



USAID
ВІД АМЕРИКАНСЬКОГО НАРОДУ

ПРОЕКТ “СПРАВЕДЛИВЕ ПРАВОСУДДЯ”

ПРАКТИЧНИЙ ПОСІБНИК ДЛЯ РОЗРОБНИКІВ ТЕСТОВИХ ЗАВДАНЬ

УПОРЯДКОВАНО НА ОСНОВІ ТРЕНІНГОВИХ МАТЕРІАЛІВ,
РЕКОМЕНДАЦІЙ ТА НАПРАЦЮВАНЬ ЕКСПЕРТІВ ПРОЕКТУ

Контракт No. AID-121-C-11-00002

Упорядник: Сергій Мудрук, експерт з тестових технологій Проекту USAID
«Справедливе правосуддя»

Вересень 2014 року

Погляди авторів, викладені у даній публікації, не обов'язково відображають офіційну точку зору Агентства США з міжнародного розвитку (USAID) або Уряду Сполучених Штатів Америки.

Зміст

ВСТУП.....	7
І ІНСТРУКЦІЇ ДЛЯ РОЗРОБКИ ТЕСТОВИХ ЗАВДАНЬ. <i>Стівен Баккер</i>.....	9
1. Передмова.....	9
2. Плани тестів	11
2.1. Вимір змісту.....	11
2.2. Вимір поведінки (розумових операцій).....	12
2.2.1. Таксономія Блума.....	12
2.2.2. Рамки FCAT	12
2.2.3. Рамки TIMSS.....	13
2.2.4. Глибини знань (DOK)	13
2.2.5. Структура інтелекту Гілфорда	15
2.2.6. «Простір для тестового завдання».....	17
2.3. Класифікація тестових завдань	18
3. Аналіз тестів і тестових завдань	19
3.1. Середній бал тесту.....	19
3.2. Стандартне відхилення	20
3.3. Складність тестів і тестових завдань	22
3.4. Співвідношення результатів тестових запитань і результатів екзамену в цілому	23
3.5. Надійність тесту	24
3.6. Стандартна похибка вимірювання.....	26
4. Розробка тестових завдань.....	27
4.1. Термінологія	27
4.2. Процес розробки тестових завдань.....	27
4.3. Формати тестових завдань.....	28
4.3.1. Формати завдань із варіантами відповідей	28
4.3.2. Тестові завдання з розширеними відповідями	30
4.4. Критерії якості когнітивних тестів	30
4.4.1. Валідність.....	30

4.4.2. Надійність	31
4.4.3. Об'єктивність	31
4.4.4. Ефективність	31
4.4.5. Прийнятність	32
4.4.6. Контрольний перелік	32
5. Розробка тестових завдань із варіантами відповідей	33
5.1. Валідність	33
5.1.1. Чітка мета	33
5.1.2. Релевантна мета	33
5.1.3. Відповідність сучасним розумінням	34
5.1.4. Релевантні дистрактори (відволікаючі варіанти)	34
5.2. Надійність	35
5.2.1. Нечіткий текст запитання чи варіантів відповіді	35
5.2.2. Маловідомі слова чи вирази	36
5.2.3. Подвійне заперечення	37
5.2.4. Заперечення тексту запитання; заплутане запитання	37
5.2.5. Відкритий текст завдання	37
5.2.6. Більш, ніж одна правильна відповідь	38
5.2.7. Ненавмисні підказки	39
5.2.8. Здоровий глузд	41
5.2.9. Розумність тесту	41
5.3. Неупередженість	42
5.4. Ефективність	42
5.5. Прийнятність	43
II. ОГЛЯД СТАНДАРТІВ У СФЕРІ ПРОФЕСІЙНОГО	
ЕКЗАМЕНУВАННЯ. <i>Сергій Мудрук</i>	44
1. Державні стандарти в Україні	44
2. Міжнародний стандарт ISO 17024	44
3. Стандарти AERA/APA/NCME	45

3.1. Перелік стандартів AERA/APA/NCME рекомендованих для першочергового ознайомлення:	46
3.2.Переклад деяких стандартів освітнього і психологічного тестування AERA/APA/NCME	47
4.Кодекс справедливої практики тестування в освіті	48
ІІІ. МОНІТОРИНГ ПРОЦЕСУ АНОНІМНОГО ТЕСТУВАННЯ КАНДИДАТІВ НА ПОСАДУ СУДДІ, ПРОВЕДЕНОГО ВККСУ 5 ЧЕРВНЯ 2012 РОКУ (витяг). <i>Сергій Мудрук</i>.....	50
1. Анонімне тестування і стандарти екзаменування	50
1.1. Застосування Стандартів до роботи ВККСУ	50
1.2. Завади відповідності стандартам екзаменування	50
2. Норма чи критерій?	51
3. На що націлене, і на що має бути націлене анонімне тестування?	51
4. Яке тіло знань і наскільки глибоко має оцінювати анонімне тестування ?	52
5. Рівень вимог та пробний тест (приклад тесту)	53
6. Відкритість vs конфіденційність	54
7. Процедура подання документів кандидатами	54
8. Аналіз відповідності Стандартам.....	55
9. Аналіз практики	60
10. Узагальнення спостережень та висновки	61
11. Рекомендації.....	63
ІV. ДВНАДЦЯТЬ КРОКІВ ЕФЕКТИВНОГО СТВОРЕННЯ ЕКЗАМЕНУВАННЯ. <i>Сергій Мудрук, за матеріалами Стівена Даунінга</i>	67
V. МІНІ-КУРС З РОЗРОБКИ ЗАВДАНЬ ДЛЯ ТЕСТІВ. <i>Слайди презентації. Стівен Баккер</i>	70
VI. ВАЛІДНІСТЬ ТЕСТУ, КЛЮЧОВЕ ПИТАННЯ РОЗРОБКИ ТЕСТІВ. ЩО МОЖНА ВИЗНАЧИТИ ЗА ДОПОМОГОЮ ЦЬОГО ТЕСТУ? <i>Слайди Презентації. Стівен Баккер</i>	82
VII. АНАЛІЗ ТЕСТУ І ЗАПИТАНЬ: ЯК СТАТИСТИКА МОЖЕ РОЗШИРИТИ ВАШІ ПОГЛЯДИ. <i>Слайди презентації. Стівен Баккер</i>	91
VIII. ЯКІСТЬ ТЕСТУ ТА КРИТЕРІЇ ЯКОСТІ ТЕСТУ. <i>Слайди презентації. Леонід Серета</i>	102

IX. ПОКАЗНИКИ ЯКОСТІ ОКРЕМОГО ЗАВДАННЯ. <i>Слайди презентації. Леонід Серета</i>	121
X. ПРАКТИЧНІ ПРАВИЛА ДЛЯ РОЗРОБНИКІВ ТЕСТОВИХ ЗАВДАНЬ. <i>Слайди презентації. Сергій Мудрук</i>	136
XI. ТИПОВА КАРТКА ТЕСТОВОГО ЗАВДАННЯ	168
XII. СПРОЩЕНИЙ КОНТРОЛЬНИЙ СПИСОК ПО ТЕСТОВИХ ЗАВДАННЯХ. <i>Сергій Мудрук</i>	170
Контент	170
Формат	171
Стиль	171
Варіанти відповідей.....	172
Неупередженість і гумор	173
XIII. ПРИКЛАДИ АНАЛІЗУ ТЕСТОВИХ ЗАВДАНЬ. <i>Сергій Мудрук, за підтримки Секретаріату Вищої кваліфікаційної комісії суддів України</i>	175
XIV. ЗАГАЛЬНЕ ПЛАНУВАННЯ ТА ЕТАПИ РОЗРОБКИ ТЕСТОВИХ ЗАВДАНЬ. РОЗПОДІЛ РЕСУРСІВ. <i>Сергій Мудрук</i>	188

ВСТУП

Проект USAID «Справедлива, незалежна та відповідальна судова влада України» (скорочена назва – Проект USAID «Справедливе Правосуддя») розпочав свою роботу в Україні 1 жовтня 2011 року. Метою Проекту є підтримка законодавчої, нормативної та інституційної реформи судових установ з метою зміцнення відповідальності та незалежності судової влади України. Його основними завданнями є підтримка діяльності USAID/Україна, спрямованої на укріплення верховенства права, демократії та управління шляхом: 1. приведення нормативно-правової бази судової реформи у відповідність до європейських та міжнародних норм задля підтримки відповідальності та незалежності судової системи; 2. підвищення рівня прозорості та відповідальності діяльності ключових судових установ; 3. підвищення професіоналізму та ефективності судової системи України; та 4. посилення ролі громадських організацій, які відстежують та популяризують судову реформу.

У відповідності до завдань, поставлених перед Компонентом 2.1, Проект USAID «Справедливе правосуддя» працює над розбудовою потенціалу Вищої кваліфікаційної комісії суддів України (ВККСУ) щодо розробки та запровадження об'єктивного, прозорого, заснованого на сучасних, науково обґрунтованих технологіях та методах добору кандидатів на посаду судді.

Забезпечення справедливості оцінювання досягається створенням однакових умов участі у тестуваннях, своєчасним інформуванням учасників щодо програми тестування/іспиту, а також забезпеченням відповідної якості тестових завдань. З метою сприяння ВККСУ та Національній школі суддів України в підготовці авторів тестових завдань та формуванні банку тестових запитань високої якості, Проект USAID «Справедливе правосуддя» періодично проводить тематичні тренінги, експертні дослідження та надає експертні рекомендації.

Цей Посібник є добіркою матеріалів, які використовувались у навчанні розробників тестових завдань, що проводилось Проектами USAID «Верховенство права» та «Справедливе правосуддя» у період з 2009 по 2014 роки.

Посібник буде корисним учасникам тренінгів для оновлення знань, але, в першу чергу, він розрахований на самостійне опрацювання та засвоєння матеріалу майбутніми розробниками тестових завдань.

Рекомендований порядок освоєння матеріалу відповідає послідовності розміщення матеріалів у цьому посібнику.

- Розділ I. В першу чергу варто освоїти «Інструкції для розробки тестових завдань» (Стівен Баккер) що надасть читачеві базове, але комплексне розуміння предмету.
- Розділ II. Стаття щодо основних стандартів, використовуваних у царині екзаменування з метою допуску особи до професії («Огляд стандартів у сфері професійного екзаменування», Сергій Мудрук) дозволить читачеві оглядово ознайомитись з загальними нормативними документами, що регламентують процес створення та проведення екзаменування.
- Розділ III. Витяг зі звіту 2012р («Моніторинг процесу анонімного тестування кандидатів на посаду судді, проведене ВККСУ 5 червня 2012 року. Рекомендації щодо покращення правил адміністрування, оцінювання та шкалювання анонімного

тестування») дозволить читачеві ознайомитись з прикладом встановлення відповідності діяльності екзаменуючої організації одному зі Стандартів.

- Розділ IV. Матеріал «Дванадцять кроків ефективного створення екзаменування» (Сергій Мудрук) структурує процес створення системи екзаменування для розробника тестових запитань, а також буде корисним для вивчення посадовими особами, що відповідають за побудову і підтримку системи конкурсного добору кандидатів на посаду судді.

Далі в Посібнику розміщено підбірку презентаційних слайдів. Ознайомлення з ними дозволить заглиблено пройти важливі для розробника тестових запитань теоретичні і практичні теми.

- Розділ V. Слайди презентації «Мінікурс зі складання завдань для тестів» (Стівен Баккер) нагадують частину термінології, поясняють процес створення екзаменаційного завдання, нададуть порівняння (недоліки і переваги) різних форматів екзаменаційних завдань і нададуть певні прості рекомендації зі складання тестових завдань.
- Розділ VI. Слайди презентації «Що можна визначити за допомогою цього тесту? Валідність тесту, ключове питання розробки тестів» (Стівен Баккер) дають читачеві ключове поняття тестології – валідність, а також поняття плану змісту тесту (матриця тесту, таксономія).
- Розділ VII. Слайди презентації «Аналіз тесту і запитань: Як статистика може розширити ваші погляди» (Стівен Баккер) надають поняття надійності і похибки вимірювання, методи оцінки якості завдань через кореляцію «питання» - «весь тест».
- Розділ VIII. Слайди презентації «Якість тесту та критерії якості тесту» (Леонід Середа) надають поняття базових характеристик тестів (надійність, валідність, стандартна похибка вимірювання).
- Розділ IX. Слайди презентації «Показники якості окремого завдання» (Леонід Середа) надають базові терміни та поняття сучасної теорії тестів (IRT).
- Розділ X. Слайди презентації «Практичні правила розробки тестових завдань» (Сергій Мудрук), разом з «Інструкціями для розробки тестових завдань» Стівена Баккера, є ключовими матеріалами Посібника. «Практичні правила» є квінтесенцією досвіду практичної роботи великої кількості розробників тестових завдань. Ця презентація містить також вступну частину з повторенням термінології і понятійного апарата розробки тестових завдань.
- Розділ XI. Приклад формалізованої «Картки тестового запитання» у випадку використання її при виконанні розробок, дозволить не втратити (не забути) важливі параметри при створенні тестового завдання.
- Розділ XII. «Спрощений контрольний список по тестових завданнях» (Сергій Мудрук) є прикладом списку контрольних запитань, яким в своїй роботі може керуватись і розробник тестових завдань (для здійснення самоконтролю якості), і рецензент чи контролер якості.
- Розділ XIII. Підбірка прикладів тестових запитань («Приклади аналізу тестових завдань», Сергій Мудрук) з коментарями та технічною аналітикою дозволяє закріпити Практичні правила.
- Розділ XIV. Стаття «Загальне планування та етапи розробки тестових завдань. Розподіл ресурсів» (Сергій Мудрук) дозволить читачеві, майбутньому розробнику тестових завдань, мати реалістичні орієнтири щодо розподілу власних сил під час здійснення розробки тестових запитань.

І ІНСТРУКЦІЇ ДЛЯ РОЗРОБКИ ТЕСТОВИХ ЗАВДАНЬ.

Стівен Баккер

1. Передмова

Когнітивні тести, що використовуються в освіті та навчанні – це інструменти вимірювання, які складають основу для прийняття рішень. Це можуть бути рішення, починаючи від відповідей на питання «чи засвоїв учень навчальну програму на достатньому рівні, аби отримати сертифікат?» або «чи цей кандидат засвоїв відповідні навички та вміння для того, щоб пройти навчання кандидата на посаду судді?», і закінчуючи такими як «чи належний рівень освіти в країні?». Такі тести слід пристосовувати до мети, з якою вони використовуються, і вони повинні надавати релевантну і надійну інформацію для прийняття відповідного рішення. Тест, який є саме таким, називається «валідним тестом».

Отож, поняття «валідності» стосується відповідності, значущості та корисності конкретних висновків, зроблених за результатами тестування. Будь-які тести, що використовуються для прийняття важливих рішень – наприклад, тести для добору та сертифікації майбутніх суддів – повинні давати валідні (надійні) результати. Іншими словами: щоб рішення стосовно того, чи видавати сертифікати про прийом студентів до університету чи на програми професійного навчання, прийняті на основі цих тестів, були обґрунтованими і – якщо необхідно – щоб їх можна було відстояти в суді. Частину «валідності» можуть складати докази того, що склад тесту відповідає узгодженому плану; ще один вагомий доказ – задовільні психометричні характеристики тесту та окремих тестових завдань.

Тестові завдання – це будівельні блоки тестів, кожен з яких повинен докладатися до валідності тесту. Добре розроблені тестові завдання відділяють кандидатів, які мають якості, необхідні для отримання сертифікату чи для допуску до подальшої освіти, від кандидатів, які таких якостей не мають. Знання та розуміння, необхідні для того, щоб дати правильну відповідь, повинні бути релевантними, завдання повинні бути сформульовані чітко і недвозначно. Правильна відповідь повинна справді означати, що учень отримав знання і досягнув розуміння, перевірити які має це тестове завдання, без випадкових підказок у тексті завдання чи – якщо йдеться про тестові запитання з варіантами відповідей – не функціональних відволікаючих варіантів (неправильних відповідей).

На випускному шкільному екзамені було запропоноване таке тестове завдання:

- Який фараон збудував найвищу піраміду?
- A. Аменхотеп IV
 - B. Рамзес I
 - C. Рамзес II
 - D. Тутанхамон

Хоча у підручниках, можливо, чітко і зазначена висота пірамід та імена фараонів, які їх збудували, однак відповідь на це тестове запитання не є свідченням розуміння студентом давньоєгипетської культури. Релевантні запитання – це чому фараони будували піраміди, значимість знахідок у пірамідах тощо, а не звуження часів епохи фараонів до змагань у будівництві пірамід.

Малюнок 1. Змагання з будівництва пірамід

Завдання про піраміду (Малюнок 1) вочевидь, не таке завдання, яке

допоможе відрізнити слабких учнів від сильних на будь-якій релевантній основі.

Автори тестових завдань повинні постійно пам'ятати про два важливі запитання, які визначають якість тестового завдання, яке вони створюють: (1) «яка мета тесту, для якого я пишу тестове завдання, і чи зміст цього тестового завдання відповідає такій меті?», а також (2) «чи формулювання цього завдання і його технічне представлення такі, що воно недвозначно і надійно визначить підготованих кандидатів і непідготованих?».

Запропоновані Інструкції допоможуть авторам тестових завдань перевіряти свої творіння і покращити свою практику завдяки формулюванню низки правил та їх ілюстрації реальними прикладами тестових завдань.

Однак перш, ніж розповісти про те, що слід робити, і те, чого не слід робити при розробці хороших тестових завдань, слід зазначити ще дві теми, які є важливими для правильного розуміння мистецтва розробки тестових завдань: це план тесту і психометричні характеристики тестів і тестових завдань

.

2. Плани тестів

Так як будинок не можливо збудувати без плану, – плану будинку, в якому буде зазначено розташування та розміри кімнат, типи будівельних матеріалів, водопровідно-каналізаційну систему тощо – так і тест слід скласти виключно на основі *плану тесту*.

План тесту – це відображення мети тесту, визначення знань та вмінь, які потрібно оцінити, очікуваний рівень засвоєння, типи і кількість тестових завдань, які повинні бути у тесті для прийняття обґрунтованого рішення на основі результатів тестування.

Зазвичай, «знання та навички» розділені на дві чи більше різних категорій. Для тестів, спрямованих на відбір кандидатів на спеціальну підготовку суддів, чи остаточний вибір кандидатів на посади суддів, звичайно ж, однією із таких категорій повинні бути знання та розуміння змісту законодавства та теорії права. Ще одна категорія визначає, які розумові операції повинен вміти виконати кандидат, використовуючи знання та розуміння змісту, чи які загальні вміння повинен продемонструвати кандидат. Такі різні категорії можна об'єднати і створити так звану тестову матрицю. Часто ці категорії називають «вимірами» тестової матриці: вимір змісту, вимір розумових операцій (чи «поведінки») і вимір вмінь.

2.1. Вимір змісту

Вимір змісту Математика

1. Арифметика і алгебра
 - 1.1. Цифри
 - 1.2. Рівняння
 - 1.3. Лінійна алгебра
 - 1.4. Підрахунки
 - 1.5. ...
2. Геометрія
 - 2.1. ...
3. Математичний аналіз
 - 3.1. ...
4. Статистика і ймовірність
 - 4.1. ...
5.

Малюнок 2. Вимір змісту тесту з математики

Вимір змісту – це всього лиш перелік тем, іноді він розподіляється на основні категорії та підкатегорії, як от у прикладі з математики на Малюнку 2.

Зазвичай, вимір змісту охоплює 15-25 тем, що дає можливість розробникам тесту легко надати окремі тестові завдання до кожної з них. У деяких випадках окремі тестові завдання охоплюватимуть кілька тем, в такому разі має бути можливість визначити, яка категорія змісту є вирішальною для визначення правильної відповіді, а яка є менш важливою.

Існування багатьох тестових завдань, які репрезентують три чи більше категорій змісту, – це показник надмірної деталізації виміру змісту, і слід докладати зусиль, щоб структурувати категорії так, аби кожне тестове завдання

однозначно належало до однієї, максимум двох категорій.

Відкриті тестові завдання, які вимагають два чи більше кроки розв'язку, можуть дати змогу виділяти часткові бали за різні категорії, таким чином вказуючи вагу декількох категорій змісту у правильній відповіді на тестове запитання.

2.2. Вимір поведінки (розумових операцій)

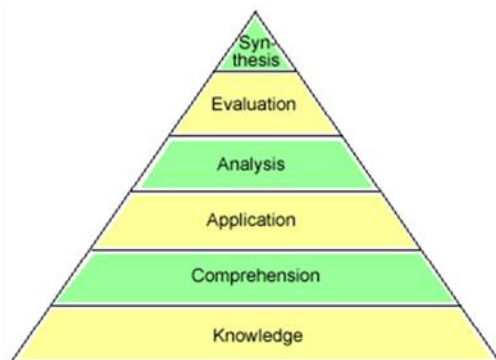
Одна справа – активізація знань та навичок, які вивчалися та набувалися в рамках навчальної програми, при визначенні відповіді на запитання, зовсім інша – їх використання для вирішення проблеми. Чи суть у простому відтворенні якоїсь частини знань, як от, наприклад, співставлення назви столиці із назвою країни, чи суть у демонстрації вивченої процедури, як от пошуки x з рівняння $x^2 + 2x - 3 = 0$? Чи це щось більше, і для того, аби визначити правильну відповідь знадобляться навички вищого рівня? Розумові процеси, які, як припускається, відбуваються в мозку кандидата, складають другий вимір тестової матриці. Оскільки ми можемо лише опосередковано спостерігати за цими процесами, судячи про них з поведінки кандидатів, то цей вимір називається «виміром поведінки при вирішенні проблем», або коротко: «виміром поведінки». Дехто називає його «виміром вирішення проблем», ще інші – «виміром навиків».

2.2.1. Таксономія Блума

Під час розробки описів когнітивних сфер використовувалося чимало теорій для розробки відповідного поведінкового виміру. Одна з найперших теорій – таксономія Блума (малюнок 3).

Складність: як, чому?

- Синтез
- Оцінка
- Аналіз
- Застосування
- Розуміння
- Знання



Низька когнітивна складність: що, коли, хто, де?

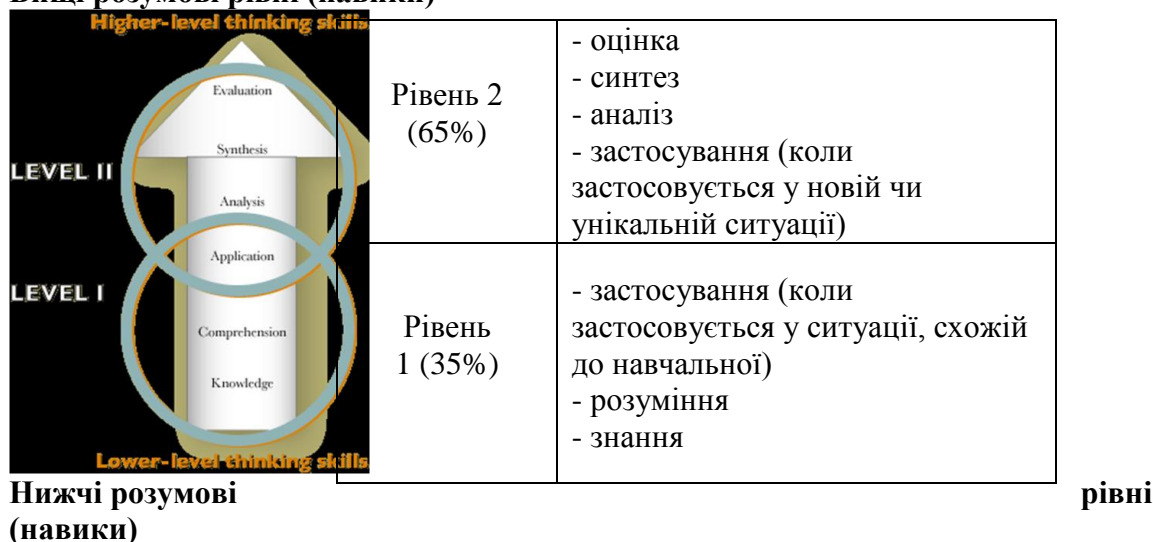
Малюнок 3. Таксономія Блума

Таксономія Блума надає хорошу і добре відому структуру, але видається, що простіше застосувати цю таксономію до оцінки класної роботи, аніж до стандартизованого тестування. На практиці досить складно поділити тестові завдання на шість різних категорій. Надто часто автори тестових завдань не сходяться у думках щодо того, яка точно навичка потрібна для вирішення конкретного тестового завдання, і це веде до тривалих і непродуктивних дискусій. Більше того, вищі категорії часто дуже складно оцінити у ті часові рамки і в тому форматі, в яких проводиться стандартизоване тестування.

2.2.2. Рамки FCAT

Зараз все частіше простежується тенденція використання лише кількох рівнів на відміну від таксономії Блума. Для прикладу, рамки Тесту загальної оцінки у Флориді (Florida Comprehensive Assessment Test) (FCAT) визначають когнітивні навички, як показано нижче (Малюнок 4), і використовують цю схему для класифікації тестових завдань за двома рівнями.

Вищі розумові рівні (навики)



Малюнок 3. Класифікаційна схема FCAT

2.2.3.Рамки TIMSS

Ще один типовий приклад перерозподілу категорій Блума наведений на схемі, що використовується проектом TIMSS – великим міжнародним дослідженням освітніх досягнень, яке було вперше проведено у 1995 році.

«TIMSS Advanced» збирає міжнародні порівняльні дані про досягнення учнів з метою приведення шкільних програм у відповідність з науково-орієнтованими програмами в університеті. Вимір поведінки складається із трьох категорій: знання, застосування і обґрунтування (Малюнок 5)

Удосконалений вимір поведінки TIMSS	
Знання	Обґрунтування
Згадати	Проаналізувати
Визнати	Узагальнити
Підрахувати	Синтезувати/ Інтегрувати
Відтворити	Обґрунтувати
	Вирішити нерутинні проблеми
Застосування	
Відібрати	
Представити	
Змодельовати	
Вирішити рутинні проблеми	

Малюнок 4. Удосконалений вимір поведінки TIMSS

2.2.4.Глибини знань (DOK)

На даний момент у США для валідації державних тестів широко використовується схема класифікації «Глибини знань» (DOK), розроблена Норманом Веббом. Це прагматичне поєднання загальних навиків і навиків, специфічних для якогось предмету.

Загальна схема DOK має чотири рівні вимог до розумових процесів, які є однаковими для усіх предметів:

Рівень 1: просте відтворення

Рівень 2: необхідна певна розумова обробка

Рівень 3: необхідне складніше/стратегічне розмірковування

Рівень 4: необхідне розширене мислення

Для кожного предмету чи сфери ці чотири рівні розробляються по-своєму, як показано нижче для математики (Малюнок 6) і фізики (Малюнок 7).

Рівні глибини знань (DOK) Математика	
<p>Рівень 1 (Згадати) відтворення факту, визначення, термін чи проста процедура здійснення простого алгоритму чи застосування формули.</p> <p>Рівень 2 (Навики та поняття) як у 1, але також рішення, як підійти до вирішення проблеми чи до діяльності. Наприклад: тлумачення простого графіка</p>	<p>Рівень 3 (стратегічне мислення) вимагає розмірковування, планування, із використанням свідчень, а також вищого рівня мислення, аніж два попередні рівні. Наприклад: Витлумачення складного графіка, яке вимагає прийняття рішень щодо того, які характеристики графіку потрібно розглянути і яку інформацію можна зібрати із цього графіка в одне ціле. У більшості випадків, коли від учнів вимагається пояснити їхнє мислення, – то це рівень 3.</p> <p>Рівень 4 (Розширене мислення) вимагає комплексного розмірковування, планування, розробки та мислення упродовж, найімовірніше, тривалішого періоду часу.</p>

Малюнок 6. Рівні глибини знань для тестів з математики

Підхід DOK є прагматичним. Більшість практиків розуміють різницю між навиками вищого та нижчого рівня. Недолік концепції DOK полягає у тому, що два вищі рівні не часто використовуються для класифікації стандартизованих тестових завдань.

Більшість завдань у тестовому форматі (MCQ) – це рівень 1 і 2. Дуже мало завдань підпадають під рівень 3, якщо взагалі такі будуть. Тестові завдання, що вимагають розширеної відповіді (пояснить, чому ...), найбільше підходять для рівня 3. Тип завдань рівня 4, зазвичай, не підходить для письмових тестів, але може складати частину аудиторних проектів.

У поєднанні з виміром змісту схема DOK може використовуватися для складання проекту тестової матриці.

У Таблиці 1 показано частину такої матриці для фізики. Перша колонка містить основні предметні категорії, у другій колонці 2 наведено їх поділ на підкатегорії. Колонки, позначені як рівень навиків (1,2,3), – це рівні DOK. У клітинах матриці позначені кількість тестових завдань чи кількість балів, що відповідають конкретному поєднанню змісту та вміння.

Тематичні блоки	Частини	Рівень навиків			Загалом
		Рівень 1	Рівень 2	Рівень 3	
Механіка	Основи кінематики	2	1		3
	Основи динаміки	2	1		3
	Закони збереження механіки	1	3		3
	Елементи механіки рідин та газів	1	2		
Молекулярна фізика і термодинаміка	Основи молекулярно-кінетичної теорії	3	1		
	Основи термодинаміки	3	2		
	Властивості газів, рідин і твердих речовин	1	2		

Малюнок 7 Рівні глибини знань у матриці тесту з фізики

2.2.5. Структура інтелекту Гілфорда

Схема ДОК тісно пов'язана з рівнями професійної кваліфікації, що використовуються у США для моніторингу ефективності викладання та навчання. Досить часто ці рівні пов'язані з рівнем складності завдань, у тому сенсі, що завдання, на які дає правильну відповідь невеликий відсоток кандидатів, як припускається, вимагають високого рівня професійної кваліфікації, і навпаки.

Однак це не завжди так. На деякі тестові завдання, які вимагають простого відтворення, можна отримати правильну відповідь всього від кількох кандидатів і такі завдання назвуть складними, але вони не відповідають опису жодного із рівнів 2-4. На такий спосіб мислення, мабуть, впливають такі таксономії, як таксономія Блума, що передбачають ієрархію розвитку навиків.

Хоча це може бути корисним для викладача в розробці плану лекції, однак думки кандидата не обов'язково слідує тій самій ієрархії, коли він відповідає на тестове запитання. З цих причин СІТО, Голландський національний інститут освітніх вимірювань, наприкінці 1970-х вирішив розробити схему класифікації, яка базуватися на теорії роботи мозку при вирішенні завдання, а не на розвитку навиків. У цій схемі використовувався операційний вимір Структури інтелекту Гілфорда (SI), як основа для виміру поведінки. Шість рівнів структури Гілфорда було поділено на 3, як показано на Малюнку 8.



Малюнок 8. Вимір поведінки у вирішенні проблеми, що базується на Операційному вимірі Структури інтелекту Гілфорда

Цей трирівневий вимір використовується для планування державних шкільних іспитів у Нідерландах. Приклад частини такого плану для екзаменів з біології наведено на Малюнку 9.

Поведінка/ Предмет	Відтворення А	Відтворення Б	Продуктування (створення)	Підсумок
Живі істоти				8
Метаболізм				6
Харчування	2	2	4	8
Дихання				6
Рух				4
Репродукція				4
Гормони				8
...				
Підсумок	10	30	60	100

Малюнок 9. Тестова матриця для екзаменів з біології

Схема на малюнку 10 показує, як автори тестових завдань можуть діяти, зараховуючи завдання до однієї із трьох категорій виміру поведінки Структури інтелекту Гілфорда.

Вирішальна відмінність полягає між створенням та відтворенням, ключові слова тут – «звична ситуація, відпрацьована та добре відома процедура» на противагу «незвичному, невідомому та новому підходу».

Слід зазначити, що останній не означає, що учні повинні виявляти якісь знання поза межами програми. Знання та вміння за програмою – це наче інструменти у скриньці столяра.

Знання щодо того, який використовувати для якої операції, це відтворення А (Ra). Його застосування з метою створення шафки за стандартним планом, – це відтворення В (Rb). Створення нестандартної шафки, що виражає «повноту природи, красу нашого ремесла і важливість нашої сім'ї» (а саме таким було завдання, яке багатий купець у 17-му столітті давав майстру-столяру), – це і є Створення.



Малюнок 10. Схема «Класифікаційна поведінка»

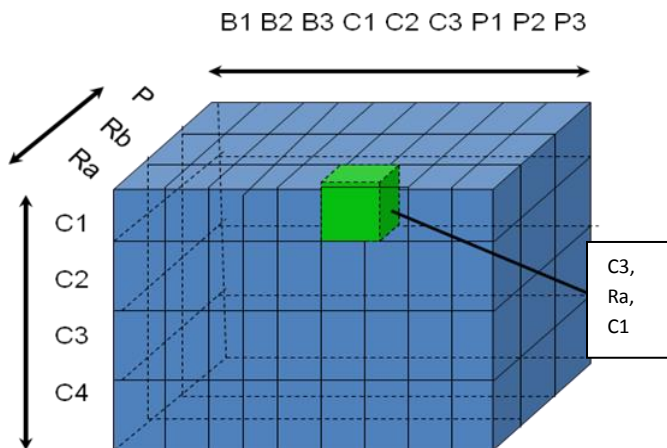
2.2.6.«Простір для тестового завдання»

Зміст, поведінку та інші виміри, які формують «простір для тестового завдання», світ, де «живуть» тестові завдання, так само, як і світ, де ми живемо, можна описати у просторових вимірах.

Візьмімо, для прикладу, науковий тест, що включає три виміри, предметний зміст (біологія «В», хімія «С», фізика «Р»), поведінку при вирішенні проблеми (P, Rb, Ra) та вміння (C1, C2, C3, C4).

Відтак, усі завдання мають три координати і «живуть» у просторі, окресленому цими трьома вимірами.

У кожній клітинці живуть завдання з однаковими координатами, для прикладу, на Малюнку 11 у клітині усі завдання з ярликами Хімія 3, Відтворення А і Вміння 1. По суті, у кожній клітинці є місце для невизначеної кількості тестових завдань.



Малюнок 11. Тривимірний простір для тестового завдання

2.3.Класифікація тестових завдань

Тести, особливо підсумкові¹ тести – це, в основному, зразки, взяті з багатьох тестових завдань, які живуть у «просторі для тестових завдань», сфері, яку, як припускається, цей тест повинен охоплювати.

Тест для добору суддів на 100 тестових завдань, до прикладу, – це всього лиш підбірка, взята з десятків тисяч тестових завдань, які можна придумати, і всі з яких тестують окремі аспекти знань та навиків у сфері судочинства, що ними повинні володіти кандидати, коли подають заявку на допуск до підготовки для майбутніх суддів. На основі такої вибірки із всього лиш 100 тестових завдань слід вирішити, чи кандидат досягнув достатнього рівня, чи ні. Відтак, цю вибірку слід відбирати надзвичайно ретельно.

План тесту визначає правила відбору тестових запитань і повинен гарантувати, що ця вибірка є репрезентативною і охоплює усі категорії знань та навиків, які слід оцінити, і настільки, наскільки це вказано у плані, який, у свою чергу, відображає важливість окремих категорій. Щоб дати можливість здійснювати такий керований відбір, усі тестові завдання повинні мати маркування, на яких повинно бути позначено, який зміст та поведінку вони оцінюють. З цієї причини, класифікація тестових завдань у відповідності до змісту та поведінки складає важливу частину процесу підготовки тестових завдань і включення їх в базу завдань.

Класифікація також відіграє важливу роль в рецензуванні тестових завдань, що розробляються, адже вони відкривають мету, яку ставив перед собою автор при створенні тестового завдання. При рецензуванні команда повинна узгоджувати не лише наступну версію тестового завдання, а й класифікації. Віднести тестове завдання до якоїсь категорії предметного змісту (як обговорювалося у розділі «Вимір змісту» на сторінці 11) – це, зазвичай, завдання не складне, на відміну від його віднесення до якоїсь категорії поведінки та вмінь. Візьмімо, для прикладу, категорії відтворення та створення Гілфорда. Відмінність між цими двома категоріями, в першу чергу, регулюється «звичним» контекстом на

¹ Тести, які проводяться наприкінці якогось тривалішого періоду навчання і охоплюють значні обсяги знань та навиків і спрямовані на оцінку того, чи засвоїв кандидат їх достатньо, аби перейти на наступний рівень (клас, курс) чи отримати сертифікат.

противагу «незвичному» чи «вивченому» підходу на противагу «невивченому, самостійно спланованому».

Природно, у прийнятті рішення щодо того, чи є даний контекст звичним чи ні для якогось конкретного кандидата і чи спосіб вирішення проблеми є по-справжньому новим, чи, можливо, він все-таки вивчений, велику роль відіграє суб'єктивний фактор. Завдання групи розробників тестів полягає у тому, щоб обговорити контекст і вирішити, що стосується більшості кандидатів, а не що справджується для груп кандидатів, звичних лише кожному з розробників.

3. Аналіз тестів і тестових завдань

Результати тестів – це бали окремих учнів за окремі тестові завдання. Їх можна підсумувати у так званій матриці даних. Приклад наведено нижче у таблиці на Малюнку 12.

Кандидат	Тестове завдання 1	Тестове завдання 2	Тестове завдання 3	Тестове завдання 4	Тестове завдання 5	Тестове завдання 6
1231001	A	D	A	C	B	3
1231002	A	C	B	C	B	1
1232003	A	C	B	C	A	2
1232004	B	D	B	B	A	0

Малюнок 12. Матриця даних, 4 кандидати, 6 тестових завдань

У цій матриці даних наведено результати 4 кандидатів, позначені цифровим кодом, і 6 тестових завдань. Завдання 1-5 – це завдання з множинними варіантами відповідей на 4 варіанти (A,B,C,D), а завдання 6 – це відкриті завдання з балами 0,1, 2 і 3. Такі матриці даних можна аналізувати за допомогою статистичних пакетів програмного забезпечення.

Окрім результатів тестів для окремих кандидатів, такий аналіз надає важливу інформацію щодо якості тесту і окремих тестових завдань, основні з яких:

- Середній бал тесту
- Стандартне відхилення результатів тесту
- Складність тесту і тестового завдання
- Співвідношення результатів тестових завдань і тесту
- Надійність тесту
- Стандартна похибка вимірювання у тесті

Розуміння цієї інформації є важливим для розробників тесту, оскільки воно допомагає їм покращити розробку тестових завдань, рецензування і збір (генерацію) екзаменаційних модулів з бази тестових завдань.

3.1. Середній бал тесту

Середній бал тесту – це сума усіх результатів кандидатів, поділена на кількість кандидатів.

Кандидат (N=4)	1-A	2- D	3-A	4-C	5-B	6	Всього
1231001	1	1	1	1	1	3	6
1231002	1	0	0	1	1	1	4
1232003	1	0	0	1	0	2	4
1232004	0	1	0	0	0	0	1
Всього/і	3	2	1	3	2	6	17
Середній бал тесту							4,25

Малюнок 13. Матриця результатів і показників

У Таблиці на Малюнку 13 наведено результати кандидатів із Таблиці Малюнку 12, отримані із застосуванням ключа (набору правильних відповідей) до їхнього вибору варіантів відповідей. У якості ключа можна взяти відповіді кандидата №1231001 (перший з чотирьох в таблиці). В останній колонці наведені сумарні результати. Всього ці кандидати набрали 17 балів на тесті. Середній бал тесту – це загальний бал усіх учнів (17), поділений на кількість учнів, які складали тест (4), отож у даному випадку виходить $17/4=4,25$.

3.2.Стандартне відхилення

Результати тесту часто представляються у вигляді розподілу частоти, яка показує, скільки кандидатів набрали якусь конкретну кількість балів. У Таблиці на Малюнку 14 наведено частину розподілу результатів державного випускного екзамену, який складали 511 кандидатів.

Результат	Част	Відс%	Сум част	Сум%
-----------	------	-------	----------	------

40	11	2,15	127	24,85
41	15	2,94	142	27,79
42	16	3,13	158	30,92
43	11	2,15	169	33,07
44	7	1,37	176	34,44
45	15	2,94	191	37,38
46	20	3,91	211	41,29
47	15	2,94	226	44,23
48	11	2,15	237	46,38
49	17	3,33	254	49,71
50	9	1,76	263	51,47
51	17	3,33	280	54,79
52	16	3,13	296	57,93
53	19	3,72	315	61,64
54	14	2,74	329	64,38
55	17	3,33	346	67,71
56	16	3,13	362	70,84
57	14	2,74	376	73,58
58	14	2,74	390	76,32
59	11	2,15	401	78,47
60	15	2,94	416	81,41

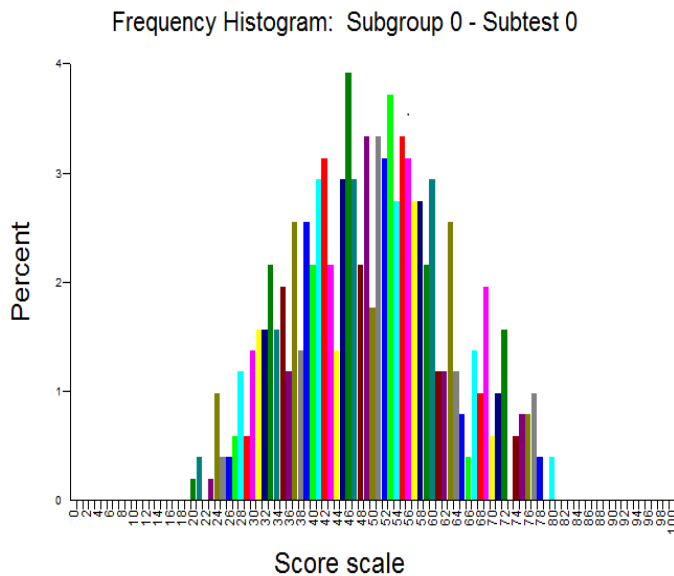
Пояснення до таблиці розподілу частотності

Результати були виражені за шкалою від 0 до 100. Найнижчий отриманий результат склав 20, а найвищий – 80. Показані лише результати від 40 до 60 (перша колонка). У другій колонці наведено кількість кандидатів, які отримали відповідну кількість балів. У третій колонці показано відсоток кандидатів, у яких був такий результат. Останні дві колонки, підсумовуючі, показують кількість та відсоток кандидатів, у яких був певний результат чи нижчий результат. Для прикладу: 9 учнів мали результат 50; тобто це 1,76 % учнів, які склали цей тест, і 263 або 51,47 % учнів набрали 50 чи менше балів. Якщо результат 51 чи більше потрібен для того, щоби скласти тест, т.з. «прохідний бал», то 263 або 51,47% кандидатів не склали б його.

Малюнок 14. Частина розподілу частотності результатів

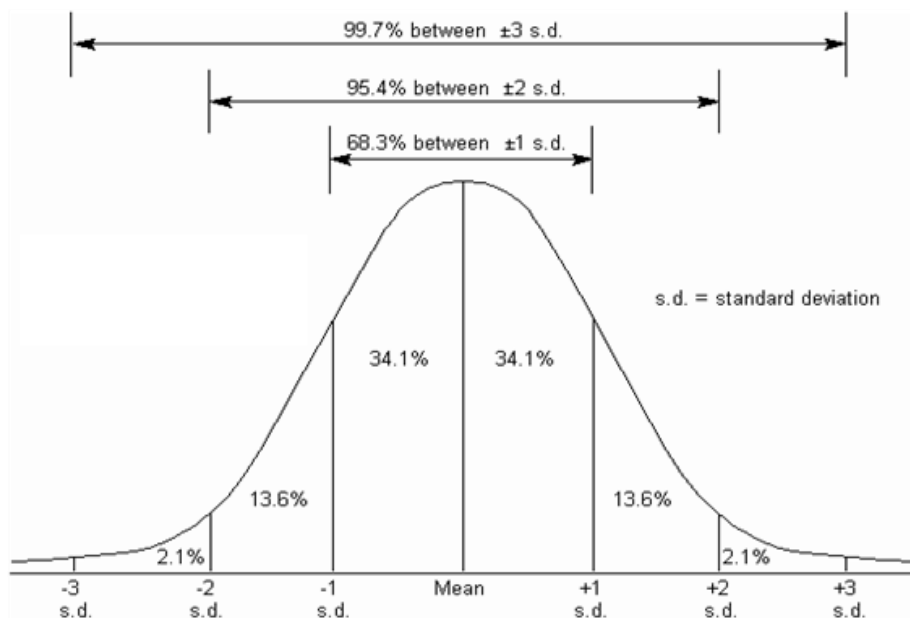
Такий розподіл частотності можна також представити у вигляді гістограми, як показано на Малюнку 14.

Тут кожний стовпчик показує кількість кандидатів з певним балом. Крива, що огинає гістограму такої форми, називається кривою Гаусса, або Нормальним розподілом (Малюнок 15).



Малюнок 15. Гістограма розподілу частотності, показана у Таблиці Малюнку 14

Стандартне відхилення (SD – standard deviation) – це міра розкиданості результатів тесту і, також показник вимірювання ширини кривої.



Нормальний розподіл чи крива Гаусса

Малюнок 16. Стандартне відхилення

SD виражається у балах на шкалі результатів. Оскільки мета тесту – розрізнити кандидатів за рівнем підготовки, і відрізнити сильних кандидатів від слабких, то одним із якісних показників тесту є стандартне відхилення. Низьке стандартне відхилення вказує на те, що більшість кандидатів отримали майже однаковий результат.

СТАТИСТИЧНІ ПОНЯТТЯ

Стандартне відхилення:

- Це рівень коливання у розподілі результатів
- Це «середнє» відхилення результатів від середнього результату

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}}$$

Стандартне відхилення від 1/5 до 1/6 за шкалою результатів може вважатися нормальним. Відтак, за бальною шкалою від 0 до 100 очікується стандартне відхилення 16-20. У прикладі, наведеному на Малюнку 11, у розподілі балів стандартне відхилення =12,7 (також див. Малюнок 14), і це означає, що тест не достатньо широко розкидав результати для цієї вибірки. Як видно із гістограми на Малюнку 11, нижчий кінець (0-20) та вищий кінець (80-100) шкали взагалі не задіяно, і, можливо, це зумовлено наявністю більших груп «крайніх» тестових завдань: надто легких і надто складних.

3.3.Складність тестів і тестових завдань

У Таблиці Малюнку 17 наведено так звані р-показники для тих самих тестових завдань, як показано у Таблиці Малюнку 13. Р-показники – це показники складності, які вказують на відсоток правильних відповідей, отриманих групою кандидатів. р-показники підраховуються шляхом додавання усіх результатів кандидатів за одне тестове завдання (Разом (i)) і діленням цієї суми на суму теоретично можливих найкращих результатів за це тестове завдання, Макс (i). Останнє – це просто кількість учнів, які набрали максимальний результат за це завдання. Для завдань MCQ (тестового формату з вибором однієї правильної відповіді серед декількох запропонованих) р-показники – це еквівалент відсотку кандидатів, які дали правильну відповідь на це завдання.

Також для цього тестового завдання в цілому можна підрахувати р-показник (р-величину), як показано в останній колонці Таблиці Малюнку 17.

Учень	1-A	2- D	3-A	4-C	5-B	6	Разом
1231001	1	1	1	1	1	3	6
1231002	1	0	0	1	1	1	4
1232003	1	0	0	1	0	2	4
1232004	0	1	0	0	0	0	1
Разом (i)	3	2	1	3	2	6	17
Макс (i)	4	4	4	4	4	12	32
P	.75	.50	.25	.75	.50	.50	.53

Малюнок 17. Р-показники

Тести спрацьовують оптимально, якщо вони складаються із тестових завдань, які не є ні надто простими, ні надто складними. У такий спосіб отримується найкраща диференціація серед кандидатів. Дуже складні тести дозволяють добре відрізнити рівні підготовки серед сильних кандидатів, але залишають більшість інших кандидатів без значущих (відмінних один від одного) результатів. Те ж саме стосується надто простих тестів: багато високих

результатів, але замало відмінностей результатів у кандидатів. Тестові завдання з р-показником 0,55 найкраще диференціюють результати кандидатів.

Відсоток кандидатів, які обрали правильну відповідь

Відсоток кандидатів, які обрали не правильну відповідь

Item nr.	Weight	Key	P- and A- val			
			A	B	C	D
1	1	B	19	78*	4	0
2	1	B	15	63*	7	11
3	1	C	7	7	70*	15
4	1	B	15	70*	4	7

У запитаннях з декількома варіантами відповідей важливими є не лише Р-показники, а й відсоток учнів, які вибрали один із неправильних варіантів відповідей, так званих дистракторів, відволікаючих варіантів. Дистрактори, обрані невеликим відсотком кандидатів, можуть бути не функціональними.

Малюнок 18: Таблиця 18. р- та а-показники

У Таблиці на Малюнку 18 показано частину результату так званого Аналізу тесту і тестових завдань (ТІА). У третій колонці показано правильну відповідь (ключ), а у колонках справа для кожного варіанту відповіді показано відсоток учнів, які їх вибрали. Відсоток учнів, що обрали правильну відповідь (р-показник), у цих результатах позначений зірочкою. Відсоток учнів, які вибрали неправильні варіанти, називаються а-показниками. Для тестового завдання 1 р-показник складає 78, а-показники – 19, 4 і 0.

3.4.Співвідношення результатів тестових запитань і результатів екзамену в цілому

Показник вимірювання того, наскільки добре ключі відповідей розмежовують сильних та слабких учнів. Може набувати величини між -1 та +1, але повинен бути позитивним

Показник вимірювання того, наскільки добре ключі відповідей розмежовують сильних та слабких учнів. Може набувати величини між -1 та +1, але повинен бути негативним

Item nr.	Weight	Rir and Rar valu			
		A	B	C	D
1	1	-37	38*	-7	0
2	1	-23	52*	-27	-18
3	1	-11	28	13*	-29
4	1	-38	49*	2	-8

Складність тестового завдання – це один аспект, а інший – його дискримінаційна здатність. Тестові завдання у тесті повинні послідовно розмежовувати сильних і слабких учнів. Спосіб перевірки цього – перевірка співвідношення результатів за окремими тестовими завданнями і результату екзамену в цілому, або – іншими словами – слід відповісти на запитання: чи кандидати, які дають правильну відповідь на тестове завдання,

Малюнок 19: Показники r_{it} та r_{ir}

мають вищі результати тестування, ніж кандидати, які дають неправильну відповідь? На це запитання дає відповідь статистична операція, до якої залучено два коефіцієнти кореляції: r_{it} та r_{ir} .

Перший, r_{it} , це показник кореляції балів за одне тестове завдання з результатами тесту в цілому, другий (r_{ir}) – кореляція балів за одне тестове завдання з результатами тесту в цілому, не включаючи це тестове завдання. r_{ig} – це точний показник.

Ці коефіцієнти кореляції коливаються від -1,00 до +1,00. Показник кореляції +1,00 означає 100% зв'язок, або, якщо виразити це словами: усі кандидати, які правильно відповіли на це тестове запитання, отримали вищий результат тестування, ніж усі кандидати, які дали неправильну відповідь на нього.

Однак на практиці так ніколи не буває. Зазвичай, показники r_{ir} коливаються від +0,10 до +0,60.

Від'ємна величина r_{ir} – це чіткий сигнал про те, що тестові завдання неправильно функціонують: більшість учнів, які правильно відповідають на це тестове завдання, отримали нижчі результати тесту, аніж ті, які неправильно на нього відповіли. Це може бути пов'язано з якоюсь непередбаченою перевагою слабших учнів в плані змісту чи способу формулювання завдання, однак, зазвичай, це зумовлено помилкою у ключах з відповідями.

Співвідношення а-показників (r_{ar}) із загальними результатами тесту має бути прямо протилежним. Тут від'ємний показник означає, що більшість учнів, які вибирають цей варіант, отримують нижчі результати тестування, а саме цього і очікують автори тесту.

У Таблиці Малюнку 19 наведено показники r_{ig} (позначені зірочкою) і r_{ar} для одних і тих же тестових завдань, як у Таблиці на Малюнку 18. Для завдання 1 величина r_{ig} складає 0,38, а величини r_{ar} – -37, -7 та 0.

Ці коефіцієнти кореляції – важливі показники якості тестових завдань у тесті. Ось деякі правила:

1. r_{ir} повинен бути позитивним, і бажано $>0,25$
2. r_{ar} повинен бути нижчим, ніж 0,10
3. r_{ir} повинен бути вищим, ніж r_{ar}

Див. тестове завдання 3: правила 1, 2 і 3 порушено. Серед кандидатів, які вибрали варіант В (неправильний), було чимало сильніших кандидатів, а серед тих, які обрали варіант С (який вважався правильною відповіддю), було надто багато слабких кандидатів. Мабуть, ключем тут є відповідь В, а не С. Однак у такому випадку також буде порушено правило 2.

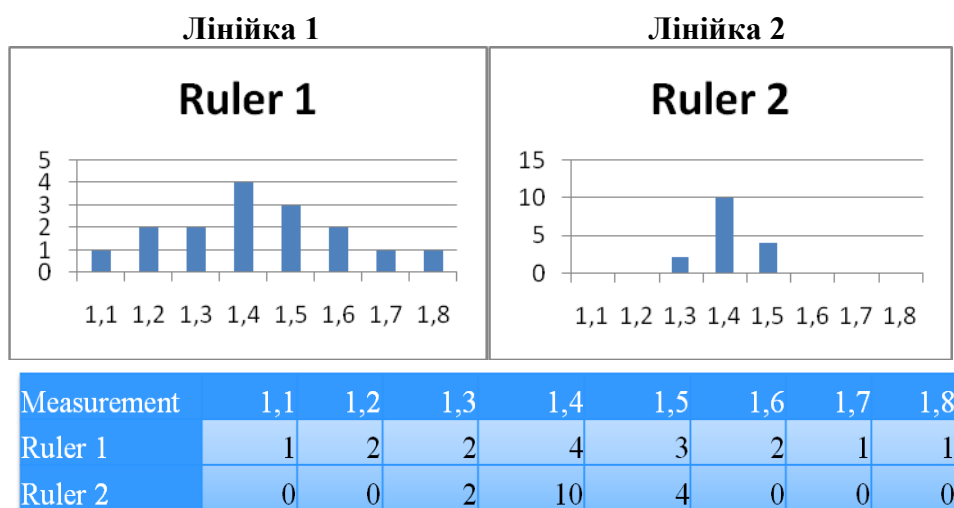
3.5. Надійність тесту

Тести – це інструменти для вимірювання, а результати – це бали кандидатів, які означають показник вимірювання їхніх знань та розуміння, здатностей, компетентності, придатності до роботи, а також інших завдань, що ставляться перед тестом.

Ці вимірювання повинні бути надійними і відповідати важливості результатів для кандидатів та тих, хто використовуватиме результати. «Надійні» можна розтлумачити як «повторювані вимірювання одного об'єкту, що ведуть до однакового результату».

Вимірювання стержня довжиною 1,4 см лінійкою повинно давати результати 1,4 см більшість разів, але деякі вимірювання відхилятимуться від такої «справжньої» довжини.

Для лінійки 2 на Малюнку 20 часто виходить по-іншому. Вимірювання менш надійною лінійкою, як от лінійкою 1 на тому ж малюнку даватимуть результати, що частіше відхиляються і відтак ведуть до менш точної оцінки довжини стержня.



Малюнок 20. Вимірювання двома різними лінійками

Статистичні методи, що підтверджують послідовність, з якою вибірка розподіляється на сильніших і слабших кандидатів, за допомогою тесту, дають оцінку надійності цього тесту. Надійність зазвичай виражається коефіцієнтом, величина якого знаходиться десь у межах між 0 і 1,00.

Часто використовується коефіцієнт α , або Альфа Кронбаха, за Лі Кронбахом, введений у 1951 році.

Низька Альфа Кронбаха показує, що тест «внутрішньо непослідовний», тобто деякі тестові завдання ділять групу осіб, які складають тест, у дуже відмінний спосіб на сильних та слабких, ніж інші тестові завдання, і, відтак, результати тестів не є надійними.

Зазвичай, показники 0,75-0,85 сприймаються як прийнятний мінімум. Надійність тесту залежить не лише від його внутрішньої послідовності, а й від його довжини (кількості завдань): надійність тесту зростає з його довжиною.

Subgroup number	: 0	SubTest number	: 0
Number of persons in test	: 511	Number of selected items	: 100
Minimum test score	: 0	Maximum test score	: 100
Average test score	: 49,67	Standard deviation	: 12,66
Average P-value	: 49,67	Std. Error of Measurement	: 4,47
Average Rit	: 0,28		
Coefficient Alpha	: 0,88	SE coeff. Alpha	: 0,01

90% Confidence limits for Coefficient Alpha: (0,86 =< 0,88 =< 0,89)

Estimated Coefficient Alpha if this test had a standard norm length of 40 items: 0,74 (Spearman-Brown)

Номер підгрупи

Кількість осіб, що складають тест

Мінімальний результат тесту

Середній результат тесту

Середня р-величина

Середня величина Rit

Коефіцієнт Альфа

Довірчий інтервал Альфа при впевненості 90% : (0,86 =< 0,88 =< 0,89)

Підрахований коефіцієнт Альфа, якщо цей тест був би стандартної тривалості і складався б із 40 тестових завдань: 0,74 (Спірман-Браун)

Номер підтесту

Кількість відібраних тестових завдань

Максимальний результат тесту

Стандартне відхилення

Стандартна похибка вимірювання

Стандартна похибка Коефіцієнта Альфа

Малюнок 21. Аналіз статистичних даних за допомогою програмного забезпечення TiaPlus на основі вибірки із 511 осіб, що складала тест

У Таблиці на Малюнку 21 наведено приклад аналізу даних кандидатів за допомогою TiaPlus² (тест на зрілість для загальної гімназії, варіант А).

Цей тест зі 100 тестових завдань має коефіцієнт Альфа 0,88, і це хороший показник. Важливий внесок до цієї надійності вносить тривалість (довжина) тесту.

Якщо ми підраховуємо (використовуючи формулу прогнозування Спірмана-Брауна) тестову надійність, якщо у тесті було б лише 40 тестових завдань – це звична довжина предметних тестів з множинними варіантами відповідей – надійність зменшилася б до 0,74, якраз нижче мінімуму (див. останній рядок у Таблиці Малюнку 21).

3.6. Стандартна похибка вимірювання

STANDARD ERROR OF MEASUREMENT

$$SEM = s_x \sqrt{1 - \rho}$$

s_x = standard deviation of observed scores

ρ = reliability coefficient

27

Стандартна похибка вимірювання

s_x - стандартне відхилення отриманих результатів

ρ - коефіцієнт надійності

Малюнок 22. Стандартна похибка вимірювання

Коли коефіцієнт надійності і стандартне відхилення тесту відомі, можна підрахувати стандартну похибку вимірювання, SEM, як показано на Малюнку 22. SEM – це вимірювання неточності результатів кандидатів, пов'язана з ненадійністю тесту.

SEM надає впевненість для тлумачення та зіставлення результатів: при 95% точності справжній результат учня лежить десь у проміжку отриманого результату \pm SEM. Для прикладу, у Таблиці на Малюнку 21 вище, SEM складає 4,47.

Це означає, що для студента, результат якого складає 60 пунктів, ми знаємо з точністю 95%, що справжній результат цього учня лежить десь у проміжку між 55,53 і 64,47. Чи іншими словами, якщо він складає той самий чи схожий тест повторно, у 95% випадків цей учень набере бали приблизно в цьому ж проміжку.

Це має цікаві наслідки для визначення прохідного балу, тобто результату, нижче якого кандидат вважається таким, що не склав тест (див. сторінку 20).

Припустимо, що прохідний бал у нашому прикладі складе 62. Кандидат з 60 балами не склав тест, але SEM показує, що це може бути пов'язано із браком точності в інструментах вимірювання і що справжній результат цього кандидата може бути 62, або навіть 63 чи 64. З таких міркувань у деяких випадках SEM береться до уваги при визначенні прохідного балу шляхом зведення цього 1 SEM до запланованого прохідного балу. У нашому випадку ця цифра склала би від 62 до 58.

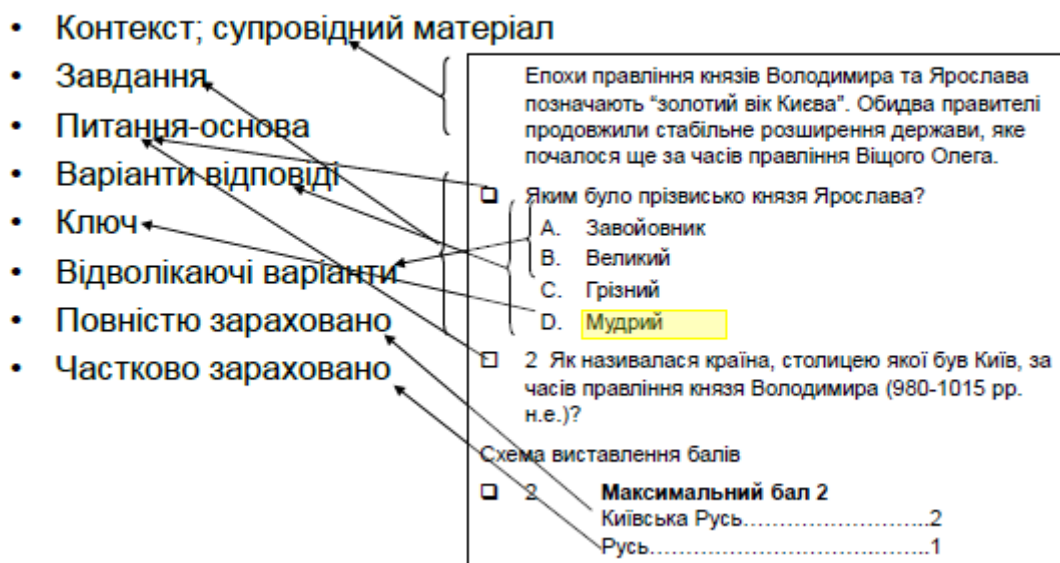
Тепер ми знаємо з точністю 95%, що усі кандидати, які не склали тест, мають справжній бал нижче 62. Кандидати з отриманими результатами 59-61 – які, можливо, мають реальний результат 62 чи вище – отримують перевагу.

² Пакет програмного забезпечення, підготований СІТО, Нідерланди, для класичного статистичного аналізу тесту і тестових завдань.

4.Розробка тестових завдань

Наступні сторінки присвячені розробці тестових завдань, і тут будуть наведені поради щодо того, що робити і чого уникати при створенні тестових завдань, за допомогою яких можна створити валідні, надійні, об'єктивні та ефективні тести. Після вступу до процесу створення і форматів тестових завдань, буде надано вказівки щодо розробки тестових завдань із варіантами відповідей (вибір вірної відповіді серед запропонованих, або *selecter response* – SR)) і розширеними відповідями (відкритих, або *constructed response* -- CR).

4.1.Термінологія



Малюнок 23. Термінологія завдань із варіантами відповідей (SR)

На малюнку 23 показано назви різних компонентів тестових завдань SR і CR. Зазвичай, тестові завдання розпочинаються з тексту, у якому наведено інформацію, яка повинна тлумачитися і використовуватися при наданні відповідей на тестові запитання. Це називається контекстом, або **стимулюючим матеріалом** (стимулом).

Іноді два чи більше тестових завдання мають один і той самий стимулюючий матеріал. Частина, у якій міститься реальне запитання, називається **текстом тестового завдання**.

Тестові завдання SR мають **варіанти відповідей**, з яких вибирає кандидат.

Варіант, в якому міститься правильна відповідь, називається **ключем**.

Інші варіанти – це **відволікаючі варіанти** (дистрактори).

Для тестових завдань CR важливою рисою є часткове зарахування: кандидати можуть відповісти на це тестове запитання не повністю правильно, але тим не менше отримати бал, оскільки відповідь є частково правильною.

4.2.Процес розробки тестових завдань

У процесі розробки тестових завдань, зазвичай, здійснюються такі кроки:

1. Вибір контекстуальних матеріалів
2. Розробка тестових завдань
3. Рецензування тестових завдань і їх перегляд автором (через деякий час)
4. Рецензування тестових завдань і їх перегляд другим рецензентом
5. Попереднє тестування
6. Перегляд третім рецензентом

Кроки 4-6, можливо, необхідно буде повторити. Нерідко під час рецензування виникає питання, чи залишити тестове завдання і далі його рецензувати, чи виключити тестове завдання і досягти мети тестування розробивши інше завдання. Розробники тестових завдань мають бути критичними і, не вагаючись, «викидати тестове завдання», замість того, щоб докладати зусилля для збереження ідеї завдання, яка є насправді недостатньо дієздатною.

4.3.Формати тестових завдань

Формати тестових завдань – це інструмент у руках розробника тесту, а не мета сама по собі.

Мета визначає формат, і не інакше.

Рішення використовувати конкретний формат залежить виключно від питання: який формат найкраще підходить для досягнення мети тесту.

Тестові завдання SR, вочевидь, мають перевагу, яка полягає у тому, що полегшується об'єктивний підрахунок балів, але вони можуть не підходити для оцінки певних якостей, коли кандидат, як припускається, повинен самостійно формулювати складний ланцюг обґрунтування, виражати і відстоювати певну точку зору, тощо.

Тестові завдання CR вимагають складних інструкцій щодо перевірки екзаменаційних робіт.

Перевірка великої кількості відповідей силами оцінників – це недешева справа, і відмінність у тлумаченні відповідей між оцінниками може бути додатковим критерієм при оцінці ненадійності тесту.

Хороша стратегія, зазвичай, полягає у тому, аби спершу спробувати поставити запитання у форматі SR і лише тоді скористатися CR, якщо використання SR виявиться неможливим без значного погіршення валідності.

4.3.1.Формати завдань із варіантами відповідей

Часто використовуються такі формати для тестових завдань із варіантами відповідей (SR):

- Варіанти відповідей
3-6 варіантів; зазвичай 4. Дихотомічні тестові завдання (з варіантами відповіді так/ні) не рекомендуються у зв'язку з високим рівнем вгадування. Ключі та відволікаючі варіанти пов'язані з однією і тією ж метою тесту, відтак відволікаючі варіанти не повинні бути випадковими.
- Підбір
Один комплекс подій, характеристик, назв тощо слід підставити/з'єднати до іншого комплексу. Перший комплекс зазвичай має на один об'єкт менше, ніж другий. Перевага такого типу полягає у тому, що в одному тестовому завданні поняття в цілому можна оцінити шляхом підбору його аспектів до назв тощо. Недолік полягає у впливі неправильного підбору на подальший підбір, і це

викликає такі питання: якщо цей підбір правильний, а той – ні, то що це нам говорить про розуміння кандидата? Якщо комплекс із 5 об’єктів слід підставити до іншого комплексу із 6, то можливими є 20 різних варіантів відповідей, і всі з них потрібно оцінити і підрахувати. Простий підхід на зразок: усе підібрано правильно = 5 балів, підібрано 4 правильних варіанти – це 4 бали, тощо, може не підійти.

Підставте країни і континенти

Країни:

1. Йорданія
2. Нова Зеландія
3. Грузія
4. Уганда
5. Перу

Столиці:

- A) Веллінгтон
- B) Тбілісі
- C) Кампала
- D) Амман
- E) Ліма

Правила визначення балів:

- 1 варіант підбору правильний 1 бал
- 2 варіанти підбору правильні 2 бали
- 3 варіанти підбору правильні 3 бали
- 4 варіанти підбору правильні 4 бали
- усі варіанти підбору правильні 5 балів

На Малюнку 24 показано тестове завдання з реального тесту з географії. Зверніть увагу на те, що «4 варіанти правильні» і «усі варіанти правильні» – це, по суті, одне і те ж: якщо у Вас 4 правильних варіанти підбору, то 5-й не може бути неправильним.

Малюнок 24 Погане тестове завдання з підбору

- Розташування за порядком

Поставте комплекс подій, дій, тощо, у правильній послідовності, для прикладу, у хронологічній послідовності. Таке завдання може бути корисним при оцінці розуміння учнем процедури чи ходу історичного процесу, але у ньому виявляється такий самий недолік, як і у тестових завданнях з підбору.

- Тест на заповнення

Процедура заповнення – це діяльність «заповніть пробіли», коли учень використовує припущення з контексту для того, щоб доставити слова, які зумисне усунуті з тексту. Процедура заповнення – це тест на розуміння при читанні. Відповіді показують і розуміння тексту, і рівень володіння мовою.

Світовий Банк	океан тонуті години не тонучий вижив	рятувальні човни холодний айсберг США половинки	1912 морська подорож підводний човен довгий океанографи

«Титанік» був висококласним океанським лайнером, який потонув у свій перший _____. Цей велетенський корабель мав майже 900 футів (269 метрів) _____ і рекламувався як «_____».

У ніч 14 квітня _____ «Титанік» зіштовхнувся з _____, коли пропливав у Північно-Атлантичному _____. Корабель плів із Британії до _____. За кілька _____ корабель _____, і загинуло близько 1500 людей.

Більшість із них потонуло у _____ воді, оскільки не було достатньо _____ для всіх. Лише 705 осіб _____.

У 1985 році _____ виявили залишки «Титаніка» на дні океану за допомогою ехолокатора і дистанційного контролю _____. Гігантський корабель, коли тонув, розвалився на дві частини, і дві _____ тепер лежать на відстані сотні футів, окрім уламків, які розкидані між передньою і задньою частиною.

Малюнок 25. Текст на заповнення

На Малюнку 25 наведено приклад тестового завдання на заповнення. Слова з бази слів потрібно вставити у відповідні пробіли. По суті, це своєрідне завдання на підбір. При підготовці інструкції з перевірки робіт слід вирішити, чи усі варіанти підбору дають однакові бали, чи слід визначати певні зразки і виставляти результати для них.

4.3.2. Тестові завдання з розширеними відповідями

Для завдань CR можна виділити такі типи:

- Коротка відповідь
Відповідь одним чи декількома словами, цитатою чи цифрою. Зазвичай, їх можна розробити краще в форматі завдання SR
- Розширена відповідь
Відповідь – (складне) речення або ж вона містить більш, ніж один елемент відповіді.
- Есе
Відповідь у вигляді есе, що відображає структуру, як цього вимагає інструкція до запитання.

4.4. Критерії якості когнітивних тестів

Кожне тестове завдання слід розробляти, пам'ятаючи про загальні критерії якості тестування. П'ять основних критеріїв такі:

1. Валідність
2. Надійність
3. Об'єктивність
4. Ефективність
5. Прийнятність

У наступних абзацах наведено загальні вказівки щодо дотримання цих критеріїв при розробці тестових завдань.

4.4.1. Валідність

Тести мають надавати можливість користувачам тестів приймати валідні рішення основі за результатами тесту. Відтак, індивідуальні тестові завдання повинні перевіряти саме те, що від них очікується, і ніщо інше.

- Упевніться, що Ви можете вказати мету тестування. Класифікацію тестових завдань у відповідності до розділів (вимірів) План тесту слід наводити разом із завданням, яке пропонується.
- Упевніться, що це тестове завдання знаходиться «в рамках програми».
- Упевніться, що це завдання стосується відповідного аспекту навчальної програми; уникайте банальних запитань чи запитань «на здоровий глузд».

- Упевніться, що тестове завдання відповідає поточному стану науки і думкам спеціалістів з цього питання.
- Кожна відволікаюча відповідь повинна бути обґрунтованою.

4.4.2.Надійність

Кожне завдання повинно забезпечувати послідовне розмежування (дискримінацію) слабких і сильних кандидатів

- Дотримуйтесь чітких і стислих формулювань; уникайте двозначності. Це основна із вимог, оскільки тексти запитань, які збивають з пантелику, ведуть до масштабного вгадування і значної втрати інформативності екзамену, і через це тест стає, по суті, коротшим, ніж планувалося, і руйнується дискримінація слабких і сильних кандидатів.
- Уникайте ненавмисних підказок.
- Упевніться, що ключі розкидані навмання серед неправильних варіантів відповіді, аби уникнути результативного використання вказівок для вгадування (розповсюджена вказівка - «якщо не знаєте відповідь, завжди вибирайте С»).

4.4.3.Об'єктивність

Усі кандидати повинні мати рівні можливості показати, що вони можуть зробити (а не що вони НЕ можуть зробити).

- Перевіряйте, щоб не було жодних гендерних, етнічних чи культурних упереджень. «Упередження у тестовому завданні» – це явище, коли певні тестові завдання можуть давати переваги (чи навпаки) окремим групам кандидатів, і такі переваги не мають нічого спільного з метою тестування. Добре відомий приклад гендерного упередження, коли хлопці набирають більше балів із запитань з фізики про механічні іграшки, ніж дівчата. Це пов'язано з тим, що хлопці краще знайомі з такими іграшками, і вони їм подобаються більше, ніж дівчатам. З цим і пов'язаний той факт, що середній бал за такі тестові завдання у хлопців вищий, ніж у дівчат, а не зі знанням фізики, перевірка яких є метою тесту. Учні з деяких релігійних груп не дуже добре пройдуть тест на розуміння читання тексту про дискотеки, оскільки їм не можна на них ходити, але це не пов'язано з їхнім рівнем розуміння читання, це приклад культурного упередження.
- Уникайте контексту чи атмосфери, яка може засмутити деяких учнів.
- Надайте якісні схеми оцінювання робіт і процедури оцінювання для того, аби звести до мінімуму відмінності між оцінками одних і тих же робіт, перевірених різними оцінщиками.

4.4.4.Ефективність

Тестові завдання мають надати якомога більше інформації про кандидата за якомога менший період часу.

- Надавайте лише ту інформацію, яка є релевантною до завдання. Уникайте зайвих відступів, картинок, таблиць тощо.

Примітка: це не стосується ситуацій, коли визначення необхідних даних складає частину завдання.

- Одна картинка може сказати більше, ніж багато слів.
- Скористайтеся усіма можливостями, які може дати текст завдання, для того, аби поставити більше запитань.

4.4.5.Прийнятність

Громадськість повинна вважати тестові завдання загалом справжнім і валідним представленням відповідних знань у соціально-політично прийнятному контексті.

- Скористайтесь контекстом реального життя, але нехай він залишається валідним.
- Уникайте контроверсійних (скандальних) питань.
- Не ставте заплутаних питань.

4.4.6.Контрольний перелік

Нижче наведено контрольний перелік, у якому підсумовано вказівки щодо дотримання п'яти основних критеріїв якості при розробці тестових завдань. Цей контрольний перелік може використовувати і автор, і рецензент тестових завдань.

Валідність Мета тестування чітка і вписується у програму; існує класифікація завдань за Програмою Використано релевантні аспекти Відповідає сучасному розумінню предмета Відволікаючі варіанти відповідей представляють реальні та релевантні помилки і хибні сприйняття Надійність Відсутність недоліків, зокрема - нечітка чи двозначна мова - надто відкрите формулювання - відсутність обмеження відповідей; - ненавмисні підказки - ключі не розкидані навмання	Об'єктивність Відсутність упереджень в запитанні (гендерних, етнічних, релігійних чи культурних) Відсутність контексту, який може засмутити особу, яка складає тест Відсутність заплутаних запитань Якісна і ефективна система та структура оцінки робіт Ефективність Увесь текст, таблиці і картини є функціональними Інформація чітко сприймається з першого разу Якщо потрібно, картини супроводжують пояснення Усі можливості сформулювати запитання, які дає контекст, використані Прийнятність Тестове завдання вважається валідним, а його контекст – функціональним та реалістичним Контроверсійних питань (гострі політичні, моральні, релігійні, етичні) слід уникати Громадськість розглядає це тестове завдання як об'єктивний тест
---	---

Малюнок 26. Контрольний перелік при розробці та рецензуванні тестових завдань

5.Розробка тестових завдань із варіантами відповідей

У цьому розділі представлено цілу низку прикладів, для ілюстрації досвіду розробки тестових завдань із варіантами відповідей. Представлено проблемні тестові завдання, оскільки легше пояснити, як краще працювати, на основі прикладу, який ілюструє проблему. Послідовність прикладів відповідає структурі контрольного переліку.

5.1.Валідність

5.1.1.Чітка мета

45. Which word below does **NOT** rhyme with the word **rainbow**?

- A. arrow
- B. bow
- C. allow
- D. now

45. Яке слово нижче НЕ римується зі словом rainbow?

- A. arrow
- B. bow
- C. allow
- D. now

Мета цього тестового завдання (яке скопійоване із тесту з англійської мови як другої іноземної і складає частину вступного іспиту до університету) – оцінити вимову, використовуючи контекст римування, а не саме римування. Однак визначення римування можуть відрізнятися, і це може перетворити тестове завдання на тестування питання про те, що таке римування, а не на тестування вимови. Додаткова проблема у цьому тестовому завданні – це те, що «bow» може вимовлятися двома різними способами, як у «now», і як у «gow». Його значення залежить від вимови. І останнє, але не менш важливе: у варіанті D є помилка: малося на увазі «gow». У наведеній версії тестове завдання мало дві правильні відповіді, C і D.

5.1.2.Релевантна мета

- Скільки суддів у Конституційному суді України?
 - A. 15 суддів
 - B. 18 суддів
 - C. 21 суддя
 - D. 23 судді

Приклад із тестового завдання, яке перевіряє не релевантний зміст, наводився на Малюнку 1 (Щодо висоти піраміди).

Ще одним прикладом може бути тестове завдання, що стосується кількості суддів у Конституційному суді України. Знання стосовно того, що є саме 18 суддів, не є

релевантним для відбору кандидатів на посаду судді для проходження спеціальної підготовки. Значення має контекст: чому там 18 суддів, а не якесь інше число? Також поглянемо на відволікаючі варіанти. Які релевантні помилки чи неправильні сприйняття вони представляють?

- Спосіб організації та функціонування вищих державних органів називається:
 - A. державний устрій
 - B. державне управління
 - C. державна система
 - D. форма держави

Ще один приклад – це тестове завдання, в якому криється більше лінгвістична проблема, а не судова, і яке, мабуть, не є релевантним для розрізнення кандидатів, яких слід допустити до наступних етапів добору, і тих, яких не варто допускати.

Однак це тестове завдання містить більше недоліків. “В” планувалося як правильна відповідь, але «управління» було синонімічним до «устрою», відтак А є також правильним. Також, «державна система» і «форма держави» – це надто нечіткі поняття, вони можуть означати будь-що, відтак відволікаючи відповіді – С та D не є «цілком неправильними».

5.1.3. Відповідність сучасним розумінням

Випадок з полярним ведмедем

Тестове завдання про полярного ведмеда нижче було в іспиті з біології в Голландії. Ключ до завдання – D. Одна з учениць, дівчинка, дала відповідь B. Ця дівчинка не склала іспит з біології не добравши одного балу. Для неї це був жакхливий результат, оскільки якби вона склала іспит, то отримала би сертифікат, що дав би їй право вступити до університету.

Білий колір хутра полярного ведмеда допомагає йому вижити в полярних умовах. Яке з наступних пояснень правдиве/вірне?

- I. Біле хутро робить ведмеда менш помітним під час полювання на інших тварин.
 - II. Біле хутро допомагає уникати втрати тепла.
- A. Жодне
 - B. Лише I є правильним
 - C. Лише II є правильним
 - D. Обидва варіанти правильні

З цієї причини вона почала шукати можливість подати апеляцію.

Її батько був фізиком, який з власних досліджень на радіаторах знав, що колір впливає на передачу тепла менше, ніж вважається, і у цьому випадку вплив зовсім незначний.

Він не тільки був фізиком, він був також членом Парламенту, і це дало йому можливість прямо поставити запитання Міністру.

Коли Міністра запитали, чи він знає, що в іспиті з біології були запитання, які «суперечили сучасному стану наукових досліджень і зовсім не узгоджувалися із сучасним розумінням питання», він відреагував тим, що вилучив це завдання із тесту і знизив прохідний бал на один бал.

В результаті дівчинка склала іспит. Менш щасливі були директори багатьох шкіл, яким довелося повернутися зі своїх літніх відпусток, аби перевірити результати усіх інших учнів, які складали біологію, і змінити їхні результати.

Звичайно, випадок із полярним ведмедем – поодинокий. Однак завжди бажано, аби завдання перевіряв експерт у даній сфері, і особливо, коли використовуються контекст і поняття, які належать до інших предметів.

5.1.4. Релевантні дистрактори (відволікаючі варіанти)

Кожен дистрактор (відволікаючий варіант) повинен бути значущим, у тому розумінні, що він повинен репрезентувати помилку в операції, яку повинен здійснити кандидат, чи типово неправильне розуміння. У завданні, розміщеному далі, правильна відповідь – А: в межах 3 днів. Однак які помилки представляють відволікаючі варіанти?

- В які строки з моменту надходження позову до суду суд повинен винести постанову?
 - A. протягом 3 днів
 - B. протягом 5 днів
 - C. протягом 7 днів
 - D. протягом 14 днів

Також, не зрозуміло, чи «дні» передбачають робочі дні, чи календарні. Тестове завдання слід переглянути, щоб відволікаючі варіанти стали значущими і репрезентували періоди часу, зазвичай, з урахуванням судових постанов, для прикладу, в межах 5 робочих днів, чи в межах одного тижня, двох тижнів, місяця тощо.

5.2.Надійність

Основна загроза для надійності і водночас валідності тесту – це таке формулювання, яке збиває кандидата з пантелику. Якщо кандидати не чітко розуміють, чого від них очікують, то визначення відповідей скеровується неправильним розумінням, неправильними припущеннями чи просто вгадуванням. У всіх цих випадках відповідь не є відображенням реальних здібностей чи знань кандидата, а радше відображає недбалість розробника тестового завдання. Нижче наведено перелік проблем і непорозумінь, які часто виникають через неправильне формулювання тестових завдань:

- Нечіткий, розмитий текст завдання
- Нечіткі малюнки
- Використовуються незрозумілі позначення
- Формулювання тексту запитання складніше, ніж саме запитання
- Використання відкритих текстів запитань
- Більше однієї правильної відповіді, залежно від інтерпретації тексту запитання

Ці приклади проілюстровані наступними тестовими завданнями.

5.2.1.Нечіткий текст запитання чи варіантів відповіді

В одному із тестів Вищої кваліфікаційної комісії суддів України є таке тестове завдання:

- Яке юридичне значення має Конституція Автономної Республіки Крим і чому?
 - A. Закон, оскільки вона прийнята Верховною Радою Автономної Республіки Крим
 - B. Закон, оскільки він затверджений Верховною Радою України шляхом прийняття законом України
 - C. Найвищий, оскільки є актом спільного волевиявлення народу
 - D. Місцевий підзаконний акт, як статут територіальної громади

Формулювання тексту є розмитим і стає чітким лише після вивчення усіх варіантів. Але залишаються ще значні сумніви. Для прикладу: що мається на увазі під «найвищим» (відволікаючий варіант C). Найвищий із чого? І що мається на увазі під «законом України»? Який закон? Конституція? І відволікаючий варіант D – це не відповідь-плюс-пояснення, як A, B і C.

Тестове завдання, як видається, стосується зв'язку між Конституцією Автономної Республіки Крим і Конституцією України, чи конкретним національним законодавством, чи положеннями чи повноваженнями Верховної Ради. Ключові питання можуть бути такими: чи Конституція Автономної Республіки Крим має вищий

статус ніж Конституція України, чи її може скасувати ВР і так далі. Це тестове завдання слід відредагувати, аби чітко вказати, що має на увазі автор.

У тестовому завданні стосовно «орендної плати» зліва можна посперечатися з приводу

- Вид орендної плати за наділ державної чи місцевої землі
 - A. відпрацьовувана рента;
 - B. натурою;
 - C. мішаний вид плати;
 - D. грошовий.

того, що відповідь “С” є не чіткою: поєднання (змішування) чого? Не допомагає зазначення того, що “С” є поєднанням “А” і “В”, оскільки це не є очевидним.

Часто автори, якщо зустрічаються із коментарями щодо нечітких тестових завдань чи відволікаючих варіантів, реагують реплікою: «Але мої студенти розуміють, що я маю на увазі!». Можливо, ваші

студенти і можуть здогадатися про цю інформацію-підтекст, але не варто цього очікувати від інших студентів, не кажучи вже про усіх кандидатів у країні. Будь-яка частина тестового завдання повинна бути чіткою і недвозначною сама по собі, щоб її розумів кандидат без необхідності подальших пояснень.

5.2.2. Маловідомі слова чи вирази



Ви хочете зробити скатертину, як показано на картинці. Ця скатертина не повинна досягати підлоги. Діаметр поверхні столу – 80 см. Висота столу – 75 см. Скатертина для столу повинна мати канту 5 см. Ви купите прямокутний шмат тканини шириною 3 м.

- Якою має бути довжина цього шмату?

Таке тестове завдання було включено до державного іспиту у Голландії. Коли аналізували результати цього завдання, результати Р-величин у хлопців були низькими, тоді як дівчата набагато краще впоралися із цим завданням, а це не узгоджується із результатами решти тестових завдань, де хлопці показали дещо кращі результати, ніж дівчата.

Проблема у розумінні цього SR завдання полягає у слові «кант». Дівчатам краще знайомі вирази, які стосуються шиття, а чимало хлопців, і слабких, і сильних, просто не змогли зробити це завдання, оскільки вони не знали, що таке «кант».

Таким чином, це завдання також є прикладом упередженості на користь дівчат. Урок такий: будьте обережними, використовуючи модні, але маловідомі слова чи розумні вирази.

Упевніться, що розуміння прочитаного не стоїть між кандидатом і тим, що пропонується оцінити, якщо, звичайно, оцінка розуміння прочитаного не складає мету тесту.

5.2.3.Подвійне заперечення

- Яке із наступних тверджень **не є правильним** для Амстердаму?
 - A. Амстердам не є столицею Нідерландів
 - B. Амстердам не є містом, де розміщено уряд Голландії
 - C. Амстердам не є найбільшим портом Нідерландів
 - D. Амстердам є найбільшим містом Нідерландів

У цьому тестовому завданні – подвійне заперечення («що НЕ є причиною НЕ функціонування») вже само по собі збиває з пантелику. В завданні про «Амстердам» вище 'А', 'В' і 'С' разом з текстом завдання складають подвійне заперечення, його розкодовування, саме по собі, може бути складнішим, ніж саме запитання, і це може збивати кандидатів. Ключем є відповідь А, оскільки твердження є неправильним, і, відтак, є правильною відповіддю!

5.2.4.Заперечення тексту запитання; заплутане запитання

Текст тестового запитання (Основи) нижче показує, що цей випадок слід розглядати за описаних в Основі обставин. Твердження у D вказує, що можливо також і те, що суперечить описаним в Основі обставинам.

Для кандидата, що проходить тестування, дуже складно мати справу із такими суперечними твердженнями і, найімовірніше, кандидат відкине цей варіант одразу ж як неправильну відповідь. D – правильна відповідь, яка перетворює це завдання на заплутане.

- За яких обставин розгляд випадку розпочинається наново після того, як третя особа, яка не має претензій до предмету спору, заявляє про свої права?
- A. У кожній такій справі;
 - B. Коли є клопотання з боку третьої сторони, яка має свої власні претензії;
 - C. Якщо суд вважає це необхідним;
 - D. Розгляд не розпочинається заново за жодних обставин.

5.2.5.Відкритий текст завдання

- Судова постанова щодо обшуку та конфіскації
 - A. не може бути оскаржена;
 - B. може бути оскаржена у звичний спосіб;
 - C. апеляції може бути оскаржена після обшуку та конфіскації;
 - D. боку може бути оскаржена лише особою, якої стосується ця постанова (яка зазначена у цій постанові).

По суті, в коректному випадку, має бути можливим відповісти на тестове завдання SR, прочитавши текст самого запитання (Основи), і лише тоді перевірити, чи серед запропонованих варіантів є правильна відповідь.

На практиці, чимало тестових завдань вимагають, щоб особа, яка складає тест, перечитала усі варіанти з тим, аби зрозуміти саме запитання.

Традиційний приклад таких завдань із відкритими текстами завдань – завдання про «судову постанову про обшук та

конфіскацію». Лише після перевірки усіх варіантів стає зрозумілим, що особа, яка складає тест, повинна визначити, який тип оскарження стосується цієї постанови.

І все ж, декілька моментів залишаються незрозумілими. Для прикладу: що мається на увазі під «звичним способом» у відволікаючому варіанті В? І чи це тестове завдання стосується того, чи можна оскаржувати, і, якщо так, то як і хто повинен її подавати? Рецензування і редагування цього завдання вимагатиме спершу чіткого переформулювання мети тесту автором, після чого текст завдання буде замінено повним реченням, яке пояснює, яким чином кандидат повинен розглядати варіанти. В ідеалі, кандидат повинен мати змогу відповісти на запитання навіть не розглядаючи варіантів.

5.2.6. Більш, ніж одна правильна відповідь

Коли текст запитання і/чи варіантів відповідей не добре розроблений, може виникнути ситуація, у якій можна буде довести правильність більш ніж однієї відповіді.

«Завдання про Ніл» праворуч – це типовий приклад такої ситуації.

Популярний спосіб – коли викладачі перевіряють зв'язки, що пов'язують об'єкти, в форматі запитання «вказіть, що не належить до групи» – досить небезпечний у тестуванні на допуск до професії, оскільки може існувати набагато більше зв'язків, ніж ті, про які думав автор.

У тестовому завданні про Ніл, для прикладу, Месопотамія може бути правильною відповіддю, оскільки вона не є рікою, або ж Ніл може бути правильною відповіддю, оскільки він розташований не у Вавилоні (завдання взято з тесту про Вавилонську культуру).

- Яке із наведених вище є зайвим?
А. Євфрат
В. Месопотамія
С. Ніл
D. Тігр

«Яке слово відрізняється?» – це дуже двозначне запитання.

Зрештою, кожне із чотирьох слів-варіантів, відрізняються одне від одного. Також, автор стверджував, що студенти зрозуміють запитання.

І все ж, якщо ви шукаєте характеристику, яка є спільною для трьох із цих слів, і відмінною для одного – а саме таким був намір – ви можете опинитися в пастці через подвійне значення деяких слів.

«Словником» може бути поняття – набір слів, які використовує особа – або реальний об'єкт книжка-словник. А «читач» може бути особою чи підбіркою уривків із книжки (друге значення (підбірка уривків) є в слові іноземною мовою в оригіналі тексту завдання, а в українській мові цього другого значення немає).

Якщо ви розумієте В і D в альтернативних, інших значеннях, то для вас є не одна група із трьох слів, які мають спільну характеристику, і одне, яке не належить до цієї групи, але дві групи до дві відповіді.

- Яке слово відрізняється?
А. Граматика
В. Словник
С. Вимова
D. Читач

- Процес реальної реалізації обов'язкових приписів права шляхом правильної поведінки суб'єктів права, які мають юридичні зобов'язання, такий:
А. використання правових норм
В. використання норм права
С. застосування норм права
D. реалізація правових норм

Що точно означає текст завдання зліва – про «реалізацію обов'язкових приписів» - не зовсім зрозуміло.

Це також стосується варіантів відповідей. Наприклад, чи є різниця між «правовими нормами» і «нормами права»?

Але найважливіше питання

полягає у тому, що можна довести правильність всіх варіантів відповіді.

С і D, по суті, синонімічні.

А реалізація передбачає використання і застосування, відтак D включає В і С.

Із кількох способів уникнення ситуації, коли правильним буде більш ніж один варіант відповіді, часто використовується додавання обмежень. Такі слова як «лише якщо» або

- Коли суд має право ставити запитання свідку в процесі допиту останнього в адміністративній справі?
 - А. після опитування свідка
 - В. до опитування свідка
 - С. у будь-який момент допиту

«принаймні», можливо, допоможуть обмежити тлумачення варіантів. Без таких обмежень варіанти можуть бути двозначними, і призвести до ситуацій, коли більше ніж один варіант можна буде відстояти як правильний.

Наведений вище приклад ілюструє як можна використати такі «обмежувачі». У своїй вихідній формі А і В були би правильними, якби С був правильним. Додавання обмежувача «лише» до А і В вирішує проблему.

Популярний формат «два твердження» – це спосіб вирішення проблеми кількох правильних відповідей. У наведеному нижче прикладі А в оригінальній версії включає В і С. Відредагована версія набагато чіткіша і має лише одну правильну відповідь.

- Як би Ви розцінили факт, коли менеджер незаконно забирає майно зі складу при виконанні своїх обов'язків:
 - А. Поєднання злочинів – пограбування з незаконним вторгненням (ч. 3 Ст. 185) і незаконним привласненням (ч. 2 Ст. 191 КК України);
 - В. Пограбування з незаконним вторгненням (ч. 3 Ст. 185);
 - С. Незаконне привласнення (ч. 2 Ст. 191 КК України);

Оригінальна версія

- Менеджер складу забирає майно зі складу до себе додому і продає його третій особі. Які з наступних обвинувачень можна висунути проти нього?
 - І. пограбування з незаконним вторгненням (ч. 3 Ст. 185 КК України)
 - ІІ. незаконне привласнення (ч. 2 Ст. 191 КК України)
 - А. І І, і ІІ
 - В. Лише І
 - С. Лише ІІ

Відредагована версія

5.2.7. Ненавмисні підказки

- Безпосередня мета проекту Геном людини полягає у
 - А. Розробці нових технологій для вивчення ДНК людини з різних частин світу
 - В. Розробці карти і послідовності всіх РНК
 - С. Розробці карти усіх генів людини і визначенні послідовності нуклеотидів всього геному людини
 - Д. Порівнянні геному великої кількості індивідів з різних частин світу

Кандидати повинні знайти правильну відповідь шляхом використання своїх предметних знань, розуміння та навиків, а не на основі підказок в самому завданні чи на основі загальних знань.

Вид ненавмисних підказок, який часто трапляється, – це слова з Основи, які повторюються в одному із варіантів відповідей, як от у випадку із завданням про геном людини. І справді, С – це правильна відповідь на це тестове завдання.

- У лікарнях простирадла завжди перуть при високих температурах, а також прасують. Є причина для цього. Яка це причина?
А. Відчистити їх
В. Довести їх до білизни
С. Вбити бактерії, оскільки ці організми не виживають при високих температурах

Ще один вид ненавмисних підказок має місце, коли один із варіантів чітко виділяється, чи то за довжиною, чи то за традиційністю та складністю формулювання. У завданні про «лікарняні простирадла» правильна відповідь – С, і це більш-менш видно за його довжиною та точністю формулювання.

В інших випадках варіанти відповіді логічно пов'язані і виключають одне одне, як от у випадку із завданням про метеостанцію. Якщо правильним є А, то і D є правильним, отже обидва, мабуть, неправильні, інакше є дві правильні відповіді. Правильна відповідь – В.

Традиційна ситуація, коли підказується правильна відповідь, створюється двома пов'язаними завданнями. Нижче наведено приклад про делегацію України, яка відвідала Лондон. Хоча це й видається мало ймовірним, однак такі ситуації трапляються!

- Дві метеостанції розміщені поблизу одна одної. Атмосферний тиск на одній станції змінюється так, що різниця між тисками зростає. Швидкість вітру між цими двома місцями, мабуть,
А. зменшиться.
В. зростає.
С. залишиться такою ж.
D. наблизиться до нуля.

Питання після удаваного тексту, в якому описується візит делегації від ВККСУ до Лондону.

- В який аеропорт прибуде українська делегація?
А. Гатвік
В. Хітроу
С. Лутон
D. Станстед
- Хто зустрічатиме українську делегацію в аеропорту Хітроу?
А. Міністр юстиції
В. Голова Верховного Суду
С. Мер Лондона
D. Віце- канцлер Лондонського Університету

5.2.8.Здоровий глузд

Іноді ненавмисно надана здогадка може стати причиною вгадування слабкими кандидатами правильної відповіді, але іноді таке може трапитися і шляхом логічного мислення, як наведено у прикладі нижче.

Ці варіанти уже показують, що «ні» не буде правильною відповіддю. В – це «здоровий глузд»: ви повинні завжди робити речі, які є обґрунтованими. Також зверніть увагу на те, що В включає С і D. Якщо В є неправильним варіантом, тоді С і D також неправильні. Відтак, мабуть, В – правильна відповідь, якщо А відкидається.

- Чи власник або орган, уповноважений ним, має право переводити працівників без їхньої згоди в межах підприємства, установи, організації, спеціалізації та кваліфікації (посади) на роботу, яка небажана при їх стані здоров'я (але не має чітких медичних заборон):
 - А. так, він робить це без жодних обмежень;
 - В. він робить це, за умови, що для цього існують добре обґрунтовані причини;
 - С. він робить це лише у випадках, коли це зазначено у колективному договорі;
 - Д. він робить це лише у випадках, коли це визначено окремим положеннями трудового законодавства

5.2.9.Розумність тесту

Ще одна загроза надійності наших тестів – це «розумність тесту», явище, яке кандидати визначають як спосіб для ідентифікації правильних відповідей без реальних знань предмету, за допомогою спеціального навчання на складання будь-яких тестів.

Навчитися тому, як помічати ненавмисні підказки, – ось, зазвичай, результат такого навчання. У США є ціла низка установ, які заробляють добрі гроші на тому, що вчать учнів складати вступні тести. Загальна інструкція така: «якщо ви не знаєте, виберіть ту відповідь, яка Вам першою спала на думку» – ось частина такого навчання, і «якщо у Вас немає підказки, виберіть варіант С».

Останній принцип пов'язаний з психологією розробника тестових завдань. Чимало з них, як видається, не хочуть ставити правильну відповідь на першому чи другому місці, адже вони хочуть, щоб студенти перечитали усі варіанти і могли вибрати неправильний варіант.

Дослідження показали, що якщо не приймаються запобіжні заходи, то й справді ймовірність правильності відповіді С – значно вища, ніж будь-яка із інших позицій правильної відповіді.

Існує кілька способів, як уникнути небажаної пріоритетності будь-яких позицій у тесті з варіантами відповідей. Один із них – застосування простих правил упорядкування послідовності відповідей в алфавітному порядку, за цифровим чи логічним принципом, як от у прикладі нижче (правильна відповідь виділена шрифтом).

Дострокове звільнення від відбування покарання можливе лише після закінчення:

- A. Принаймні однієї третьої терміну для звільнення від відбування кримінального покарання;
- B. Принаймні двох третіх терміну для звільнення від відбування кримінального покарання;
- C. Принаймні половини терміну для звільнення від відбування кримінального покарання;**
- D. Принаймні трьох четвертих терміну для звільнення від відбування кримінального покарання.

Після упорядкування

- A. Дострокове звільнення від відбування покарання можливе попереднього засудження дозволяється лише після закінчення: Принаймні однієї третьої терміну для анулювання кримінального покарання;
- B. Принаймні половина терміну для анулювання кримінального покарання;**
- C. Принаймні двох третіх терміну для анулювання кримінального покарання;
- D. Принаймні трьох четвертих терміну для анулювання кримінального покарання.

5.3. Неупередженість

Тестове завдання про «скатертину на стіл», описане раніше, – це хороший приклад гендерного упередження. З причин, які не мають стосунку до мети тестування, дівчата набагато краще впоралися з цим завданням.

5.4. Ефективність

Економіка Нідерландів процвітає, і вона суттєво залежить від зовнішньої торгівлі. Її економіка відома через свої стабільні промислові відносини, помірний рівень незайнятості та інфляції, значний надлишок на поточних рахунках, а також важливу роль як транспортного вузла Європи. Промислова діяльність – це переважно харчова, хімічна промисловість, нафтопереробна промисловість та електротехнічна промисловість. Надзвичайно сильно механізований сільськогосподарський сектор залучає не більш, ніж 4% робочої сили, але дає значний надлишок для харчової промисловості і для експорту. Нідерланди, разом із 11 партнерами по ЄС, почала користуватися валютою євро 1 січня 2002 р. Країна продовжує бути однією з провідних країн Європи в плані залучення прямих іноземних інвестицій. Економічний ріст значно сповільнився у 2001-04, в зв'язку зі світовою економічною кризою, однак чотири роки до того середній рівень річного зростання складав близько 4%, що значно перевищувало середній рівень по країнах ЄС.

- Яка столиця Нідерландів?

Це традиційний приклад не функціонального тексту. Багато інформації для особи, яка складає тест, але ця інформація є абсолютно зайвою для визначення правильної відповіді на питання. Це також може стосуватись ілюстрацій, які не додають жодної інформації, необхідної для того, аби дати відповідь. Автори часто роблять це, аби їхнє тестове завдання виглядало привабливішим, і забувають, що це може заплутати

кандидатів, оскільки вони вважають, що вся інформація є необхідною і повинна якимось чином використовуватися для визначення правильної відповіді. Не функціональний текст і ілюстрації збивають їх з пантелику і можуть навіть привести їх до того, що вони не дадуть відповіді чи виберуть неправильну відповідь.

5.5.Прийнятність

Тестові завдання повинні бути прийнятними, і з погляду психометрії, і з етичного погляду. Дистрактори низької якості, які додаються, оскільки складно знайти якісні, є неприйнятними для тесту. Такі контрверсійні чи суперечливі питання, як релігія та політика, секс

і наркотики, чи сильний емоційний контекст, можуть засмучувати і їх краще не використовувати для тестових завдань. Вони можуть настільки змінити емоційний стан кандидата, що він не зможе дати правильну відповідь на запитання. Нещодавній приклад тестового завдання, яке довелося вилучити із державного іспиту в Голландії після тестування, яке пройшли десятки тисяч кандидатів, наведено нижче.

- У денний час температура на рівні 5см вище землі зазвичай
 - A. вища, ніж температура на рівні 2 м.
 - B. нижча, ніж температура на рівні 2 м.
 - C. така сама, як температура на рівні 2 м.
 - D. жоден із варіантів

Одне з тестових завдань державного випускного завдання в Голландії у 2009 році на розуміння читання, має контекст (стимулюючий матеріал), запозичений із веб-сторінки VVD (центристсько-правої політичної партії). Після стимулюючого матеріалу давалось завдання такого змісту:

- Використання мови у цьому тексті показує негативне і навіть зверхне ставлення до людей, які розраховують на соціальні пільги. Ось чотири цитати:

1; 2; 3; 4.....

У якій із цих чотирьох цитат виражено це ставлення?

Після різких протестів з боку VVD і дебатів у Парламенті Державний секретар наказав вилучити це тестове завдання із тесту.

II. ОГЛЯД СТАНДАРТІВ У СФЕРІ ПРОФЕСІЙНОГО ЕКЗАМЕНУВАННЯ. *Сергій Мудрук*

1. Державні стандарти в Україні

Історично першим стандартом в Україні в царині оцінки персоналу (щодо дорослого населення, зайнятого в професійній діяльності) був Міждержавний ГОСТ 30488-97 «Органи з сертифікації персоналу. Загальні вимоги. Міждержавний стандарт», що є еквівалентом Європейського стандарту EN 45013-89 «Загальні критерії з сертифікації для органів, які проводять сертифікацію персоналу»

Цей стандарт було введено в дію в якості Державного стандарту з 1 січня 2000 року. Пізніше, Наказом Держспоживстандарту №90 від 14 квітня 2005 року стандарт ГОСТ 30488-97 було скасовано з наступним введенням в дію з 1 жовтня 2006 року нового стандарту ДСТУ ISO/IEC 17024:2005 «Оцінювання відповідності. Загальні вимоги до органів, що здійснюють сертифікацію персоналу», що діє дотепер і є еквівалентом міжнародного стандарту ISO 17024:2003.

2. Міжнародний стандарт ISO 17024

Перевагою стандартів ДСТУ/ISO 17024 є надання кращого, визнаного в усьому світі, перевіреного десятиліттями досвіду щодо побудови систем надання доступу до професійної діяльності.

Простота і загальний характер формулювань статей діючих стандартів (ДСТУ і ISO) дозволяє задовольнити основним вимогам стандартів, за умови поетапної деталізованої ревізії пов'язаних з іспитами процедур, норм та інституційних засад. Найпроблемнішими є саме інституційні засади.

На даний момент чітко не визначено, до якого суб'єкта чи групи суб'єктів слід застосовувати стандарти групи ДСТУ/ISO 17024 в процедурі добору кандидатів на посаду судді.

З одного боку, виглядає, що Вища кваліфікаційна комісія суддів України (далі – ВККСУ) на сьогодні є «органом, що здійснює сертифікацію персоналу», а саме здійснює добір кандидатів на посаду судді, тому може підпадати під дію відповідного стандарту (щодо застосування стандартів див. п.2 ст.23 Закону України «Про стандартизацію» від 5 червня 2014 року №1315-18). Але, з іншого боку, потрібно дослідити комплекс взаємовідносин Вищої ради юстиції (далі - ВРЮ, ВККСУ та Національної школи суддів (далі – НШСУ) щодо процедур допуску осіб до професійної діяльності судді.

За інтерпретацією ВРЮ норм Закону про судоустрій і статус суддів, перевіряти практичні роботи кваліфікаційного іспиту мають безпосередньо члени ВККСУ. Якщо розглядати ВККСУ як орган сертифікації, це суперечить певним пунктам стандарту ДСТУ/ISO 17024, за яким особи, що приймають рішення по сертифікації, не можуть безпосередньо перевіряти екзаменаційні роботи. Роз'яснення щодо цього дав у 2013 році Голова ВРЮ, акцентуючи увагу на тому, що рішення про допуск до професійної діяльності судді приймає не ВККСУ, а ВРЮ, відповідно ВККСУ не є органом сертифікації.

В стандартах є норма щодо неприпустимості для органу сертифікації пропонувати підготовку кандидатів (це є очевидним конфліктом інтересів). Зокрема, орган навчання і орган оцінки (що безпосередньо виносить рішення про допуск до професійної діяльності) не можуть бути у відносинах підпорядкування.

В свою чергу, НШСУ за багатьма розділами ДСТУ/ISO 17024 не може розглядатись і виконувати суттєву частину функцій «органу, що здійснює сертифікацію», оскільки здійснює спеціальну підготовку кандидатів на посаду судді.

З іншого боку, ВРЮ у відповідності до деяких розділів ДСТУ/ISO 17024, також не може самостійно виступати в якості «органу, що здійснює сертифікацію».

Таким чином ВККСУ залишається основним органом, до якого слід застосовувати більшість норм стандартів ДСТУ/ISO 17024.

Короткий огляд відповідності функцій ВККСУ стандарту ДСТУ/ISO 17024 можна знайти у Розділі IV – витяг зі Звіту «Моніторинг процесу анонімного тестування кандидатів на посаду судді», листопад 2012р.

3. Стандарти AERA/APA/NCME

Стандарт ISO 17024 є досить загальним, тому його важко застосувати як керівництво до дії для організації конкретних деталей робочих процесів організації професійного добору.

Для цього існують набагато більш об'ємні стандарти, найбільш визнаними з яких є «Стандарти освітнього і психологічного тестування», розроблені трьома провідними американськими установами - Американською асоціацією освітніх досліджень (American Educational Research Association (AERA)), Американською психологічною асоціацією (American Psychological Association (APA)), та Національною радою з оцінювання в освіті (National Council on Measurement in Education (NCME)).

«Стандарти освітнього і психологічного тестування» визнано фахівцями з екзаменування в усьому світі.

Ці стандарти є всеосяжним об'ємним документом (близько 200 сторінок англійською мовою), що містить 15 груп стандартів, об'єднаних у 3 частини:

Розділ I Побудова тесту, оцінка, документування

1. Валідність
2. Надійність та похибка вимірювання
3. Розробка та перегляд тесту
4. Шкали, норми та порівняння оцінок
5. Адміністрування іспиту, виставлення оцінок та повідомлення результатів
6. Супровідна документація для тестів

Розділ II Справедливість у тестуванні

7. Справедливість у тестуванні та використання тесту
8. Права і обов'язки кандидатів
9. Тестування осіб з різномірною мовною підготовкою
10. Тестування осіб з особливими потребами

Розділ III Застосування тестів

11. Обов'язки користувачів тестів
12. Психологічне тестування і оцінка
13. Освітнє тестування і оцінка
14. Тестування з метою атестації і працевлаштування
15. Тестування в оцінці програм і державній політиці

Переклад деяких зі стандартів AERA/APA/NCME, відібраних для супроводу тренінгу, наведено в кінці цього розділу.

3.1. Перелік стандартів AERA/APA/NCME рекомендованих для першочергового ознайомлення:

- Стандарт 1.1
- Стандарт 1.2
- Стандарт 1.6
- Стандарт 1.12
- Стандарт 2.2
- Стандарт 2.7
- Стандарт 2.8
- Стандарт 3.1
- Стандарт 3.2
- Стандарт 3.3
- Стандарт 3.4
- Стандарт 3.6
- Стандарт 3.7
- Стандарт 3.8
- Стандарт 3.9
- Стандарт 3.11
- Стандарт 3.17
- Стандарт 3.18
- Стандарт 3.19
- Стандарт 3.20
- Стандарт 3.21
- Стандарт 3.22
- Стандарт 4.10
- Стандарт 4.11
- Стандарт 4.19
- Стандарт 4.20
- Стандарт 4.21
- Стандарт 6.4
- Стандарт 6.5
- Стандарт 7.2
- Стандарт 8.13
- Стандарт 11.6
- Стандарт 11.12
- Стандарт 11.15
- Стандарт 13.18
- Стандарт 13.19
- Стандарт 14.8
- Стандарт 15.10

3.2.Переклад деяких стандартів освітнього і психологічного тестування AERA/APA/NCME

Стандарт 1.1

Повинно бути представлено обґрунтування для кожного рекомендованого способу тлумачення і використання оцінки за екзамен (бал), разом з повним резюме доказів та теоретичним підґрунтям планованого використання та тлумачення оцінок (балів) за екзамен.

Стандарт 1.2

Розробники екзаменів мають чітко пояснювати як передбачається використовувати і тлумачити оцінки за екзамен. Має бути ясно визначено, для якої аудиторії (категорії кандидатів) призначено екзамен і яку концепцію (конструкт) передбачається оцінювати.

Стандарт 1.6

Коли валідація частково покладається на відповідність змісту тесту, мають бути описані процедури генерації змісту тесту. Такі процедури мають бути обґрунтовані щодо того, наскільки результат генерації відповідає Конструкту, що планується вимірювати, чи тілу знань, що має представляти тест.

Стандарт 1.12

Якщо пропонується тлумачення складових екзаменаційної оцінки, чи відмінностей оцінок, чи екзаменаційних профілів, має бути надано пояснення і відповідні обґрунтування для такого тлумачення. Якщо використовується складена (багатокомпонентна) екзаменаційна оцінка, має бути надано обґрунтування і пояснення того, як і чому саме таким способом формується складена оцінка.

Стандарт 2.2

Стандартна похибка вимірювання повинна вказуватись і в одиницях первинного бала і в одиницях остаточного бала, який рекомендовано для використання в якості офіційного результату екзамену.

Стандарт 2.7

У випадку, якщо з приводу частини запитань екзамену, що відповідають специфікації (екзамену), можна припустити, що вони вимірюють частково незалежні особливості чи здібності, то, при оцінюванні надійності, необхідно враховувати мультифакторний характер екзамену, як інструмента вимірювання.

Стандарт 2.8

Сторона, що використовує екзамен, має бути поінформована щодо ступеня впливу швидкості роботи кандидата на екзамені на його екзаменаційний результат.

Стандарт 3.1

Тести і програми тестування повинні розроблятись на міцній науковій основі. Розробники тестів мають упорядковувати їх і документувати достатні підтвердження їх актуальності, що є основою розробки тесту.

4. Кодекс справедливої практики тестування в освіті

Кодекс було напрацьовано «Спільним комітетом екзаменаційної практики», в який увійшли представники AERA, APA, NCME (див вище) та Американської асоціації з консультування (American Counseling Association ACA), Американської мовної асоціації для людей з вадами слуху (American Speech-Language-Hearing Association (ASHA)), Національної асоціації шкільних психологів (National Association of School Psychologists (NASP)), Національної асоціації тестових директорів (National Association of Test Directors (NATD)).

Мета Кодексу – надати путівник професіоналам для створення справедливого стосовно всіх кандидатів тестування, незалежно від віку, статі, етнічної належності, громадянства, релігії і т.ін.

Кодекс є добре структурованим лаконічним документом, основні розділи якого (див.нижче) об'єднано у секції:

- зобов'язання організації, що розробляє тест (Розробник)
- зобов'язання організації, що використовує тест (Користувач)

Розділи Кодексу:

A -- Розробка та вибір відповідних (надійних) тестів

B -- Адміністрування тесту та виставлення оцінок

C -- Повідомлення результатів, та інтерпретація результатів тесту

D -- Інформування кандидатів

З метою загального ознайомлення з Кодексом, далі наведено переклади довільно обраних пунктів Кодексу:

A Розробка та вибір відповідних (надійних) тестів

Зобов'язання Розробника:

...

A.1 Надати докази того, що саме вимірює тест. Надати інформацію щодо рекомендованого використання тесту, переваги та обмеження тесту, включно з інформацією щодо точності (рівня похибки) підсумкового балу за тест.

A.2 Надати опис того, як було обрано зміст та вміння, що оцінюються тестом, а також опис процесу розробки тесту.

A.5 Надати докази того, що технічні якості, включно з надійністю та валідністю тесту, відповідають поставленим цілям.

Зобов'язання Користувача:

A.1 Встановити цілі тестування, зміст та вміння, що мають бути оцінені. Визначити аудиторію, на яку розраховано тестування. Обрати та використовувати найбільш валідний (надійний) тест, базуючись на ретельній перевірці доступної інформації.

A.4 Обирати тести через процедури, виконувані особами з необхідними знаннями, уміннями, що пройшли відповідні тренінги

B Адміністрування тесту та виставлення оцінок

Зобов'язання Розробника:

В.3 Надати Користувачу і кандидатам інформацію щодо формату тестових завдань і процедур надання кандидатами відповідей, включно з інформацією щодо необхідності використання будь-яких додаткових матеріалів чи обладнання

Зобов'язання Користувача:

В.3 Надати кандидатам можливість ознайомитись з форматом завдань та додатковими матеріалами чи обладнанням, що можуть бути необхідні під час тестування

С Повідомлення результатів, та інтерпретація результатів тесту

Зобов'язання Розробника:

С.3 Точно вказати допустиме використання результатів тесту і попередити Користувача про можливе неправильне використання.

Зобов'язання Користувача:

С.5 Уникати використання оцінки за одне тестування в якості єдиного визначника для прийняття рішення щодо кандидата. Використовувати оцінку за тест в поєднанні з іншою інформацією по кандидату для прийняття рішень.

Д Інформування кандидатів

Д.3 Розробники і користувачі тесту мають інформувати кандидатів про природу тесту (*маються на увазі параметри тесту*), про права та обов'язки кандидатів, про належне використання (інтерпретацію) отриманої оцінки, і про процедури розв'язання спорів щодо оцінок.

Д.6 Опишіть процедури для дослідження і обставин врегулювання ситуацій, що можуть призвести до відміни чи відкликання оцінки за тест. Зокрема, у випадку недотримання кандидатом процедур тестування.

Д.7 Опишіть процедури, які допоможуть кандидату чи іншій зацікавленій стороні отримати більше інформації про тест, зареєструвати апеляцію, і вирішити проблему.

ІІІ. МОНІТОРИНГ ПРОЦЕСУ АНОНІМНОГО ТЕСТУВАННЯ КАНДИДАТІВ НА ПОСАДУ СУДДІ, ПРОВЕДЕНОГО ВККСУ 5 ЧЕРВНЯ 2012 РОКУ (витяг).

Сергій Мудрук

1. Анонімне тестування і стандарти екзаменування

1.1. Застосування Стандартів до роботи ВККСУ

Вища кваліфікаційна комісія суддів України на сьогодні очевидно є «органом, що здійснює сертифікацію персоналу», тому підпадає під дію відповідного стандарту.

Статті ДСТУ ISO/IEC 17024:2005 сформульовано у досить загальний і простий спосіб. Це дозволяє ВККСУ в порівняно короткі строки виконати всі вимоги цих стандартів, але за умови поетапної акуратної ревізії пов'язаних з іспитами процедур, норм і інституційних засад. Особливо, якщо виконувати ревізію за підтримки партнерів/експертів, досвідчених у сфері виконання норм цих стандартів.

На сьогодні стандарт ДСТУ ISO/IEC 17024:2005 в значній мірі не виконується Комісією, і ми детально показуємо це далі в тексті. Чи є це невиконання Державного стандарту правопорушенням? Відповідь – ні, тому, що:

-- виконання стандартів в Україні є справою добровільною, ст.11 Закону України «Про стандартизацію» стверджує: «Стандарти застосовуються на добровільній основі»;
-- ВККСУ виконує перш за все вимоги Закону «Про судоустрій та статус суддів» та інші релевантні закони і підзаконні акти, тому в контексті про можливе правопорушення слід оцінювати в першу чергу саме таке виконання.

Ми поверхнево (через брак ресурсів) дослідили питання виконання міжнародного стандарту ISO 17024 ліцензуючими та сертифікуючими державними органами низки країн ЄС та СНД. Виконання вимог ISO 17024 державними органами (навіть так званих розвинутих країн) в своїй власній роботі з ліцензування професійної діяльності, виявилось явищем не повсюдним. В багатьох випадках ліцензуючі державні органи делегують досить складну екзаменаційну діяльність професійним недержавним установам, екзаменаційна робота яких повністю відповідає ISO 17024. Прецеденти такого делегування державними органами складної екзаменаційної діяльності є і в Україні (наприклад – Держгірпромнагляд).

Відповідно до відомої світової практики, на побудову (створення) досконалої та відкритої системи сертифікаційного/ліцензійного екзаменування у державних органів чи саморегулюючих професій може піти багато років, іноді десятиліть. Наразі можна констатувати, що система екзаменування ВККСУ (щонайменше в частині анонімного тестування) прогресує швидко і успішно. Тому, судячи з досягнутого в порівнянні з 2011 роком прогресу, наші українські перспективи є дуже оптимістичними.

1.2. Завади відповідності стандартам екзаменування

При всьому бажанні удосконалювати екзаменаційну роботу, що демонструє наразі ВККСУ, на заваді відповідності роботи ВККСУ стандартам екзаменування стоять, в тому числі, певні недоліки законодавчо-нормативного поля, в якому працює Комісія. Надання можливості для ВККСУ працювати у відповідності до міжнародних та українських стандартів є питанням політичної волі і законодавчих ініціатив. І в цьому питанні допомогли б надання Комісією роз'яснень та пропозицій законодавцю. Детальніша інформація щодо відповідності/невідповідності стандартам екзаменування наведена, разом з рекомендаціями в наступних розділах. Якщо вдасться не загальмувати швидкість позитивних змін (включно з законодавчо-нормативними змінами), то протягом одного-двох років відбір кандидатів на посаду судді вперше в Україні зможе стати зразком дотримання міжнародних стандартів для аналогічних систем в інших країнах.

2. Норма чи критерій?

Об'єктом подальшої роботи над нормативною та законодавчою базою іспитів має стати поступовий перехід від нормо-орієнтованого підходу до критерій-орієнтованого підходу у визначенні рівня вимог до кандидатів, і, відповідно, до визначення прохідних балів.

Зараз при визначенні прохідних балів виконується приблизне планування конкретної кількості кандидатів, що мають скласти іспит, незалежно від загальної кількості допущених до іспиту кандидатів, і, найголовніше, незалежно від рівня їх підготовки. Застосування нормо-орієнтованого підходу не перешкоджає змагальності у відборі, але не дозволяє управляти рівнем компетентності кандидатів, що пройшли відбір, ясно і прозоро встановлювати мінімальні стандарти компетентності для призначення вперше, і, в підсумку, не націлене на безпосереднє пряме стимулювання зростання рівня компетентності осіб, що призначаються на посаду судді вперше.

Для забезпечення щонайменше мінімально-допустимого по якості рівня відправлення правосуддя призначеними вперше суддями, необхідно забезпечити не просто відбір кандидатів, успішних відносно групи допущених до іспитів кандидатів, натомість, необхідно здійснювати відбір кандидатів, успішних відносно критеріїв компетентності.

Поточна екзаменаційна робота ВККСУ повністю відповідає закріпленому в нормативній базі цього екзаменування нормо-орієнтованому підходу, тому для стратегічного покращення результатів діяльності слід поступово переглянути законодавчу і нормативну базу для іспитів.

3. На що націлене, і на що має бути націлене анонімне тестування?

У відповідності до чинного законодавства, метою проведення анонімного тестування є оцінка рівня загальних теоретичних знань кандидатів у галузі права. Іншими словами, якщо в склад анонімного тесту буде включено запитання для визначення рівня здатності застосовувати правові знання чи рівня здатності проводити аналіз ситуацій в розрізі правових знань, то це формально буде порушенням вимог законодавства, оскільки такі запитання оцінюють не знання, а більш високі когнітивні рівні досягнень і здібності кандидата, ніж просто знання. Таким чином, наявність такого роду запитання

в тесті дехто може розглядати як незаконне, як пред'явлення не передбачених Законом «Про судоустрій та статус суддів» завищених вимог до кандидатів.

Поточна офіційна постановка задачі для анонімного тестування є очевидно звуженою, і практика цього екзаменування напевне *de facto* виходить за рамки перевірки лише знань. Зокрема, опубліковані заздалегідь на офіційному сайті ВККСУ приклади п'яти екзаменаційних питань, формально кажучи, перевіряють не стільки знання (тобто просто запам'ятовування), як розуміння і застосування матеріалу. Крім того, виконавши аналіз грейдингової (оцінювальної) інформації тестування 22.05.2011 і 05.06.2012 ми можемо стверджувати, що велика кількість запитань цього тесту відповідає не рівню «знання», а щонайменше рівню «розуміння» і більш високим рівням.

Визначення рівня загальних теоретичних знань кандидата у галузі права є не більше ніж оцінка підсумку формального освітнього процесу, що особисто пройшов кандидат. Така задача є дуже доречною в учбовому процесі. Але у випадку із ВККСУ мова йде не про учбовий процес, а про кваліфікаційний процес, про оцінку професійної компетенції, в якій знання (особливо загальні) є лише однією із багатьох складових.

Зараз Комісією відповідно до законодавства вибудовується наступний ланцюжок: попередній відбір (анонімне тестування) → спеціальна підготовка в Національній школі суддів України (надалі - НШСУ) → кваліфікаційний іспит.

Оптимальна функція анонімного тестування в такому ланцюжку повинна була би полягати виключно у відборі найбільш перспективних кандидатів, з точки зору вкладення в них державних коштів на їх спеціальну підготовку, з метою успішного складення ними кваліфікаційного іспиту. В такому випадку на анонімному тестуванні слід оцінювати не лише формальні загальні знання, але й правниче мислення кандидата, його здатність до професійного судження та аналізу у застосуванні права. Такі речі є невід'ємною складовою професійної компетентності судді, і, на відміну від спеціальних знань, не можуть бути сформовані у кандидата в результаті кілька місячної спеціальної підготовки в НШСУ. Тобто, вони або є у кандидата на момент успішного складання анонімного тестування (і тоді спеціальна підготовка такого кандидата в НШСУ є доцільною з точки зору державних інтересів), або їх немає, як і шансів на те, що за декілька місяців до кваліфікаційного іспиту вони з'являться, тому вкладення державних коштів в такого кандидата є недоцільним.

Таким чином, ми настійливо рекомендуємо спонукати зміни формулювань законодавчо-нормативної бази, щоб в задачі анонімного тестування включались оцінки правничих когнітивних здібностей кандидата не лише на рівні «знання», але й на більш високих рівнях, включно з «розумінням», «застосуванням», «аналізом», і, навіть, «оцінкою». Після налагодження вказаного вище ланцюжка «Попередній відбір (анонімне тестування) → спеціальна підготовка в НШС → кваліфікаційний іспит» анонімне тестування за своїм змістом має стати близьким до того, що зараз планується як зміст тестової частини кваліфікаційного іспиту, в якому тепер з'являється запитання на вищі когнітивні рівні.

4. Яке тіло знань і наскільки глибоко має оцінювати анонімне тестування ?

За загальноприйнятими нормами, технічне завдання (таксономія) на публічне екзаменування високого рівня повинно бути:

1. загальнодоступним публічним документом;
2. містити структурований опис тіла знань/умінь/навичок/компетенцій, які покликаний оцінювати іспит – тобто перелік розділів/підрозділів;
3. щонайменше за основними розділами опис повинен містити параметри ширини охоплення матеріалу (відсоток в загальному тілі знань тесту), та глибини опанування (тобто когнітивні рівні: застосування, аналіз, оцінка та інші);
4. іспит повинен оцінювати виключно те, що описано в технічному завданні (таксономії).

Опублікована ВККСУ Програма анонімного тестування (іспиту) кандидатів на посаду судді на виявлення рівня загальних теоретичних знань у галузі права (надалі – Програма анонімного тестування) не є таксономією, оскільки не містить в собі двох важливих елементів, згаданих в п.3 вище. Проте, разом з нормами 3.3 та 3.7 Положення про порядок складення кандидатами на посаду судді анонімного тестування (іспиту) та методичку його оцінювання від 18.05.2012 (надалі - Положення про анонімне тестування) Програма анонімного тестування умовно може бути визнана в якості неповноцінної (неповної) таксономії, в якій прямо задано перелік тем (Програма), непрямо задано ширину охоплення (п.3.7 Положення) і ніяк не задано глибину опанування.

Експерти з моніторингу рекомендують:

- підготувати і опублікувати першу робочу версію таксономії анонімного тестування на основі рекомендацій внутрішньої робочої групи, яку варто організувати для створення цієї таксономії;
- в подальшому формувати таксономію у відповідності до поточних результатів Аналізу практики;
- виключити з Положення про анонімне тестування норми, релевантні для таксономії.

Ідея відносин двох документів (Положення і Таксономія) полягає в тому, що Положення стає постійним незмінним документом після налагодження всіх екзаменаційних процедур. Що стосується Таксономії, то періодично (раз на рік чи рідше) вона змінюється відповідно до зміни вимог щодо компетенції за результатами Аналізу практики. Мати Положення, що постійно змінюється, і незмінну Програму – найгірша з можливих ідей.

5. Рівень вимог та пробний тест (приклад тесту)

Ми не мали доступу до текстів питань анонімного тестування, але провели (з нагоди публікації вищевказаного зразка з п'яти запитань на сайті ВККСУ) експеримент в офісі МГО «Універсальна екзаменаційна мережа». Декільком співробітникам, що мають вищу не юридичну освіту, було надано 8-9 хвилин на відповідь на 5 опублікованих запитань. Середня успішність виявилась 3,5 правильних відповідей з 5 (або еквівалент 70 балів зі 100). Це свідчить про те, що, з одного боку, рівень вимог анонімного тестування до кандидатів не є справді жорстким, але, найголовніше, зміст запитань анонімного тестування не є специфічним і не є орієнтованим саме на юристів-практиків (адже до кандидатів висувається вимога щодо юридичного стажу), і не є занадто специфічним саме щодо освіти, отриманої за юридичною спеціальністю.

Зрозуміло, що в дійсності такі висновки формально не можна робити лише на основі дрібного локального експерименту по невеликій кількості питань. Проте, вищевикладене є підґрунтям для нашої наполегливої рекомендації опублікувати не окремий набір питань, а повний приклад всього іспиту, щоб кандидати і громадськість змогли ознайомитись з прикладом цілісного інструменту для оцінки кандидата. Враховуючи проведені апробації запитань, а також наявність якісної аналітики по двох тестуваннях 2012 і 2011 років, ми вважаємо, що нескладно буде скомпонувати такий «пробний тест», чи приклад тесту, який був би збалансований, і не містив би проблемних запитань, які викликали б публічну критику. Іншими словами, зараз є з чого кваліфіковано вибрати 100 безпроблемних перевірених запитань.

6. Відкритість vs конфіденційність

Комісія продемонструвала глибоку секторальну відкритість при проведенні анонімного тестування. Вірогідно це є на часі і необхідно суспільству для поступового подолання кризи недовіри до системи призначення суддів. В майбутньому, при досягненні стабілізації позитивної суспільної репутації системи призначення суддів, рекомендується делікатніше відноситись до конфіденційної персональної інформації кандидатів, і, можливо, трохи зсунути баланс з бажання ВККСУ до максимальної відкритості первинних видів інформації в бік конфіденційності певних даних кандидатів, зокрема екзаменаційних невдач (статус «НЕ склав») кандидатів і таких персональних даних, як точна дата народження.

Вище згадано, що продемонстрована відкритість була секторальною. Це означає, що при активній відкритості процедур і інформації в одних областях, в інших допустимих для відкритості областях була продемонстрована повна закритість. Див., наприклад, розділ 4.8 в частині Звіту про відповідність Стандартам, щодо відсутності будь-яких згадувань процедур надійності і безпеки по генерації екзаменаційних модулів (наборів зі 100 питань) і ключів для перевірки. Можна вимагати перевірки/експертизи програмного забезпечення автоматизованої системи, що вибирає тестові запитання і генерує ключ для перевірки. Але, в дійсності, ми не бачимо в цьому сенсу, тому, що більше значення має не якість використовуваного програмного забезпечення, а режим його використання конкретними посадовими особами і режими доступу разом з розумними процедурами внутрішнього контролю якості і безпеки.

7. Процедура подання документів кандидатами

Щодо процедури подання документів кандидатами.

Зі спілкування спостерігачів з кандидатами складається враження про:

- недостатність проведеної ВККСУ роз'яснювальної роботи із кандидатами щодо процедури подання документів;
- певну затримку ВККСУ з організації черги через web.

Очевидно, що в завершальний період подання документів висновки було зроблено, тому очікуємо, що наступні кампанії прийому документів пройдуть без недоліків.

8. Аналіз відповідності Стандартам

Відповідно до стандарту ДСТУ/ISO 17024:2005, впевненість у відповідних схемах (в нашому випадку – в процедурі допуску майбутнього судді до професійної діяльності) досягається шляхом:

1. загальноприйнятого процесу оцінки (анонімного тестування та кваліфікаційного екзаменування, в нашому випадку);
2. наступного нагляду; та
3. періодичного контролю компетентності ліцензованих/сертифікованих спеціалістів (в нашому випадку – призначених/обраних).

Стандарт накладає вимоги на широкий перелік аспектів роботи органу ліцензування/сертифікації професійної діяльності, а не лише на екзаменаційний компонент. Тому, здійснюючи спостереження (моніторинг), ми розглядали діяльність ВККСУ в ширшому, ніж спосіб проведення екзаменів, контексті.

В цій частині звіту, для зручності, у викладі наших зауважень та рекомендацій ми будемо дотримуватись структури тексту Стандарту (ДСТУ/ISO 17024:2005) і йти по його розділах.

Розділ 4.2. Організаційна структура

Стандарт накладає певні вимоги на організаційну структуру органу сертифікації та види взаємодії підрозділів. Зокрема, у перших трьох пунктах розділу 4.2 Стандарту мова йде про вимоги до організаційної структури для забезпечення компетентності, неупередженості та цілісності, а також про так званий «Комітет схеми». Ми бачимо на сьогодні складнощі з відповідністю ВККСУ вимогам пунктів 4.2.1-4.2.3 Стандарту стосовно внутрішньої організаційної структури ВККСУ. Проте, з особистого спілкування із ВККСУ нам відомо про усвідомлення складності і важливості екзаменаційної і пов'язаних з нею діяльностей ВККСУ. А, найголовніше, ми сподіваємось на усвідомлення необхідності певних змін і доповнень організаційної структури ВККСУ. **Звертаємо особливу увагу на необхідність створення (чи реорганізації з наданням необхідних ресурсів і повноважень) підрозділу, що, користуючись термінологією Стандарту, взяв би на себе суттєву частину функцій «Комітету по схемі сертифікації».** Зокрема, мова йде про відносно ресурсо місткий напрямок «Аналіз практики» (practice analysis) та напрямок актуалізації екзаменаційної таксономії, як основного елементу «Схеми сертифікації» в термінах Стандарту (див. окремий розділ цього Звіту – «Аналіз практики»).

Звертаємо також увагу на можливу проблематичність з відповідністю вимогам п.4.2.5 Стандарту щодо певних аспектів взаємовідносин ВККСУ та НШСУ. В п.4.2.5 Стандарту мова йде про незалежність процесу підготовки тестів та процесу оцінки.

Аналіз нормативної документації та спостереження фактичного стану справ у виробничих відносинах ВККСУ та НШСУ, НШСУ та ВНЗ України не виконувався. Хоч таке завдання можливо не є на часі як високопріоритетне, проте, ми рекомендуємо не відкладати на довго таке дослідження щодо специфічних відносин «організатори навчання» ↔ «орган сертифікації».

Ми констатуємо також певні проблеми щодо відповідності вимогам п.4.2.6 Стандарту (в цьому пункті Стандарту мова йде про роботу зі скаргами та апеляціями). Ми

спостерігаємо налагоджену роботу ВККСУ зі скаргами в основних сферах роботи ВККСУ крім однієї: апеляції кандидатів щодо результатів іспитів. Законодавство, підзаконні та нормативні акти не передбачають право кандидата на апеляцію за результатами анонімного тестування (іспиту) і відповідно роботу ВККСУ з апеляціями кандидатів. Тому, поточний стан справ можна вважати задовільним лише з точки зору формальної відповідності законодавству.

Наша рекомендація полягає в тому, щоб ВККСУ створив прозору систему роботи з апеляціями. Це не вступає в протиріччя з законодавством, яке не забороняє мати систему розгляду апеляцій для письмового тестування. Враховуючи безкоштовність цього процесу для кандидатів та реалістичність загрози очікування масової подачі апеляцій, ми готові надати консультаційну підтримку ВККСУ у створенні системи апелювання. Варто зазначити, що певна сурогатна/часткова система апеляцій вже існує *de facto*, оскільки:

1. кандидатам надана можливість звертатись з проханням про ознайомлення зі своєю екзаменаційною роботою;
2. ВККСУ фактично превентивно проводить розгляд потенційно спірних ситуацій (наприклад: випадки можливої невідповідності коду роботи);
3. ВККСУ фактично розглядає скарги кандидатів по суті екзаменаційних запитань.

Дещо, з нашої точки зору, можна покращити вже зараз, ще до створення системи апеляцій. В нормативній документації (Положенні про анонімне тестування) вказано про можливість для кандидата вмотивовано звертатись до ВККСУ з проханням про ознайомлення з його роботою. Але в нормативній документації, на жаль, нічого не вказано щодо критеріїв, за якими ВККСУ розглядає такі звернення кандидата і нічого не вказано про необхідність для ВККСУ вмотивовувати своє рішення про ненадання можливості для кандидата ознайомитись зі своєю роботою. З нашої точки зору варто врегулювати таку несправедливу несиметричність відносин ВККСУ ↔ кандидат.

Щодо п.4.2.7 Стандарту (в ньому мова йде про необхідність залучення достатньої кількості фахівців з необхідним досвідом, підготовкою і технічними знаннями), то на даний момент ця норма Стандарту частково виконується завдяки підтримці донорського Проекту USAID «Справедливе правосудді». Наша рекомендація полягає в тому, щоб ВККСУ, з одного боку, поступово налагоджувала інституційні зв'язки з організаціями в Україні, що можуть надавати необхідні для якісного екзаменування ресурси. З іншого боку, рекомендуємо Комісії посприяти змінам законодавчої і нормативної бази з метою зробити можливим і полегшити залучення сторонніх ресурсів (експертів, фахівців, колективів, якщо необхідно) у такий спосіб, щоб неможливо було б поставити під сумнів неупередженість ВККСУ до окремих кандидатів чи їх груп. Кінцева мета: мати можливість задовольнити будь-які потреби Комісії в експертизі (expertise, кваліфікованому ресурсі) закінчення навіть за умови відсутності допомоги з боку будь-яких донорських організацій. Це важливо, враховуючи те, що певні види ресурсів (наприклад – кваліфікованих досвідчених психометристів; колективи для адміністрування іспиту) ВККСУ не зможе мати в своєму складі через короткостроковість потреби в них і їх специфічність.

Розділ 4.3. Розробка і підтримка схеми сертифікації

Очевидні певні складнощі з відповідністю розділу 4.3 Стандарту вирішяться, якщо буде враховано висловлені вище зауваження щодо відповідності пп.4.2.1-4.2.3. Тобто, якщо в складі ВККСУ буде створено департамент, якому буде надано необхідні ресурси і

повноваження, що виконуватиме функції Комітету по «схемі» (користуючись термінологією Стандарту).

Розділ 4.4. Система управління

Конкретних зауважень по цьому розділу Стандарту у нас немає, в контексті очікування згаданих вище змін організаційної структури.

Розділ 4.5. Субпідряд

Аналізуючи відповідність даному стандарту висловлюємо не пов'язані одне з одним зауваження і рекомендацію.

Зауваження: відносини між ВККСУ та НШСУ по розробці екзаменаційних запитань є субпідрядом. В контексті стандартів екзаменування такий субпідряд не може розглядатись як «науково-методичне забезпечення діяльності».

Поточний стан цього субпідряду, а саме те, як врегульовано деталі взаємовідносин, цілком не узгоджується з п.4.5.2.b Стандарту. Над даним узгодженням слід попрацювати не пізніше ніж під час підготовки згаданої вище зміни в організаційній структурі ВККСУ. При цьому, узгодження з вимогами Стандарту може виявитись простою технічною формальністю. Узгодження може полягати як в легкій зміні відносин ВККСУ з НШСУ щодо розробки екзаменаційних питань, так і в роботі над текстовим повідомленнями (messages) суспільству, що міститимуться в нормативних документах щодо проведення іспитів. Мова йде про закріплення незалежності підготовки і оцінки через констатацію абсолютного контролю Комісією процесу розробки іспитів.

Рекомендація: в контексті розділу 4.5 Стандарту запропонувати зміни до законодавства і нормативних документів, щоб зробити теоретично можливими субпідряди або залучення сторонніх фахівців, на основні види робіт. Лише для прикладу – пару очевидних необхідностей:

1. можливість залучення суддів-практиків до перевірки практичних завдань;
2. можливість залучення добре підготованих прокторів для проведення адміністрування іспитів.

Розділ 4.6. Записи

Зауважень на відповідність цьому Стандарту немає.

Розділ 4.7. Конфіденційність

Рішення ВККСУ про публікацію точних дат народжень зареєстрованих кандидатів не можна вважати вдалим. З одного боку, П.І.П. і точна дата народження досить точно (унікально) ідентифікує кандидата, що, судячи з усього, і слугувало причиною прийняття рішення про публікацію дат народжень. Але, з іншого боку, в нашому суспільстві є помітний великий прошарок громадян, що вважають точну дату народження чутливою персональною інформацією. Крім того, наприклад, публікація віку (на момент складення іспиту) та населеного пункту проживання кандидата замість точної дати народження дала б той же рівень точності ідентифікації кандидата, але додатково давала б можливість громадськості побачити кандидата, що живе поруч. В свою чергу, це дозволить ВККСУ вчасно отримувати сигнали від громадян, щодо кандидатів з їх місця проживання в особливо резонансних ситуаціях небажаних особистих якостей кандидата. Публікація віку на момент складення іспиту, не менш

успішно, ніж публікація дати народження, дозволить також громадськості контролювати виконання конституційної вимоги щодо вікового цензу в 25 років.

Розділ 4.8. Безпека

Немає справді великих і проблемних зауважень до безпеки екзаменаційних матеріалів в контексті організації адміністрування анонімного тестування (детальніше зауваження – див. в частині Звіту щодо адміністрування). Проте, режим безпеки на шляху від розробників екзаменаційних питань через генерацію екзаменаційного модуля, до друку екзаменаційних матеріалів **не описано в публічних документах навіть в загальних рисах** (а цього було б досить). Тому поки що розробка завдань, а, особливо, процедури, пов'язані з генерацією екзаменаційного модуля можуть бути об'єктом підозр аж до наступного екзаменаційного циклу, перед яким ми настійливо рекомендуємо дати опис режиму безпеки в публічних документах (в тому обсязі, який допустимий для відкритих документів)

Розділ 5.1. Загальні вимоги до спеціалістів.

Оскільки зараз постійно і регулярно відбувається підвищення кваліфікації фахівців ВККСУ та НШСУ, то єдиним зауваженням є необхідність обліку відповідного (до процесу екзаменування) підвищення кваліфікації (і самої кваліфікації) співробітників. Цього, зокрема, вимагає п.5.4.1 Стандарту.

Розділ 5.2. Вимоги до екзаменаторів.

Під «екзаменаторами» Стандарт розуміє і тих, хто адмініструє іспит (тобто виконує технічну роботу, прокторів), і тих, хто безпосередньо приймає рішення по результатах екзаменування і проводить оцінку. Наші зауваження до адміністрування іспиту 05.06.2012р. надані в іншій частині звіту. В контексті розділу 5.2 Стандарту відзначимо лише те, що нам не вдалось віднайти ні письмових згадувань про забезпечення уникнення конфліктів інтересів, ні поспостерігати якісь неписані методи/процедури в дії. Тому ми пропонуємо запровадити відображені в публічних документах процедури виявлення та уникнення конфлікту інтересів персоналу ВККСУ. Самий лише факт наявності цих процедур в нормативних документах зніме більшість потенційних чи можливих претензій/питань громадськості щодо конфлікту інтересів. Навіть якщо ці процедури не будуть детальними і досконалими.

Розділ 6.1. Подання заяви (процес сертифікації).

Враховуючи виконані Комісією прямо в період проведення прийому документів корекції цього процесу, та реакцію на критику в пресі і на форумах, можна вважати процес реєстрації кандидатів умовно задовільним. Відповідні для тестування 05.06.2012р. документи публікувались оперативно, рішення приймались практично вчасно. Проте, ми надаємо формальне зауваження в контексті розділу 6.2 Стандарту: **весь пакет нормативних документів і рішень, що описують процес екзаменування належить готувати і публікувати до моменту відкриття реєстрації на тестування.**

Розділ 6.2. Оцінювання

Глобальних зауважень в контексті розділу 6.2 Стандарту ми не маємо. Конкретні зауваження та рекомендації надано в інших розділах звіту. В інших розділах звіту не згадано лише одне зауваження по п.6.2.1.с Стандарту: в публічних документах відсутня специфікація (будь-який опис) суті спеціальної перевірки кандидатів / їх документів.

Розділ 6.3. Рішення щодо сертифікації

В зв'язку з цим розділом Стандарту, ми маємо два зауваження.

Пункт 6.3.3.с вимагає згадування (посилання на) версії вимог до кандидата (версії таксономії). Тобто, якщо до кваліфікаційного іспиту допускається певний список кандидатів, то в ньому має бути зазначено, якій версії попередніх вимог (анонімного тестування) відповідає кандидат. Тобто, в якому році (за якою програмою) він складав анонімне тестування. Відповідно, якщо вважати еквівалентом до сертифікації (в термінах Стандарту) факт включення кандидата в кадровий резерв з конкретною числовою величиною рейтингу, то Стандарт вимагає, щоб в цьому ж реєстрі обов'язково вказувалась версія кваліфікаційних вимог, яким кожний конкретний кандидат відповідає, тобто має бути вказано, в якому році кандидат успішно склав кваліфікаційний іспит.

Забігаючи вперед, до кваліфікаційного іспиту, звертаємо увагу на вимогу п.6.3.1 Стандарту, з якого випливає, що особи, які приймають рішення «щодо сертифікації» (тобто безпосередньо 11 членів ВККСУ) не можуть брати участь в навчанні кандидатів, чи безпосередньо в прийнятті іспиту у кандидата. Дотриманням цієї норми Стандарту уникають цілу низку складних конфліктів інтересів. Для контексту практичної роботи ВККСУ це означає, що безпосередньою перевіркою практичних завдань кандидатів, в рамках кваліфікаційного іспиту, не може займатись голосуючий склад ВККСУ, щоб мати можливість приймати остаточне рішення (в т.ч. голосувати) по кандидату погоджуючись чи не погоджуючись з оцінками, виставленими спеціалістами, що проводили оцінку.

Норми п.6.3.1 Стандарту є обґрунтованими і загальноприйнятими в усьому світі, включно з державними органами ліцензування в великій кількості країн. Тому ми рекомендуємо ВККСУ пропонувати зміни до законодавства, які зроблять недопустимим виконання планової перевірки екзаменаційних робіт безпосередньо членами ВККСУ.

Розділи 6.4 та 6.5 Нагляд та Ресертифікація

Зауважень по цих розділах Стандарту нема, враховуючи те, що спільну роботу ВККСУ, НШСУ, ВРЮ та ДСА в окреслених напрямках можна вважати за таку, що ще не є остаточно налагодженою. Моніторинг цієї діяльності є окремою об'ємною задачею.

Розділ 6.6 Використання сертифікатів та знаків

Цей розділ Стандарту не повністю доречний, щоб бути застосованим до ВККСУ, тому зауважень у нас немає. Проте, ВККСУ варто вивчити доцільність рекомендованої в пп.6.6.2 с і d вимоги до успішних кандидатів про підписання угоди (згоди) про, в тому числі:

- недопустимість участі кандидата у доборі таким способом, що це може скомпрометувати ВККСУ чи зашкодити репутації Комісії;
- недопустимість заяв стосовно добору, які ВККСУ може розцінити як такі, що вводять в оману чи як не уповноважені;
- припинення використання будь-яких рекламацій щодо включення до кадрового резерву, у випадку прийняття Комісією рішення про виключення кандидата з нього.

9. Аналіз практики

«Аналіз практики» (practice analysis) повинен стати одним з центральних елементів роботи виконавчого органу ВККСУ. Practice analysis не є разовим заходом, а постійною діяльністю, складовими якої є:

- періодично виконувати опитування різних категорій суддів, та користувачів професійних послуг судді, а також прокурорів, адвокатів і працівників апарату судів;
- постійний аналіз судової статистики, аналіз стану справ в судах з кваліфікацією та навичками суддів;
- формування експертних груп і аналіз їх висновків;
- інші заходи для визначення якісних і кількісних характеристик професійної компетенції судді, що відповідає поточному стану вимог до представників цієї професії.

Як правило, мають видаватись (публікуватись) щорічні звіти по аналізу практики. Одним з основних результатів аналізу практики має стати постійно оновлюваний і деталізований перелік кваліфікаційних вимог до кандидатів, що вступають в професійну діяльність судді. Тобто – екзаменаційні таксономії, які слугують технічним завданням на проведення екзаменаційного відбору.

Додатковими результатами є визначення пріоритетів та акцентів для освітньої діяльності. Мова йде не лише про задачі Національної школи суддів України, але й про можливі певні рекомендації школі вищим навчальним закладам.

Планування діяльності НШСУ (визначення нових напрямків для підготовки; визначення тематичних пропорцій між учбовими планами, пріоритетів і т.ін.) повинно здійснюватись на основі періодично оновлюваних стратегічних рішень спеціалізованого колегіального органу суддівського самоврядування (наразі – ВККСУ) на базі і з урахуванням Аналізу практики.

Оскільки згідно зі ст.54 Закону «Про судоустрій і статус суддів» кожен суддя має періодично проходити двотижневу підготовку в Національній школі суддів України (надалі – НШСУ), цей інститут (НШСУ) природно може стати однією з баз для постійного проведення аналізу практики.

Наприклад, підготовка кожного з суддів в НШСУ може починатись з формалізованого (заповнення форм і анкет) і неформального (колоквиум/інтерв'ю/співбесіда) аналізу практики конкретного судді/суду. З одного боку, це надасть матеріал для загального Аналізу практики, а з іншого боку, дозволить НШСУ будувати індивідуалізовану роботу з судьями та працівниками апарату суддів. Тобто, участь НШСУ в Аналізі практики виглядає органічною, оскільки крім безпосереднього отримання НШСУ загальної картини практики «з перших вуст», така участь дозволить правильно формувати індивідуальні частини підготовки суддів, які доповнять стандартні планові заходи і теми підготовки/перепідготовки суддів.

Необхідно окремо звернути увагу на те, що з міркувань «протиवाг та стримувань», а також задля уникнення конфлікту інтересів, центром ваги діяльності з Аналізу практики не може стати НШСУ (хоч це і може здатись дуже зручним і практичним). Мова йде про те, що результати Аналізу практики покладаються в основу політичних

рішень, що формують задачі НШСУ. Тому слід уникнути ситуацій, коли спеціалізована науково-методична освітня структура, що формується адміністративними рішеннями, а не механізмами самоврядування, вносить визначальний за обсягом внесок в формування власних стратегічних задач, функціонуючи як самодостатня саморегульована організація.

Заходи Аналізу практики, що їх проводитиме НШСУ, слід виконувати за планом і під керівництвом відповідного підрозділу ВККСУ, на який буде покладено роботу з аналізу практики і формування «схеми сертифікації», користуючись термінологією Стандарту ДСТУ/ISO 17024.

10. Узагальнення спостережень та висновки

В цьому розділі зроблена спроба систематизувати і повторно згадати основні спостереження та висновки, що надано в тексті цього Звіту.

ВККСУ продемонструвала відкриту і конструктивну співпрацю в контексті проведення моніторингу. Критика і пропозиції оперативно приймалися до розгляду. Велику кількість пропозицій і зауважень було враховано. Не залишається сумнівів в бажанні Комісії максимально прозоро проводити анонімне тестування і удосконалювати діяльність.

Комісія прийняла сміливе і відповідальне рішення про централізоване одночасне екзаменування всіх кандидатів в одному місці в один час. Такий спосіб адміністрування екзаменів надзвичайно складний і є професійним викликом навіть для досвідчених фахівців з адміністрування екзаменів. Перехід на новий метод адміністрування, при його правильній реалізації, дозволяє витримати єдині високі стандарти екзаменування. Разом з тим, було виявлено низку суттєвих недоліків в адмініструванні, про що див. нижче (Загальні висновки щодо адміністрування екзамену).

Є доцільним розглядати відповідність екзаменаційної діяльності ВККСУ до вимог ДСТУ ISO/IEC 17024:2005.

На сьогодні цей стандарт в значній мірі не виконується при проведенні екзаменування ВККСУ. На заваді відповідності роботи ВККСУ стандартам екзаменування стоять недоліки законодавчо-нормативного поля, в якому працює Комісія. Надання Комісії можливості працювати у відповідності до міжнародних та українських стандартів є питанням політичної волі і законодавчих ініціатив.

Однією з перешкод відповідності стандарту є невідповідність вимогам до організаційної структури ВККСУ для забезпечення компетентності, неупередженості та цілісності, а також відсутність в організаційній структурі ВККСУ підрозділу, що, мовою Стандарту, виконував би функції «Комітету схеми», для чого займався би:

- аналізом практики;
- формуванням і періодичним оновленням таксономій
- процедурами екзаменування
- іншими принциповими для екзаменування функціями, зокрема контролем і координацією роботи над базою екзаменаційних запитань, після екзаменаційним аналізом і т.ін..

Екзаменаційна робота ВККСУ на сьогодні повністю відповідає закріпленому в нормативній базі нормо-орієнтованому підходу до визначення рівня мінімальних вимог до успішного кандидата на посаду судді.

Поточна постановка задачі для анонімного тестування є очевидно звуженою, але добре, що практика цього екзаменування *de facto* виходить за рамки перевірки лише знань (тобто просто здатності кандидата відтворити запам'ятовану інформацію).

Не створено і не опубліковано таксономію (технічне завдання) анонімного тестування. Опубліковану Програму анонімного тестування, разом з нормами п.3.3 та п.3.7 Положення про анонімне тестування від 18.05.2012р можна вважати частковою заміною Таксономії.

Рівень вимог анонімного тестування до кандидатів не є справді високим, вимоги анонімного тестування не орієнтовані саме на юристів-практиків.

В пресі і на інтернет-ресурсах відмічались проблеми з організацією процесу подання кандидатами документів, Заключний період подання документів, після врахування Комісією критики, пройшов без зауважень.

На сьогодні не існує адекватної системи апеляцій щодо результатів іспитів ВККСУ. Законодавством не передбачено обслуговування апеляцій Комісією. А оскарження через ВРЮ чи Адмін.суд по своїй суті є оскарженням чи самої системи екзаменування, чи дій ВККСУ, а не самої оцінки кандидата. В фокусі такого оскарження є дії посадових осіб, а не суть оцінки кандидата. За загально прийнятими в професійному екзаменуванні правилами апеляцію належить розглянути органу, що затверджує оцінку, і лише після цього, якщо необхідно, оскаржувати саму систему проведення чи оцінки іспиту, чи дії посадових осіб.

Не вдалим було рішення про відкриту публікацію точних дат народжень всіх кандидатів.

В публічних документах не згадується навіть в загальних рисах режим безпеки на шляху від розробників екзаменаційних питань через генерацію екзаменаційного модуля, до друку екзаменаційних матеріалів.

Не вдається віднайти письмові згадування в публічних документах забезпечення уникнення конфліктів інтересів співробітників (і членів Комісії і працівників Секретаріату). Процес професійного (кваліфікаційного в широкому сенсі) екзаменування – специфічний. Крім можливих корупційних аспектів в ньому є велика кількість потенційних етичних проблем, тому загальних норм законодавства про протидію корупції тут поза будь-якими сумнівами не досить.

Те, що на сьогодні голосуючий склад ВККСУ особисто займається перевіркою кваліфікаційних екзаменаційних робіт (відповідно до трактування законодавства Вищою радою юстиції) розходиться з вимогами загальноприйнятих екзаменаційних стандартів (зокрема п.6.3.1 міжнародного стандарту ISO 17024 і відповідного національного стандарту ДСТУ) і вимагає уточнення вимог законодавства, якщо необхідно, то і його змін.

Аналіз результатів оцінювання екзаменаційних робіт свідчить про **надійність** (в сенсі надійності, як внутрішньої узгодженості) тесту 05.06.2012р.

Присутні зовнішні ознаки **валідності** тесту 05.06.2012р. Остаточні висновки щодо валідності мають робити фахівці-правники, що мають доступ до текстів екзаменаційних запитань і до аналітичних звітів, що ми надаємо (див. Стандартизовані звіти UEnet в Додатках)

11. Рекомендації

Наведені в цьому розділі рекомендації більш детально викладено разом з поясненнями в попередніх розділах Звіту, в цьому ж розділі наведено лише основні з них.

Перш за все варто звернути увагу на зміни, які потребують внесення відповідних змін до вітчизняного законодавства, а саме.

1. Привести екзаменаційну діяльність ВККСУ у відповідність до вимог ДСТУ ISO/IEC 17024:2005. Ініціювати (через надання детальних роз'яснень і пропозицій законодавцю) зміни в законодавстві в таких напрямках:
 - перехід від нормо-орієнтованого підходу до критерій-орієнтованого підходу у визначенні рівня вимог до кандидатів на посаду судді.
 - зміни формулювань законодавчо-нормативної бази, щоб в задачі анонімного тестування включались оцінки правничих когнітивних здібностей кандидата не лише на рівні “знання”, але й на більш високих рівнях, включно з “розумінням”, “застосуванням”, “аналізом”, і, навіть, “оцінкою”.
2. Запропонувати зміни до законодавства, які зроблять недопустимим виконання планової перевірки екзаменаційних робіт безпосередньо членами ВККСУ, для того, щоб вони мали можливість у відповідності до загальноприйнятих екзаменаційних стандартів, без будь-яких конфліктів інтересів і упередженості, приймати остаточні рішення (голосувати) по кандидату, тобто встановлювати остаточну оцінку кандидата.
3. Зробити анонімне тестування (тобто перше відбіркове екзаменування) за своїм змістом близьким до того, що зараз планується як зміст тестової частини кваліфікаційного іспиту.
4. Проводити анонімне тестування декілька разів на рік з введенням обмеження для кандидатів на повторну участь в анонімному тестуванні не раніше, ніж через рік після попередньої участі.
5. Сприяти змінам законодавчої і нормативної бази з метою зробити можливим і полегшити залучення сторонніх висококваліфікованих ресурсів (експертів, фахівців, колективів, підрядників) у такий спосіб, щоб неможливо було б поставити під сумнів неупередженість Комісії до окремих кандидатів чи їх груп. Зробити можливим законність субпідрядів на основні види робіт, крім принципових функцій (наприклад, функції «комітету схеми», прийняття рішень по оцінці та інше). Це дозволить виконувати специфічні роботи професійного екзаменування силами висококваліфікованих фахівців.

Окрім того, варто зауважити на потребі у змінах, які можливо запровадити на підзаконному рівні:

1. Підготувати і опублікувати першу робочу версію таксономії анонімного тестування.

2. В подальшому формувати таксономію у відповідності до поточних результатів Аналізу практики;
3. Виключити з Положення про анонімне тестування норми, які є частиною таксономії.
4. Опублікувати приклад повного екзаменаційного варіанта (набору запитань) анонімного тестування, а не обмежений набір питань, щоб кандидати і громадськість змогли ознайомитись з прикладом цілісного інструмента для оцінки кандидата.
5. Створити в складі Секретаріату ВККСУ департамент, що взяв би на себе функції “Комітету схеми” (користуючись мовою ISO 17024), надати йому достатньо ресурсів та повноважень для створення та регулярного оновлення екзаменаційних таксономій, для проведення аналізу практики, та виконання інших важливих для якісного проведення екзаменаційної діяльності функцій, включно з координацією розробки екзаменаційних модулів, організації адміністрування іспитів, організації перевірки екзаменаційних робіт.
6. Запровадити так званий аналіз практики (practice analysis), який має стати постійною діяльністю відповідного департаменту ВККСУ. Публічними результатами цієї роботи мають стати щонайменше періодично оновлювані кваліфікаційні вимоги до кандидатів (екзаменаційні таксономії) та публікування щорічних звітів з аналізу практики.
7. Створити публічну систему роботи ВККСУ з апеляціями.
8. В розділі Положення про анонімне тестування, в якому згадується право кандидата вмотивовано звертатись з проханням про ознайомлення зі своєю екзаменаційною роботою, включити хоча б мінімальний обсяг інформації про принципи чи порядок розгляду такого прохання Комісією. Щонайменше, необхідно включити симетричну норму, у відповідності до якої Комісія зобов’язана вмотивовувати рішення (якщо таке приймається) про відмову в доступі до роботи, якщо кандидату (відповідно до тексту Положення) належить вмотивовувати своє прохання про можливість ознайомитись зі своєю роботою.
9. Для точної ідентифікації кандидатів в публічних документах використовувати крім П.І.П. не точну дату народження, а комбінацію з віку на момент складення екзамену та населеного пункту проживання кандидата.
10. Дати опис режиму безпеки розробки екзаменаційних питань, генерації екзаменаційних модулів (разом з процедурами впевненості і контролю якості) в публічних документах, в тому обсязі, який допустимий для відкритих документів.
11. Налагодити облік підвищення кваліфікації співробітників щодо екзаменаційних технологій.
12. Запровадити відображені в публічних документах процедури виявлення та уникнення конфліктів інтересів персоналу ВККСУ.
13. Дати інформацію в публічних документах щодо процедури спеціальної перевірки, особливо якщо суть спеціальної перевірки ВККСУ виходить за рамки простої і очевидної імплементації норм Закону про протидію корупції.
14. В зв’язку з тим, що вимоги до кандидатів змінюються з часом, в загальних реєстрах кандидатів, що вже склали екзаменування ВККСУ (приклади: список кадрового резерву; список допущених до кваліфікаційного іспиту) необхідно вказувати в якому доборі (році) кандидат успішно склав відповідний іспит.
15. Допускати до участі в адмініструванні лише відібраний і добре підготований персонал, що пройшов низку тренінгів, включно з психологічною підготовкою та підготовкою до толерантного спілкування з кандидатами. Перед доступом до

роботи, адмініструючий персонал має проходити контроль рівня підготованості (екзамен/залік чи, щонайменше, співбесіду з незацікавленим співробітником, здатним об'єктивно оцінити рівень підготовки проктора (уповноваженого представника) для виконання ним функцій по адмініструванню іспиту). Якщо власного підготованого і перевіреного трудового ресурсу виявиться не досить, залучати додатково сторонній досвідчений трудовий ресурс для адміністрування екзаменів.

16. Заздалегідь інформувати зареєстрованих кандидатів щодо очікуваних процедур, в яких вони прийматимуть участь. Для цього публікувати популярні роз'яснення щодо послідовності дій кандидатів і адмініструючого персоналу в зв'язку з іспитом, а не тільки нормативні документи, положення і офіційні об'яви.
17. Приміщення екзаменаційного центру та прилегла територія мають бути забезпечені достатньою кількістю правильно розміщених інформативних вказівників і плакатів.
18. Для прискорення процесу первинної реєстрації рекомендуємо планувати в півтора-два рази більшу кількість столів реєстрації (і реєстраторів).
19. Переглянути загальну організацію процесу реєстрації з метою виключення помітного очікування кандидатами між моментом отримання картки та заведенням в сектор. Для цього по мірі отримання кандидатами карток формувати групи, що заводились би на отримання коду, після чого також по мірі накопичення формувати одночасно декілька секторів, з випадковим розподіленням кандидатів по секторах і по робочих місцях в секторах.
20. Робочі місця в секторах має бути пронумеровано. Розподіл кандидатів по їх робочих місцях має відбуватись жеребкуванням, а не за розпорядженнями конкретних уповноважених представників на місцях, і не хаотично по волі самих кандидатів. Бажано, щоб жеребкування відбувалось за участю самих кандидатів, хоча, при певних умовах, допустиме і комп'ютерне жеребкування.
21. При плануванні адміністрування іспиту необхідно мати набір резервних планів на випадок природних (град, заметіль і т.ін.), техногенних (відсутність електроенергії, аварія водогону і т.ін.) та громадських (заворушення і т.ін.) форс-мажорів.
22. При плануванні адміністрування іспиту необхідно пам'ятати, що організатор такого масового заходу несе відповідальність (незалежно від того, що про це говорить Закон) за безпеку і неушкодження здоров'я запрошених на захід кандидатів. Тому варто крім залучення чергових медичних працівників залучати додатковий персонал для забезпечення охорони і безпеки, і не допускати ситуацій, що загрожують здоров'ю всіх чи окремих кандидатів (навіть якщо результатом такого роду «несуттєвої» загрози буде нежить всього у декількох кандидатів).
23. У випадку неможливості забезпечення належного рівня підготовки адмініструючого персоналу до запобігання використанню засобів зв'язку, рекомендуємо санкціоноване застосування обладнання, що перешкоджає використанню мобільного GSM/CDMA зв'язку в приміщенні проведення іспиту.
24. Налаштувати систему супроводу кандидатів на «перерву», забезпечивши регламентований порядок (чергу) виходу на перерву, достатні трудові ресурси для забезпечення супроводу, підготовку (тренінг) таких працівників, а також підготованих чергових в туалетних кімнатах.
25. Виконання процедур початку іспиту і його завершення має бути уніфіковано і синхронізовано, для чого слід провести додаткові тренінги з уповноваженими представниками по секторах.

26. При опломбовуванні пакетів з екзаменаційними матеріалами після закінчення іспиту на пакеті слід вказувати номер сектора, ППП і підписи уповноваженого представника і представника з числа кандидатів. Опціонально: склад вкладення (скільки робіт).
27. Виконувати публічний облік невикористаних екзаменаційних матеріалів.
28. Повністю реалізувати раніше задекларовану ідею супроводу (безпосередньо до місця сканування) опломбованих пакетів з екзаменаційними матеріалами представниками кандидатів.
29. Залучити кваліфікованих спостерігачів для безпосереднього контролю (спостереження) за розкриттям пакетів і обробкою робіт кандидатів.
30. Передбачити штатний спосіб корекції відповіді кандидатом під час виконання ним екзаменаційної роботи.
31. При виконанні змін чи доповнень до опублікованої інформації щодо результатів екзаменування, чи при публікації нових версій даних/документів слід публікувати також зрозумілі для стороннього спостерігача пояснення про суть змін і їх причини. Слід уникати необхідність та можливість самостійних припущень (з боку кандидатів і будь-яких сторонніх спостерігачів) щодо суті і причин змін, через відсутність прямого і простого їх пояснення з боку ВККСУ.
32. Змінити спосіб прийняття рішень щодо дискваліфікації неякісних запитань. На етапі анонімного тестування слід знімати з розгляду (з урахування) відповіді на запитання, що визнано неприйнятними для цілей тестування, а не зараховувати всім кандидатам будь-які відповіді на такі питання як вірні відповіді. Такий підхід слід безпосередньо застосовувати лише при першому анонімному тестуванні, коли набраний кандидатом бал має значення лише для статусу «Склав» чи «Не склав» і не дає внесок в значення рейтингу кандидата для конкурсу.
33. Після проведення кожного іспиту фахівцям ВККСУ рекомендується проводити контентний аналіз екзаменаційних запитань з урахуванням IRT аналізу (див. Додатки до цього звіту) на предмет контентної валідності і якості кожного запитання як інструмента оцінки. З результатом аналізу слід ознайомлювати розробників відповідних запитань для його врахування в подальших розробках.

IV. ДВНАДЦЯТЬ КРОКІВ ЕФЕКТИВНОГО СТВОРЕННЯ ЕКЗАМЕНУВАННЯ. *Сергій Мудрук, за матеріалами Стівена Даунінга*

Стівен Даунінг з Іллінойського університету (Чикаго) запропонував структурувати в 12 груп всі роботи і функції зі створення і проведення екзаменування. Оскільки між виконанням цих груп функцій є хронологічний зв'язок, їх назвали «12 кроків». Концепція «12 кроків» визнана великою кількістю фахівців як методологічно і практично цінна. «12 кроків» наочно і лаконічно, в одній концепції, представляє всі різноманітні етапи і проблеми процесу створення і використання системи екзаменування. Концепція «12 кроків» дозволяє використовувати її як «дорожню карту» при плануванні і здійсненні серйозної системи екзаменування.

Нижче наведено стислий виклад 12 кроків з посиланнями на Стандарти освітнього та психологічного тестування, розроблені трьома американськими організаціями - AERA/APA/NCSE, і визнані у всьому світі найбільш повним і серйозним зібранням вимог, виконання яких дозволяє уникнути неналежного створення, виконання і використання екзаменування (включно з неналежним використанням результатів екзаменів).

Крок №1 Загальне планування

Задачі:

Планування, зокрема, системи управління всіма аспектами розробки тестів:

- загальне формулювання Конструкту;
- формулювання способу інтерпретації результатів тестування;
- вибір форматів тестів; визначення очікуваних доказів валідності тестування;
- ясне формулювання цілей тестування;
- бажаний тип/формат висновків щодо кандидатів за результатами тестування;
- психометрична модель;
- планування часу, терміни;
- застосовувані заходи безпеки;
- заходи контролю якості.

Стандарти:

1.1, 3.2, 3.9

Крок №2 Визначення контенту

Задачі:

- План змісту екзамену, як документ, з якого обирається зміст конкретного тесту;
- розгляд різних методів визначення змісту, пов'язаних з метою оцінювання;
- найважливіші джерела доказу змістової валідності;
- специфікація Конструкту.

Стандарти:

1.6, 3.2, 3.11

Крок №3 Специфікація тесту (технічне завдання)

Задачі:

- операційне визначення змісту;
- структура доказів валідності, що базуються на обґрунтованій системі вибірки з контенту екзаменування (План змісту екзамену);
- нормо-орієнтоване чи критерій-орієнтоване тестування;

-- бажані параметри екзаменаційних завдань.

Стандарти:

1.6, 3.2, 3.3, 3.4, 3.11

Крок №4 Розробка екзаменаційних завдань

Задачі:

- розробка ефективного Стимулу (Основи);
- формати;
- докази валідності, що ґрунтуються на фактичних даних;
- тренінг розробників завдань, рецензентів завдань;
- ефективне редагування завдань;
- невідповідності до Конструкту, пов'язані з недоліками розроблюваних завдань.

Стандарти:

3.6, 3.7, 3.17, 7.2, 13.18,

Крок №5 Дизайн і збірка екзаменаційних матеріалів

Задачі:

- створення і верстка екзаменаційних форм (модулів);
- вибір завдань для запланованих екзаменаційних форм;
- оперативний відбір у відповідності до Плану змісту екзамену;
- включення завдань для апробації

Стандарти:

3.7, 3.8

Крок №6 Виготовлення тесту (екзаменаційних матеріалів)

Задачі:

- видавнича підготовка;
- друк (для традиційного тестування) чи програмне оформлення (для комп'ютерного тестування);
- пов'язані з контролем якості питання валідності.

Крок №7 Адміністрування екзамену

Задачі:

- питання валідності, пов'язані зі стандартизацією умов проведення екзамену;
- питання, пов'язані з наданням умов для осіб з особливими потребами;
- робота адміністративного персоналу (proctoring, робота «уповноважених представників»);
- питання безпеки (як екзаменаційних матеріалів, так і умов для кандидатів)
- координація часу, хронометраж

Стандарти:

3.18, 3.19, 3.20, 3.21

Крок №8 Виставлення оцінок за відповідями кандидатів

Основна задача - питання валідності:

- контроль якості;
- валідація правильних відповідей;
- аналіз екзаменаційних завдань (предметний, технічний (IRT), процедури вирівнювання (equating), шкалювання)

Стандарти:

3.6, 3.22

Крок №9 Встановлення прохідних балів

Задачі:

- встановлення обґрунтованих прохідних балів;
- відносні критерії проти абсолютних;
- питання валідності відносно прохідного балу;
- співставність прохідних балів;
- підтримка сталості шкали балів (вирівнювання/пов'язаність/відповідність (equating/linking))

Стандарти:

4.10, 4.11, 4.19, 4.20, 4.21,

Крок №10 Повідомлення результатів екзаменів

Основна задача - питання валідності:

- точність (похибка);
- контроль якості;
- своєчасність;
- повнота, корисність і зрозумілість документа з результатами;
- питання неправильного розуміння і використання (інтерпретації) результатів;
- опротестування результатів;
- переєкзаменування, повторне складення іспиту.

Стандарти:

8.13, 11.6, 11.12, 11.15, 13.19, 15.10, 15.11

Крок №11 Ведення бази даних екзаменаційних завдань

Задачі:

- питання безпеки;
- корисність, гнучкість бази;
- принципи ефективності ведення бази даних екзаменаційних завдань.

Стандарти:

6.4

Крок №12 Технічний звіт по тестуванню

Задачі:

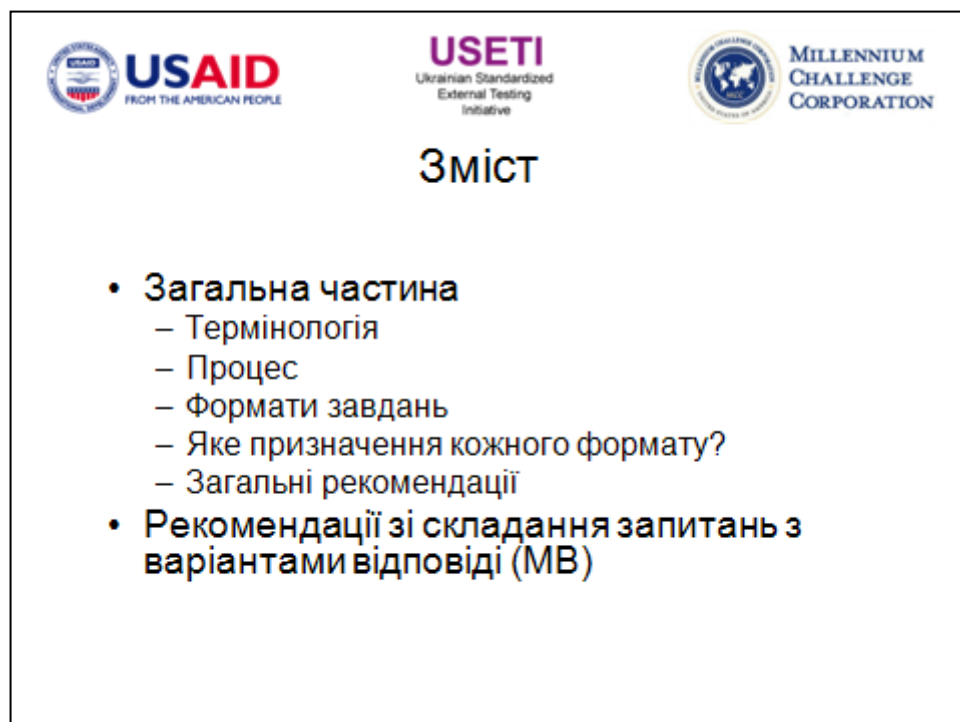
- систематичне, ретельне документування доказів валідності;
- архітектура документування – відповідає 12-ти крокам;
- документування висновків і рекомендацій щодо наступних тестувань і їх покращення.

Стандарти:

3.1, 6.5

V. МІНІ-КУРС З РОЗРОБКИ ЗАВДАНЬ ДЛЯ ТЕСТІВ.

Слайди презентації. Стівен Баккер



Термінологія

- Контекст; супровідний матеріал
- Завдання
- Питання-основа
- Варіанти відповіді
- Ключ
- Відволікаючі варіанти
- Повністю зараховано
- Частково зараховано

Епохи правління князів Володимира та Ярослава позначають "золотий вік Києва". Обидва правителі продовжили стабільне розширення держави, яке почалося ще за часів правління Віщого Олега.

Яким було прізвисько князя Ярослава?

- A. Завойовник
- B. Великий
- C. Грізний
- D. Мудрий

2. Як називалася країна, столицею якої був Київ, за часів правління князя Володимира (980-1015 рр. н.е.)?

Схема виставлення балів

2	Максимальний бал	2
Київська Русь	2
Русь	1

Процес створення завдання

1. Відбір контекстного матеріалу
 2. Складання завдання
 3. Перевірка та перегляд завдань автором (але нехай перед цим мине деякий час)
 4. Перевірка та рецензування завдання іншим експертом
 5. Апробація
 6. Рецензування третім рецензентом
- Не вагайтеся на будь-якій стадії виключати з тесту окремі завдання!

Формати завдань

- Запитання з варіантами відповіді (іноді також називаються “об’єктивні запитання”)
 - Множинний вибір
 - Встановлення відповідності
 - Визначення послідовності
 - Заповнення пропусків в тексті (клоуз-тест)
- Відкриті запитання
 - Коротка відповідь
 - Сконструйована відповідь
 - Твір/ есе /модельна справа

Встановлення відповідності

Тест для відвідувачів музею

В музеї експонувалися п’ять абстрактних картин відомого художника. Відвідувачам запропонували п’ять назв та попросили встановити, якій картині відповідає кожна з цих п’яти назв.

- Результати обробки заповнених бланків відповідей:
 - 6480 відвідувачів встановили 0 правильних відповідностей
 - 5750 відвідувачів встановили 1 правильну відповідність
 - 5020 відвідувачів встановили 2 правильних відповідностей
 - 4290 відвідувачів встановили 3 правильних відповідностей
 - 3560 відвідувачів встановили 4 правильних відповідностей
- ☐ Скільки відвідувачів встановили 5 правильних відповідностей?

То не вигадки!

Встановіть відповідність між країнами та їх столицями:

Країни

- 1) Йорданія
- 2) Нова Зеландія
- 3) Грузія
- 4) Уганда
- 5) Перу

Столиці

- A) Веллінгтон
- B) Тбілісі
- C) Кампала
- D) Амман
- E) Ліма

Правила виставлення балів

1 правильна відповідність	1 бал
2 правильних відповідностей	2 бали
3 правильних відповідностей	3 бали
4 правильних відповідностей	4 бали
Всі відповідності правильні	5 балів

Клоуз-тест

- Клоуз-тест являє собою завдання, що передбачає заповнення пропусків, коли учень, керуючись контекстом, здогадується, які навмисно вилучені з тексту слова треба повернути на місце.
- Клоуз-тест перевіряє рівень розуміння прочитаного матеріалу. Відповіді демонструють як рівень розуміння тексту, так і рівень володіння мовою.



USAID
FROM THE AMERICAN PEOPLE

USETI

Ukrainian Standardized
External Testing
Initiative



**MILLENNIUM
CHALLENGE
CORPORATION**

**Світовий
банк**

океану	рятувальних човнів	1912 року
затонув	холодний	крузу
годин	айсбергом	підводний човен
напотоплюваний	США	довжину
викили	половини	океанографи

"Титанік" був комфортабельним океанським лайнером, який затонув під час свого першого _____. Цей величезний корабель сягав майже 900 футів (269 метрів) в _____ та рекламувався як "_____."

В ніч на 14 квітня _____ "Титанік" зіштовхнувся з _____, коли він плів північною частиною Атлантичного _____. Корабель прямував з Великої Британії до _____. Через декілька _____ корабель _____, зібравши з собою в безодню близько 1500 життів. Більшість пасажирів потонули в _____ воді, тому що на кораблі було недостатньо _____, щоб урятувати всіх. Тільки 705 чоловік _____ в цій катастрофі.

У 1985 році, _____ знайшли залишки "Титаніка" на океанському дні, використовуючи гідрооператорську станцію та дистанційно керований _____. Гігантський корабель розколовся навпіл, коли тонував, та дві його _____ зараз лежать на відстані декількох сот футів одна від одної серед уламків, розкиданих між носом та кормою.



USAID
FROM THE AMERICAN PEOPLE

USETI

Ukrainian Standardized
External Testing
Initiative



**MILLENNIUM
CHALLENGE
CORPORATION**

Порівняння відкритих запитань та запитань з варіантами відповіді

	Запитання з варіантами відповіді	Відкриті запитання
Від кандидата вимагають	визначити правильну відповідь	дати свою відповідь
Охоплення матеріалу	обмежене	гнучке
Часу на відповідь	відносно мало	більше
Загалом питань в тесті	багато	менше



USAID
FROM THE AMERICAN PEOPLE

USETI
Ukrainian Standardized
External Testing
Initiative



**MILLENNIUM
CHALLENGE
CORPORATION**

Порівняння відкритих запитань та запитань з варіантами відповіді (продовження)

	Запитання з варіантами відповіді	Відкриті запитання
Що вони тестують?	знання, розуміння, застосування та інші розумові здібності вищого рівня	все вказане в клітинці зліва, а також вміння оригінально мислити та писати
Головний внесок автора завдання	складання запитання	розробка схеми виставлення балів та оцінювання тестів (залежно від кількості студентів)
Оцінювання завдання	легке, швидке, об'єктивне	потребує міркування, суб'єктивне



USAID
FROM THE AMERICAN PEOPLE

USETI
Ukrainian Standardized
External Testing
Initiative



**MILLENNIUM
CHALLENGE
CORPORATION**

Загальні рекомендації зі складання завдань

1. **Випишіть процес розв'язання проблеми, який, на вашу думку, відбувається, коли учень шукає правильну відповідь.**
 - Проаналізуйте цей процес, щоб переконатися що на шляху до правильної відповіді немає інших перепон, окрім тих, що ви запланували.
 - Подумайте, чи не містить завдання ненавмисних підказок, що вказують на правильну відповідь
 - Зробіть список логічних та інших вірогідних помилок, які може зробити учень
 - Використовуючи свої аналітичні здібності, розробіть схеми виставлення балів, відволікаючі варіанти відповідей та сам тест в ІНТЕРАКТИВНІЙ формі



USAID
FROM THE AMERICAN PEOPLE

USETI
Ukrainian Standardized
External Testing
Initiative



**MILLENNIUM
CHALLENGE
CORPORATION**

Загальні рекомендації зі складання завдань (продовження)

2. Використовуйте чіткі, лаконічні формулювання та просту мову та уникайте будь-якої сторонньої інформації



USAID
FROM THE AMERICAN PEOPLE

USETI
Ukrainian Standardized
External Testing
Initiative



**MILLENNIUM
CHALLENGE
CORPORATION**

Загальні рекомендації зі складання завдань: стороння інформація про мармури Парфенону

Подивіться на цей малюнок; це частина знаменитих мармурів Парфенону. Вони були переправлені з афінського Парфенону до Лондону лордом Елджином більш ніж століття



тому, що, вірогідно, врятувало їх від подальшого руйнування через забруднення. Мармур псується під дією кислотних дощів, проблема яких все більше турбує сучасне людство. Спричиняє кислотні дощі забруднення атмосфери оксидами азоту та сірки, які вступають в хімічну реакцію з водою, в результаті чого утворюються азотна та сірчана кислоти.

А запитанням екзамену було:

- Наведіть формулу азотної та сірчаної кислот



USAID
FROM THE AMERICAN PEOPLE

USETI
Ukrainian Standardized
External Testing
Initiative



**MILLENNIUM
CHALLENGE
CORPORATION**

Рекомендації зі складання запитань множинного вибору

- Зосередьтеся на одному питанні
- Намагайтеся сформулювати питання-основу як пряме запитання, а не як незавершене речення (відкрите питання-основа).
- Складіть доступні відволікаючі варіанти відповіді, які будуть прийнятними навіть для тих учнів, які недостатньо оволоділи матеріалом.
- Переконайтеся, що усі альтернативні варіанти відповіді узгоджуються з питанням-основою за граматикою та змістом.



USAID
FROM THE AMERICAN PEOPLE

USETI
Ukrainian Standardized
External Testing
Initiative



**MILLENNIUM
CHALLENGE
CORPORATION**

Рекомендації зі складання запитань множинного вибору

- Переконайтеся, що запитання не містить мимовільних підказок, які вказують на правильну відповідь.
- Уникайте використання комбінованих альтернатив або альтернативи "всі з наведених варіантів є вірними" або "жоден з варіантів не є вірним".
- Розташовуйте варіанти відповідей у довільному порядку, наприклад завжди використовуючи алфавітний або числовий порядок. Логічна послідовність може виявитися надмірною:
 - A. Обидва варіанти правильні
 - B. Тільки 1 варіант правильний
 - C. Тільки 2 варіант правильний
 - D. Обидва варіанти неправильні



USAID
FROM THE AMERICAN PEOPLE

USETI
Ukrainian Standardized
External Testing
Initiative



**MILLENNIUM
CHALLENGE
CORPORATION**

Відкрите питання-основа

Коли того дня Вадим прийшов зі школи додому, він був весь спітнілий та відчував, що йому жарко. Він не хотів їсти та рано ліг спати. Його батьки почали хвилюватися чи не викликати лікаря.

- ☐ Вадим був
- A. учнем
 - B. хворим
 - C. дома у своїх батьків
 - D. перевтомлений.



USAID
FROM THE AMERICAN PEOPLE

USETI
Ukrainian Standardized
External Testing
Initiative



**MILLENNIUM
CHALLENGE
CORPORATION**

Формулювання питання-основи як прямого запитання

- Тромбоцити в крові
 - A. є меншими, ніж еритроцити
 - B. викликають гемофілію
 - C. виробляють гемоглобін
 - D. переносять кисень
 - Краще
 - Тромбоцити в крові беруть участь у
 - A. виробництві гемоглобіну
 - B. виробництві тромбіну
 - C. перенесенні кисню до місця інфекції
 - D. розщепленні згустків крові
- ...але найкращий варіант формулювання це:
- В якому процесі в крові беруть участь тромбоцити?
 - A. У виробництві гемоглобіну
 - B. У виробництві тромбіну
 - C. ...тощо

Відволікаючі варіанти (дистрактори)

- Відволікаючі варіанти мають бути схожими на “ключ” за довжиною, структурою та стилем.
“Третій зайвий”
- Відволікаючі варіанти не повинні містити екстремальні або безапеляційні твердження (“все”, “завжди”, “ніколи”)
- Відволікаючі варіанти не повинні брати під сумнів або заперечувати припущення питання-основи (“пункт <питання-основа> не відповідає дійсності”)

“Третій зайвий”

У лікарнях простирадла завжди перуть при високій температурі і також прасують. Для цього є особлива причина:

- ☐ Яка це причина?
- А.Почистити їх
- В.Зробити їх білими
- С.Знищити всі бактерії, завдяки тому, що ці організми гинуть при високій температурі

“Третій не зайвий”

У лікарнях простирадла завжди перуть при високій температурі и також прасують. Для цього є особлива причина:

☐ Яка це причина?

- A. При високій температурі мило стає більш активним, та через це простирадла стають чистішими
- B. При високій температурі простирадла відбілюються та через це виглядають чистішими
- C. При високій температурі більшість бактерій гине

Інші питання: всі/жоден з вищенаведених варіантів

Уникайте фраз “всі з вищенаведених варіантів” та “жоден з вищенаведених варіантів”

тому що

- Зазвичай це не є правильною відповіддю.
- Якщо ви знаєте, що один з відволікаючих варіантів є неправильним, формулювання “всі з вищенаведених варіантів” автоматично відпадає.



USAID
FROM THE AMERICAN PEOPLE

USETI
Ukrainian Standardized
External Testing
Initiative



**MILLENNIUM
CHALLENGE
CORPORATION**

Інші питання: римські цифри

- Два пронумерованих твердження – яке з них є вірним, або чи є вірними обидва?
Прочитайте наступні два твердження:
I Амстердам – столиця Нідерландів
II В Амстердамі розташована резиденція голландського уряду
Яке з цих тверджень є вірним?
 - A. Обидва – I та II
 - B. Тільки I
 - C. Тільки II
 - D. Жодне
- Використовуйте римські цифри тільки з двома твердженнями, зв'язаними одним тестовим завданням
- Ніколи не використовуйте римські цифри, коли питаєте, який варіант є найкращим, найменшим, найбільшим тощо

VI. ВАЛІДНІСТЬ ТЕСТУ, КЛЮЧОВЕ ПИТАННЯ РОЗРОБКИ ТЕСТІВ. ЩО МОЖНА ВИЗНАЧИТИ ЗА ДОПОМОГОЮ ЦЬОГО ТЕСТУ? *Слайди Презентації. Стівен Баккер*

**USAID**
ВІД АМЕРИКАНСЬКОГО НАРОДУ

ПРОЕКТ
“СПРАВЕДЛИВЕ ПРАВОСУДДЯ”

**USAID**
ВІД АМЕРИКАНСЬКОГО НАРОДУ

ПРОЕКТ
“СПРАВЕДЛИВЕ ПРАВОСУДДЯ”

Валідність на практиці

- Перш, ніж запропонувати тести кандидатам, вони повинні бути валідовані
- Валідація тесту передбачає надання відповідей на такі запитання :
 1. Які рішення ми хочемо прийняти на підставі результатів цього тесту?
 2. Які докази нам потрібні для підтримання цих рішень?
 3. Яким чином і якою мірою цей тест надає такі докази?

3

Приклад: Вступ до університету Питання 1

1. Які рішення ми хочемо прийняти на підставі результатів цього тесту?
 - Ми хочемо вирішити, яких абітурієнтів прийняти на навчання, а яких – ні.
 - Це має бути здійснено на підставі їх здібностей: потрібно прийняти тих, які найімовірніше будуть успішними в навчанні, і відмовити тим, які, ймовірно, не будуть успішними.

4

Приклад: Вступ до університету

Питання 2

2. Як можна отримати докази того, хто володіє цими навичками, а хто - ні?
 - Опис предметного змісту і навичок, необхідних для успіху в навчанні.
 - Процедура та інструменти, які надійним чином поділяють абітурієнтів на дві групи: ту, яка володіє цими навичками, і ту, яка ними не володіє.
 - Встановлення зв'язку між результатами цієї процедури та описом необхідного змісту та навичок.

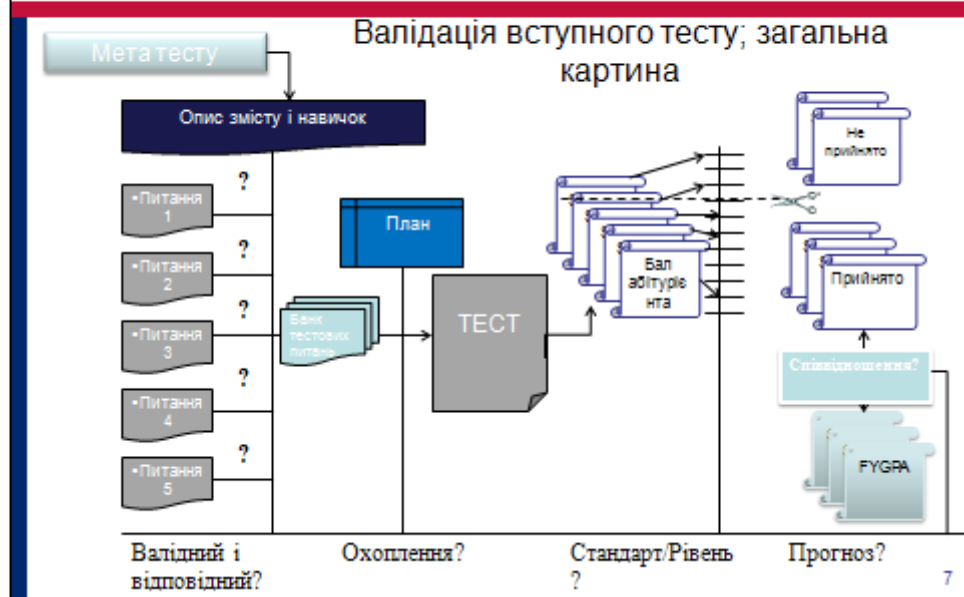
5

Приклад: Вступ до університету

Питання 3

- Як можна встановити зв'язок між результатами нашого тесту і рішенням щодо наявності/відсутності відповідних навичок?
 - Чи є надання правильної відповіді на те чи інше запитання доказом володіння будь-якою з відповідних навичок?
 - Чи відображає тест усі відповідні навички та їх відносну важливість?
 - На скільки питань – або на які питання – потрібно відповісти правильно, щоб це слугувало доказом наявності відповідних навичок?
 - Чи свідчить високий бал за цей тест про те, що абітурієнт буде успішним в університеті, а низький бал – що абітурієнт буде неуспішним?

6



Забезпечення валідності авторами тестів

- Організація, що розробляє тести, відповідає за
 - Розробку описів відповідного змісту та навичок
 - Розробку планів тестів, що відображають зазначене вище, а також відносну важливість тестів
 - Розробку процедури визначення прохідного бала (критеріїв проходження/не проходження)
 - Проведення дослідження прогнозової валідності

Забезпечення валідності авторами завдань

- Автори завдань відповідають за:
 1. Розуміння і ознайомлення з описом змісту екзамену і когнітивних навичок
 2. Надання інформації по кожному розробленому завданню того, що правильна відповідь на нього слугує доказом наявності знань змісту екзамену і когнітивних навичок, шляхом зазначення, які саме з них є важливими для надання правильної відповіді.

9

Опис змісту екзамену (знань)

- Зміст: перелік тем для вивчення, понять для розуміння, технік для застосування
- Приклад переліку тем із математики
 - Знання формули для обчислення площі круга
 - Здатність розв'язати квадратне рівняння
 - Розуміння поняття диференціювання

10

Опис когнітивних навичок

- Навички: застосування знань і розуміння змісту
 - У літературі з тестування наведено багато прикладів опису навичок
 - З одного боку, розрізняють відтворення знань і відтворення застосування, а з іншого боку – певну кількість навичок вищого порядку
 - Узагальнення часто використовуваних описів у навчальному тестуванні див. Інструкції для розробки тестових завдань, ст. 12-17

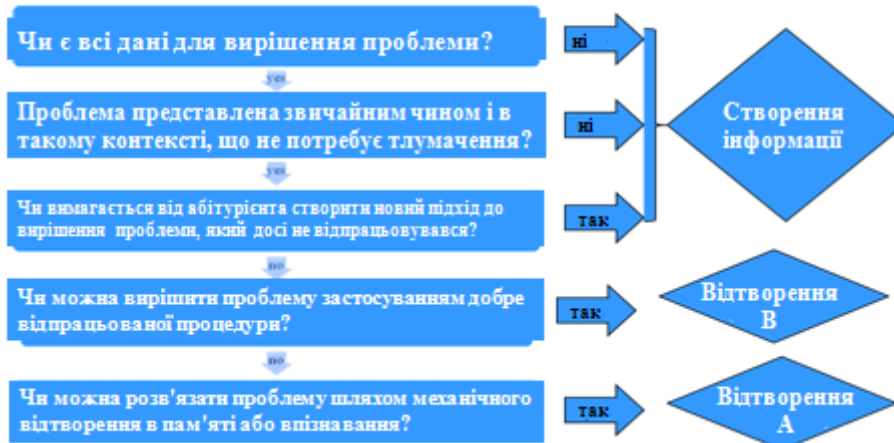
11

Класифікація питань

- Автори питань повинні поділяти проекти питань на відповідні категорії змісту і навичок
 - Зміст: зазвичай поділяється на 10-20 основних категорій
 - Навички: залежно від предмету. Приклади:
 - Навички з розв'язання проблем (відтворення/створення)
 - Функціональні навички (комунікація, співпраця, вироблення точки зору, прийняття рішень, рефлексія)
 - Предметні навички, наприклад, розуміння прочитаного: поверхнєве ознайомлення, сканування, інтенсивне читання, екстенсивне читання

12

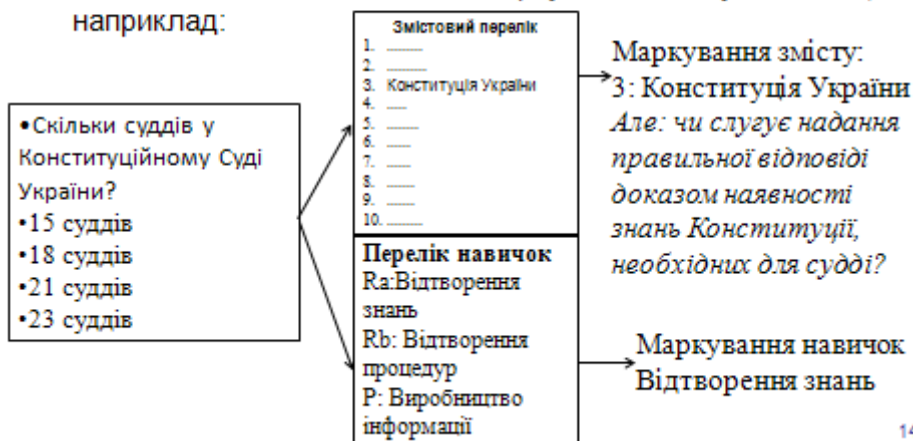
Схема класифікації навичок із розв'язання проблем



13

Марковані питання: складові валідації тесту

- Автори питань повинні розробити “марковані питання”
- Кожне питання повинно мати маркування змісту і навичок, наприклад:



14

Зміст і навички: виміри матриці тесту

Дії Предметний зміст↓	Відтворення А	Відтворення В	Виконання	
Живі організми				8
Метаболізм				6
Живлення	2	2	4	8
Дихання				6
Рух				4
Розмноження				4
Гормони				8
...				
Всього	10	30	60	100

15

План тесту

- План тесту необхідний для встановлення валідності тесту
- Матриці тесту є важливим компонентом Планів тесту. Вони скеровують процес складання тесту. Якщо склад тесту відповідає плану, тест може вважатися таким, що має “валідність змісту”.
- Іншими компонентами плану тесту є критерії складності та складання/не складання

16

Висновки

- Плани допомагають забезпечити валідність змісту тестів.
- Авторам питань необхідно розуміти виміри плану для того, щоб створити необхідні складові тесту.
- Авторам тестів необхідно розробити плани та навчити авторів запитань розуміти і застосовувати виміри плану.
- Авторам тестів також необхідно розробити процедури оцінювання складності питань і стандарти складання/не складання тесту (прохідний бал).
- Авторам тестів необхідно перевірити, чи мають тести, окрім валідності змісту, прогнозу валідність.

17

VII. АНАЛІЗ ТЕСТУ І ЗАПИТАНЬ: ЯК СТАТИСТИКА МОЖЕ РОЗШИРИТИ ВАШІ ПОГЛЯДИ. Слайди презентації. Стівен Баккер

**USAID**
ВІД АМЕРИКАНСЬКОГО НАРОДУ

ПРОЕКТ
“СПРАВЕДЛИВЕ ПРАВОСУДДЯ”

Аналіз тесту і запитань: Як статистика може розширити ваші погляди

Д-р Стівен Баккер
Семінар Проекту “Справедливе правосуддя”,
травень, 2012

1

**USAID**
ВІД АМЕРИКАНСЬКОГО НАРОДУ

ПРОЕКТ
“СПРАВЕДЛИВЕ ПРАВОСУДДЯ”

Вимірювання і надійність

- Тести – це інструменти вимірювання.
- Результати вимірювання – це певні показники вимірюваних величин, а не самі величини

Лінійка 1



Вимірювання	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
Лінійка 1	1	2	2	4	3	2	1	1
Лінійка 2	0	0	2	10	4	0	0	0

Лінійка 2



Вимірювання	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
Лінійка 1	1	2	2	4	3	2	1	1
Лінійка 2	0	0	2	10	4	0	0	0

Вимірювання	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8
Лінійка 1	1	2	2	4	3	2	1	1
Лінійка 2	0	0	2	10	4	0	0	0

2

Надійність і валідність

- Тести повинні бути надійними
 - Без надійності немає валідності
 - Як можна визначити надійність наших тестів?
 - Що можна зробити для підвищення надійності наших тестів?
- Відповідь треба шукати в статистичному аналізі відповідей кандидатів!

3

Надійність і бали осіб, що проходять тестування

Наскільки надійними є наші вимірювання?
 Чи свідчать “однакові бали” про
 “однаковий рівень підготовки”?

	Бал
Особа 1	5
Особа 2	5

4

4

Надійність і бали осіб, що проходять тестування

Відповідь: не обов'язково...!

Питання	1	2	3	4	5	6	7	8	9	10	Бал
Особа 1	+	+	+	+	+	-	-	-	-	-	5
Особа 2	-	-	-	-	-	+	+	+	+	+	5

- Наскільки надійним є цей тест?

5

5

Статистичні поняття: р-значення

Особа	1-A	2-D	3-A	4-C	5-B	6	Всього
1231001	1	1	1	1	1	3	6
1231002	1	0	0	1	1	1	4
1232003	1	0	0	1	0	2	4
1232004	0	1	0	0	0	0	1
Всього(i)	3	2	1	3	2	6	17
Макс.(i)	4	4	4	4	4	12	32
p	.75	.50	.25	.75	.50	.50	.53

6

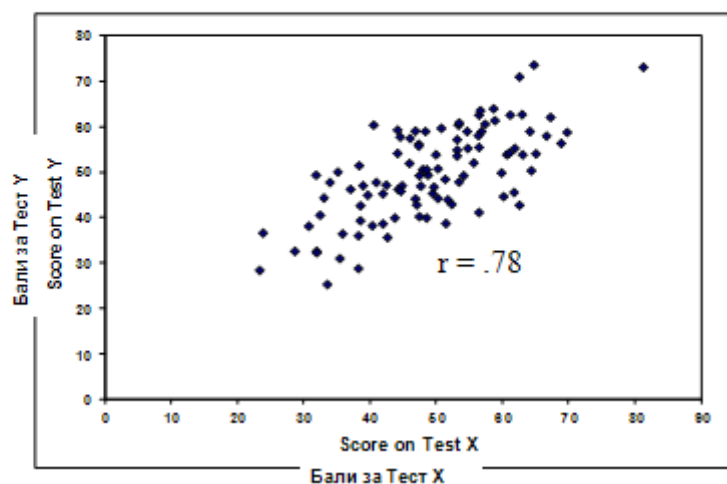
Статистичні поняття: кореляція

Кореляція:

- вимір сили лінійної залежності між двома наборами балів
- має значення від -1 до $+1$
- позитивна кореляція показує, що при зростанні однієї змінної, інша змінна також зростає

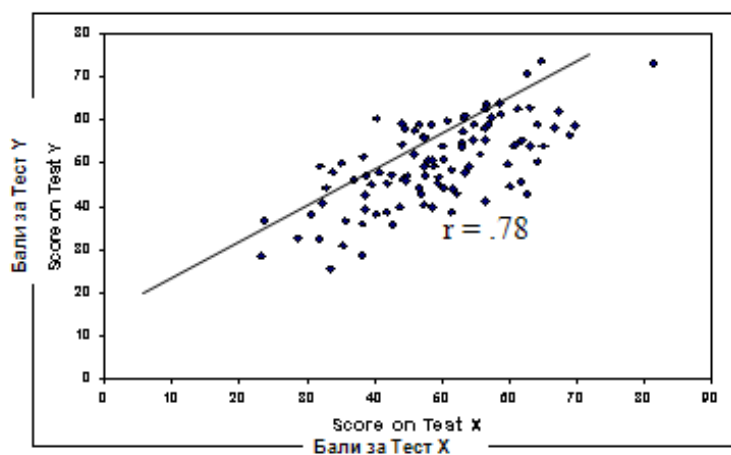
7

Лінійна залежність між балами двох тестів



8

Лінійна залежність між балами двох тестів



9

Кореляція між балами за тести і р-значеннями: ідеальне

	Питання 1 P=.90	Питання 3 P=.72	Питання 5 P=.64	Питання 6 P=.58	Питання 4 P=.42	Питання 3 P=.21
Особа 1 S=35	X					
Особа 2 S=45	X	X				
Особа 3 S=60	X	X	X			
Особа 4 S=75	X	X	X	X		
Особа 5 S=85	X	X	X	X	X	
Особа 6 S=100	X	X	X	X	X	X

10

Кореляція між балами за тести і р-значеннями: більш реалістичне

	Питання 1 P=.90	Питання 3 P=.72	Питання 5 P=.64	Питання 6 P=.58	Питання 4 P=.42	Питання 3 P=.21
Особа 1 S=35	X		X			
Особа 2 S=45	X		X	X		
Особа 3 S=60	X	X	X	X		
Особа 4 S=75	X	X			X	X
Особа 5 S=85	X	X	X	X	X	
Особа 6 S=90		X	X	X	X	X

Надійність тесту: в теорії і на практиці

- Теоретично, при повторному проведенні надійний тест повинен показувати одні й ті ж бали для одних і тих же осіб, що його проходять.
- Практично, це нездійсненно. Проходження тестів змінює осіб, що їх проходять.
- Оцінювання надійності тесту шляхом оцінювання його узгодженості: якою мірою тестові питання допомагають однаковим чином визначити рейтинг осіб, що складають тест.

12

Надійність і бали осіб, що проходять тестування

- Два тестові питання, за допомогою яких однаковий рейтинг осіб 1 і 2 визначається не однаковим чином:

Питання	1	2	3	4	5	6	7	8	9	10	Бал
Особа 1	+	+	+	+	+	-	-	-	-	-	5
Особа 2	-	-	-	-	-	+	+	+	+	+	5

- Якщо тест містить багато питань, що суперечать одне одному, тест вважається ненадійним.

13

13

Альфа Кронбаха: вимір надійності тесту

- Названий на честь Лі Кронбаха, який вперше використав цей коефіцієнт у 1951 р.
- Обчислюється шляхом зіставлення того, яким чином визначається рейтинг осіб, що проходять тест, за допомогою окремих питань
- Альфа Кронбаха допускає значення від 0.00 до 1.00
- Низьке значення Альфи Кронбаха свідчить про те, що тест не є “внутрішньо узгодженим”
- Значення .75-.85 є мінімально допустимими в тесті з вузьким предметом.

14

Похибка у балах кандидатів

- α^2 є виміром відсотку досліджуваного балу, що показує дійсні знання і навички кандидата.
- Наприклад: $\alpha = .80 \rightarrow \alpha^2 = .64 \rightarrow 64\%$ балу кандидата є гарантовано правдивим, а в 36% - міститься похибка
- Альфа (α) використовується для обчислення **стандартної похибки вимірювання (СПВ)**: сумнівності балів окремих кандидатів.
- СПВ надає довірчий інтервал для тлумачення і порівняння балів: з впевненістю у 95% можна сказати, що правдивий бал кандидата знаходиться межах досліджуваного балу \pm СПВ.

15

СПВ надає довірчий інтервал

- Наприклад, у тесті зі шкалою балів 0-100, СПВ дорівнює 4.47.
- Це означає, що для особи, яка набрала 55 балів, з упевненістю в 95% ми можемо сказати, що її правдивий бал знаходиться в межах від 50.53 до 59,47

16

СПВ і визначення прохідного балу

Score Freq Abs% Cum Freq Cum%
кумулятивна таблиця

46	20	3,91	211	41,29
47	15	2,94	226	44,23
48	11	2,15	237	46,38
49	17	3,33	254	49,71
50	9	1,76	263	51,47
51	17	3,33	280	54,79
52	16	3,13	296	57,93
53	19	3,72	315	61,64
54	14	2,74	329	64,38
55	17	3,33	346	67,71
56	16	3,13	362	70,84
57	14	2,74	376	73,58
58	14	2,74	390	76,32
59	11	2,15	401	78,47
60	15	2,94	416	81,41

Відсоток осіб, що
не склали
тестування

Прохідний бал,
якщо
досліджуваний бал
виявиться
правдивим

17

СПВ і визначення прохідного балу

Score Freq Abs% Cum Freq Cum%
кумулятивна таблиця

46	20	3,91	211	41,29
47	15	2,94	226	44,23
48	11	2,15	237	46,38
49	17	3,33	254	49,71
50	9	1,76	263	51,47
51	17	3,33	280	54,79
52	16	3,13	296	57,93
53	19	3,72	315	61,64
54	14	2,74	329	64,38
55	17	3,33	346	67,71
56	16	3,13	362	70,84
57	14	2,74	376	73,58
58	14	2,74	390	76,32
59	11	2,15	401	78,47
60	15	2,94	416	81,41

Відсоток осіб, що
не склали
тестування

Пропонований
прохідний бал

Довірчий
інтервал

Запланований
прохідний бал

18

Статистичний аналіз окремих питань (1)

Статистичний аналіз перших 4 питань тесту з 50

- колонка 3: ключ (key)
- колонка 4 - 8: кількість осіб, які вибрали варіанти A-D

Item nr.	Weight	Key	----- P- and A- val			
			A	B	C	D
1	1	B	19	78*	4	0
2	1	B	15	63*	7	11
3	1	C	7	7	70*	15
4	1	B	15	70*	4	7
-	-	-	---	-	--	-

Відсоток осіб, які вибрали правильну відповідь

Відсоток осіб, які вибрали неправильну відповідь

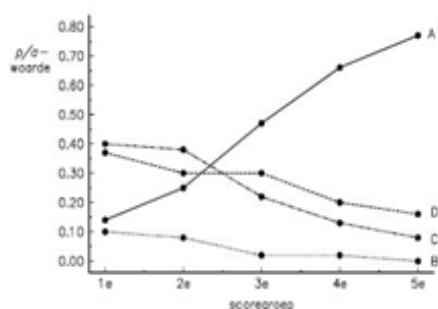
19

Кореляція питання-тест: загальна картина

Tabel 3.5

 Per scoregroep de p - en a -waarden van een item

score	n	A*	B	C	D
0 - 18	123	0.14	0.10	0.40	0.37
19 - 22	124	0.25	0.08	0.38	0.30
23 - 29	124	0.47	0.02	0.22	0.30
30 - 35	124	0.66	0.02	0.13	0.20
36 - 47	124	0.77	0.00	0.08	0.16
0 - 47	619	0.46*	0.04	0.24	0.26
gem. score	26.0	30.8	18.8	21.0	23.5



Візуалізація результатів

20

Кореляція питання-тест : Rir and Rar

Кореляція питання-тест перших 4 питань тесту з 40

- Питання №2 має хороші характеристики: ключову відповідь обрали сильні слухачі, а відволікаючі відповіді – слабкі
- Питання №3 проблематичне: ключ відображає слабкі кореляції, відволікаючу відповідь В обирають переважно сильні слухачі

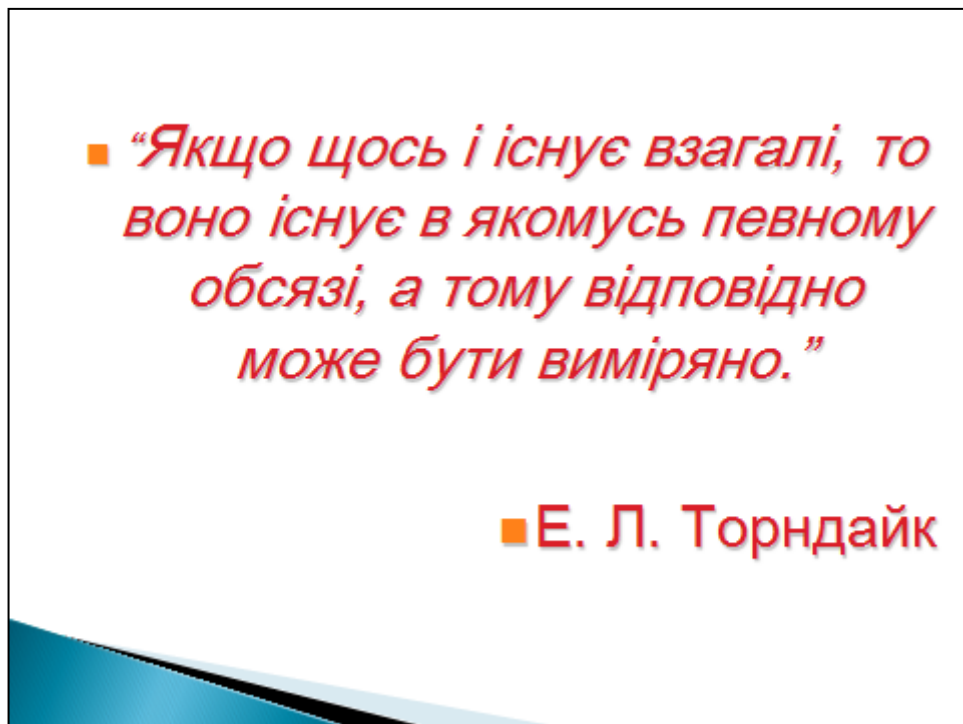
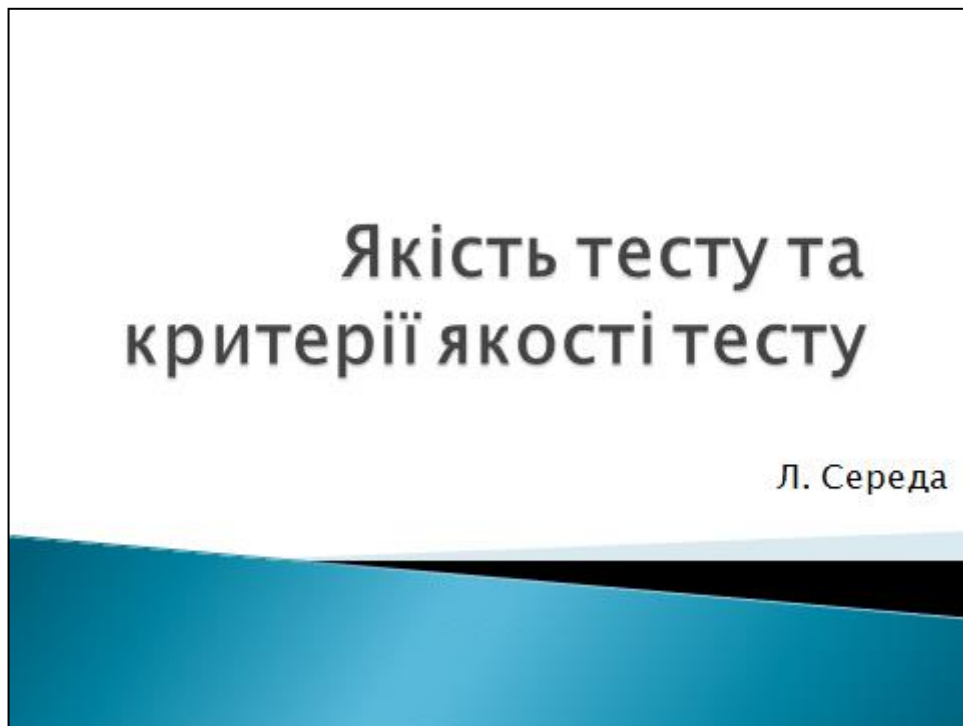
Вимір того, наскільки добре ключова відповідь допомагає відокремити сильних від слабких слухачів. Може набувати значень від -1 до +1, але повинен бути позитивним

Вимір того, наскільки добре відволікаюча відповідь допомагає відокремити слабких від сильних слухачів. Може набувати значень від -1.00 and +1.00 , але повинен бути негативним

Item nr.	Weight	----- Rir and Rar valu			
		A	B	C	D
1	1	-37	38*	-7	0
2	1	-23	52*	-27	-18
3	1	-11	28	13*	-29
4	1	-38	49*	2	-8

21

VIII. ЯКІСТЬ ТЕСТУ ТА КРИТЕРІЇ ЯКОСТІ ТЕСТУ.
Слайди презентації. Леонід Середа



Перш ніж говорити про якість тесту

- ▶ Що саме Ми хочемо дізнатись про кандидатів, організовуючи для них тестування?
- ▶ Навіщо ми хочемо про це знати?
- ▶ Як це пов'язано з реальним життям?
- ▶ Як Ми дізнаємось, що тестові бали відображають саме це?
- ▶ Кого ми включаємо в «МИ» відповідаючи на ці запитання?

3

Чому складно відповідати на такі запитання?

- ▶ Домени знань брудні – зміст, процес, зміст процесу
- ▶ Вимірювання того, що люди знають і вміють робити – у порівнянні з можливістю спрогнозувати їх досягнення або здатності долати перешкоди – дуже просто
- ▶ Те, що насправді важливо – дуже складно оцінювати
- ▶ Те, що легко оцінити – не дуже важливо

4

Тест, іспит, оцінювання тільки вимірювання

- ▶ Вимірювання сили струму, знань, здібностей не інструмент для боротьби з корупцією
 - ▶ Вимірювання – процес збору інформації
 - ▶ Найбільша проблема – інтерпретація результатів
- Інформація за результатами вимірювання може використовуватись лише в тому сенсі в якому проводилось вимірювання

5

Характеристики тесту

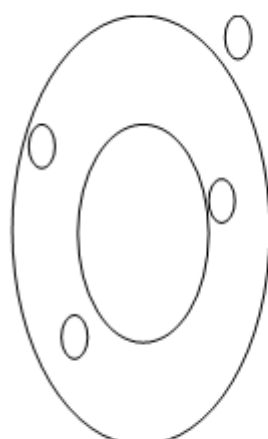
Надійність

Надійність показує якою мірою різниці індивідуальних тестових результатів можна вважати “істинними” і якою мірою такими, що виникли випадково

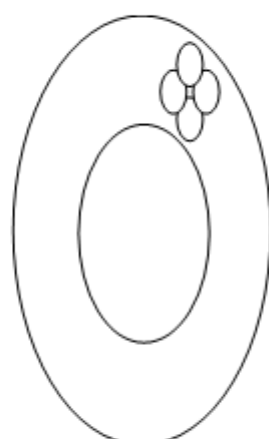
Валідність (від англ. valid – придатний) – *адекватність, придатність та повноцінність висновків, які зроблено на основі отриманих тестових балів.*

Валідність тесту стосується того, що тест вимірює і як добре він це робить.

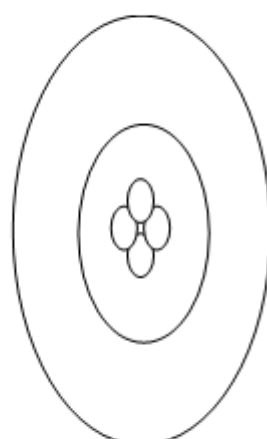
ЯКІСТЬ ТЕСТУ



Результати не надійні
та не валідні



Результати надійні,
але не валідні



Результати і надійні,
і валідні

Валідність: визначення

“Поняття валідності відноситься до відповідності, змістовності та користі певних висновків, зроблених за результатами тесту. **Валідація тесту – це процес накопичення доказів на підтвердження цих висновків**”.

Стандарти освітнього та психологічного тестування AERA/APA/NCME

8

8

ВАЛІДНІСТЬ

ЗМІСТОВА ВАЛІДНІСТЬ – визначає повноту охоплення навчальної програми тестовими завданнями та показує наскільки тестові завдання відповідають обсягові викладеного за час навчання матеріалу. Вона суб'єктивна, її визначають експертним шляхом.

КРИТЕРІАЛЬНА ВІНІСТЬ – ступінь відповідності між результатами тестування і зовнішнім критерієм.

ВАЛІДНІСТЬ ПРОГНОЗУ – Придатність тесту для прогнозування майбутньої поведінки (результатів)

Особливу увагу слід звернути на прогностичну валідність тесту

- ▶ Оскільки наші тестування проводяться з метою добору
- ▶ на навчання у Національній школі суддів
- ▶ на зайняття посад

Дуже важливо запланувати і провести масштабне дослідження прогностичної валідності тестів

-----Надати докази того, що кандидат, який отримав вищу оцінку на тестуванні, буде успішніше опановувати програму школи суддів

Можлива база для порівняння Результати оцінювання прогностичної валідності критеріїв добору (США, 2008р.)

Прогностичні інструменти	Коефіцієнт кореляції
1. HSGPA (середній бал атестата)	0.36
2. SAT-CR (критичне читання)	0.29
3. SAT-M (математика)	0.26
4. SAT-W (аналітичне письмо)	0.33
5. SAT-M, SAT-CR (математика + критичне читання)	0.32
6. HSGPA, SAT-M, SAT-CR (середній бал атестату + математика + критичне читання)	0.44
7. SAT-CR, SAT-M, SAT-W (критичне читання + математика + аналітичне письмо)	0.35
8. HSGPA, SAT-CR, SAT-M, SAT-W (середній бал атестату + середній бал за тестом)	0.46

Два типи шкалювання

- Критеріальне (встановлюється перед тестуванням і залежить від стандартів досягнень)



- Нормативне (встановлюється після тестування і залежить від загального розподілу популяції)



Прості статистики для оцінки якості тестування

- ▶ Міри центральної тенденції
- ▶ Міри варіювання
- ▶ Міри форми

Міри центральної тенденції

- ▶ Середній бал

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$$

- ▶ Структурні середні
- ▶ Медіана
- ▶ Мода
- ▶ Співпадіння по величині середнього, медіани і моди – свідчить про нормальність розподілу

Міри варіювання

Дисперсія
$$S_x^2 = \frac{\sum_1^n (X_i - \bar{X})^2}{n}$$

Стандартне відхилення
$$S = \sqrt{\frac{\sum_1^n (X_i - \bar{X})^2}{n}}$$

Для більш якісного тесту зазвичай характерна більша дисперсія балів

Міри форми

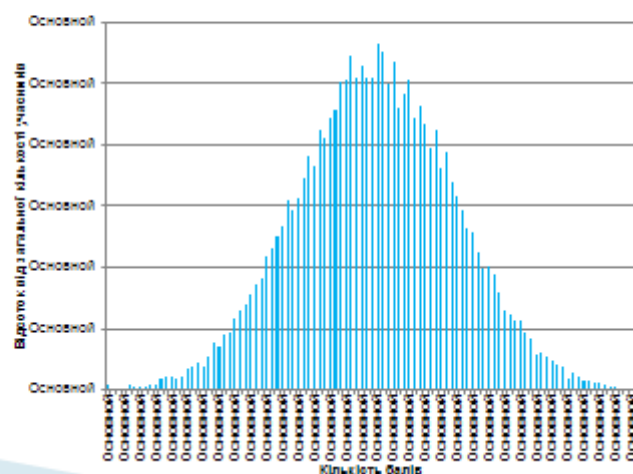
► Асиметрія
$$A_s = \frac{\frac{1}{n} \sum_1^n (X_i - \bar{X})^3}{S_x^3}$$

► Ексцес
$$Ex = \left[\left(\frac{\frac{1}{n} \sum_1^n (X_i - \bar{X})^4}{S_x^4} \right) - 3 \right]$$

Якщо ці показники відмінні від 0 слід шукати причини такої відмінності

Найпростіший спосіб оцінки якості тестування – побудова графіку розподілу тестових балів (візуалізація результатів)

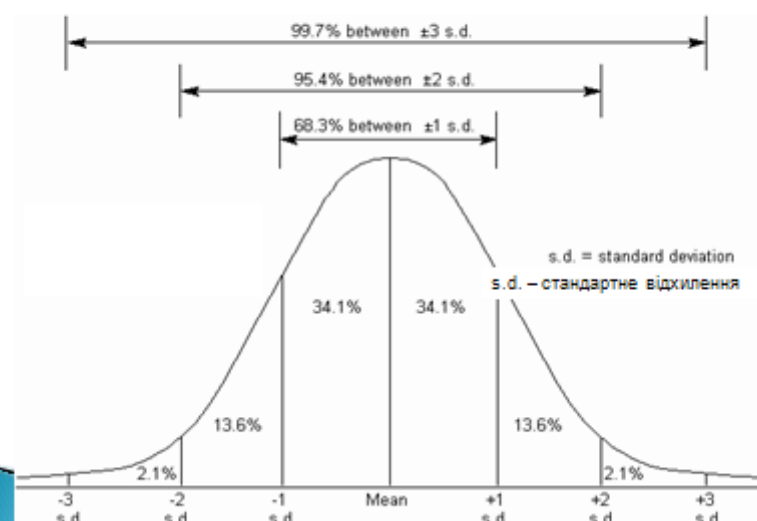
Нормальний розподіл результатів



17

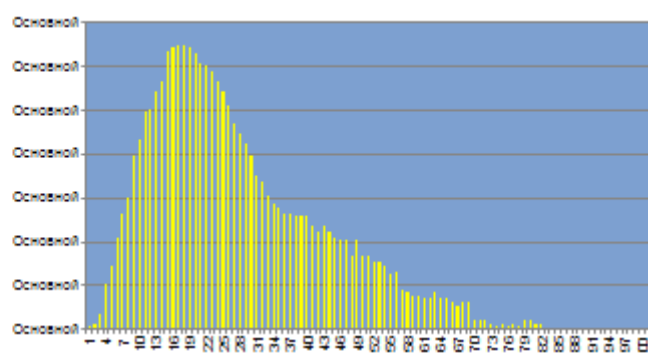
Статистичні поняття

нормальний розподіл

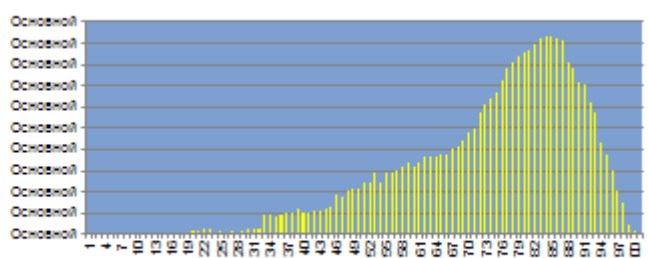


18

Занадто
складний тест



Занадто
легкий тест



Надійність тесту

Надійність показує якою мірою різниці індивідуальних тестових результатів можна вважати “істинними” і якою мірою такими, що виникли випадково

За класичною теорією надійності

$$T_a = X + E$$

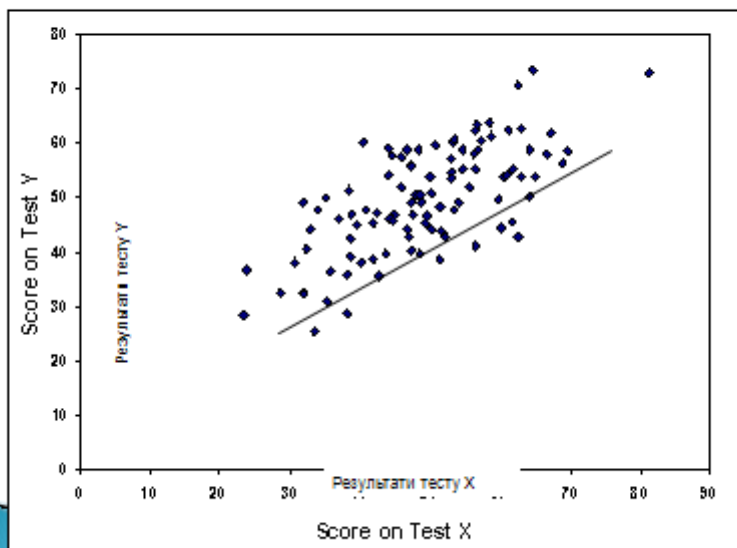
- ▶ де T_a – істинна оцінка учасника a ,
- ▶ X – результат отриманий при проведенні тесту, те саме, що кількість взятих завдань,
- ▶ E – компонент випадкової похибки.
- ▶ Концепцією класичної теорії тестів є намагання зменшити випадкову помилку
- ▶ Чим менша випадкова помилка, тим краще сирі бали відображають істинні бали

Статистичні поняття

- Кореляція
 - ✓ Вимірювання лінійної залежності двох сукупностей балів
 - ✓ Можливі значення від -1 до +1
 - ✓ Позитивна кореляція свідчить про те, що якщо одна зі змінних величин зростає, так само зростає й інша
 - ✓ Коефіцієнт детермінації

22

Лінійна залежність між результатами двох тестів



23

- ▶ Що це означає ?
- ▶ Очікуване значення (EV)
 - I. $EV(E) = 0$.
 - II. $Korr(E, T) = 0$

Більше того значення помилки не
корелює з жодною іншою змінною:

- III. $Korr(E_x, T_y) = 0$
- IV. $Korr(E_x, E_y) = 0$

Ці припущення називають аксіомами
класичної теорії тестів

Коефіцієнт надійності

$$rel = 1 - \frac{S_e^2}{S_x^2}$$

S_e^2 дисперсія помилки

S_x^2 дисперсія результатів

Максимальне значення коефіцієнта надійності = 1, отже слід мінімізувати

$$\frac{S_e^2}{S_x^2}$$

Зменшуючи значення S_e^2

Збільшуючи значення S_x^2

Надійність

- Помилки учасників обумовлені такими факторами, як умови тестування, втома, відповіді навімання (спроба вгадати відповідь), емоційний та фізичний стан, тощо
- Різні типи коефіцієнтів надійності відображають різні інтерпретації помилок

Типи коефіцієнтів надійності

1. Ретестова надійність
2. Еквівалентні форми
3. Внутрішня погодженість
 - Метод розщеплення
 - Коефіцієнти KR20, KR21
 - Альфа Кронбаха

27

СТТ: Внутрішня надійність (узгодженість)

- Коефіцієнт Альфа визначається як:

$$rel = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{s_x^2} \right)$$

- композиція дисперсій (якщо вибірки додавалися)
- дисперсія між i -ми та j -ми випробуваннями де i не дорівнює j
- k – кількість елементів

Внутрішня погодженість

- Вимірює наскільки усі тестові завдання вимірюють одне і те саме
- Потребує адміністрування одного тесту
- Тест розглядається як низка паралельних форм до тих самих завдань

29

Фактори, які впливають на надійність тесту

1. Тривалість тесту
 - Більш тривалі тести більш надійні
2. Розкид результатів
 - Надійність менша у однорідних групах

30

Фактори, які впливають на надійність тесту

3. Складність та наявність дискримінаційних завдань
- Відповіді на вгадування (спроби вгадати відповідь) зменшують надійність
 - Наслідком надлегких тестів є менша дисперсія
 - Більше дискримінаційних завдань – більша надійність

81

Фактори, які впливають на надійність тесту

4. Об'єктивність оцінювання
5. Швидкість
- швидкість виконання тесту підвищує надійність та внутрішню погодженість тесту

82

Стандартна похибка вимірювання

- ✓ Вказує на величину помилки, які можна очікувати в окремому результаті
- ✓ Залежить від надійності тесту та мінливості складу групи студентів

33

Стандартна похибка вимірювання (SEM)

$$SEM = s_x \sqrt{1 - \rho}$$

s_x = стандартне відхилення отриманих результатів

ρ = коефіцієнт надійності

34

Використання SEM

Можна побудувати довірчий інтервал, в якому з певною імовірністю знаходиться істинна оцінка учасника тестування

Наприклад:

Якщо учасник отримав 70 балів за тестом, що має коефіцієнт надійності 0,7 і дисперсія оцінок становить 15 то з імовірністю 95% можна стверджувати, що істинна оцінка є в межах
Від 68 до 72 балів

Стандартна похибка вимірювання

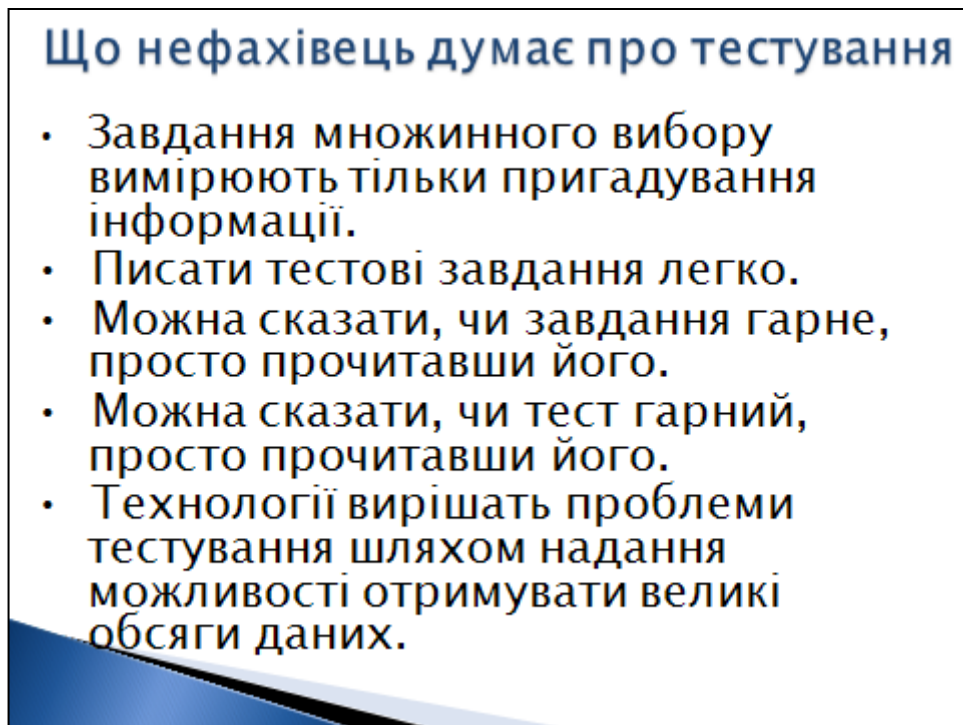
- | | |
|-----------------------------------|----------------|
| ▶ Величина коефіцієнта надійності | ▶ Оцінка |
| ◦ 0,90 – 0,99 | ◦ Відмінна |
| ◦ 0,80 – 0,89 | ◦ Хороша |
| ◦ 0,70 – 0,79 | ◦ Задовільна |
| ◦ менше 0,69 | ◦ Незадовільна |

Від хорошого до найкращого або що нового у галузі тестування

- ▶ Доказовий дизайн тесту
- ▶ Оцінки на основі мереж Баєса

88

ІХ. ПОКАЗНИКИ ЯКОСТІ ОКРЕМОГО ЗАВДАННЯ. *Слайди презентації. Леонід Середа*



Що потрібно знати про те, як людина розуміє світ

- ▶ Перше – Ми створюємо образи, або у термінах психолога Джерома Брунера (1990) – наративи.
- ▶ Друге – Ми створюємо ці образи незалежно від того, чи насправді розуміємо, що відбувається.

3

- ▶ Інтуїтивна фізика достатньо добре **працює**, коли граєш із собакою, кидаючи йому палицю, або будуєш шпаківню.
- ▶ Інтуїтивна фізика **не працює**, коли йдеться про будівництво мосту або запуск космічної ракети на місяць.

4

Створити справді якісний інструмент можна з завдань які пройшли апробацію і для яких встановлено психометричні характеристики

- ▶ Складність завдання
- ▶ Розподільна здатність завдання
- ▶ Крім того перевіряють чи конкретне завдання не диференціює учасників за ознакою, яка не належить до конструкту

5

Завдання з вибором однієї правильної відповіді

Інструкція – коротка, чітка вказівка того, як потрібно діяти особі під час робити над тестовими завданнями. Пишеться лише один раз на початку кожної конкретної форми завдань.

Порядковий номер завдання (арабські цифри).

Завдання з вибором однієї правильної відповіді

Зміст завдання – формулюється максимально точно, стисло, зрозуміло та бажано одним підрядним реченням. Зміст повинен читатися лише один раз перед виконанням.
Мета – мінімізація зусиль особи для його розуміння.

Завдання з вибором однієї правильної відповіді

Відповіді до завдання – це сукупність правильної та неправильних, але максимально правдоподібних, відповідей – дистракторів (від англійського **to distract** – відволікати).

Їхня кількість визначається за змістовим і формальним критеріями. Перший визначає максимальне число правдоподібних відповідей, другий – ймовірність вгадування правильної відповіді. Тому там, де це можливо, слід прагнути мати 4–6 варіантів відповідей.

Складність завдання

- ▶ Для дихотомічних завдань, складність завдання визначається як відношення кількості правильних відповідей усіх кандидатів на це завдання до загальної кількості кандидатів.
- ▶ Для політомічних – середній бал набраний учасниками за конкретне завдання.
- ▶ Наприклад відкрите завдання оцінюється
- ▶ 0 – 3, середнє 2.7 надто легке завдання.

Навіщо наперед знати складність завдань?

- ▶ Середня складність завдань хорошого тесту має складати 0.5 або 50%
- ▶ Разом з тим, оскільки маємо учасників з різною підготовкою, в тесті мають бути завдання з складностями від 0.05 to 0.95.
- ▶ Тести діляться на блоки – необхідно збалансувати їх за показником складності

Відмінності у Сучасній теорії тестів (IRT)

- ▶ 1) IRT зосереджується на виконанні кожного завдання а не лише усього тесту;
- ▶ 2) IRT описує результат виконання кожного завдання на кожному рівні підготовки екзаменованих
- ▶ 3) IRT ґрунтується на ряді взаємопов'язаних математичних моделей.

Основні теоретичні поняття

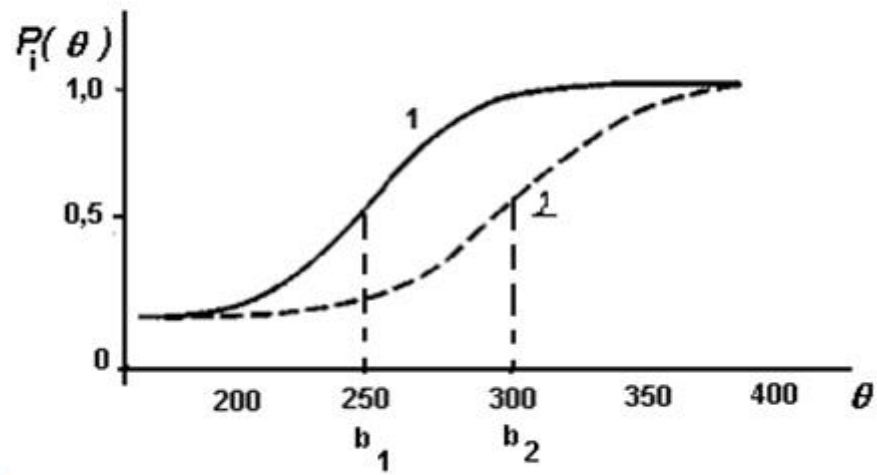
- ▶ Ймовірність P правильної відповіді респондента з рівнем досягнень θ на тестове завдання з параметрами a , b , c :

$$P(\theta) = \frac{1}{1 + e^{-(\theta - b)}}$$

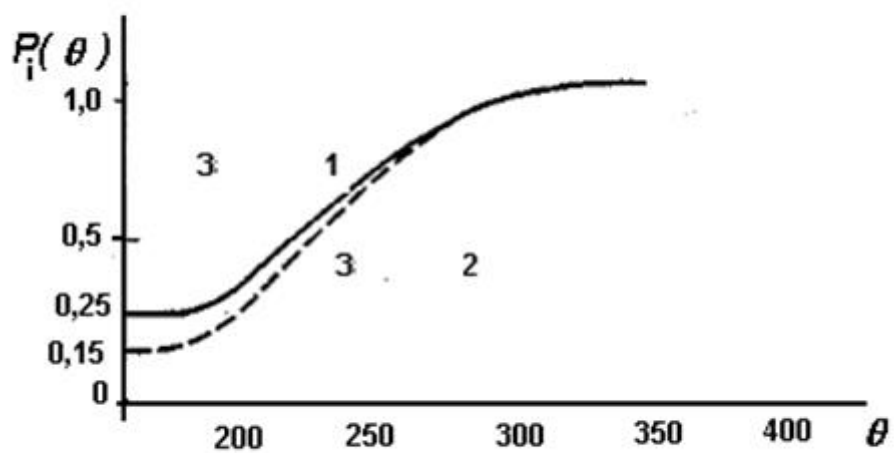
$$P(\theta) = \frac{1}{1 + e^{-a(\theta - b)}}$$

$$P(\theta) = \frac{1}{1 + e^{-a(\theta - b)}} \times (1 - c) + c$$

Складність завдання IRT



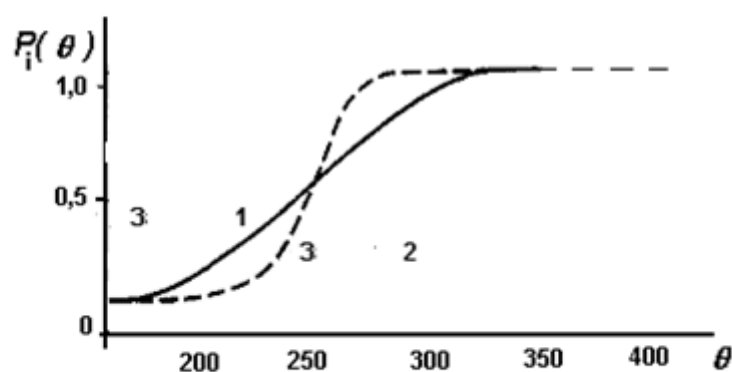
Вгадування c_i



Розподільна здатність (*Discrimination*)

- ▶ Чим вище розподільна здатність тим краще конкретне завдання відрізняє тих, що знають більше і навпаки
- ▶ Існує кілька способів визначення індексу розподільної здатності

Найбільш вживаним є бісеріальний коефіцієнт кореляції між тим, як учасники відповідають на конкретне завдання і їх загальним результатом



На практиці вважається:

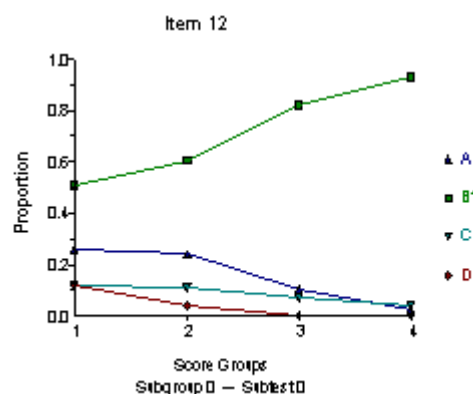
- ▶ Коефіцієнт дискримінаційної здатності має бути позитивним. (якщо негативний або близький до 0 – завдання працювало навпаки. Такі завдання не враховують у результатах тестування.
- ▶ Якщо коефіцієнт розподільної здатності $0,15 < r < 0,25$ – завдання вимагає доопрацювання
- Якщо $r > 0,25$ – хороше завдання

Аналіз дистракторів (відволікаючих варіантів)

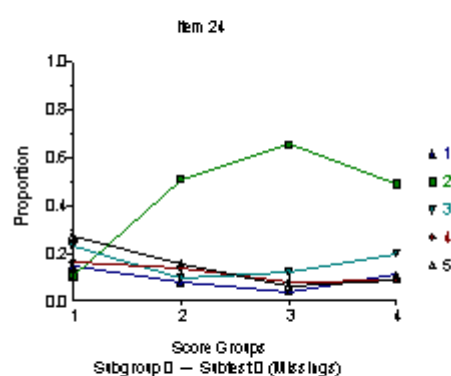
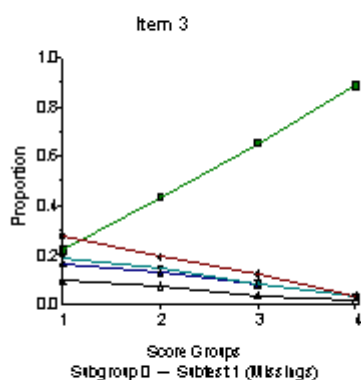
- ▶ Бажано щоб запропоновані дистрактори мали схожу складність (тобто відсоток учасників, що обирають ту чи іншу неправильну відповідь був схожим)
- ▶ Зрозуміло, що кількість учасників, що обирають неправильну відповідь – менша, ніж правильну
- ▶ Аналог коефіцієнта дискримінації для неправильних відповідей має бути негативним

Візуалізація у вигляді графіка завдання

- Визначте, до якої групи прав людини належить право на участь у всеукраїнському референдумі:
- A до особистих прав;
- B до політичних прав;
- C до соціально-економічних прав;
- D до культурних прав.

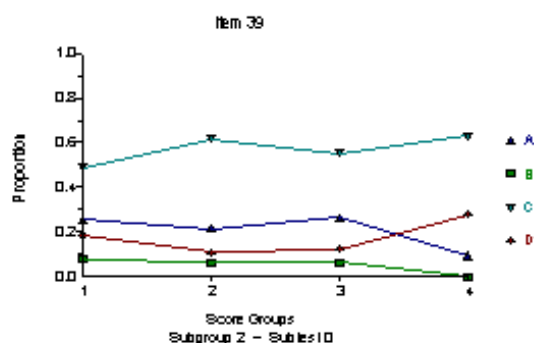


Більш наглядно. Картинка завдання



Графік завдання

- З метою заволодіння грошима, Ф запропонував узяти участь в нападі на касира фірми своєму другові Д. Ф чекав приятеля в машині. Д, погрожуючи ножем, відібрав у касира гроші та поїхав з місця події. Визначте, як треба кваліфікувати дії Д.
- A скоєно грабіж, Д є співвиконавцем;
- B скоєно розбій, Д є посібником;
- C скоєно грабіж, Д є виконавцем;
- D скоєно розбій, Д є виконавцем;



Як працює IRT

- Визначення рівня досягнень респондента за алгоритмом А. Birnbaum включає ітерації з біжучим номером s за рекуррентною формулою:

$$\theta_{s+1} = \theta_s + \frac{\sum_{i=1}^n a_i [u_i - P_i(\theta_s)]}{\sum_{i=1}^n a_i^2 P_i(\theta_s) [1 - P_i(\theta_s)]}$$

: Теорія характеризує знайдені шляхом ітерацій значення θ стандартною похибкою SE , обчисленою за формулою

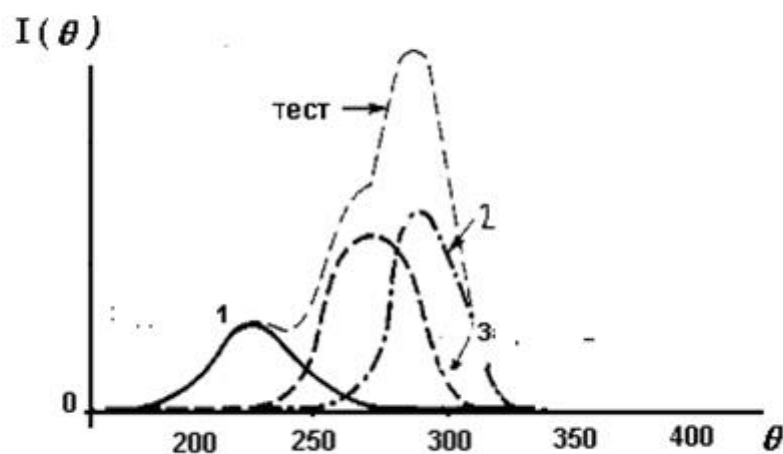
$$SE(\theta) = \frac{1}{\sqrt{\sum_{i=1}^n a_i^2 P(\theta)[1 - P(\theta)]}}$$

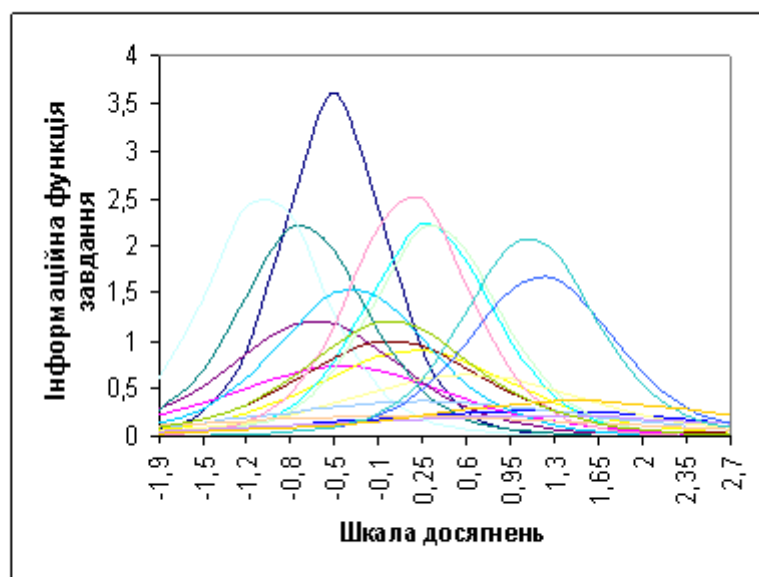
Якщо параметри запитань відомі, то кожне окреме завдання та весь тест можна охарактеризувати інформаційними функціями:

$$I_i(\theta) = a_i^2 P(\theta)[1 - P(\theta)]$$

$$I(\theta) = \sum_{i=1}^n I_i(\theta)$$

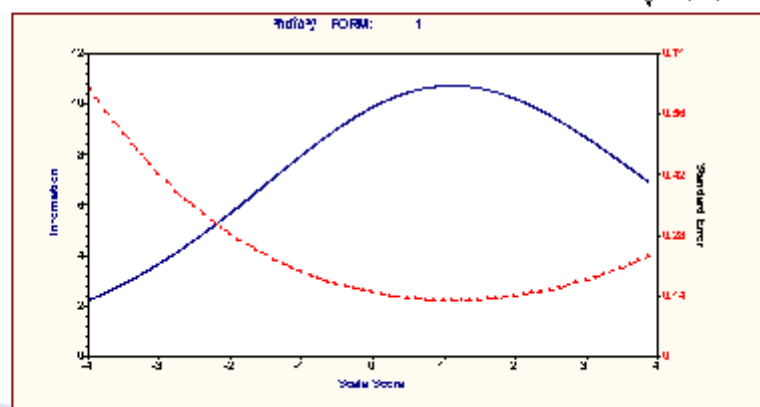
Інформаційна функція тесту:





Стандартна помилка вимірювання

standard error of measurement – SEM $SEM(\theta) = \frac{1}{\sqrt{I(\theta)}}$



Упередженість

- Завдання несе упередженість, якщо одна група має несправедливу перевагу над іншою
- «Професійні, освітні або психологічні тести є упередженими, якщо результати тестів у однаково здібних учасників систематично відрізняються.
- Це можуть бути расові, етнічні, культурні та професійні групи

Розрахунок

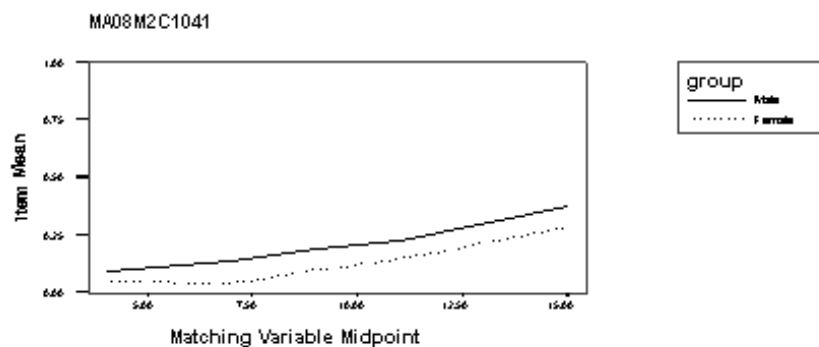
- **DIF– different item functioning /відмінне функціонування завдання**

DIF розраховують з 2 x 2 таблиць супряженості

$$\hat{\alpha}_{MH} = \frac{\sum_{k=1}^m A_k D_k / N_k}{\sum_{k=1}^m B_k C_k / N_k} \quad (2)$$

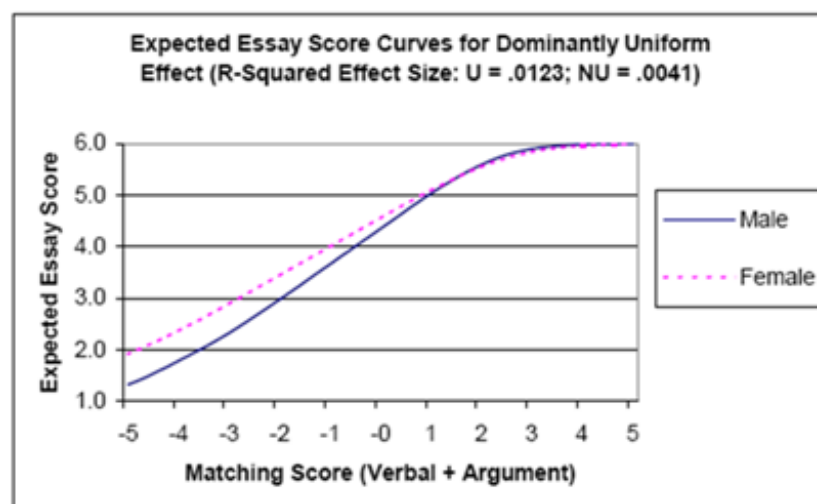
► Два типи DIF функції

1.



► Два типи DIF функції

2.



**Х. ПРАКТИЧНІ ПРАВИЛА ДЛЯ РОЗРОБНИКІВ
ТЕСТОВИХ ЗАВДАНЬ.** *Слайди презентації. Сергій Мудрук*

**Практичні правила для
розробників тестових
завдань**

Сергій Мудрук

Термінологія

- **Конструкт** (англ. Construct – концепція, узагальнений образ)
- **Таксономія екзамену** (План змісту іспиту)
- **Пункт таксономії**
- **Глибина по таксономії**
- **Таксономія Б.Блума**

Термінологія

- Контекст запитання (лат.: *contextus*, «переплетення, тісний зв'язок, з'єднання»)
- Контент запитання (англ.: *content* «зміст»)

Термінологія

- Стимул (лат.: *stimulus*)
- Основа (англ.: *stem* стовбур, стержень, лінія запитання)
- Матеріал (супровідний матеріал)
- Головне речення (англ.: *lead line*)

Термінологія

- **Ключ (англ.: key) -- правильна відповідь**
- **Дистрактор (англ.: distractor) – неправильна відповідь**

Термінологія

СТТ (classical test theory)

- **Складність**
- **Дискримінація**
- **Надійність**

Термінологія

IRT (Item Response Theory)

- Складність
- Дискримінація
- Вгадування
- Характеристична крива
- Інформаційна крива

Правила. Основа

Акцент - на суттєві елементи Конструкту

- Питання має бути направлене на суттєві елементи предмета екзаменування
- Неправильно: розробляти завдання по питаннях, які простіше формулювати

Правила. Основа

Лише один предмет завдання

Не створюйте завдання на більше ніж одну концепцію, ідею, проблему

Сформулювавши запитання впевніться, що:

- Для надання правильної відповіді достатньо знань, що ви запланували вимірювати
- Для надання правильної відповіді не допоможуть інші знання, ніж ті, що ви запланували вимірювати

Правила. Основа

Комплексні тестові запитання
- лише для дуже досвідчених
розробників, це «вищий
пілотаж» психометрії

Правила. Основа

Перевіряйте не пам'ять, а знання і компетенції

Не створюйте завдання на знання матеріалу лише з однієї групи підручників

Не створюйте завдання на перевірку засвоєння лише одного способу подачі матеріалу

Правила. Основа

Актуальність

- Не створюйте завдання, що не відповідають сучасним досягненням предмета тестування
- Не створюйте завдання, що використовують застарілу інформацію
- Питання повинні відповідати актуальному стану нормативної та регулятивної бази
- Забезпечуйте довговічність питання ще на етапі його розробки

Правила. Основа

Приклад використання застарілої інформації

Населення України
складає:

- А) 33 млн.
- Б) 52 млн.
- В) 61 млн

Правила. Основа

Для забезпечення довговічності
запитання не використовуйте в Основі,
Ключі і Дистракторах:

- ***дати***
- ***конкретні події***
- ***особистості***
- ***дані чи іншу інформацію, що дозволяє прив'язати питання до місця, того чи іншого часу, історичного періоду***

Правила. Основа

Релевантність

Питання треба ставити виключно в рамках Програми екзамену

Кожне створене питання має вимагати від кандидата:

- в обов'язковому порядку - знань/компетенцій в межах Програми
- опціонально - самих базових здібностей і знань (лише базових, щоб не перетворити предметний тест в загальний тест IQ)

Правила. Основа

Рівень розумових процесів

Розробник має чітко формулювати, який мінімальний рівень розумових процесів має бути задіяний кандидатом для визначення правильної відповіді. Орієнтир - екзаменаційна таксономія

Знання типу розумових процесів, на який розраховано питання, допомагає в процесі розробки, а, головне, необхідне на етапі перевірки робіт чи розгляду скарг (апеляцій) кандидатів

Правила. Основа

Гранична стислість

Питання має бути
сформульовано гранично
стисло

Правила. Основа

Без стороннього матеріалу

- Питання має бути звільнене від всякого стороннього (щодо предмету запитання) матеріалу
- Кандидат має вміти відділити інформацію, що має відношення до питання, від такої, що не має відношення ?
- Ви знаєте, що робити! (див. «Лише один предмет запитання»)

Правила. Основа

Без вступних речень і фраз

Звільніть запитання від вступних фраз, які мають мало зв'язку з головною думкою запитання

Правила. Основа

Проста структура речень

- Текст питання повинен мати гранично просту синтаксичну структуру.
- Рекомендується використовувати прості, не складнопідрядні речення без складних оборотів.

Правила. Основа

Граматична завершеність запитання

Не рекомендується використовувати незавершену форму Основи, в якій варіанти відповіді є продовженням речення Основи.

Правила. Основа

Повнота Основи

Вся важлива інформація повинна міститись в Основі

Чи достатньо кандидату прочитати Основу, щоб вже формулювати відповідь?

Чи можна повністю зрозуміти питання без того, щоб прочитати варіанти відповіді?

Правила. Основа

«Що з перерахованого нижче
є вірним відносно ХХХХ ?»

Це запитання - свідоцтво
повної творчої кризи у
розробника!

Правила. Основа

На практиці, неповна Основа
найчастіше дає такі недоліки:

- *варіанти відповідей
неоднорідні, гетерогенні;*
- *торкаються різнорідних за
свою класифікацією фактів;*
- *не можуть знаходитись «в
одному вимірі»*

Правила. Основа

Небажані слова

- Деякі слова несуть в собі оціночні якості.
- Не варто використовувати такі слова, що можуть по різному сприйматись кандидатами, залежно від того, в якому життєвому, емоційному, професійному, побутовому контексті знаходиться кандидат.
- Досить часто слова, що несуть в собі оціночні якості, є непрямыми підказками.

Правила. Основа

Приклади небажаних слів

Корисно	Завжди,	Невеликий,
для,	Ніколи,	Малий,
Важливо	Деякі,	Багато,
для,	Всі,	Мало,
Може,	Повсюдно,	Менше,
Можливо,	Зрідка,	Більше
Іноді,	Зазвичай,	і т.ін.
Часто,	Великий,	

Правила. Основа

«Небажані слова» іноді відображають суть запитання, тому бувають необхідними

приклад: менше/більше в якості ключових слів

Правила. Основа

Багато деталей - ймовірна підказка

Уникайте занадто деталізованих тверджень.

Ви маєте бути впевнені, що деталі не підштовхнуть кандидата до певної відповіді завдяки асоціативному ряду, чи лише логічним висновкам без знань по суті питання.

Не створюйте кандидату спокусу лише завдяки синтезу і резюмуванню Основи дійти до формулювання Ключа.

Але Основа має бути повною, без необхідності додумування ненаписаних речей, які мають на увазі.

Правила. Основа

Позитивне формулювання

НЕ рекомендується ставити запитання так, щоб вимагалось вибрати одну НЕправильну відповідь серед правильних.

Статистика рекламаций: багато забракованих запитань містять формулювання:

- Що з наступного вірно КРІМ ?
- Яка з наступних тверджень є НЕ вірним?

Правила. Основа

Без прямих цитат

Намагайтесь не включати в Основу фрагменти з підручників, нормативних документів, законодавства.

Правила. Основа

Термінологія

В нормо-орієнтовному тестуванні термінологія не повинна виходити за рамки офіційно затвердженої учбової літератури

В критерій-орієнтовному тестуванні має використовуватись лише найбільш використовувана на практиці (а не в академічному середовищі) термінологія.

Правила. Основа

Словесні асоціації підштовхують до здогадки

- Не використовуйте провокаційні запитання
- Не використовуйте запитання - пастки
- Не вимірюйте загальний IQ

Правила. Основа

Візуальний акцент на запереченні

Без подвійних заперечень

«чому не можна не допустити...»

Кожне питання - окреме

«Як відомо //твердження-
підказка// ...»

Правила. Відповіді

Максималізм

Варіанти відповіді мають бути
або абсолютно правильними,
або безперечно невірними.

Напівтони - не вітаються

Правила. Відповіді

Тільки одна правильна відповідь

Якщо через недогляд при розробці чи рецензуванні один з Дистракторів виявляється правильним, це збиває подальший ритм роботи кандидата на екзамені

Правила. Відповіді

Стислість

Відповіді мають бути короткими, не містити довгих пояснень, які можна замінити на лаконічні вирази

Правила. Відповіді

Співпадіння слів в Ключі і в Основі

Характерне слово не повинно повторюватись і в тексті Основи і в тексті Ключа.

Повторення характерного слова є однією найрозповсюдженіших «підказок» від недосвідченого розробника, спричинених непередуманою конструкцією питання

Правила. Відповіді

«Конгруентність» відповідей

Всі відповіді мають бути узгоджені, відповідні, співмірні одна з одною за зовнішніми та внутрішніми структурними характеристиками.

- Подібність граматичної структури
- Аналогічність по довжині
- Схожість по дизайну (верстці)

Типова помилка: одна з відповідей помітно довша чи коротша за інші

Правила. Відповіді

Без багатокomпонентності

Суть відповідей (і Ключа і Дистракторів) має формулюватись з участю одного-двох важливих слів/понять, що є ключовими для сенсу відповіді.

Таких важливих (для суті варіанта відповіді) понять не повинно бути багато.

Краще, якщо кількість ключових понять в усіх варіантах відповіді буде однаковою, щоб жодна з відповідей не виділялась

Правила. Відповіді

Граматичне узгодження

Всі варіанти відповідей мають бути граматично узгоджені з Основою.

У недосвідчених розробників частіше всього спостерігається неспівпадіння родів і відмінків.

Неспівпадіння числа (однина/множина)

Приклад: називається/називаються

Правила. Відповіді

Виносимо повторення з варіантів в Основу

Якщо в варіантах відповіді повторюється група слів, то треба переписати Основу, перенести повторювану групу слів в неї, і переписати варіанти відповідей.

Правила. Відповіді

Заборонені вирази

- Ні один з перерахованих
- Все вищеперераховане
- Все, окрім
- Всі вище/нижче перераховані
- Жоден з вище/нижче перерахованих

Правила. Відповіді

Впорядкування послідовності відповідей

Якщо між варіантами відповіді можна встановити відносини упорядкованості (більше/менше, алфавітний порядок, раніше/пізніше, і т.ін.) то в завданні їх треба розміщувати упорядковано (в порядку зростання, в алфавітному порядку і т.ін.)

Правила. Ключ

Абсолютна правильність

- Правильна відповідь має бути абсолютно правильною без виключень.
- Ключ не повинен ставати частково правильним чи неправильним в особливих умовах, які можливо допускає Основа

Правила. Ключ

Абсолютна правильність

приклад:

Путь до серця чоловіка лежить через:

- а. аорту
- б. легеневі артерії
- в. легеневі судини
- г. шлунок

Правила. Дистрактори

Дистрактор - породження типових помилок

- Використовуйте лише правдоподібні помилкові відповіді, взяті з професійного досвіду (чи викладацького досвіду).
- Автор повинен бути здатен дати відповідь на запитання, яку з поширених помилок виражає кожен з дистракторів

Правила. Дистрактори

Дистрактор - розумне твердження

- Дистрактор повинен виглядати як розумне, позбавлене очевидних недоліків твердження.
- Здаватись правдоподібним, реалістичним, цілком переконливим, привабливим для кандидата з нетвердими знаннями

Правила. Дистрактори

Без неточностей

- Дистрактор не повинен містити явних неточностей.
- Неточність - підказка для уважного кандидата

Правила. Дистрактори

Без сторонніх понять

- Дистрактор не повинен містити понять, що не мають відношення до предмету тестування
- Стороннє поняття - підказка для уважного кандидата

Правила. Дистрактори

Незнайомий термін - вказівка на неправильність відповідь

Навіть слабкі кандидати здатні ідентифікувати термін, як необов'язковий для опанування в рамках програми предмета.

Правила. Дистрактори

... від третьої особи

І в Основі і в варіантах відповіді слід використовувати безособові формулювання, чи формулювання від третьої особи.

Правила. Дистрактори

Якщо у вашій присутності відбувається епілептичний напад у людини, яку першу допомогу ви маєте надати?

- А. Ви маєте покласти людину на спину.
- Б. Ви маєте покласти пацієнта на бік.
- В. Ви маєте дати людині щось попити.
- Г. Ви маєте вкласти людині щось між щелеп

Правила. Дистрактори

Яка перша допомога має бути надана хворому під час епілептичного нападу?

- А. хворого треба покласти на спину.
- Б. хворого треба повернути на бік.
- В. хворому треба дати випити води.
- Г. Вкласти хворому щось між щелеп.

Правила. Дистрактори

Використання курсиву і жирного накреслення

Курсив - написання іноземних слів, термінів, назв творів, періодичних видань, суден і т.ін.

Жирне накреслення - звичайні слова, на які робиться акцент (ключові слова).

«не», «менше», «ефективно/менш ефективно»

Але не завжди ключові слова слід виділяти

Правила. Дистрактори

Що важить менше:

А. унція

Б. фунт

В. кілограм

Г. золотник

Правила. Дистрактори

Використання взаємозамінних слів

«який / котрий»

Використовуйте одноманітно!

Приклад:

Який - коли відповідь абсолютна

Котрий - коли відповідь з обмеженою
кількістю варіантів

Правила. Дистрактори

Яка одиниця служить для вимірювання довжини?

- А. верста
- Б. міра
- В. берковець
- Г. вузол

Котра з фігур має більшу площу?

Недоліки тестів, пов'язані з досвідом кандидата в екзаменуванні

Граматичні підказки

- Один чи більше дистракторів граматично не відповідають Основі
- Неузгодження родів/відмінків/числа
- Різна довжина

Недоліки тестів, пов'язані з досвідом кандидата в екзаменуванні

60-річний алкоголік доставлений міліцією до медичного закладу, оскільки був виявлений на центральній вулиці міста в стані буйного божевілля з закривавленою головою, раптовим паралічем нижніх кінцівок та безліччю кровоточивих ран. Першим кроком для пацієнта буде внутрішньовенне введення:

- а. кровозупиняючого засобу
- б. дослідження спинномозкової рідини
- в. заспокійливого засобу
- г. комп'ютерна томографія голови та хребта

Недоліки тестів, пов'язані з досвідом кандидата в екзаменуванні

Логічні підказки

- Підказка повноти
- Підказка “загального виміру”

“підвищується” / “знижується” /
“залишається на попередньому рівні”

Недоліки тестів, пов'язані з досвідом кандидата в екзаменуванні

Злочинність

- а. однаково поширена серед усіх соціальних верств населення*
- б. найбільш широко представлена серед бідного населення*
- в. найбільш широко представлена серед людей середнього достатку і багатих*
- г. головним чином вказує на психосексуальну дезадаптацію*
- д. досягає рівня терпимості суспільства*

Недоліки тестів, пов'язані з досвідом кандидата в екзаменуванні

- Абсолютність тверджень
“завжди”, “ніколи”
- Довгий ключ
- Повторення слів

**Недоліки тестів, пов'язані з
досвідом кандидата в екзаменуванні**

60-річний чоловік, який зловживає алкоголем і раніше був госпіталізований до психіатричної лікарні, розгублений і збуджений. Він сприймає оточуючий світ як нереальний. Цей симптом називається

- а. дереалізація
- б. деперсоналізація
- в. дезорієнтація
- г. часткова втрата пам'яті

XI. ТИПОВА КАРТКА ТЕСТОВОГО ЗАВДАННЯ

Типова картка тестового завдання

Автор: _____

Дата надання розробки: « ____ / ____ / ____ »

Рецензент №1: _____

Дата надання рецензії №1: « ____ / ____ / ____ »

Рецензент №2: _____

Дата надання рецензії №2: « ____ / ____ / ____ »

Яка психометрична характеристика вимірюється запитанням?

☐ Знання ☐ Розумове вміння / навик ☐ Розумова здатність / здібність

Який тип контенту вимірюється?

☐ Факт ☐ Концепція ☐ Принцип ☐ Процедура

Запитання відповідає такому розділу таксономії: _____

Який тип ментальної поведінки кандидата вимірюється?

(нижче наведено приклади трьох альтернативних шкал ментальної поведінки. В залежності від обраної моделі таксономії, обирається лише одна зі шкал)

☐ Згадування ☐ Розуміння ☐ Критичне мишлення ☐ Вирішення проблеми

☐ Згадування ☐ Розуміння ☐ Застосування ☐ Аналізування ☐ Створення ☐ Оцінювання

☐ Інтерпретація ☐ Порівняння ☐ Класифікація ☐ Узагальнення ☐ Логічний висновок

☐ Аналізування ☐ Синтезування ☐ Побудова гіпотези ☐ Передбачення ☐ Оцінка

===== ЗАПИТАННЯ =====

Супровідний матеріал (контекст): _____

Основа: _____

Основне (питальне) речення Основи: _____

===== ВІДПОВІДІ =====

Ключ: _____

Дистрактор №1: _____

Пояснення типової помилки закладеної в Дистрактор №1 _____

Дистрактор №2: _____

Пояснення типової помилки закладеної в Дистрактор №3 _____

Дистрактор №3: _____

Пояснення типової помилки закладеної в Дистрактор №3 _____

ХІІ. СПРОЩЕНИЙ КОНТРОЛЬНИЙ СПИСОК ПО ТЕСТОВИХ ЗАВДАННЯХ. *Сергій Мудрук*

В цьому розділі наведено стислий перелік правил (подекуди з коротким супровідним текстом) яким мають відповідати якісні тестові завдання.

Правила об'єднано в п'ять груп:

- Контент
- Формат
- Стиль
- Варіанти відповідей
- Неупередженість і гумор

Цей скорочений перелік необхідно використовувати розробникам тестових завдань, для самоконтролю якості своєї роботи.

Експерти, що професійно виконують обов'язки контролю якості, можуть взяти цей контрольний список за основу, але повинні мати (створити) більш детальний і специфікований під конкретне тестування контрольний список.

Контент

1. ЛИШЕ ОДИН ЗМІСТОВНИЙ ЕЛЕМЕНТ.

Кожне запитання повинно відображати лише один змістовний елемент програми екзамену і лише один конкретний вид ментальної поведінки (розумових процесів) у відповідності до технічного завдання (екзаменаційної таксономії, плану екзамену, матриці тесту).

В дійсності, які саме знання, уміння компетенції вимірює тестове завдання?

Чи ці знання, уміння, компетенції можуть бути чітко ідентифіковано в екзаменаційній таксономії (технічному завданні на розробку екзамену)?

2. ЛИШЕ ВАЖЛИВИЙ ЗМІСТ.

Кожне запитання повинно торкатись важливого, не другорядного, змісту. Слід уникати тривіального змісту.

3. ЛИШЕ ОРИГІНАЛЬНИЙ ТЕКСТ.

Має створюватись оригінальний текст для питання. Не повинні використовуватись фрагменти тексту з підручників, нормативних документів, щоб уникнути простого тестування здатності згадати знайомі фрази чи текстові конструкції.

4. ЗМІСТИ ЗАПИТАНЬ НЕ ПОВ'ЯЗАНІ ОДИН З ОДНИМ.

Зміст кожного запитання має бути окремим один від одного і не перегукуватись один з одним.

5. БЕЗ ЗАНАДТО КОНКРЕТНОГО І ЗАНАДТО ЗАГАЛЬНОГО.

Необхідно уникати як занадто конкретного так і занадто узагальненого контенту.

6. БЕЗ ОЦІНОЧНИХ СУДЖЕНЬ.

Слід уникати використання питань, правильна відповідь на які базується на переконаннях, поглядах, оціночних категоріях.

7. БЕЗ ХИТРОЩІВ.

Слід уникати нечистих прийомів (трюків, вивертів, пасток, обманних ходів) по відношенню до кандидатів.

Не намагайтесь підступно вчиняти, використовуючи другорядні та неважливі (не ключові) деталі, щоб зробити варіант відповіді невірним.

Не намагайтесь обдурити кандидатів, щоб вони обрали невірну відповідь.

Не пробуйте тестувати на невеликі розбіжності, окрім випадку, якщо такі невеликі розбіжності є дуже важливими.

8. ДОСТУПНА МОВА І СЛОВНИКОВИЙ ЗАПАС.

Словниковий запас і мова завдання повинні бути достатньо простими для кандидатів, на яких розраховано іспит.

Чи рівень лексики і словникового складу завдання дійсно відповідає аудиторії, якій це завдання буде запропоновано?

Формат

9. УНИКАЙТЕ КОМПЛЕКСНОГО MCQ ФОРМАТУ.

10. ВЕРТИКАЛЬНИЙ ФОРМАТ.

Форматуйте запитання і варіанти відповідей вертикально, не горизонтально.

Стиль

11. РЕДАГУВАННЯ.

Запитання повинні проходити наукове редагування, коректуру і вичитку.

12. ПРАВИЛЬНА ПИСЬМОВА МОВА.

Коректна граматики, пунктуація, орфографія, використання заголовних літер.

13. МІНІМУМ ЧИТАННЯ ДЛЯ КАНДИДАТА.

Кількість тексту в кожному запитанні повинна бути мінімальною.

Тестове завдання коротке? Його можна скоротити без шкоди для суті вимірювання?

Тестове завдання лаконічне? Чи є в ньому слова чи речення, що можуть бути видалені без шкоди для основного, важливого для екзаменування сенсу завдання?

14. ЯСНІСТЬ ВКАЗІВОК.

Необхідно впевнитись, що вказівки (в т.ч. інформація) в Основі абсолютно ясні.

Тестове завдання виражене точно і однозначно зрозуміле?

15. ІДЕЯ – В ОСНОВІ, НЕ У ВІДПОВІДЯХ.

Центральна ідея запитання повинна міститись в Основі, а не у варіантах відповіді.

Чи завершеною і достатньою є Основа?

Зокрема:

- чи Основа містить завершене запитання?
- чи можна отримати правильну відповідь з представленої лише в Основі інформації?
- чи достатньо в Основі інформації, щоб добре підготовлений кандидат був здатен самостійно сформулювати очікувану правильну відповідь на поставлене запитання, без того, щоб прочитати запропоновані варіанти відповідей?

16. БЕЗ БАГАТОСЛІВНОСТІ.

Повинні бути виключені зайві текстові прикраси (показовий характер тексту) і надмірна багатослівність.

17. ПОЗИТИВНА ОСНОВА.

Основу слід формулювати позитивно: уникати заперечень («не» чи «виключаючи»). Якщо ж слова заперечень використовуються, це слід робити обережно і завжди давати такі слова жирним накресленням чи великими літерами.

Варіанти відповідей

18. МАКСИМУМ ЕФФЕКТИВНИХ АЛЬТЕРНАТИВ.

Повинно бути напрацьовано стільки ефективних альтернативних відповідей, скільки можливо. Але дослідження вказують на адекватність трьох якісних альтернативних відповідей.

19. ТІЛЬКИ ОДИН ПРАВИЛЬНИЙ.

Необхідно впевнитись, що лише один з запропонованих варіантів відповіді є повністю вірним.

20. РОЗМІЩЕННЯ КЛЮЧА.

Варіюйте розміщення ключа в послідовності відповідей. Балансуйте розміщення Ключа, щоб кількості разів, яку правильна відповідь знаходиться на кожному з можливих місць (в послідовності відповідей) були приблизно однаковими.

21. ВПОРЯДКОВАНІСТЬ РОЗМІЩЕННЯ ВАРІАНТІВ.

Якщо це можливо, розміщуйте відповіді в логічному, числовому, чи алфавітному порядку.

22. НЕЗАЛЕЖНІСТЬ.

Варіанти відповідей мають бути незалежними, і такими, що не перетинаються по змісту (значенню). Виключенням є організована шкала повноти, коли кожен наступний варіант є підмножиною попереднього.

Чи є серед варіантів відповідей випадки, коли одна відповідь за своїм значенням є підмножиною іншої?

23. ГОМОГЕННІСТЬ

Варіанти відповіді мають бути однорідними (гомогенними) і по змісту і по граматичній структурі.

Чи є варіанти відповіді подібними і значущими? (Чи належать вони до одного домену)

24. ОДНАКОВА ДОВЖИНА.

ПОСІБНИК ДЛЯ РОЗРОБНИКІВ ТЕСТОВИХ ЗАВДАНЬ

Довжина тексту всіх варіантів відповіді має бути приблизно однаковою.
Чи виділяється Ключ своєю довжиною (самий короткий чи найдовший текст серед варіантів відповідей) чи граматичною структурою від інших варіантів відповіді?

25. Відповідь «ЖОДЕН З ВКАЗАНИХ»

слід використовувати з великою обережністю і в виключних випадках.

26. Відповідь «ВСІ ВКАЗАНІ»

ніколи не повинна використовуватись.

27. ПОЗИТИВ.

Формулюйте відповіді позитивно, уникаючи «не» та інших заперечень.

28. БЕЗ ПІДКАЗОК.

Слід уникати непрямих ненавмисних підказок, таких, як:

- А) специфічні крайні визначення, включно з: «завжди», «ніколи», «цілком» і «абсолютно»;
- Б) асоціації типу «відлуння», варіанти відповідей, отримані комбінуванням фрагментів Основи;
- В) граматичні неузгодженості, що підводять кандидатів до певного вибору;
- Г) явно правильні відповіді;
- Д) пари чи трійки відповідей, що логічно підводять кандидата до правильної відповіді;
- Е) безглузді, смішні, абсурдні відповіді.

Чи немає серед Дистракторів прямого протиставлення значенню Ключа?

Якщо один варіант відповіді – «належить до компетенції» а другий «не належить до компетенції», то один з цих варіантів напевно є ключем.

Чи спостерігається частковий збіг текстів варіантів відповіді?

Чи спостерігається частковий збіг тексту Ключа і Основи?

29. ПРАВДОПОДІБНІСТЬ.

Всі Дистрактори повинні бути правдоподібними.

30 . ТИПОВІ ПОМИЛКИ.

Саме типові помилки представників професії (чи студентів) мають використовуватись для створення Дистракторів. Бажано, щоб кожен Дистрактор мав у якості відповідника якусь типову помилку.

Неупередженість і гумор

31. ГУМОР.

Використання гумору допустиме безпосередньо в учбовому процесі, якщо викладач керує орієнтирами, настроями і контекстом, в якому знаходяться студенти.

Використання гумору в тестуванні допустимо лише якщо тестування проводиться співзвучно (сумісно) з особистістю викладача. Оскільки екзаменування для доступу до професійної діяльності проходить у відриві від будь-якого учбового процесу, використання гумору в такому екзаменуванні вважається недоречним.

32. УПЕРЕЖДЕНІСТЬ

Кандидати застосовують різноманітні соціокультурні, емоційні, та ментальні підходи в розуміння тесту. Жоден кандидат не повинен мати переваги, чи бути покараним внаслідок особливостей свого життєвого досвіду, що не пов'язаний з предметом оцінки тестом.

Стимул (Основа) не повинен торкатись (чи просто згадувати) етичні, культурні, расові, регіональні чутливі питання (окрім випадків, коли згадане безпосередньо є предметом вимірювання).

Тестові завдання повинні бути сконструйовані таким чином, щоб не надавати перевагу кандидатам, що володіють розвиненими стратегіями складання тестів (завдяки досвіду, чи спеціалізованій підготовці по стратегіях складання тестів)

Чи надає тестове завдання більше шансів на правильну відповідь кандидатам, що володіють спеціальними знаннями, що виходять за рамки таксономії?

Якщо очікувана аудиторія кандидатів має помітні статистичні відмінності за якимись з загальновідомих (в першу чергу – демографічних) параметрів, то текст кожного тестового завдання обов'язково необхідно особливо уважно оцінювати з точки зору неупередженості, справедливості за відповідними параметрами (стать, національність, вік, соціальний статус і т.ін.)

Чи не базується це тестове завдання на спірних питаннях, які засмучують, або поглядах, які виходять за рамки таксономії (специфікації) тесту?

Приклад: згадування абортів, медичних проблем, чи згадування регіону країни, в якому склались особливі економічні чи небажані суспільно-політичні умови, якщо запитання безпосередньо стосується абсолютно іншої теми, наприклад, господарського процесу. Матеріал тесту не повинен засмучувати, ображати чи лякати кандидата (нагадувати про щось дуже неприємне).

ХІІІ. ПРИКЛАДИ АНАЛІЗУ ТЕСТОВИХ ЗАВДАНЬ. *Сергій Мудрук, за підтримки Секретаріату Вищої кваліфікаційної комісії суддів України*

В цьому розділі представлено низку тестових завдань, більшість з яких мають ті чи інші недоліки, що розглянуто в коментарях до завдань.

Частина представлених завдань використовувалась у тестуванні великої кількості фахівців, що дало можливість отримати представлені тут характеристичні криві.

№ 0.1 Чи можна свідку в цивільному процесі заявити відвід?

А) так

Б) ні

В) тільки за ініціативою головуєчого

Г) тільки коли чиниться незаконний тиск чи інший вплив з метою зміни його позиції у справі на користь тої чи іншої особи

Завдання №0.1 має кілька очевидних недоліків, через які це завдання не може бути допущене до апробації:

- Основа є двозначною: з її формулювання незрозуміло, про можливість чого саме питається: 1) «чи може свідок заявити (сам)?», або ж 2) «чи можна заявити свідку (тобто провести відвід самого свідка у судовому процесі)?»;
- запитання сформульовано як дихотомічне, тобто таке, що очікує в першу чергу на відповідь «Так» чи «Ні»;
- додавання крім «Так» чи «Ні» ще й варіантів відповідей, що містять певні пояснення, аргументації (чому «Так» чи «Ні») чи додаткові умови (при виконанні яких відповідь буде «Так» чи «Ні»), вимагає іншого питального речення, не «чи може...?» а «чи може ... і чому?» або ж «чи може ... і за яких обставин?»;
- запропонований перелік альтернатив (варіантів відповідей) є неоднорідним граматично, а, головне, є неоднорідним за охопленням (деякі з варіантів відповідей знаходяться у відносинах включеності (коли одна відповідь є підмножиною іншої) а деякі – ні.

№ 0.2 Рішення іноземного суду може бути пред'явлено до примусового виконання в Україні протягом:

А) шести місяців з дня набрання ним законної сили;

Б) одного року з дня набрання ним законної сили

В) трьох років з дня набрання ним законної сили

Г) двох років з дня набрання ним законної сили

Єдиною правильною відповіддю розробники вважають «В».

Завдання №0.2 визнано неякісним і не проходило апробацію. Деякі з недоліків:

- при запропонованому формулюванні Основи всі варіанти відповіді з формальної точки зору є вірними, оскільки будь-який строк пред'явлення, що не перевищує три роки, є допустимим, тобто відповідає словам «МОЖЕ бути пред'явлено...»;
- формулювання Основи є граматично неповним. Основа не містить власне повністю сформульованого запитання. Суттєва частина змісту запитання винесена в альтернативи (варіанти відповідей).

№ 0.3 Суддя не залишає позов без розгляду якщо:

А) позивач не звертався до банківської установи за одержанням з відповідача заборгованості, коли вона відповідно до законодавства мала бути одержана через банк;

Б) спір не підлягає вирішенню в господарських судах України

В) позовну заяву підписано особою, посадове становище якої не вказано

Г) у провадженні господарського суду або іншого органу, який в межах своєї компетенції вирішує господарський спір, є справа зі спору між тими ж сторонами, про той же предмет і з тих же підстав або є рішення цих органів з такого спору

Це також приклад тестового завдання з суттєвими недоліками. Деякі з недоліків:

- Основа є не повною. Частина суті запитання міститься в альтернативах (варіантах відповіді);
- формулювання «не залишає» не є коректним в контексті предмету цього завдання, а також відсутнє як правова конструкція загалом;
- використано подвійне заперечення;
- альтернативи (варіанти відповіді) не є однорідними, присвячені потенційно різним контекстам і різним типам знань, що тестуються
- альтернативи не є граматично подібними

№ 0.4 Клопотання судового експерта про надання додаткових матеріалів розглядається господарським судом:

А) після поновлення провадження у справі;

Б) до поновлення провадження у справі;

В) після або до поновлення провадження у справі;

Г) після зупинення провадження у справі.

Недоліки цього завдання, які можна відразу помітити:

- Основа не є повною, ні граматично (немає завершеного питального речення), ні за змістом (не зрозуміло, про що питається в завданні, якщо не прочитати варіанти відповіді);
- дистрактори «Б» та «Г» - завідомо нереальні з правової точки зору і виглядають як повністю нелогічні навіть без застосування правничих знань;
- на ключ вказує граматична підказка – повторюваність ключової фрази «після поновлення провадження у справі»;
- варіанти відповідей не є однорідними

№ 0.5 До адміністративного суду України надійшло доручення від суду Франції про надання правової допомоги щодо допиту громадянина А. Який порядок виконання судового доручення іноземного суду?

А) відповідно до міжнародного договору, згода на обов'язковість якого надана Верховною Радою України, а якщо міжнародний договір не укладено, - дипломатичними каналами

Б) надсилається Міністерству юстиції України;

В) надсилає доручення Міністерству закордонних справ України для передачі дипломатичними каналами;

Г) надсилає Міністерству юстиції України, яке надсилає доручення Міністерству закордонних справ України для передачі дипломатичними каналами.

Правильна відповідь «А».

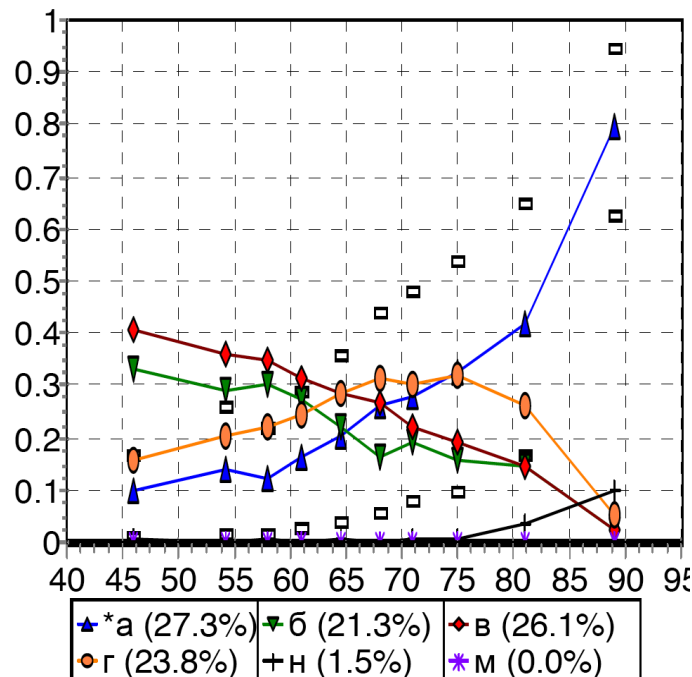
- варіанти відповідей не є однорідними, відрізняються граматичною формою і довжиною;
- ключ «А» складено некоректно оскільки в такому вигляді він визначає порядок отримання доручення (ст. 115-1), а порядок виконання (про який питається в завданні) встановлюється ст.ст. 115-3 та 115-4 Кодексу адміністративного судочинства, окрім цього самі дистрактори є не однорідними, та не відповідають одній граматичній формі.

№1.22. У разі усиновлення повнолітньої особи, різниця у віці між нею та усиновлювачем:

- А) не може бути меншою ніж 18 років
 Б) не встановлена Сімейним кодексом України
 В) усиновлення повнолітньої особи неможливе
 Г) не може бути меншою, ніж 15 років

Серед недоліків цього завдання:

- неоднорідність дистракторів;
- можливе застосування кандидатами асоціації між «повнолітня особа» і «вік 15 років»



Вище наведено характеристичні криві цього завдання. З аналізу видно, що:

- це завдання має підвищену складність і якісну (високу) дискримінацію;
- ключ «А» і дистрактори «Б» і «В» функціонують цілком коректно;
- дистрактор «Г» повноцінно функціонує саме як дистрактор лише для більш підготовленої половини кандидатів. Серед слабо підготовлених кандидатів роботу цього дистрактора щось спотворює (вірогідно вказана вище асоціація).

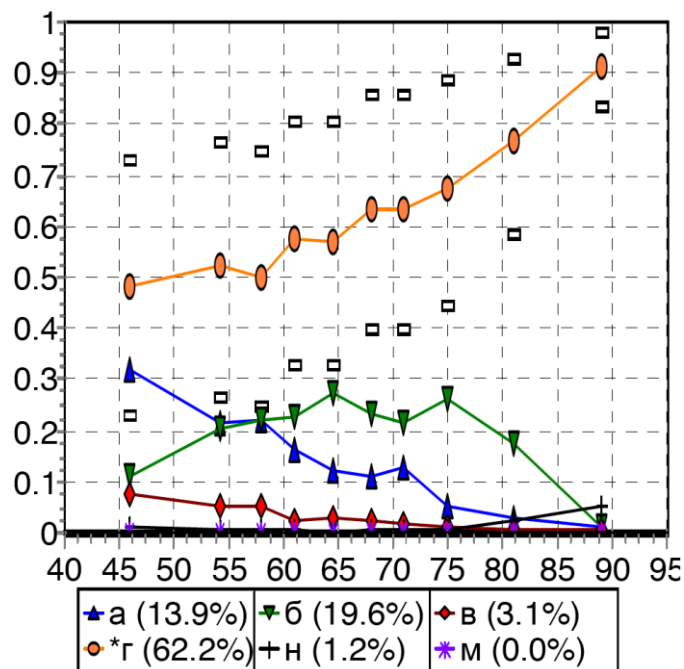
№1.57. Призначення експертизи є обов'язковим для:

- А) визначення психічного стану потерпілого, при наявності в справі даних, які свідчать, що він хворіє на психічну хворобу

Б) встановлення віку потерпілого, якщо про це немає відповідних документів і неможливо їх одержати

В) встановлення характеру і розміру майнової шкоди, заподіяної злочином

Г) встановлення причин смерті



Не дивлячись на певні граматичні недоліки цього Завдання (неповнота Основи, неоднорідність альтернатив), судячи з аналітики (див. характеристичні криві) в основному це завдання працює як належить:

- ключ «Г» і дистрактори «А» і «В» функціонують правильно;
- питання є простим і має прийнятну (посередню) дискримінацію;
- дистрактор «Б» працює як належить лише для краще підготовленої половини аудиторії.

Ймовірно для слабкої частини аудиторії працювала асоціація з передбаченим чинним законодавством (на момент проведення тестування – КПК 1960 року) аналогічної умови для обвинуваченого і потерпілого «якщо про це немає відповідних документів і неможливо їх одержати» -- п. 5 ч.1 статті 76 КПК 1960 року

№2.14. У разі порушення політичною партією Конституції України та законів України до неї можуть бути вжиті такі заходи:

А) попередження про недопущення незаконної діяльності, штраф

Б) заборона політичної партії, штраф

В) попередження про недопущення незаконної діяльності, заборона політичної партії, штраф

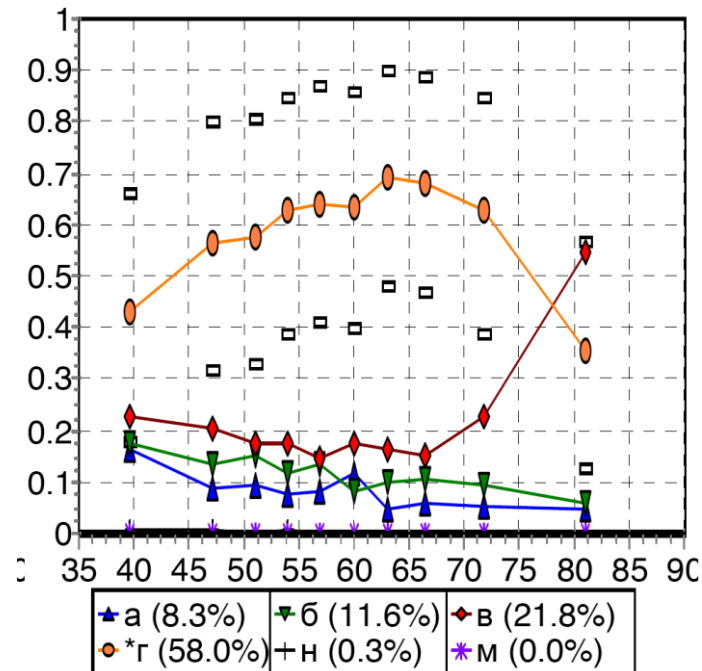
Г) попередження про недопущення незаконної діяльності, заборона політичної партії

Правильна відповідь – «Г».

Характеристичні криві за результатами використання цього завдання в 2012р наведено нижче.

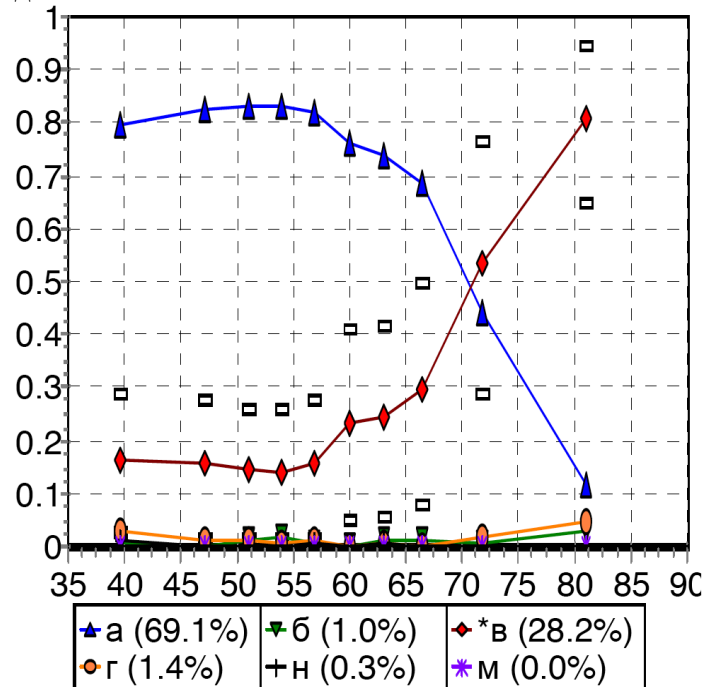
Завдання стосується теми, по якій широковідомого досвіду правозастосування на момент проведення тестування не було. Велика кількість кандидатів не маючи твердих

знань з відповідної теми, спробували застосовувати асоціації, логіку і «народні методи»: «правильна відповідь – сама повна і найдовша». Незвичайна поведінка характеристичної кривої дистрактора «В» очевидно викликана неоднорідністю варіантів відповідей: серед чотирьох відповідей три є комбінацією двох елементів з трьох можливих, і лише відповідь «В» містить в собі всі три елементи.



№2.82. Вкажіть до якого виду належать посади членів Кабінету Міністрів України:

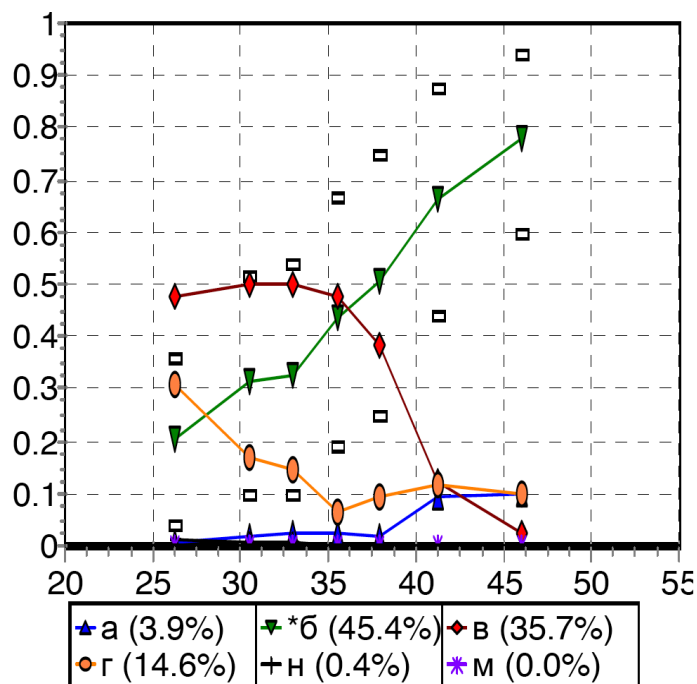
- А) посади державних службовців
- Б) громадські посади
- В) політичні посади
- Г) виробничі посади



В цілому, завдання виконує свою дискримінаційну функцію. Але проблемами цього завдання є неякісні дистрактори. Дистрактори «Б» і «Г» - нелогічні, їх обрала несуттєва частина кандидатів. Ці варіанти відповідей обирали швидше за все в процесі випадкового вибору

№3.17. Позначте, у який із перелічених способів суд може забезпечити адміністративний позов:

- А) зупинення постанови Верховної Ради України
- Б) зупинення дії постанови Кабінету Міністрів України
- В) заборона Вищій Раді юстиції вчиняти певні дії
- Г) зупинення рішення Національного банку України щодо призначення та здійснення тимчасової адміністрації банку



Завдання добре виконує дискримінаційну функцію.

Два дистрактори («В» і «Г») адекватно працюють.

Дистрактор «А» містить редакційний граматичний недолік, який повністю нівелює роботу дистрактора і фактично дає підказку своєї неправильності, можливим правильним формулюванням було б «зупинення дії постанови Верховної Ради України», але у такому випадку дистрактор перетворився б у другий ключ.

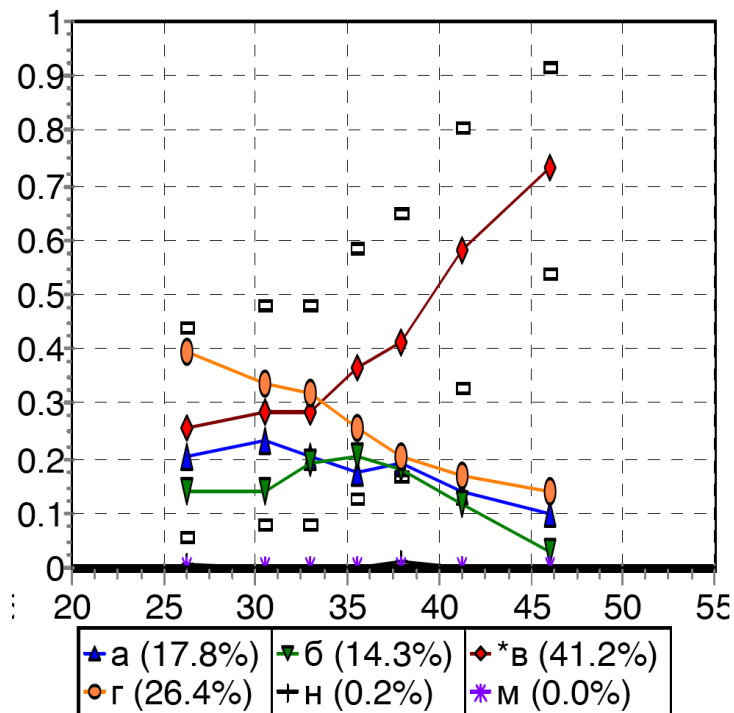
№3.23 Суд розглядає позов фізичної особи А до органу місцевого самоврядування про поновлення на посаді в зв'язку із звільненням із публічної служби й забезпечив позов шляхом зупинення дії правового акта індивідуальної дії про звільнення позивача з посади. Чи правомірним є забезпечення позову в такий спосіб та чому?

А) неправомірним, оскільки адміністративне судочинство не передбачає такого способу забезпечення позову саме для позову про поновлення на посаді в органі місцевого самоврядування;

Б) правомірним, оскільки без його вжиття посада позивача може бути зайнята й у разі задоволення позову необхідно буде докласти значних зусиль для захисту прав позивача;

В) неправомірним, оскільки такий спосіб не відповідає меті застосування правового інституту забезпечення позову;

Г) правомірним, оскільки способом забезпечення адміністративного позову якраз і є зупинення дії рішення суб'єкта владних повноважень;



Слід звернути увагу на формулювання питального речення: “Чи правомірним є забезпечення позову в такий спосіб та чому?”. Це питальне речення є комплексним, оскільки одночасно:

- в ньому питається «чи правомірним є...?», що вимагає відповіді «так» чи «ні»;
- в ньому питається про причини правомірності чи неправомірності забезпечення позову, що вимагає розгорнутих варіантів відповідей (пояснень).

На щастя ці «два запитання в одному» коректно перевіряють один і той самий концепт (одну й ту ж саму здатність кандидата) тому тестове запитання в цілому не є комплексним.

Не дивлячись на частково проблематичне формулювання основного речення, запитання зарекомендувало себе як достатньо якісне запитання підвищеної складності.

Дискримінаційна функція виконується цим питанням добре (див. круто зростаючу криву «В», що відповідає правильній відповіді («ключу»)). Всі три дистрактори в цьому питанні, як і належить, мають спадаючі характеристичні криві (негативна дискримінація) і виявились привабливими для великої кількості кандидатів з невисоким рівнем підготовки (від 14.3% до 26.4% кожний).

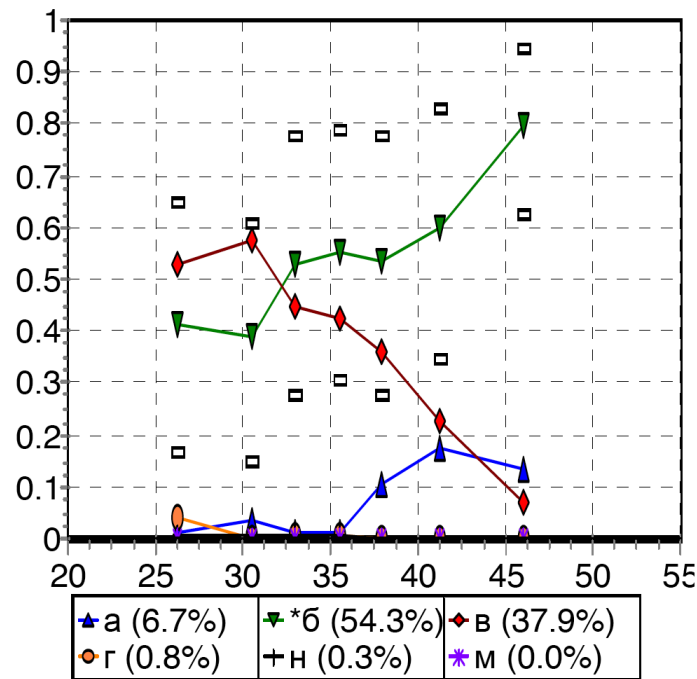
№3.32. Покарання у вигляді виправних робіт може бути призначене неповнолітньому у віці:

А) від 15 до 18 років

Б) від 16 до 18 років

В) від 14 до 18 років, якщо він працює

Г) від 14 до 16 років



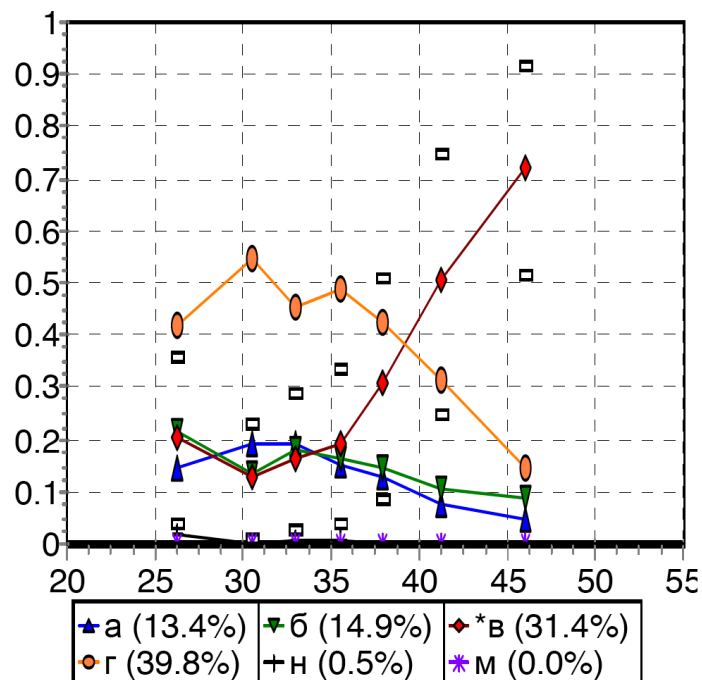
Завдання посередньо виконує дискримінаційні функції. Зміна дистракторів могла би покращити завдання.

Розробником застосовано неоднорідні за формою дистрактори.

Вибір правильного варіанту швидше за все інтуїтивно пов'язувався з мінімальним віком працевлаштування неповнолітніх за загальним правилом.

№3.53 Виконання зобов'язання зі вступом третьої особи в права кредитора отримало назву:

- А) новації;
- Б) регресу;
- В) суброгації;
- Г) цесії.



В цілому це завдання добре виконує дискримінаційні функції.

Характеристичні криві Ключа і дистракторів, з урахуванням похибки, є цілком задовільними.

Покращити це завдання можна лише сформулювавши Основу як повну (сформулювати повне питальне речення)

№4.49. Злочин визнається таким, що вчинений групою осіб, якщо:

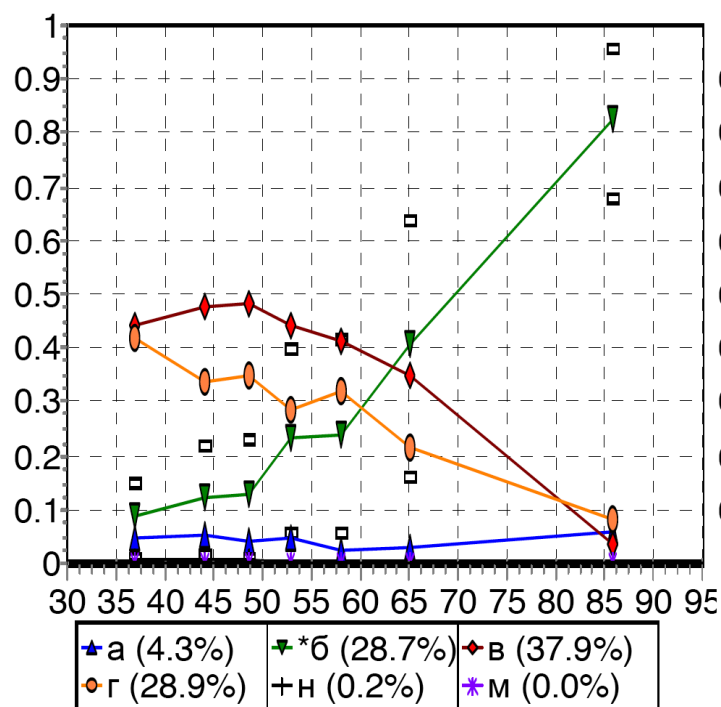
А) в його вчиненні брали участь три і більше особи, які попередньо

зорганізувалися у стійке об'єднання для вчинення цього злочину

Б) у ньому брали участь два або більше виконавців без попередньої змови між собою

В) у ньому брали участь дві або більше особи, незалежно від ролі кожної особи у вчиненні злочину

Г) його спільно вчинили дві або більше особи, які заздалегідь, тобто до початку злочину, домовилися про спільне його вчинення

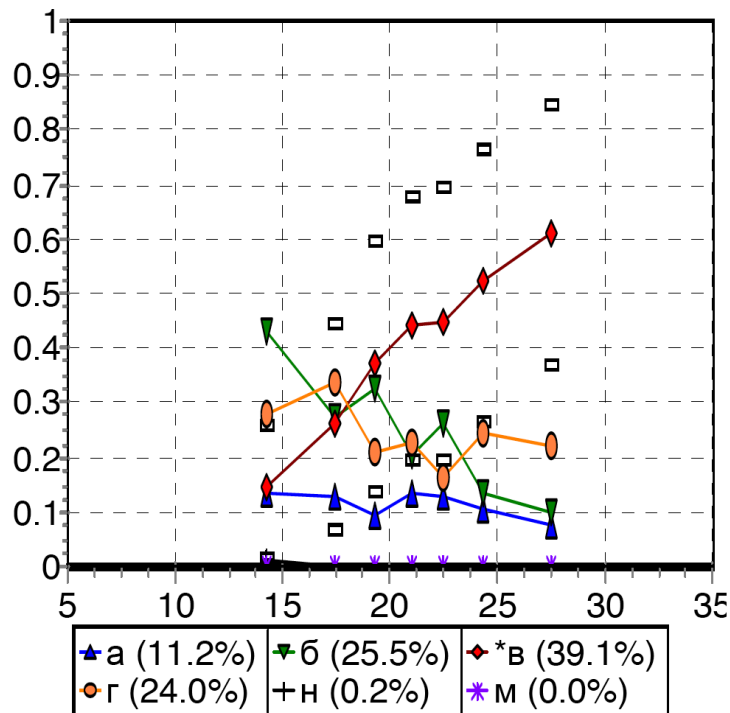


Це завдання дуже добре виконує дискримінаційну функцію. Ключ («Б») і два дистрактори («В» і «Г») працюють адекватно. Дистрактор «А» - не виконує свою функцію.

Причини цього в тому, що в структурі варіантів відповідей є логічна підказка: дистрактори «В» та «Г» і ключ «Б» повторюють кількість осіб (2), а дистрактор «А» окрім відмінності в три дійові особи, має граматичну форму «стійке», що характерне для банди.

№5.12 Колегією суддів адміністративного суду у порядку скороченого провадження прийнято рішення у справі. Вказане рішення суду вступило в законну силу і було виконане. Відповідач по справі звернувся з заявою про роз'яснення вказаного судового рішення. Вкажіть наступні дії суду.

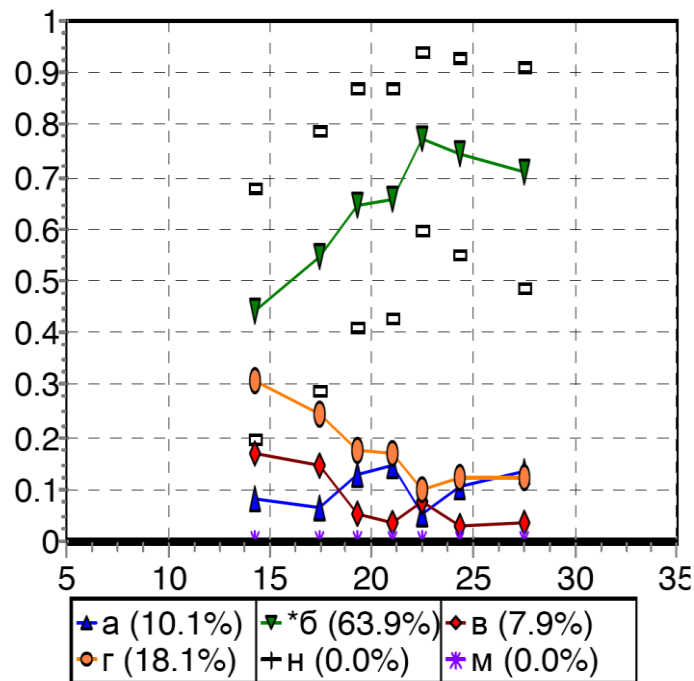
- А) суддя адміністративного суду одноособово у порядку скороченого провадження постановляє ухвалу про задоволення або відмову у задоволенні заяви;
- Б) колегія суддів, яка розглядала справу, у скороченому провадженні розглядає заяву та постановляє ухвалу про її задоволення або відмову у задоволенні;
- В) суддя одноособово приймає ухвалу про залишення заяви про роз'яснення судового рішення без розгляду;
- Г) колегія суддів, яка розглядала вказану справу, або інша колегія, приймає ухвалу про залишення заяви без розгляду.



В цілому це завдання добре виконує дискримінаційні функції. Характеристичні криві Ключа і дистракторів, з урахуванням похибки, є цілком задовільними.

№5.17 Орган виконавчої влади, який не є стороною у справі, не надав без поважних причин витребуваний судом за клопотанням позивача письмовий доказ та не повідомив причин його неподання. Суд постановив ухвалу про тимчасове вилучення цього доказу для дослідження судом. Чи правильно вчинив суд?

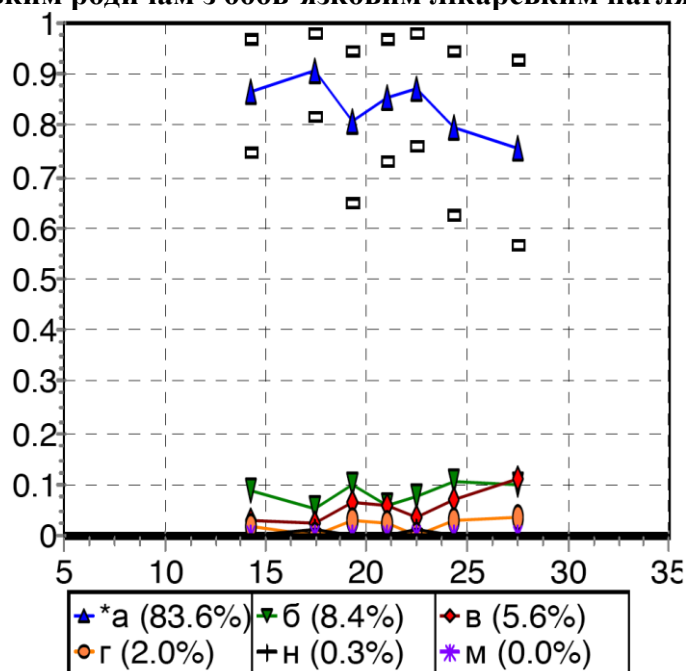
- А) ні, суд мав зупинити провадження у справі до подання доказів до суду;
- Б) так, суд вчинив правильно;
- В) ні, суд мав розглянути справу на основі наявних доказів;
- Г) ні, суд мав оголосити перерву у судовому засіданні і повторно витребувати доказ.



Завдання задовільно виконує дискримінаційну функцію, Ключ («Б») та два дистрактори («В» та «Г») працюють, але завдяки не правому формулюванню дистрактора «А» він не працює.

№5.29. У кримінальному провадженні щодо застосування примусових заходів медичного характеру судом встановлено факт розладу психічної діяльності особи. Вкажіть запобіжний захід, який НЕ може бути застосований судом у даному випадку:

- А) домашній арешт
- Б) поміщення до психіатричного закладу в умовах, що виключають небезпечну поведінку особи
- В) передача членам сім'ї з обов'язковим лікарським наглядом
- Г) передача близьким родичам з обов'язковим лікарським наглядом



Це завдання має слабку негативну дискримінацію. Тобто, замість того, щоб правильну відповідь частіше давали добре підготовлені кандидати, у випадку цього завдання погано підготовлені кандидати з трошки більшою ймовірністю, ніж добре підготовані кандидати, надають правильну відповідь.

Завдання не працює через наявність в Основі очевидної граматичної підказки – «примусовий захід медичного характеру», який вказує на неможливість застосування судом домашнього арешту, який є загальним видом запобіжних заходів і відповідає варіанту «А»,

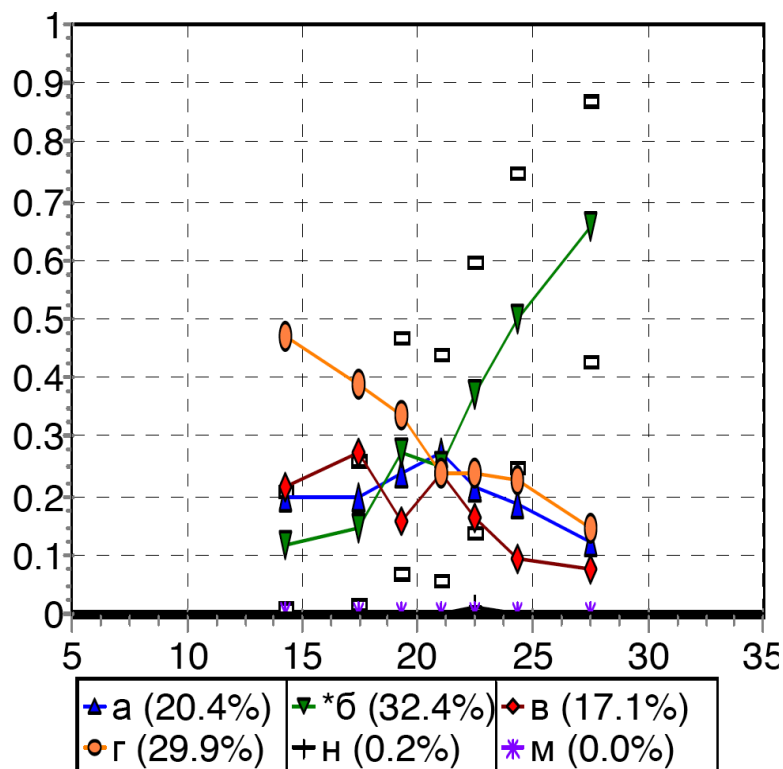
№5.35. Вкажіть вид процесуальної дії, яку повинен вчинити суд у випадку, коли під час судового провадження до суду надійшла угода про визнання винуватості, відповідно до якої між сторонами провадження узгоджено питання звільнення неповнолітнього обвинуваченого від кримінальної відповідальності із застосуванням до нього заходів виховного характеру:

А) невідкладно зупинити проведення процесуальних дій та затвердити угоду про визнання винуватості

Б) відмовити у затвердженні вказаної угоди

В) у обов'язковому порядку допитати усіх свідків у справі, завершити дослідження інших доказів, після чого прийняти рішення про затвердження угоди про визнання винуватості

Г) зупинити проведення процесуальних дій, визначити дату та час розгляду угоди та повідомити усіх учасників судового провадження про дату та час проведення судового засідання щодо розгляду угоди



Завдання добре виконує дискримінаційну функцію. Ключ («Б») на характеристичній кривій зростає, а дистрактори «В» і «Г» спадають.

Дистрактор «А» має особливість: спочатку він зростає, від найслабших кандидатів, до середньо підготовлених. Тобто дистрактор «А» частково повторює поведінку Ключа в групі з підготовкою нижче середньої. Лише починаючи з групи середніх і сильніше підготовлених кандидатів цей дистрактор спадає. Такий «горб» можливо спричинено використанням граматичної «підказки» (слова «Невідкладно зупинити провадження»). Частина слабких кандидатів орієнтується не на суть варіанта відповіді, на граматичну форму, формулювання з «невідкладно» більше очікується в правильних відповідях.

XIV. ЗАГАЛЬНЕ ПЛАНУВАННЯ ТА ЕТАПИ РОЗРОБКИ ТЕСТОВИХ ЗАВДАНЬ. РОЗПОДІЛ РЕСУРСІВ. *Сергій Мудрук*

У процес розробки тестових питань повинні бути залучені такі експертні ресурси:

- Розробники;
- Рецензенти;
- Координатори (координатор розробки та координатор бази екзаменаційних завдань в деяких випадках може бути однією особою)

Координатор розробки планує замовлення на розробку завдань за розділами.

Координатор розробки виконує фінальне рецензування тестових питань, що надійшли для включення в базу даних. У процесі рецензування координатор приймає остаточне рішення про:

- 1) відхилення питання;
- 2) повернення його на доопрацювання;
- 3) включення його до Баз екзаменаційних завдань. Ця робота виконується періодично, відповідно з узгодженим розкладом/планом.

Координатор, що відповідає за ведення Баз екзаменаційних завдань, організовує виконання наступних видів робіт і експертизи :

- 1) остаточну експертизу питань, що включаються в Базу. Для цього можуть залучатися зовнішні експерти -практики, або ж координатор виконує цю експертизу самостійно.
- 2) періодичну експертизу актуальності розміщених в Базі питань. Цю роботу необхідно проводити з відповідними розділами Баз після кожної зміни нормативної бази, законодавства, діючих стандартів. Актуалізація Баз виконується з великим інтервалом навіть у разі відсутності будь-яких згаданих вище змін, з метою виявлення питань з застарілими даними.
- 3) реформатування і перебудову Баз в разі змін у Програмі іспиту. Питання, присвячені виключеним з Програми розділам мають бути вилучені з Баз . Формуються замовлення на розробку питань по нових або розширених розділах програми. Питання з розділів, вагу яких в новій версії Програми зменшено, проводять через критичну експертизу з метою залишити в Базі тільки найбільш якісні питання.

Загальний розподіл ресурсів (бюджет) розробки тестових завдань

Бюджет розробки складається з:

А) власне розробка, плюс авторський супровід питання - 80% бюджету

Б) два незалежних рецензування (огляди, перегляди) - 2x8% бюджету.

Рецензування виконуються послідовно двома різними експертами, з двома (або більше) вивченнями матеріалу кожним з рецензентів (перше вивчення → передача (через координатора) зауважень автору → робота автора над зауваженнями → розгляд перероблених автором матеріалів і відповідей → завершальне резюме (для передачі матеріалів координатору, а через нього наступному рецензенту або в базу екзаменаційних питань)

За результатами кожного рецензування приймається одне з трьох рішень:

- зберегти (схвалити) виправлене завдання;
- повернути завдання на доопрацювання;

- усунути завдання з робочої Базы і відмовитися від його подальшого розгляду.

В) супровід питання в Базі екзаменаційних завдань - 4% бюджету

У супровід входить, зокрема, періодичні експрес-перегляди на предмет підтвердження актуальності питань що знаходяться в базі протягом не менше певного періоду.

Зусилля (робочий час) Координаторів зазвичай розглядають як частину адміністративного бюджету а не бюджету розробки.

Апробація стоїть трохи осторонь в бюджеті здійснення розробки.

Апробація є одним з найважливіших компонентів системи контролю якості, і абсолютно необхідна для масштабних систем екзаменування і систем допуску до відповідальної професійної діяльності.

Організація апробації може здійснюватися абсолютно різними способами:

- 1) або як захід, що окремо організовується і бюджетується,
- 2) або як додаткова операція в штатному навчальному процесі в якості рубіжного контролю,
- 3) або безпосередньо в складі реального екзаменування, але без нарахування балів (про що мають бути попереджені кандидати).

Бюджет апробації найчастіше не включають безпосередньо до бюджету розробки.

У разі виявлення під час апробації будь-яких індикаторів проблем якості тестового запитання, Координатор повинен виконати цикл рецензування, можливо (на власний розсуд) без участі авторів і попередніх рецензентів.

Типовий розподіл ресурсів (часу) розробника

Створення Основи

Вимагає найчастіше від чверті до третини (20 - 35%) загального обсягу зусиль розробника.

Вимагає креативних зусиль і здатності вербалізації (verb - дієслово) ідей.

Складним творчим завданням є розробка тестових питань на вищі рівні розумових процесів (аналіз, оцінка, особливо - синтез).

Ще одна складність для розробників, що ще не мають великого досвіду - адекватна оцінка мінімального рівня розумових процесів, необхідних кандидату для надання правильної відповіді на розроблене завдання.

Створення Ключа

При добре опрацьованій ідеї Основи та адекватній оцінці відповідності запитання таксономії, розробка Ключа вимагає зовсім небагато зусиль розробника - до 10-15% від загального їх обсягу.

Створення Дистракторів

Основна частина (більше половини загального обсягу) зусиль розробника припадає на створення ідей дистракторів, їх вербалізацію і письмову реалізацію.

Створення ефективних дистракторів вимагає або величезного практичного досвіду (досвіду великого числа усвідомлених власних помилок, або спостереження великого числа помилок інших професіоналів), або великого досвіду перевірки знань в процесі навчання чи атестації фахівців.

У середньому, на формулювання **кожного** дистрактора йде 15 - 20 % загального обсягу зусиль розробника, тобто помітно більше, ніж на формулювання Ключа

Фінальне редагування, самоконтроль якості

- Верстка у відповідності зі стандартизованими формальними вимогами замовника розробки.
- Перевірка граматичної узгодженості.
- Вирівнювання довжин і досягнення граматичної однорідності всіх відповідей
- Пошуки і видалення можливих словесних асоціацій та інших «підказок» і «пасток». Це найбільш «нетехнологічна» частина зусиль розробника, що погано формалізується, вимагає і відстороненої оцінки, та евристики, і свіжого погляду на власний текст.

Зазвичай «доведення» і «полірування» питання, складеного за всіма рекомендаціями, не займає багато зусиль. Однак у складних випадках удосконалення питання може зайняти до третини загального обсягу зусиль розробника. У середньому, перегляд екзаменаційного питання в цілому, контекстне та стильове редагування займає близько 10% зусиль розробника

Виправлення розробленого тестового завдання за результатами рецензування

Ресурси, необхідні на цей опціональний етап розробки, зазвичай не специфікуються.

Найчастіше умовою контракту з розробником є зарахування подібних робіт до категорії відпрацювання реклаमाції, що виконується як авторський супровід в рамках затвердженого бюджету на розробку. Тобто, на такі роботи не виділяється додатковий бюджет (робочого часу або коштів).