поняття, а так само виявлення тих чинників, які визначають смислову значущість об'єктів для кожної людини. Простір, що утворюються цими факторами, і є там самим семантичним простором, до якого респондент немов би вміщує об'єкт, оцінюючи його таким чином. Цей метод визначає різницю у сприйнятті людиною різних об'єктів (диференціює). Метод Осгуда спрямований на виділення типів людей, що мають схожу картину досліджуваних змістів. Відповідні усереднені змісти інтерпретуються як значення об'єктів для субкультури, ототожнюється з розглядуваним типом людей. Цей метод опирається на вивчення явища синтестезиса – мислення за аналогією. Фактично це інтервальна шкала. Полягає цей метод у наступному: респондентам по черзі пред'являються для оцінювання розглянуті об'єкти, і пропонують співвіднести інтенсивність свого внутрішнього відчуття по черзі зі всіма оціночними шкалами. Але так як при вирішенні конкретних задач методом семантичного диференціала виникає маса методичних труднощів (наприклад, важко обмежитися використанням тільки когнітивних ознак), зазвичай використовується ослаблений, більш простий метод проєктивних процедур.

11. Шкала самооцінки.

Це найбільш простий вигляд шкали вимірювання установки. Вона може бути сконструйована у формі звичайного запитання або у вигляді деякої числової осі з позитивними та негативними градаціями. При конструюванні шкали самооцінки у формі "традиційного" питання, її позиції обов'язково розташовуються симетрично і складаються з рівного числа позитивних і негативних оцінок, розділених

"нейтральною" позицією. У соціологічній практиці найчастіше застосовується шкала самооцінки з п'ятьма позиціями. Перші дві позиції реєструють відповідний ступінь позитивної установки, останні дві – негативної. Нульова позиція є нейтральним станом або станом складності відповіді. Основний недолік шкали самооцінки – високий ступінь її суб'єктивності. Справа в тому, що респонденти, які мають різний ступінь інтенсивності установок, повніше можуть відзначити одну й ту ж саму позицію на шкалі вимірювання, а ті, що володіють однаковою інтенсивністю установки – різні позиції.

12. Одномірне розгортання.

Цей метод вперше був запропонований Кумбса. Він глибоко проаналізував аспекти інтерпретації даних, намагаючись зрозуміти ті положення, без яких немислимий жоден метод шкалювання, намагаючись уявити собі, яким має бути метод шкалювання, заснований тільки на таких припущеннях. Він намагався зрозуміти, наскільки адекватні ті шкали, які до того моменту в соціології були традиційними. Дійшов висновку, що вони абсолютно неадекватні і це вимагає розроблення нового методу, заснованого на більш чіткому розгляді особистості респондента, щоб він дав більш чітку відповідь.

Одним із найбільш поширених завдань в емпіричній соціології є пошук таких числових характеристик, які б відображали усереднену думку усіх респондентів про розглянуті об'єкти. Кумбс поставив перед собою питання: чи можна знайти оціночну шкалу, не нав'язуючи респонденту чисел, не вкладаючи у відповіді респондента того, чого він не говорив? Саме метод одновимірного розгортання і став відповіддю на це питання.

Таким чином, основна мета зазначеного методу – побудова оцінної шкали на базі ранжування досліджуваних об'єктів з використанням прийнятної моделі сприйняття, не опирається на підміну рангів числами.

Кумбс запропонував модель, яка відображає інтерпретацію ранжувань. Модель ідеальної точки. Звертаючись до експертів із проханням проранжувати об'єкти, дослідник не говорить про те, по якому конкретному якості

ранжування не повинні здійснюватися. Базуючись на припущеннях про те, які об'єкти подобаються і не подобаються респонденту, складається шкала. Побудовану оціночну шкалу можна вважати результатом усереднення вихідних ранжувань. На цій прямій розташовуються ідеальні точки – думки. Їх має бути рівно 2, як крайніх думок з даного питання, ну а серединна відстань ділиться навпіл, ставлячись до близлежачої точки. Шукається саме те положення точок, яке виражає найбільш повну картину дії. Якщо кількість об'єктів які шкалюються більше трьох, то розглянутий підхід враховує не лише порядок розташування об'єктів на осі, але і співвідношення інтервалів між ними.

7.2. Стандартизація та норми

Стандартизація шкали. Варто розрізняти дві форми стандартизації. Під стандартизацією тесту розуміють, насамперед, стандартизацію самої процедури проведення, інструкцій, бланків, способу реєстрації, умов і т.п. Без стандартизації тесту неможливо одержати нормативного розподілу тестових балів, а тому і тестових норм.

Під стандартизацією шкали розуміють лінійне перетворення масштабу нормальної (або штучно нормалізованої) шкали.

Застосування стандартних шкал дозволяє прибїгати на практиці до більш грубих, наближених способів перевірки типу розподілу тестових балів.

Застосування стандартних шкал є необхідним для співставлення результатів по різних тестах, для побудови «діагностичних профілів» подібних цілей.

Кожному з нас доводиться оцінювати інших. Природно, нелегко робити це об'єктивно, особливо, коли під рукою немає необхідних психодіагностичних методик.

Об'єктивність тесту означає, що первинні показники, їх оцінка й інтерпретація не залежать від поведінки і суб'єктивних суджень експериментатора. Первинними називають показники, отримані після опрацювання даних виконання респондентом

тестових завдань. Виражаються первинні показники в так званих "сирих балах".

Домогтися об'єктивності тесту можна при виконанні таких умов: однаковість процедури проведення тесту для одержання результатів, зрівняних з нормою; однаковість оцінки виконання тесту; визначення норми виконання тесту для зіставлення з ними показників, отриманих у результаті опрацювання даних тестування. Ці три умови називають **етапами стандартизації тесту**.

Перший етап стандартизації тесту полягає у створенні однакової процедури тестування. Вона включає визначення наступних моментів діагностичної ситуації:

- умови тестування (приміщення, освітлення та ін. зовнішні фактори). Очевидно, що тестування краще проводити коли немає зовнішніх подразників, таких як сторонні звуки, голоси і т.д.;
- зміст інструкції й особливості її пред'явлення (тон голосу, паузи, швидкість мовлення і т.д.);
- наявність стандартного стимульного матеріалу;
- тимчасові обмеження виконання даного тесту;
- стандартний бланк для виконання даного тесту. Використання стандартного бланка полегшує процедуру опрацювання;
- врахування впливу ситуаційних змінних на процес і результат тестування. Під змінними розуміємо стан випробуваного (втома, перенапруга і т.д.), нестандартні умови тестування (погане освітлення, відсутність вентиляції та ін.), переривання тестування;
- врахування впливу поведінки діагноста на процес і результат тестування. Наприклад, що схвально – заохочуюча поведінка експериментатора під час тестування може сприйматися респондентом як підказка "правильної відповіді" та ін.
- врахування впливу досвіду респондента в тестуванні. Природно, що респондент, який уже не в перший раз проходить процедуру тестування, переборів почуття невідомості і виробив певне ставлення до тестової ситуації. Наприклад, якщо респондент уже виконував тест

Равена, то, швидше за все, не варто його пропонувати йому в другий раз.

Другий етап стандартизації тесту полягає в створенні однакової оцінки виконання тесту: стандартної інтерпретації отриманих результатів і попередньої стандартної обробки. Цей етап передбачає також порівняння отриманих показників з нормою виконання цього тесту для даного віку (наприклад, у тестах інтелекту), статі і т.д.

Третій етап стандартизації тесту полягає у визначенні норм виконання тесту. Норми розробляються для різного віку, професій та ін. Ось деякі з існуючих видів норм:

- шкільні норми розробляються на основі тестів шкільних досягнень або тестів шкільних здібностей. Вони встановлюються для кожного шкільного ступеня і діють на всій території країни;
- професійні норми встановлюються на основі тестів для різних професійних груп (наприклад, учителів, менеджерів, юристів та ін.);
- локальні норми встановлюються і застосовуються для вузьких категорій людей, які вирізняються наявністю загальної ознаки – віку, статі, географічного району, соціо-економічного статусу та ін.;
- національні норми розробляються для представників даної народності, нації, країни в цілому. Необхідність таких норм визначається конкретною культурою, моральними вимогами і традиціями кожної нації.

Наявність нормативних даних (норм) у стандартизованих методах психодіагностики є їх істотною характеристикою. Норми необхідні для інтерпретації тестових результатів (первинних показників) як еталон, з яким порівнюються результати тестування.

Усі методи психологічної діагностики, починаючи зі стандартизованих і закінчуючи нестандартизованими, спрямовані на вимірювання властивостей особистості й особливостей інтелекту. Кожен метод призначений для вимірювання якоїсь властивості, що і визначає зміст цього методу. Відомості про ступінь, за якого тест дійсно вимірює те, для чого він призначений, входять у поняття валідності.

Крім загального значення до поняття валідності входять багато інших відомостей. Існують різні типи і види валідності, у яких відбиті ці відомості. Нижче наведені деякі типи.

1. Практична валідність характеризує тест не стільки з боку його психологічного змісту, скільки з погляду його цінності відносно досягнення певної практичної мети (прогнозування, діагностики). Виділяють підтипи практичної валідності: прогностична і співпадаюча. Тест із високою прогностичною валідністю дозволяє зробити прогноз, на скільки успішно випробуваний виконуватиме певну діяльність згодом. Тест із високою співпадаючою валідністю дозволяє відповісти на запитання: "Якою є імовірність того, що індивід X має властивість Y на даний момент часу?"

2. Валідність за обсягом відображає ступінь, в якому обсяг тесту вибірково репрезентує той клас ситуацій або навчального матеріалу, відносно якого повинні бути зроблені висновки. Валідність за обсягом показує, у якому обсязі, в якій мірі певна властивість відображена в методиці.

3. Оцінна валідність виражає ступінь кореляції (відповідності) між показниками, отриманими випробуваними за даним методом, і оцінками вимірювальної властивості з боку експертів. Наприклад, при валідації (перевірці тесту на валідність) тестів оцінки розумових здібностей школярів в якості експертів зазвичай виступають вчителі.

4. Зовнішня валідність означає, що враження, яке створюється у недосвідчених фахівців при знайомстві зі стимульним матеріалом і зовнішньою стороною тестування, відповідає заявленому змісту методу.

Ефективна методика може бути створена, якщо вона пройшла всі етапи валідації, тобто коли прийняті міри для надання їй змістовної валідності.

Про високу надійність методу говорять у тому випадку, коли метод точно вимірює ту властивість, для виміру якої він призначений. В якості **критеріїв точності** можна відзначити наступні:

1. При повторному застосуванні методу до тих же випробуваних у тих самих умовах через певний інтервал часу результати обох тестувань істотно не відрізняються між собою.

2. Дії випадкових сторонніх факторів не здійснюють істотного впливу на результати тестування. В якості сторонніх факторів можна назвати наступні: емоційний стан і втому, якщо вони не входять до кола досліджуваних характеристик, температуру, освітленість приміщення та ін. Такі сторонні випадкові фактори ще називають факторами нестабільності вимірювальної процедури.

3. При повторному застосуванні методу до тих же самих випробуваних через певний інтервал часу у змінених умовах результати обох тестувань істотно не відрізняються між собою. Під зміненими маються на увазі наступні умови: інший експериментатор, стан респондента та ін.

Тест може бути надійним, але не валідним. Це означає, що він вимірює якусь властивість дуже точно, але яке саме – залишається під сумнівом. У такій ситуації необхідна більш точна валідізація тесту, як змістовна, так і практична.

До безперечних переваг тестів відноситься об'єктивний характер процедури, можливість повторного огляду результатів, що ґрунтується на виконанні укладачем тестів вимог психометрії, тобто вимог репрезентативності, надійності, валідності.

Репрезентативність – відповідність тестових норм вибірки стандартизації тестовим нормам популяції, на якій застосовується тест. Надійність – стійкість процедури вимірювання, незалежність від випадкових величин. Валідність – відповідність методики вимірюваному концепту.

Якщо всі ці психологічні процедури не виконані відповідно до тесту, то його використання не носить науково обґрунтованого характеру, а є так само довільним, як і будь-яка життєва ситуація розуміння людьми один одного. Результат буде залежати від безлічі випадків.

Знання психометрики дає педагогу необхідний матеріал для розуміння обмеженості методик, тих припущень, які були зроблені їх авторами для цілей дослідження. Людина при роботі з тестом активна і може застосовувати такий спосіб дії, таке ставлення до ситуації завдання, які розроблені методики зовсім не передбачали. Діапазон же застосування тестів дуже

вузкий і часто зводиться до тієї вибірки, на якій відбувалася розробка тесту.

7.3. Основні типи шкал в педагогічних вимірюваннях



Шкала – засіб фіксації результатів вимірювання властивостей об'єктів шляхом впорядкування їх в певну систему чисел, в якій відношення між окремими результатами виражено у відповідних числах.

У процесі впорядкування кожному елементу сукупності емпіричних даних, які спостерігаються, ставиться у відповідність певний бал (шкальний індекс), який встановлює положення результату на шкалі.

Операція впорядкування вихідних емпіричних даних шляхом переводу їх в шкальовані носить назву **шкалювання**.

Процес шкалювання складається з конструювання шкали за певними правилами і включає два етапи. Перший – збір емпіричних даних. Другий етап – опрацювання і аналіз, за результатами якого будується числова система, що являє собою шкалу.

В педагогічних вимірюваннях шкали розрізняються залежно від характеру властивостей, що лежать в основі їх побудови. В якості таких властивостей виділяють:

ідентичність, що дозволяє однозначно відносити об'єкти до однієї з виділених категорій;

транзитивність, що сприяє ранжуванню об'єктів в певному порядку;

матричність, що забезпечує єдину одиницю вимірювання; наявність абсолютного нуля.

Найбільш загальна класифікація шкал запропонована С. Стівенсом, в якій подана з подробицями характеристика рівнів вимірювання.

Відповідно цій класифікації чотири основні способи вимірювання, пов'язані з різноманітними правилами, називають вимірювальними шкалами.

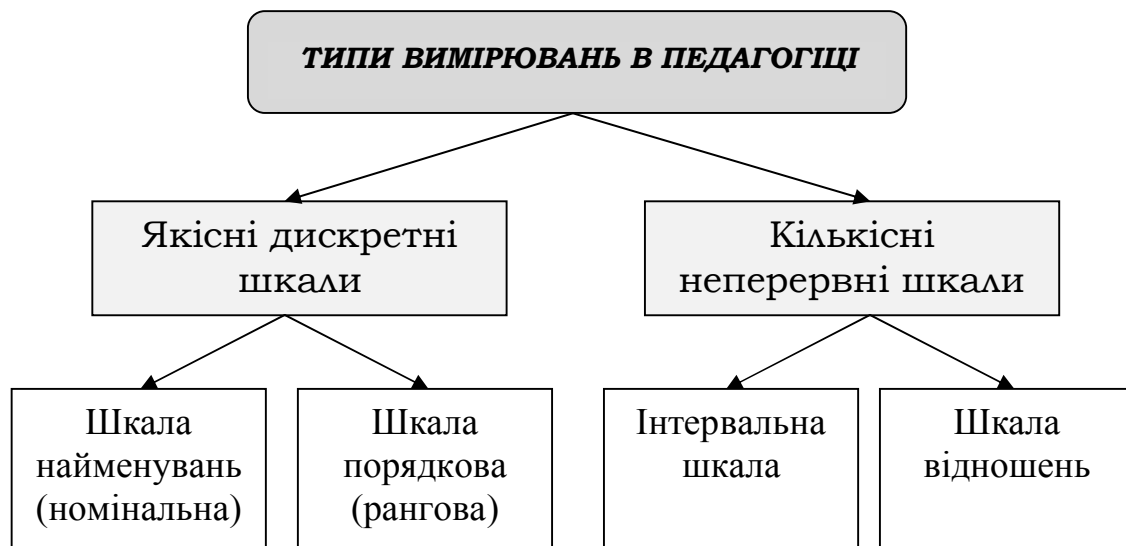


Рис. 7.1. Класифікація шкал вимірювання за С. Стівенсом

Найбільш елементарна форма вимірювання – це **номінальна шкала**, або шкала найменувань, що володіє лише першою властивістю.

Номінальна шкала (шкала найменувань), яку правильніше було б вважати класифікацією, а не виміром, поділяє усі об'єкти на групи за будь-якою ознакою (відмінністю). У цій шкалі кожному оцінюваному об'єкту відповідно до яких-небудь властивостей приписується найменування або число. Арифметичні дії з числами в номінальній шкалі не мають сенсу, між ними не встановлюється відношення порядку, числа застосовуються тільки для найменування об'єктів.

Існують два типи номінальної шкали.

У шкалі першого типу кожному об'єкту приписується число. Кожне число представляє окрему людину (об'єкт), і між об'єктами є відмінність, оскільки числа присвоюються певним чином.

Другий тип номінальної шкали особливо широко застосовується в педагогічній практиці. Він пов'язаний з класифікацією учнів за групами відповідно до яких-небудь ознак, після чого число або найменування приписується не окремому учню, а групі учнів. Наприклад, в процесі перевірки відповідності підготовки випускників школи до вимог освітніх стандартів з'являється група атестованих і група неатестованих учнів.

Взагалі, при застосуванні шкали класифікації можна обійтися і без чисел. Об'єктам приписують букви або якісь інші

позначення, наприклад оцінки в заліковій книжці студентів «Залік» – «незалік».

За допомогою номінальної шкали можна вимірювати тільки якісні ознаки, тому опрацювання кількісних даних слід проводити не з самими цими числами, а з питомою вагою кількості об'єктів даного класу.

У цій шкалі припустимі наступні статистичні операції:

- розрахунок частот (питомих ваг) об'єктів даного класу;
- визначення моди ознаки яка вивчається.

Шкала порядку (порядкова, рангова, ординальна) призначена для вимірювання (позначення) ступеню відмінності якої-небудь ознаки або властивості різних об'єктів. У **порядковій шкалі** вводяться числа і відношення «Більше-менше», тому за числом, що відповідає оцінюваному об'єкту, можна дізнатися про місце об'єкта в сукупності.

Наприклад, за результатами іспитів можна приписати ранги кожному учню, але тільки в межах тієї групи, де проводився екзамен.

Найяскравішим прикладом порядкової шкали є п'ятибальна система оцінки ЗУН учнів. Для неї розроблені критерії і різні методи вимірювання. Значно важче застосовувати порядкову шкалу для кількісних оцінок інших якостей особистості (у виховному процесі). Є декілька різновидів порядкового шкалювання (вимірювання):

- ранжування (в ряд),
- групування (ранжування по групах),
- парне порівняння,
- метод рейтингу,
- метод полярних профілів.

Ранжування. Досліджувані об'єкти розташовуються в ряд (упорядковуються) за ступенем вираженості якої-небудь властивості. Перше місце в цьому ряді займає об'єкт з найбільш високим рівнем зазначеної властивості, і йому присвоюється найвищий бал (числове значення вибирається довільно). Потім кожному об'єкту рангового ряду присвоюються більш низькі оцінки, відповідні займаним місцям.

Наприклад, при ранжуванні спортсменів кращий якраз отримує місце з мінімальним номером один. Цілком зрозуміло,

що не слід проводити додавання і віднімання номерів місць зважаючи на невизначеність одержуваного результату. Однак цієї великої істини дотримуються далеко не завжди. У сфері освіти був період, коли директору будь-якої школи чимало клопоту додавав так званий середній бал, який розглядався як важливий показник якості роботи школи. При цьому якимось забувається про те, що арифметичні дії з номерами місць груп школярів неприпустимі, і тому середній бал аніскільки не відображає об'єктивних закономірностей результатів навчального процесу. Нікому і ніколи не спадало на думку шукати середній результат команди спортсменів, проте протягом ряду років безвідповідально за середнім балом порівнювалися результати роботи педагогічних колективів, що, зазвичай, нерідко призводило до невиправданих висновків, що завдають шкоди якості навчального процесу. При вимірюванні ознак в порядковій шкалі можливі тільки монотонні перетворення, що допускають множення на постійний множник, піднесення до степеня і добування кореня, та деякі статистичні операції. Зокрема, в порядковій шкалі як середню оцінку використовують медіану, міру розсіювання – квантилі, в якості міри зв'язку двох ознак – ранговий коефіцієнт кореляції.

У порядковій шкалі вимірюються тільки якісні ознаки. Об'єкти оцінюються з точки зору відношення рівності між ними або відношення «більше-менше», а відстані між об'єктами не мають жодного сенсу. За приблизними оцінками, порядковою шкалою можна ранжувати учнів, але робити висновок про те, наскільки один краще іншого, не можна в силу відсутності одиниці вимірювання. Порівнюваність результатів учнів досягається в інтервальній шкалі.

Групування усієї сукупності об'єктів спостереження в кілька рангів, досить чітко відрізняються один від одного за ступенем вимірюваної ознаки.

Приклад: учні класу згідно п'ятибальної системи оцінювання ЗУН діляться на відмінників, хорошистів і т. д.

Парне порівняння. Учні зіставляються один з іншим (кожен з кожним) з якої-небудь властивості. Якщо вони однакові, то кожен отримує по балу. Якщо у одного учня цієї властивості більше, ніж у іншого, перший отримує два бали,

другий – 0 (як при спортивних іграх за круговою системою). Підсумовуючи отримані кожним бали, отримуємо кількісний вираз рівня розвитку даної якості у кожного учня (його ранг).

Рейтинг. У цьому прийомі оцінювання об'єкта здійснюється шляхом усереднення оціночних суджень групою компетентних експертів. Маючи загальні критерії оцінки (у порядковій шкалі, у балах), експерти незалежно один від одного (в усній або письмовій формі) виносять свої судження. Усереднений результат експертної оцінки є досить об'єктивним і називається рейтингом.

Метод полярних профілів. Цей прийом передбачає застосування для оцінювання умовної шкали, крайніми точками якої є протилежні значення ознаки (наприклад, добрий – злий, теплий – холодний і т.п.). Проміжок між полюсами ділиться на довільну кількість частин (балів).

Приклад. Оцінка ступеня довіри до кандидата на посаду дається в полярній шкалі:

(Довіряю повністю) 10–9–8–7–6–5–4–3–2–1(Зовсім не довіряю)

Інтервальна шкала, що включає перші три властивості, дозволяє подолати недоліки номінальної і порядкової шкал, оскільки в ній визначено відстань між об'єктами і передбачена спільна для всіх об'єктів постійна одиниця виміру.

Інтервальна шкала – кількісна. У ній можливі усі арифметичні дії над числами, крім ділення. Таким чином, в інтервальній шкалі не можна визначити, у скільки разів один об'єкт більший або менший за інший. Наприклад, якщо учень відповів правильно на 30 завдань, то це не означає, що він знає вдвічі більше учня, який відповів на 15 завдань тесту.

Недоліком цієї шкали є невідоме значення абсолютного нуля. Наприклад, при оцінюванні виконання учнем тесту, нуль вірних відповідей не означає повної відсутності знань. У інтервальній шкалі допустимі майже усі статистичні операції, крім тих, які передбачають знання «істинно» нульової точки шкали. Тому в інтервальній шкалі не можна використовувати такі характеристики, як середнє геометричне і коефіцієнт варіації досліджуваної ознаки.

Приклади: температурні шкали; шкали стандартизованого тестування інтелекту.

Шкала відношень, що задовольняє усім чотирьом властивостям, дозволяє отримати найвищий рівень вимірювання. Тут можна визначити відношення чисел, що приписуються об'єктам. У шкалі відношень за точку відліку обрано абсолютний нуль. У ній можна виконувати усі арифметичні і статистичні операції. Так само як і інтервальна, вона дозволяє проводити кількісні вимірювання. Рівні вимірювання і числові характеристики, що використовуються на даних рівнях, наведено в табл.11. З цієї таблиці видно, що перехід від одного рівня до іншого супроводжується розширенням класу допустимих математико-статистичних операцій.

Як видно з табл.11, найкращою є шкала відношень, яку на сьогоднішній день вдалося реалізувати тільки в межах фізичних вимірювань. Виходячи з наведених вище міркувань можна зробити висновок: не слід прагнути до традиційних засобів контролю, що не забезпечує порівняльних кількісних оцінок у інтервальній шкалі. Хоча і з тестами справа складається не так благополучно, як здавалося в ті роки, коли за кордоном створювалися перші педагогічні тести для масового впровадження в навчальний процес.

Таблиця 11

Рівні вимірювання та їх характеристики

Рівень вимірювання (шкала)	Основна операція, що визначає рівень	Математичні і статистичні величини, обчислення яких допустиме на даному рівні
Номінальний рівень (номінальна)	Приписування однакових чисел об'єктам, які мають спільну ознаку	Мода, процентні частоти, долі, коефіцієнт зв'язку, коефіцієнт відмінності
Ординальний рівень (рангова, порядкова)	Ранжування об'єктів за спільною ознакою	Мода, медіана, квантилі, квартилі (процентиль, дециль, квартиль та ін.), рангові коефіцієнти кореляції, дисперсійний аналіз
Інтервальний рівень (інтервальна)	Визначення величини відмінностей між об'єктами	Мода, медіана, квантилі, рангові критерії, вибіркоче середнє, дисперсія, стандартне квадратичне відхилення, коефіцієнт кореляції.

Рівень вимірювання (шкала)	Основна операція, що визначає рівень	Математичні і статистичні величини, обчислення яких допустиме на даному рівні
Вимірювання відношень (відношень)	Визначення рівності відношень величин	Усі арифметичні операції, усі поняття та методи математичної статистики.

Говорячи про переваги тестових методів, пов'язаних з можливістю отримання кількісних порівняльних оцінок, слід зазначити, що вони виявляються не завжди і не скрізь.

По-перше, не кожен тест краще іспиту, а лише той, який має високу надійність і дозволяє побудувати по емпіричних результатах його виконання стійку шкалу.

По-друге, спостерігаються результати виконання тесту (сирі бали випробовуваних) не забезпечують порівняння. Для порівняння необхідно зробити шкалювання сирих балів шляхом переведення їх в одну із спеціальним чином підібраних стандартних похідних шкал.

По-третє, при переході до шкальованих показників випробовуваних бажано користуватися методами сучасної теорії тестів – теорії IRT, оскільки саме теорія IRT забезпечує переведення сирих балів у інтервальну шкалу.

7.4. Шкалювання результатів тестових вимірювань

Щоб визначити положення випробуваного щодо вибірки стандартизації, його первинний результат (кількість правильно виконаних завдань) переводять у якусь відносну міру. В основі такого перетворення лежить прагнення підвищити рівень вимірів, перейти, якщо це можливо, від порядкової до інтервальної шкали, що допускає впорядкування випробовуваних з урахуванням відмінностей у їхньому рівні підготовки шляхом введення фіксованої одиниці виміру. Щоправда, далеко не завжди перетворення первинних результатів призводить до інтервальної або хоча б до квазіінтервальної шкали. В окремих випадках факт переходу вважається досить суперечливим і спростовується аргументованою критикою на підставі результатів експертизи якості шкали і самого вимірювача.

Слід зазначити, що спроби підвищення рівня вимірювань шляхом шкалювання, спрямовані цілком на розширення можливостей інтерпретації і співставлення тестових балів, але аж ніяк не пов'язані з підвищенням точності вимірювань. Не порушуючи загальності, можна вважати, що перетворені (похідні) показники слугують двом цілям. По-перше, вони дозволяють уточнити місце, зайняте випробуваним в нормативній вибірці, і, по-друге, порівняти результати досліджуваних, отримані за різними тестами.

Шкали процентильних рангів

Похідні показники можна розділити за призначенням. Одні з них слугують для визначення досягнутого рівня підготовки в певній шкалі, а інші – для встановлення відносного положення випробуваного в деякій нормативній групі. Зокрема, вирішенням другого завдання слугують процентилі, що дозволяють встановити ранг первинного показника випробуваного в нормативній групі. Ранг показника в процентилях визначається відсотковою часткою випробовуваних з вибірки стандартизації, результати яких не вищі даного первинного показника. Процес побудови шкали процентилів полягає у визначенні процентильних рангів первинних показників нормативної групи.

Метод побудови шкали процентилів можна розглянути на невеликому прикладі результатів виконання тесту групою з 25 випробовуваних, хоча, звичайно, мало ймовірно зустріти таку вибірку на практиці. Зазвичай, побудова цих шкал виконується на великих масивах. Нехай, наприклад, 25 тестованих учнів по одному із предметів отримали первинні результати, представлені в таблиці:

Таблиця 12

Результати випробовуваних

Сирі бали випробовуваних	3	4	5	6	7	8	9	10	11	12	13	14	15
Частоти	1	1	2	2	3	4	3	4	3	2	2	1	1
Кумульовані частоти	1	2	4	6	9	13	16	19	21	22	23	24	25
Процентильні ранги	1	6	12	20	30	44	58	70	80	86	90	94	98

Перший рядок в таблиці 12 містить бали, які спостерігаються у випробуваних вибірки, впорядковані від меншого значення до більшого (зліва направо). Зазвичай для великих груп просте упорядкування є малоефективним і зручніше використовувати згруповані дані, що припускають введення розрядів оцінок для окремих груп.

У другому рядку представлені результати підрахунку числа випробовуваних, що мають один і той же тестовий бал. Кожен елемент другого рядка показує кількість повторень балу і тому називається частотою спостережуваних сирих балів випробовуваних. Якщо частоти підсумувати зліва направо, то вийдуть значення накопичених (кумуляованих) частот. Кумульовані частоти являють собою сумарну кількість частот, які спостерігаються на цьому балі або нижче нього. Наприклад, існує 9 досліджуваних, що отримали бал 7 або нижче, тому що кумульовані частоти для балів 7 – число 9.

Обчислення процентильних рангів для заповнення четвертого рядка таблиці ускладнюється необхідністю визначення фактичних меж довірчого інтервалу, який містить справжній бал кожного випробовуваного вибірки.

Фактична довжина інтервалу залежить від значення стандартної похибки вимірювання. Проте, зазвичай для визначення меж інтервалу використовують 0,5 одиниці виміру сирих балів. У цьому випадку, якщо випробуваний отримав бал 5, справжнє значення його балу лежить в інтервалі від 4,5 до 5,5, тобто $(4,5;5,5)$, а числа 4,5 і 5,5 називаються відповідно нижньою та верхньою межею одиничного інтервалу оцінок.

Поняття «верхня» і «нижня» межа використовуються для побудови шкали процентильних рангів в припущенні рівномірності розподілу результатів випробовуваних всередині довірчого інтервалу. Наприклад, при обчисленні процентного рангу для тестового балу 5 приймають, що результати двох випробовуваних розташовуються в інтервалі $(4,5;5,5)$ рівномірно (табл. 13).

Таблиця 13

Побудова шкали процентильних рангів

Крок	Етап обчислення	Приклад обчислення
1	Підрахунок кумульованих частот	Для досліджуваного з балом 4 кумульована частота $1+1=2$, для досліджуваного з балом 5: $2+2=4$, для досліджуваного з балом 6: $4+2=6$
2	Визначення точних верхніх та нижніх меж одиничних інтервалів оцінок	Для досліджуваного з балом 4 істинний бал лежить в інтервалі $(3,5;4,5)$, для досліджуваного з балом 5 – в інтервалі $(4,5;5,5)$ і т.д.
3	Підрахунок процентильних рангів	Для досліджуваного з балом 4: $(1,5/25)*100\%=6\%$, для досліджуваного з балом 5: $(3/25)*100\%=12\%$ і т.д.

Найімовірніше, один результат буде нижче точки, що відповідає 5, а один – вище цієї точки. Таким чином, до числа випробовуваних, справжній бал яких менше 5, можна віднести трьох учнів, один з яких має бал 3, другий – бал 4 і третій – один з двох, що одержали бал 5, що у відсотковому відношенні складе $(3/25) \cdot 100\% = 12\%$. Це і є процентильний ранг, що відповідає 5, який забезпечує зручну інтерпретацію результатів учня: 12% учнів з нормативної вибірки зробили 5 або менше завдань тесту. Відповідно до введеного раніше визначення 12-й процентиль в групі з 25 випробовуваних дорівнює 5. Звертаючись до отриманих даних в третьому стовпці табл. 12, можна впевнено сказати, що первинний результат 5 балів є поганим, тому що він перевищує результати тільки 12% випробовуваних вибірки стандартизації. Це конкретний результат який легко сприймається, зручний в першу чергу для учнів при порівнянні досягнень за рядом тестів. Первинний результат, який нижче будь-якого показника вибірки стандартизації, має нульовий процентильний ранг.

Результат, який перевищує будь-який інший у вибірці, отримає процентильний ранг 100. Звичайно, ні нульовий ранг, ні ранг 100% не говорять про нульове або абсолютне знання контрольованого предмета.

Можливе розв'язання зворотної задачі, коли необхідно визначити р-й центиль, вірніше, точку, нижче якої лежать р% результатів. Для визначення р-го центиля необхідно виконати 5 кроків, які отримуються за допомогою табл. 14 і наводяться в табл. 15.

Таблиця 14

Зв'язок між сирими балами і частотами

Сирі бали досліджуваних	1	15	14	13	12	11	10	9	8	7	6	5	4	3
Частота	2	1	1	1	1	2	3	3	4	3	2	2	1	1
Кумульовані частоти	3	25	24	23	22	21	19	16	13	9	6	4	2	1

Таблиця 15

Визначення центилів

Крок	Етап обчислення	Приклад обчислення
1	Обчислення $(pn)/100\%$, де n – накопичена частота в групі оцінок	$p_{12} = \frac{25 \cdot 12}{100} = 3$
2	Визначення фактичної нижньої межі L розряду оцінок, які містять результат 1-го кроку	(4,5;5,5), L=4,5
3	Віднімання накопиченої до L частоти (f) із результату 1-го кроку (визначення частот, що лежать нижче $(pn)/100\%$)	3-2=1
4	Визначення долі інтервалу розрядів, що лежить під частотою $(pn)/100\%$	$\frac{1}{2}=0,5$
5	Додавання результатів 4-го кроку до результатів 2-го кроку. Результуюча формула: $P_p = L + \frac{(pn)/100\% - (f)}{f}$	4,5+0,5=5

Центилі не слід змішувати з процентними показниками, що представляють відсоток правильно виконаних завдань випробуваним групи. На відміну від останнього – первинного – центиль є похідним показником, що вказує на частку від загального числа випробовуваних групи.

Крім переваг, пов'язаних з простотою інтерпретації, центильні ранги мають два істотні недоліки. Перший полягає в тому, що центильні ранги є значеннями

порядкової шкали, оскільки показують відносне становище кожного індивіда в нормативній вибірці, а не виявляють відмінність між результатами окремих випробуваних групи.

Другий недолік певною мірою посилює перший – відсотки не лише не відображають, а навіть спотворюють реальні відмінності результату виконання тесту. Це пов'язано з особливостями розподілу процентилів, що має прямокутний характер.

Розподіл первинних показників істотно відрізняється від прямокутного і для хороших нормативно орієнтованих тестів наближається до нормальної кривої. У зв'язку з цим невеликі відхилення від середнього в центрі розподілу спостережуваних результатів значно збільшуються процентилями, у той час як відносно великі відхилення на краях кривих нормального розподілу будуть зменшені.

Згадані недоліки – головна причина того, що використання процентилів досить обмежена. У силу зручності і простоти їх застосовують в основному в тестах для самооцінки знань учнів.

§ 8. Методика і технологія тестування

8.1. Розробка інструкцій для проведення тестування

Інструкція до тесту – керівна вказівка, настанова, у якій описано, як виконувати завдання тесту має бути простою й стислою, але водночас містити інформацію щодо кожного з таких пунктів:

- 1) мета тесту;
- 2) час, відведений для виконання тесту;
- 3) як позначати правильну відповідь;
- 4) чи вдаватися до вгадування, якщо є сумніви щодо правильної відповіді.

Інструкція повинна містити вказівки на те, що випробуваний повинен зробити, яким чином виконувати завдання. Це може бути фраза «Вибрати правильну відповідь», «найбільш правильну відповідь», «усі правильні відповіді», «хоча б одну правильну відповідь» та ін. Інструкція повинна бути складена так, щоб завдання і спосіб його виконання були абсолютно зрозумілі будь-якому з випробуваних і не призводили до помилок.

Випробуваному важливо зрозуміти, що від нього вимагається, як він повинен виконувати завдання. Мало зрозуміти те, що необхідно встановити правильну послідовність, тобто виконати інтелектуальну операцію, але й те, як власне її здійснити. Крім цього, для багатьох завдань важливо і те, в якому порядку цю правильну послідовність відновлювати. Наприклад, завдання на встановлення правильної послідовності, в якій в якості елементів для впорядкування запропоновані прізвища авторів фізичних законів стає незрозумілим, якщо не вказати в якому саме порядку слід їх розташувати: в алфавітному, в хронологічному, за тривалістю життя, по тому, хто скільки законів відкрив, по географії проживання (зі сходу на захід) та ін. Для розробників тестових завдань підготовка інструкції для випробуваних у кожному завданні є необхідною, оскільки це дозволяє поглянути на завдання з точки зору випробуваного. Текст завдання або питання представляє собою змістовне наповнення

завдання. У деяких авторів можна зустріти виділені частини тексту завдання. Наприклад: Стимулюючий матеріал: матеріал, про який йдеться в питанні, представлений переважно у вигляді тексту, малюнка, таблиці або іншого подання даних.

8.2. Стандартизація процедури та умов тестування

Стандартизація умов тестування складається з:

- об'єктивності проведення тестування;
- об'єктивності опрацювання даних;
- об'єктивності інтерпретації результатів.

Принцип об'єктивності педагогічного контролю найбільш часто порушується у традиційних педагогічних системах, які наочно віддзеркалюють помилки і вади суспільства, що побудоване на егалітарному (зрівнювальному) принципі. Такому суспільству тестування не потрібно і навіть шкідливо, оскільки воно підриває підвалини особистої відданості і дає об'єктивну відповідь на питання - "хто є хто?". Яскравим прикладом цього є сама історія розвитку тестового контролю у більшості навчальних закладів світу.

Вимірювання вважаються об'єктивними, якщо вплив суб'єктивних дій тих, хто вимірює, є мінімальним. Іншими словами, вимірювання об'єктивні у тому випадку, коли результати цих вимірів максимально незалежні від дослідників. Це означає, що різні дослідники, вимірюючи одну й ту ж характеристику, мають одержувати однаковий результат. Щоб звести до мінімуму вплив суб'єктивних факторів, необхідна максимальна стандартизація умов проведення вимірювання та аналізу його результатів. До стандартизації умов експерименту ставляться такі вимоги:

- об'єктивність проведення вимірювання, що вимагає для всіх, хто підлягає іспиту, однакових умов його проведення. Для цього стандартизується сам процес вимірювання. Тільки тоді, коли витримано однакові умови експерименту, результати вимірювання можна порівнювати між собою;
- об'єктивність опрацювання даних, що вимагає фіксації даних, їх накопичення, аналізу та збереження;

- об'єктивність інтерпретації результатів, що вимагає однакової інтерпретації одних і тих же результатів вимірів різними дослідниками.

Це можливо лише за умови існування теоретичної концепції, яка лежить в основі даного методу вимірювання та оцінювання. В межах цієї теорії існують алгоритми аналізу та критерії, які є основою інтерпретації результатів. Таким чином, досліднику надається можливість прийняти єдине вірне рішення, яке є оптимальним у межах прийнятої теоретичної концепції.

Надійність методу вимірювання, – це ступінь стійкості результатів, що впливає на точність, з якою можна виміряти ту чи іншу конкретну ознаку. Перевірка надійності методу стосується насамперед відновлення результатів при повторних вимірюваннях. Ступінь надійності методу визначається за допомогою коефіцієнта надійності.

Коефіцієнт надійності дорівнює коефіцієнту кореляції між результатами, які одержані однакою методом та за однакових умов. Він показує наскільки співпадають результати вимірювань, які проведені в однакових умовах.

Ступінь надійності методу залежить від:

- об'єктивності методу,
- параметрів засобу вимірювання,
- стабільності характеристики, яку вимірюють.

Отже, при забезпеченні об'єктивності методу коефіцієнт надійності пропорційно залежить від характеристик засобу вимірювання, які визначають точність, з якою може вимірюватися той чи інший параметр. Точність вимірювання, згідно з теорією похибок, визначає значення похибки величини, яка вимірюється. Так, чим більша точність методу, тим менша похибка при вимірюванні і тим більша надійність цього методу.

Не можна не звернути уваги на необхідність стабільності характеристики, яка вимірюється. Особливо це стосується педагогічних вимірювань. Вони певною мірою залежать від багатьох внутрішніх та зовнішніх чинників, зокрема фізичного стану особи, яка проходить вимірювання.

Для контролю надійності тесту і якості засвоєння знань створюються батареї тестів, тобто сукупності тестів одного і того ж рівня.

8.3. Зразки інструкцій для проведення тестування

Приклади інструкцій

1. Для поточного (тематичного) контролю знань:

Уважно прочитайте умову завдання і виберіть одну правильну відповідь. Номер правильної відповіді впишіть в бланк у вільне місце під номером питання для відповіді.

Виправлення виключаються. Час виконання всіх завдань - 15 хвилин.

2. Для підсумкового контролю знань:

На виконання екзаменаційної роботи з інформатики відводиться 4 години (240 хвилин), включаючи роботу за комп'ютером.

Робота складається з 3 частин, які містять 37 завдань.

Частина 1 містить 24 завдання з вибором відповіді. До кожного завдання дається чотири відповіді, з яких тільки одна – правильна. Завдання виконуються на спеціальному бланку або на комп'ютері.

Частина 2 складається з 8 завдань з короткою відповіддю (до цих завдань ви повинні самостійно сформулювати та ввести відповідь). Завдання виконуються на спеціальному бланку або на комп'ютері.

Частина 3 складається з 5 завдань. Для їх виконання необхідно написати розгорнуту відповідь у довільній формі. При виконанні завдань частини 3 ви можете (за бажанням) використовувати комп'ютер, наданий Центром тестування та оснащений відповідним програмним забезпеченням.

Виконуйте завдання в тому порядку, в якому вони дані.

Якщо будь-яке завдання викликає у вас проблему, то пропустіть його і постарайтеся виконати завдання, у відповідях на які ви впевнені. До пропущених завдань можна буде повернутися пізніше.

За кожну правильну відповідь, залежно від складності завдання, дається один або більше балів. Бали, отримані вами за всі виконані завдання, сумуються. Постарайтеся виконати якомога більше завдань і набрати якомога більше балів.

Бажаємо успіху!

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

Нормативно-правова база

1. Закон України “Про вищу освіту” // Законодавство України про освіту. Збірник законів. – К. : Парламентське вид-во, 2002.
2. Закон України “Про загальну середню освіту” // Законодавство України про освіту. Збірник законів. – К. : Парламентське вид-во, 2002.
3. Закон України “Про освіту” // Законодавство України про освіту. Збірник законів. – К. : Парламентське вид-во, 2002.
4. Наказ МОН від 29 січня 2007 року № 64 “ Про організацію та проведення міжнародних порівняльних досліджень якості освіти учнів загальноосвітніх навчальних закладів”.
5. Указ Президента України від 17 квітні 2002 р. № 347 «Про Національну доктрину розвитку освіти» // У кн.: Законодавчі акти України з питань освіти. – К. : Парламентське вид-во, 2004.
6. Указ Президента України від 4 липня 2005 року №1013 “Про невідкладні заходи щодо забезпечення функціонування та розвитку освіти в Україні”.
7. Указ Президента України «Про невідкладні заходи щодо забезпечення функціонування та розвитку освіти в Україні» // <http://www.president.gov.ua/documents/2928.html>.
8. Закон України «Про вищу освіту» // <http://zakon.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=2984-14>.

Література

1. Аванесов В. С. Композиция тестовых заданий / Вадим Сергеевич Аванесов. – М. : Центр тестирования, 2002. – 240 с.
2. Адаптивное тестирование : учеб.-метод. пос. / [Н. М. Опарина, Н. Г. Полина, Р. М. Файзулин, И. Г. Шрамкова].– Хабаровск, 2007. – 95 с.
3. Анастаси А. Психологическое тестирование / А. Анастаси, С. Урбина. – СПб. : Питер, 2005. – 688 с.
4. Беспалько В. П. Основы теории педагогических систем (Проблемы и методы психолого-педагогического обеспечения технических обучающих систем) / Владимир Павлович Беспалько. – Изд-во Воронежского ун-та, 1977. – 304 с.
5. Білоусова Л. І. Потенціал комп'ютерного тестування / Л. І. Білоусова // Вісник ТІМО. – 2008. – № 10. – С. 40 – 44.
6. Булах І. Є. Створюємо якісний тест: навч. посіб. / І. Є. Булах, М. Р. Мруга. – К. : Майстер-клас, 2006 – 160 с.
7. Вісник ТІМО, 2005 – 2010 р.

8. Гавриленко І. М. Соціологічний моніторинг і діагностика в освіті / І. М. Гавриленко // Освіта і управління. – 1998. – № 2. – Т. 2. – С. 7 – 9.
9. Енциклопедія освіти / Акад. пед. наук України ; головний ред. В. Г. Кремень. – К. : Юрінком Інтер, 2008. – 1040 с.
10. Клайн П. Справочное руководство по конструированию тестов / Пол Клайн. – К., 1994. – 283 с.
11. Локшина О. І. Моніторинг якості освіти: світовий досвід / О. І. Локшина // Педагогіка і психологія. – 2003. – № 1. – С. 108 – 116.
12. Майоров А. Н. Мониторинг как практическая система / А. Н. Майоров // [http://www.mto.ru/ children/monitoring/system.html](http://www.mto.ru/children/monitoring/system.html).
13. Майоров А. Н. Теория и практика создания тестов для системы образования / Алексей Николаевич Майоров. – М. : Интеллект-Центр, 2001. – 296 с.
14. Майоров А. Н. Тесты школьных достижений : конструирование, проведение, использование / Алексей Николаевич Майоров. – СПб : Образование и культура, 1996. – 304 с.
15. Малафіїк І. В. Дидактика : навч. посіб. / Іван Васильович Малафіїк. – К. : Кондор, 2009. – С. 173 – 175.
16. Моніторинг якості освіти: становлення та розвиток в Україні : Рекомендації з освітньої політики / [під заг. ред. О. І. Локшиної]. – К. : “К.І.С”, 2004. – 160 с.
17. Педагогічне оцінювання і тестування. Правила, стандарти, відповідність. Наукове видання / [Болюбаш Я. Я., Булах І. Є., Мруга М. Р., Філончук І. В.] – К. : Майстер-клас, 2007. – 272 с.
18. Поддубный А. В. Методические основы разработки и использования педагогических тестов / Поддубный А. В., Панина И. К., Ащепкова Л. Я. – Владивосток, 2003. – 296 с.
19. Равен Д. Педагогическое тестирование: Проблемы, заблуждения, перспективы / Джон Равен. – М. : «Когито-Центр», 2001. – 142 с.
20. Рудницька О. П. Основи педагогічних досліджень / Рудницька О. П., Болгарський А. Г., Свистельнікова Т. Ю. – К. : Експрес-об'ява, 1998. – 144 с.
21. Самылкина Н. Н. Современные средства оценивания результатов обучения / Надежда Николаевна Самылкина. – М. : БИНОМ, 2007. – 172 с.
22. Чельшкова М. Б. Теория и практика конструирования педагогических тестов : учеб. пособ. / Марина Борисовна Чельшкова. – М. : Логос, 2002. – 432 с.

Навчальне видання

КОНСТРУЮВАННЯ ТЕСТІВ
КУРС ЛЕКЦІЙ

Навчальний посібник

TEMPUS PROJECT
Educational Measurements
Adapted to EU Standards



Видано за підтримки програми **TEMPUS IV** на виконання міжнародного проекту «Освітні вимірювання, адаптовані до стандартів ЄС»

Відповідальний за випуск Гадяк Ж.В.

Надруковано з оригінал-макету, наданого авторами

Підписано до друку 09. 09. 2010 р.
Формат 60x84/16. Папір офсетний. Друк цифровий.
Гарнітура Times New Roman.
Умовно-друк. аркушів –
Тираж 100 примірників.
Замовлення
