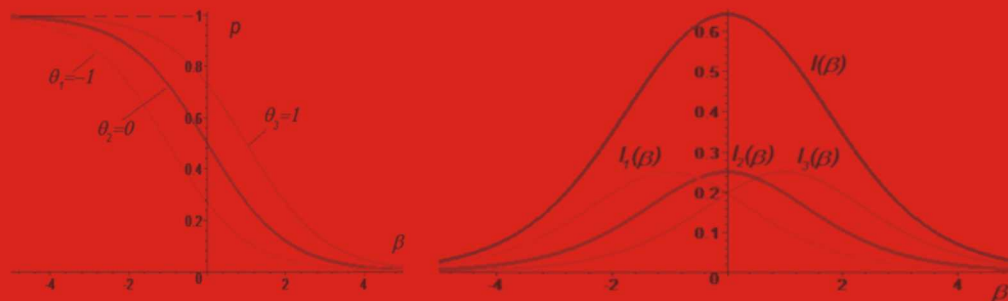




This project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

ВИМІРЮВАННЯ В ОСВІТІ

ВИМІРЮВАННЯ В ОСВІТІ



Міністерство освіти і науки, молоді та спорту України
Кіровоградський державний педагогічний університет
імені Володимира Винниченка

ВИМІРЮВАННЯ В ОСВІТІ

підручник



This project has been funded with support from the European Commission.

This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Кіровоград
Видавець Лисенко В.Ф.
2011

УДК 371
ББК 74.04(4Укр)я73
В-47

В-47 Вимірювання в освіті: Підручник / За редакцією
О.В. Авраменко.– Кіровоград: Лисенко В.Ф., 2011. – 360 с.
ISBN 978-966-2570-21-2

Підручник призначений для систематизації знань з теоретичних основ освітніх вимірювань, класичних та сучасних моделей тестування, статистичних методів параметризації тестів, конструювання тестів та тестових завдань, комп'ютерних засобів тестування, моніторингу якості освіти в Україні та світі, а також методики навчання освітніх вимірювань.

Рекомендується студентам спеціальності «Освітні вимірювання» при вивченні дисциплін «Вимірювання в освіті», «Основи педагогічного оцінювання», «Класичні тестові моделі», «Математично-статистичні методи в освітніх вимірюваннях», «Конструювання тестів», «Комп'ютерні технології у тестуванні», «Моделі та методи ІРТ», «Моніторинг якості освіти», «Прикладна статистика», «Методика навчання освітніх вимірювань», а також при підготовці до державної атестації. Рекомендується студентам спеціальності «Статистика» при вивченні дисципліни «Математично-статистичні методи в освітніх вимірюваннях», а також при підготовці до державної атестації. Рекомендується студентам спеціальностей «Інформатика», «Математика» та «Фізика» при вивченні дисциплін спеціалізації «Освітні вимірювання».

Підручник може бути корисним магістрантам, аспірантам, викладачам університетів і працівникам інших навчальних закладів та установ сфери освіти, які цікавляться теорією та практикою вимірювання рівня підготовки учасників тестування та розробкою ефективних тестів.

ББК 74.04(4Укр)я73

Рецензенти: КУШНІР В.А. доктор педагогічних наук, професор,
 ПЛІЧКО А.М. доктор фізико-математичних наук, професор.

Затверджено до друку методичною радою КДПУ імені Володимира Винниченка
(протокол № 4 від 21 грудня 2011)

ISBN 978-966-2570-21-2

© Кіровоградський державний педагогічний університет імені Володимира Винниченка, 2011
© Авраменко О.В., Котяк В.В., Лупан І.В., та інші, 2011
© Видавець Лисенко В.Ф., 2011

ПЕРЕДМОВА

Підручник підготовлений за Проектом 145029-TEMPUS-2008-SE-JPCR «Освітні вимірювання, адаптовані до стандартів ЄС». У підготовці до друку підручника брали участь такі виконавці Проекту:

- Авраменко О.В. (доктор фіз.-мат. наук, проф.) – статистичне моделювання тестів, шкали вимірювання, особливості навчання магістрантів спеціальності 8.18010022 «Освітні вимірювання», інформаційні джерела освітніх вимірювань;*
- Котяк В.В. (ст. викл.) - комп'ютерні технології у тестуванні, комп'ютерні засоби дистанційного навчання, адаптивне тестування;*
- Лупан І.В. (канд. пед. наук, доц.) – методика навчання освітніх вимірювань, дистанційне навчання, комп'ютерні статистичні пакети;*
- Лутченко Л.І. (канд. пед. наук, доц.) - математично-статистичні методи вимірювань в освіті;*
- Пасічник Н.О. (канд. пед. наук, доц.) – педагогічне оцінювання, моніторинг якості освіти;*
- Паращук С.Д. (канд. фіз.-мат. наук, доц.) – класичні підходи до моделювання тестів, конструювання тестів, комп'ютерні моделі тестування;*
- Ріжняк Р.Я. (канд. пед. наук, проф.) - організація навчання магістрантів спеціальності 8.18010022 «Освітні вимірювання»;*
- Резіна О.В. (канд. пед. наук, доц.) – понятійний апарат педагогічних вимірювань, форми тестових завдань, основи моніторингу якості освіти;*
- Янчукова Н.В. (асистент) – приклади математично-статистичного аналізу результатів тестування.*

При підготовці матеріалів були опрацьовані різноманітні наукові та науково-методичні джерела, списки яких подано наприкінці кожного розділу підручника.

Матеріал підручника широко відображає основні напрями теорії та практики освітніх вимірювань, оскільки він планувався як допомога випускникам магістратури спеціальності «Освітні вимірювання» при підготовці до державної атестації. Для випускників магістратури зазначу, що користування виключно даним посібником не може вважатися достатнім при підготовці до державного екзамену. Сподіваюсь, що підручник допоможе краще орієнтуватися у великому обсязі теоретичного матеріалу з освітніх вимірювань.

Бажаю читачам успіхів в опануванні сучасною теорією освітніх вимірювань та практичними прийомами підготовки, організації та проведення тестування. А випускникам магістратури «Освітні вимірювання» - ні пуху, ні пера на державному екзамені!

О.Авраменко

Грудень 2011 р.

ПЕРЕДМОВА

Підручник підготовлений за Проектом 145029-TEMPUS-2008-SE-JPCR «Освітні вимірювання, адаптовані до стандартів ЄС». У підготовці до друку підручника брали участь такі виконавці Проекту:

- Авраменко О.В. (доктор фіз.-мат. наук, проф.) – статистичне моделювання тестів, шкали вимірювання, особливості навчання магістрантів спеціальності 8.18010022 «Освітні вимірювання», інформаційні джерела освітніх вимірювань;*
- Котяк В.В. (ст. викл.) - комп'ютерні технології у тестуванні, комп'ютерні засоби дистанційного навчання, адаптивне тестування;*
- Лупан І.В. (канд. пед. наук, доц.) – методика навчання освітніх вимірювань, дистанційне навчання, комп'ютерні статистичні пакети;*
- Лутченко Л.В. (канд. пед. наук, доц.) - математично-статистичні методи вимірювань в освіті;*
- Пасічник Н.О. (канд. пед. наук, доц.) – педагогічне оцінювання, моніторинг якості освіти;*
- Паращук С.Д. (канд. фіз.-мат. наук, доц.) – класичні підходи до моделювання тестів, конструювання тестів, комп'ютерні моделі тестування;*
- Ріжняк Р.Я. (канд. пед. наук, проф.) - організація навчання магістрантів спеціальності 8.18010022 «Освітні вимірювання»;*
- Резіна О.В. (канд. пед. наук, доц.) – понятійний апарат педагогічних вимірювань, форми тестових завдань, основи моніторингу якості освіти;*
- Янчукова Н.В. (асистент) – приклади математично-статистичного аналізу результатів тестування.*

При підготовці матеріалів були опрацьовані різноманітні наукові та науково-методичні джерела, списки яких подано наприкінці кожного роділу підручника.

Матеріал підручника широко відображає основні напрями теорії та практики освітніх вимірювань, оскільки він планувався як допомога випускникам магістратури спеціальності «Освітні вимірювання» при підготовці до державної атестації. Для випускників магістратури зазначу, що користування виключно даним посібником не може вважатися достатнім при підготовці до державного екзамену. Сподіваюсь, що підручник допоможе краще орієнтуватися у великому обсязі теоретичного матеріалу з освітніх вимірювань.

Бажаю читачам успіхів в опануванні сучасною теорією освітніх вимірювань та практичними прийомами підготовки, організації та проведення тестування. А випускникам магістратури «Освітні вимірювання» - ні пуху, ні пера на державному екзамені!

О.Авраменко

Грудень 2011 р.

З М І С Т

I. Педагогічне оцінювання (Пасічник Н.О.)	11
II. Вступ до освітніх вимірювань: <i>понятійний апарат та статистичні методи</i> (Авраменко О.В., Лутченко Л.І., Паращук С.Д., Резіна О.В.)	30
III. Конструювання тестів (Котяк В.В., Паращук С.Д., Резіна О.В.) ..	83
IV. Класична теорія тестування (Авраменко О.В., Паращук С.Д., Янчукова Н.В.)	119
V. Статистичне моделювання тестів (Авраменко О.В.)	184
VI. Комп'ютерні технології у тестуванні (Котяк В.В., Паращук С.Д.).	239
VII. Моніторинг якості освіти (Пасічник Н.О., Резіна О.В.)	285
VIII. Освітні вимірювання: <i>стандарти, форми та засоби навчання</i> (Авраменко О.В., Лупан І.В., Ріжняк Р.Я.)	309

РОЗГОРНУТИЙ ЗМІСТ

I. ПЕДАГОГІЧНЕ ОЦІНЮВАННЯ	11
1.1. Основні види і методи педагогічного оцінювання	11
1.2. Вимоги до оцінювання в освіті, рівні навчальних досягнень	14
1.2.1. Загальнопедагогічні вимоги до оцінювання в освіті	14
1.2.2. Рівні навчальних досягнень учнів та критерії їх визначення	15
1.2.3. Визначення навчальних досягнень учнів з математики	17
1.3. Історія, сучасний стан та тенденції розвитку систем оцінювання в освіті	20
1.3.1. Системи оцінювання в освіті: історичний аспект	20
1.3.2. Тенденції розвитку систем оцінювання в освіті	22
1.4. Зовнішнє незалежне оцінювання (ЗНО): зміст, процедура та організаційно-технологічне забезпечення	25
1.4.1. Зміст та процедура ЗНО	25
1.4.2. Організаційно-технологічне забезпечення ЗНО	26
Питання для самоконтролю	27
Список джерел, використаних у розділі I	29
II. ВСТУП ДО ОСВІТНІХ ВИМІРЮВАНЬ: ПОНЯТІЙНИЙ АПАРАТ ТА СТАТИСТИЧНІ МЕТОДИ	30
2.1. Основні складові вимірювання в освіті	30
2.1.1. Поняття вимірювання в освіті та латентні змінні	30
2.1.2. Компоненти та інструменти процесу вимірювань в освіті	31
2.2. Класифікація шкал: основні кількісні та якісні шкали	32
2.3. Два типи вимірювання та два типи тестів: нормативно-орієнтовані та критеріально-орієнтовані підходи в освіті	35
2.3.1. Класифікація типів вимірювання в освіті за метою	35
2.3.2. Класифікація тестів за метою вимірювання: нормативно-орієнтовані та критеріально-орієнтовані тести	37
2.3.3. Інші типи класифікації тестів	39
2.4. Цілі навчання та таксономія Блума	40
2.4.1. Класифікація цілей навчання	40
2.4.2. Таксономія Блума	40
2.5. Випадкові величини та їх числові характеристики	42
2.5.1. Випадкова величина. Закон розподілу випадкової величини	42
2.5.2. Диференціальна та інтегральна функції розподілу випадкової величини та їх властивості	44
2.5.3. Числові характеристики центральної тенденції для випадкових величин	45
2.5.4. Числові характеристики варіації для випадкових величин	46
2.5.5. Числові характеристики форми для випадкових величин	48
2.6. Основні розподіли дискретних випадкових величин	49
2.6.1. Рівномірний закон розподілу ймовірностей	49
2.6.2. Гіпергеометричний закон розподілу ймовірностей	49

2.6.3. Геометричний закон розподілу ймовірностей	50
2.6.4. Біноміальний закон розподілу ймовірностей	50
2.6.5. Пуассонівський закон розподілу ймовірностей	51
2.7. Основні розподіли неперервних випадкових величин	51
2.7.1. Рівномірний закон розподілу	51
2.7.2. Показниковий (експоненціальний) закон розподілу	52
2.7.3. Нормальний закон розподілу	53
2.8. Закон великих чисел, його застосування в теорії оцінювання	55
2.9. Центральна гранична теорема та її застосування в освітніх вимірюваннях	58
2.10. Емпіричні дані вимірювань в освіті та методи визначення законів їх розподілу	59
2.11. Вибірковий метод та оцінка результатів вибіркового спостереження у педагогіці й психології	62
2.11.1. Вибірковий метод	62
2.11.2. Оцінка результатів вибіркового спостереження	62
2.11.3. Обробка малих вибірок	63
2.12. Загальні методи вивчення зв'язку соціально-педагогічних явищ і процесів . .	64
2.12.1. Типи зв'язків між величинами	64
2.12.2. Достовірність дослідження кореляційного зв'язку	65
2.12.3. Лінійна кореляція	66
2.12.4. Рангова кореляція	69
Практичні завдання	71
Питання для самоконтролю	79
Список джерел, використаних у розділі II	82
III. КОНСТРУЮВАННЯ ТЕСТІВ	83
3.1. Тестові завдання та тест	83
3.1.1. Поняття завдання у тестовій формі, тестового завдання, тесту	83
3.1.2. Порівняльний аналіз різних форм тестових завдань	84
3.2. Завдання закритої форми	87
3.2.1. Завдання з вибором однієї правильної відповіді	87
3.2.2. Завдання з вибором декількох правильних відповідей	88
3.3. Завдання на встановлення відповідності та послідовності	89
3.3.1. Завдання на встановлення відповідності (логічні пари)	89
3.3.2. Завдання на встановлення правильної послідовності	90
3.4. Завдання відкритої форми	91
3.4.1. Завдання відкритої форми з короткою відповіддю	91
3.4.2. Завдання відкритої форми з розгорнутою відповіддю	92
3.5. Основні етапи конструювання педагогічного тесту	92
3.5.1. Алгоритм створення тесту	92
3.5.2. Модифіковані алгоритми створення тестів різних типів	94
3.6. Планування змісту тесту та загальні принципи його відбору	95
3.6.1. Розробка плану тесту	95
3.6.2. Два підходи до розробки короткої специфікації тесту	95
3.6.3. Загальні принципи правильного відбору змісту тестів	97
3.7. Експертиза якості змісту тесту	97
3.7.1. Експертиза окремих завдань тесту	98
3.7.2. Експертиза якості змісту паралельних варіантів тесту	100

3.7.3. <i>Заключний етап експертизи: висновки та рекомендації</i>	100
3.8. Розробка супровідної документації тесту та створення інструкцій для учня і для викладача	101
3.8.1. <i>Методичне оснащення тесту</i>	101
3.8.2. <i>Пояснювальні та інструктивні матеріали до тесту</i>	101
3.9. Моделі педагогічного тестування	102
3.9.1. <i>Класична модель</i>	102
3.9.2. <i>Класична модель з урахуванням складності завдань</i>	104
3.9.3. <i>Модель за зростанням складності</i>	104
3.9.4. <i>Модель з розділенням завдань за рівнями засвоєння</i>	105
3.9.5. <i>Модель з урахуванням часу відповіді на завдання</i>	105
3.9.6. <i>Модель з обмеженням часу на тест</i>	106
3.10. <i>Адаптивне тестування - сучасний метод вимірювання в освіті</i>	107
3.10.1. <i>Поняття адаптивного тестування та методики його організації</i>	107
3.10.2. <i>Алгоритми формування тесту при адаптивному тестуванні</i>	109
Практичні завдання	111
<i>Питання для самоконтролю</i>	115
Список джерел, використаних у розділі III	118
IV. КЛАСИЧНА ТЕОРІЯ ТЕСТУВАННЯ	119
4.1. Основні положення класичної теорії тестування	119
4.2. Первинний аналіз результатів тестування	122
4.2.1. <i>Матриця результатів тестування та її редукування</i>	122
4.2.2. <i>Побудова рядів результатів тестування та графічна інтерпретація</i>	126
4.3. Основні статистичні характеристики результатів тестування	128
4.3.1. <i>Міри центральної тенденції тестових балів</i>	128
4.3.2. <i>Міри мінливості тестових балів</i>	129
4.3.3. <i>Перевірка гіпотези про нормальний закон розподілу результатів тестування</i>	131
4.3.4. <i>Асиметрія та ексцес, їх інтерпретація</i>	132
4.4. Кореляційний аналіз результатів тестування	133
4.4.1. <i>Коефіцієнт кореляції Пірсона</i>	133
4.4.2. <i>ϕ-коефіцієнт кореляції. Кореляційна матриця тестових завдань</i>	134
4.4.3. <i>Коефіцієнт бісеріальної кореляції</i>	136
4.5. Основні характеристики тестових завдань	138
4.5.1. <i>Трудність тестових завдань</i>	138
4.5.2. <i>Правдоподібність дистракторів</i>	139
4.5.3. <i>Дискримінативність тестового завдання</i>	143
4.5.4. <i>Надійність завдань і показники валідності</i>	147
4.6. Надійність тесту та її числові характеристики	148
4.6.1. <i>Надійність тесту в класичній теорії</i>	148
4.6.2. <i>Показник надійності</i>	148
4.6.3. <i>Коефіцієнт надійності</i>	150
4.6.4. <i>Надійність складених тестових оцінок</i>	152
4.7. Методи оцінки надійності та її прогнозування	152
4.7.1. <i>Методи двократного тестування</i>	152
4.7.2. <i>Методи однократного тестування</i>	153

4.8. Фактори впливу на надійність та похибки її обчислення	158
4.8.1. Фактори, що впливають на коефіцієнти надійності.	158
4.8.2. Похибка вимірювання надійності.	160
4.9. Валідність тесту у класичній теорії	163
4.9.1. Валідність для критеріально-орієнтованих тестів	163
4.9.2. Інтерпретація результатів валідації	165
4.9.3. Метод контрастних груп	169
Практичні завдання	170
Питання для самоконтролю	182
Список джерел, використаних у розділі IV	182
V. СТАТИСТИЧНЕ МОДЕЛЮВАННЯ ТЕСТІВ	184
5.1. Основні припущення та поняття сучасної теорії тестування	184
5.2. Основні параметричні моделі в освітніх вимірюваннях	185
5.2.1. Однопараметрична модель Раша	185
5.2.2. Двох-параметрична модель Бірнбаума	187
5.2.3. Трьох-параметрична модель Бірнбаума	188
5.3. Первинні бали та їх статистичні характеристики	189
5.3.1. Первинні бали та їх розподіл	189
5.3.2. Оцінки ймовірностей первинного балу учасників тестування.	191
5.3.3. Оцінки ймовірностей первинного балу тестових завдань	192
5.3.4. Важливі властивості первинних балів	193
5.4. Оцінки латентних параметрів	194
5.4.1. Стандартні оцінки параметрів та помилки вимірювання	194
5.4.2. Стійкі оцінки латентних параметрів	196
5.5. Характеристичні функції та криві для емпіричних даних	197
5.5.1. Побудова характеристичних функцій та кривих	197
5.5.2. Два підходи теорії параметризації тестів	198
5.6. Інформаційна функція	200
5.6.1. Інформаційна функція тестового завдання	200
5.6.2. Інформаційна функція системи тестових завдань	201
5.7. Основні характеристики тесту за сучасною теорією	202
5.7.1. Роздільна здатність тесту (PCT)	202
5.7.2. Оцінка дискримінаційної здатності тесту	203
5.8. Визначення надійності тесту сучасною теорією	204
5.8.1. Загальні відомості про надійність тесту	204
5.8.2. Дисперсійний аналіз тесту, що містить декілька розділів	206
5.8.3. Визначення коефіцієнта надійності дисперсійним аналізом	207
5.8.4. Кореляційна методика оцінювання надійності. Ретестова надійність	208
5.9. Валідність тесту та її визначення сучасною теорією	209
5.10. Порядкові шкали у педагогічному тестуванні	211
5.11. Побудова єдиної шкали	213
5.11.1. Побудова єдиної метричної шкали	213
5.11.2. Побудова єдиної нормованої шкали	215
Практичні завдання	217
Питання для самоконтролю	231
Список джерел, використаних у розділі V	238

VI. КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ у ТЕСТУВАННІ	239
6.1. Основні відомості про комп'ютерне тестування	239
6.1.1. <i>Види автоматизованого контролю</i>	239
6.1.2. <i>Етапи комп'ютерної атестації</i>	240
6.1.3. <i>Вимоги до комп'ютерної атестації</i>	241
6.1.4. <i>Типи сценаріїв комп'ютерного тестування</i>	242
6.2. Інформаційна безпека процедур комп'ютерного тестування	244
6.2.1. <i>Основні вимоги інформаційної безпеки масових тестувань</i>	244
6.2.2. <i>Шляхи забезпечення інформаційної безпеки</i>	245
6.2.3. <i>Кроки для досягнення безпечної технології тестування</i>	247
6.3. Сервіси Web 2.0 в освіті та в освітніх вимірюваннях	248
6.3.1. <i>Короткі відомості про Web 2.0</i>	248
6.3.2. <i>Можливості використання Web 2.0</i>	249
6.4. Hot Potatos - програмний засіб для локального тестування	252
6.4.1. <i>Загальні відомості про систему Hot Potatoes</i>	252
6.4.2. <i>Блоки системи Hot Potatoes</i>	253
6.5. Система MyTestX – засіб для проведення мережевого тестування	256
6.6. Система TCExam - веб орієнтована система тестування	259
6.7. Освітнє середовище Moodle: тестування та оцінювання	262
6.7.1. <i>Загальна характеристика VLE Moodle, створення курсів</i>	262
6.7.2. <i>Середовище Moodle, банк тестових завдань, створення тесту</i>	266
6.7.3. <i>Аналіз тестових завдань та засоби оцінювання в середовищі Moodle</i>	270
6.8. Комп'ютерні моделі педагогічного тестування	273
6.8.1. <i>Основні вимоги та класифікація</i>	273
6.8.2. <i>Лінійні моделі педагогічного тестування</i>	274
6.8.3. <i>Ієрархічні моделі педагогічного тестування</i>	276
6.8.4. <i>Мережні моделі педагогічного тестування</i>	278
Питання для самоконтролю	281
Список джерел, використаних у розділі VI	283
VII. МОНІТОРИНГ ЯКОСТІ ОСВІТИ	285
7.1. Показники моніторингу та характеристики якості освіти	285
7.1.1. <i>Моніторинг та моніторингові показники</i>	285
7.1.2. <i>Якість освіти та її основні характеристики</i>	286
7.2. Моніторинг якості освіти як кваліметричний інструмент	288
7.3. Моніторинг якості освіти та моніторингові дослідження	290
7.3.1. <i>Об'єкти і суб'єкти моніторингу якості освіти</i>	290
7.3.2. <i>Принципи моніторингових досліджень</i>	292
7.3.3. <i>Види та напрямки моніторингу</i>	292
7.4. Історія розвитку моніторингу якості освіти	294
7.5. Міжнародні моніторингові дослідження якості освіти	294
7.5.1. <i>Організація перших міжнародних моніторингових досліджень</i>	297
7.5.2. <i>Міжнародне моніторингове дослідження TIMSS та інші моніторингові дослідження кінця ХХ ст.</i>	298
7.5.3. <i>Міжнародна програма оцінки знань та умінь учнів PISA</i>	299
7.5.4. <i>Міжнародний проект PIRLS</i>	302
7.6. Моніторинг якості освіти як системне дослідження освітнього процесу	303
Питання для самоконтролю	307

Список джерел, використаних у розділі VII	308
VIII. ОСВІТНІ ВИМІРЮВАННЯ: стандарти, форми та засоби навчання	309
8.1. Освітні вимірювання як галузь	309
8.1.1. Розвиток освітніх вимірювань як окремої галузі	309
8.1.2. Проблеми освітніх вимірювань	311
8.2. Методи досліджень у галузі освітніх вимірювань	312
8.2.1. Методи емпіричного дослідження	312
8.2.2. Педагогічний експеримент	314
8.3. Методична система навчання освітніх вимірювань	316
8.3.1. Складові методичної системи	316
8.3.2. Цілі навчання освітніх вимірювань	316
8.3.3. Типові задачі діяльності фахівця з освітніх вимірювань	317
8.3.4. Зміст підготовки фахівця з освітніх вимірювань	318
8.3.5. Принципи, методи та організаційні форми навчання освітніх вимірювань	319
8.4. Галузевий стандарт та програми підготовки фахівців з освітніх вимірювань . .	321
8.4.1. Галузевий стандарт підготовки магістра	321
8.4.2. Ключові компетенції фахівця з освітніх вимірювань	325
8.4.3. Програма підготовки бакалавра	327
8.4.4. Програма підготовки магістра	329
8.5. Вимоги до професійної відповідальності при оцінюванні	333
8.5.1. Стандарти освітнього та психологічного тестування	333
8.5.2. Зарубіжні стандарти	334
8.5.3. Кодекс справедливого тестування в освіті	336
8.5.4. Стандарти компетентності вчителів з оцінювання учнів	338
8.5.5. Відповідальність при тестуванні та оцінюванні в Україні	339
8.6. Дистанційне навчання	340
8.6.1. Ознаки дистанційного навчання	341
8.6.2. Характерні риси дистанційного навчання	342
8.6.3. Принципи дистанційного навчання	343
8.6.4. Педагогічні вимоги до дистанційного навчання	345
8.6.5. Рівні дистанційного навчання	346
8.6.6. Забезпечення дистанційного навчання	348
8.6.7. Вимоги до викладачів у дистанційному навчанні	348
8.7. Засоби навчання освітніх вимірювань	349
8.7.1. Інформаційні джерела з освітніх вимірювань	359
8.7.2. Комп'ютерні статистичні пакети	356
Питання для самоконтролю	357
Список джерел, використаних у розділі VIII	359

I. ПЕДАГОГІЧНЕ ОЦІНЮВАННЯ

На початку розділу наведені основні види і методи педагогічного оцінювання. Тут містяться відомості про оцінювання досягнень учнів та загальнопедагогічні вимоги, що покладено в основу оцінювання навчальних досягнень учнів. Наведені рівні навчальних досягнень учнів, якісна характеристика рівнів і загальні критерії їх оцінювання. Коротко описані історія, сучасний стан та тенденції розвитку систем оцінювання в освіті, а також принципи і функції педагогічного оцінювання. Подана також інформація про зовнішнє незалежне оцінювання - процес (або комплекс організаційних процедур) визначення рівня навчальних досягнень випускників загальноосвітніх навчальних закладів України.

Звертаємо увагу читача, що у розділі I даного підручника з *вимірювань* в освіті розкривається сутність педагогічного *оцінювання*. Докладне тлумачення термінів *вимірювання* (англ. - measurement) та *оцінювання* (англ. - assessment, evaluation) наведене у розділі VIII (див. пункт 8.1.2).

1.1. Основні види і методи педагогічного оцінювання

Охарактеризуємо основні види і методи педагогічного оцінювання навчальних досягнень учнів в системі загальної середньої освіти і студентів вищих навчальних закладів.

Основними видами оцінювання навчальних досягнень учнів загальноосвітніх навчальних закладів і студентів вищих навчальних закладів I – II рівня акредитації є *поточне, тематичне й підсумкове (семестрове, річне оцінювання та державна підсумкова атестація), а основною його одиницею - навчальна тема.*

Поточне оцінювання здійснюється у процесі поурочного вивчення теми. Його основними завдання є: встановлення й оцінювання рівнів розуміння і первинного засвоєння окремих елементів змісту теми, встановлення зв'язків між ними та засвоєним змістом попередніх тем, закріплення знань, умінь і навичок.

Формами поточного оцінювання є індивідуальне та фронтальне опитування; робота з діаграмами, графіками, схемами; зарисовки біологічних об'єктів; робота з контурними картами; виконання учнями різних видів письмових робіт; взаємоконтроль учнів у парах і групах; самоконтроль тощо.

В умовах упровадження зовнішнього незалежного оцінювання особливого значення набуває тестова форма контролю та оцінювання навчальних досягнень учнів. Поточне оцінювання на кожному уроці в традиційному розумінні недоцільне, тому що воно не узгоджується з індивідуальним для кожного учня темпом засвоєння навчального матеріалу, націлюється не стільки на осмислення, скільки на просте запам'ятовування навчального матеріалу. Відповідно, поточне оцінювання має відігравати допоміжну роль, виконуючи, зокрема, заохочувальну, стимулюючу та діагностичну функції. Його результати не обов'язково відображаються в балах і фіксуються в журналі. Інформація, отримана на підставі поточного контролю, є основною для коригування роботи вчителя на уроці.

Тематичному оцінюванню навчальних досягнень підлягають основні результати вивчення теми (розділу). Перед початком вивчення чергової теми (розділу) всі учні мають бути ознайомлені з тривалістю вивчення теми (кількість занять); кількістю й тематикою обов'язкових робіт і термінами їх проведення; умовами оцінювання.

Тематичне оцінювання навчальних досягнень учнів забезпечує:

- усунення безсистемності в оцінюванні;
- підвищення об'єктивності оцінки знань, навичок і вмінь;
- індивідуальний та диференційований підхід до організації навчання;
- систематизацію й узагальнення навчального матеріалу;
- концентрацію уваги учнів до найсуттєвішого в системі знань з кожного предмета.

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

У процесі вивчення значних за обсягом тем дозволяється проводити декілька проміжних тематичних оцінювань. І, навпаки, якщо на опанування матеріалу теми передбачено, наприклад, 2-3 навчальні години, вони можуть об'єднуватися для проведення тематичного оцінювання.

Оцінка за семестр виставляється за результатами тематичного оцінювання, а за рік - на основі семестрових оцінок.

Учень (учениця) має право на підвищення семестрової оцінки. При цьому потрібно мати на увазі, що відповідно до Положення про золоту

медаль “За високі досягнення в навчанні” та срібну медаль “За досягнення в навчанні”, затвердженого наказом Міністерства освіти і науки, молоді та спорту України, підвищення результатів семестрового оцінювання через переатестацію не дає підстав для нагородження випускників золотою або срібною медалями.

Існує ще один вид оцінювання – самооцінювання, коли учні самостійно здійснюють контроль за своїми успіхами та невдачами, виявляють проблеми.

Більш гнучкої, різнопланової системи оцінювання потребує профільна старша школа, яка на основі диференційованого навчання повинна враховувати не лише навчальні досягнення, але і творчі, проектно - дослідницькі, соціально значущі результати.

Ефективною у старшій школі є рейтингова система оцінювання, яка сприяє формуванню ключових компетентностей і створює можливості для:

- визначення рівня підготовленості учнів на кожному етапі навчального процесу;
- отримання об’єктивних показників щодо засвоєння знань та сформованості умінь учнів не лише протягом навчального року, а й за весь період навчання у старшій школі;
- градації значущості балів, які отримують учні за виконання різних видів робіт (самостійна робота, підсумкова робота, творча робота, олімпіади, виставки, конкурси творчих робіт, науково–дослідні й художні проекти тощо);
- підвищення навчальної мотивації й відповідальності;
- підвищення об’єктивності оцінювання.

Упровадження рейтингу досягнень передбачає побудову учнем (ученицею) індивідуальної освітньої програми, яка дозволить учителям і батькам учнів аналізувати їхній освітній поступ та його (її) досягнення, виявляти помилки, а також регулювати форми й види освітньої діяльності.

У вищих навчальних закладах України III – IV рівнів акредитації з 2008 – 2009 навчального року впроваджено європейську систему оцінювання знань ECTS. Європейська кредитна трансферна система (ECTS) – система, створена для забезпечення єдиної міждержавної процедури оцінювання навчання. Вона базується на трьох ключових елементах: на інформації про навчальні програми й здобутки студентів, на взаємній угоді між закладами-партнерами та студентом і на використанні кредитів ECTS.

Основними видами оцінювання навчальних досягнень студентів у вищих навчальних закладах є поточне, модульне, семестрове та підсумкове (державна атестація). За рішенням адміністрації університету може здійснюватися також незалежне оцінювання знань студентів у вигляді ректорських контрольних робіт та контрольних зрізів. В системі вищої освіти України використовується накопичувальна і середньозважена системи оцінювання навчальних досягнень

Оцінювання якості навчальних досягнень можливе лише за умови оптимального поєднання різноманітних методів педагогічного діагностування: методів усної, письмової й практичної перевірки. Методи усної перевірки результатів навчання – бесіда, розповідь, пояснення, читання тексту, географічних, історичних карт, схем та ін. Активними формами перевірки знань є залік і усний екзамен. Методи письмової перевірки результатів навчання – контрольні роботи (завдання), твір, диктант, реферат, тестування тощо. Вони дають можливість перевірити теоретичні знання, уміння їх застосовувати, ступінь оволодіння письмовою мовою. Методи практичної перевірки – проведення лабораторних дослідів, створення виробів та моделей тощо.

1.2. Вимоги до оцінювання в освіті, рівні навчальних досягнень

1.2.1. Загальнопедагогічні вимоги до оцінювання в освіті

При оцінюванні навчальних досягнень учнів мають ураховуватися такі загальнопедагогічні вимоги:

1. *Характеристики відповіді учня:* осмисленість, цілісність, логічність, обґрунтованість, правильність.
2. *Якість знань:* повнота, глибина, гнучкість, дієвість, системність, узагальненість, міцність. Характеристики якості знань взаємопов'язані між собою і доповнюють одна одну. Наведемо визначення деяких з них:
 - а) повнота знань – кількість знань, визначених навчальною програмою;
 - б) глибина знань – усвідомленість існуючих зв'язків між групами знань;
 - в) гнучкість знань – уміння учнів застосовувати набуті знання у стандартних і нестандартних ситуаціях; знаходити варіативні способи використання знань; уміння комбінувати новий спосіб діяльності із вже відомих;
 - г) системність знань – усвідомлення структури знань, їх ієрархії і

- послідовності, тобто усвідомлення одних знань як базових для інших;
д) міцність знань – тривалість збереження в пам'яті та відтворення.
3. *Ступінь сформованості загальнонавчальних та предметних умінь і навичок.* Знання є складовою умінь учнів діяти. Уміння виявляються в різних видах діяльності і поділяються на розумові і практичні. Навички – дії доведені до автоматизму у результаті виконання вправ. Для сформованих навичок характерні швидкість і точність відтворення.
 4. *Рівень володіння розумовими операціями:* вміння аналізувати, синтезувати, порівнювати, абстрагувати, класифікувати, узагальнювати, робити висновки тощо.
 5. *Ступінь самостійності учнів в навчальній діяльності.*
 6. *Досвід творчої діяльності* (вміння виявляти проблеми та розв'язувати їх, формулювати гіпотези).
 7. *Самостійність оцінних суджень.*

1.2.2. Рівні навчальних досягнень учнів та критерії їх визначення

Вказані орієнтири покладено в основу чотирьох рівнів навчальних досягнень учнів: *початкового, середнього, достатнього, високого.* У загальнодидактичному плані рівні (див. таблицю 1.1) визначаються так:

Перший рівень – початковий. Відповідь учня при відтворенні навчального матеріалу елементарна, фрагментарна, характеризується початковими (нечіткими, розрізненими) уявленнями про предмет вивчення; уміння не сформовані, рівень самостійності навчальної діяльності низький.

Другий рівень – середній. Знання неповні, поверхові. Учень відтворює основний навчальний матеріал, але недостатньо осмислено, не вміє самостійно аналізувати, робити висновки. Здатний виконувати завдання за зразком, володіє елементарними вміннями навчальної діяльності.

Третій рівень – достатній. Характеризується знаннями істотних ознак понять, явищ, закономірностей, зв'язків між ними. Учень також самостійно застосовує знання в стандартних ситуаціях, володіє розумовими операціями (аналізом, абстрагуванням, узагальненням тощо), вміє робити висновки, виправляти допущені помилки. Відповідь учня (учениця) правильна, логічна, обґрунтована, хоча їм бракує власних суджень. Він здатний самостійно здійснювати основні види навчальної діяльності.

Таблиця 1.1. Загальні критерії оцінювання навчальних досягнень учнів

Рівні навчальних досягнень	Бали	Критерії оцінювання
I. Початковий	1	Учень (учениця) може розрізнити об'єкти вивчення
	2	Учень (учениця) відтворює незначну частину навчального матеріалу, має нечіткі уявлення про об'єкт вивчення
	3	Учень (учениця) відтворює частину навчального матеріалу; з допомогою вчителя виконує елементарні завдання
II. Середній	4	Учень (учениця) з допомогою вчителя відтворює основний навчальний матеріал, може повторити за зразком певну операцію, дію
	5	Учень (учениця) відтворює основний навчальний матеріал, здатний з помилками й неточностями дати визначення понять, сформулювати правило
	6	Учень (учениця) виявляє знання й розуміння основних положень навчального матеріалу. Відповідь його(її) правильна, але недостатньо осмислена. Вміє застосовувати знання при виконанні завдань за зразком
III. Достатній	7	Учень (учениця) правильно відтворює навчальний матеріал, знає основоположні теорії і факти, вміє наводити окремі власні приклади на підтвердження певних думок, частково контролює власні навчальні дії
	8	Знання учня (учениці) є достатніми, він (вона) застосовує вивчений матеріал у стандартних ситуаціях, намагається аналізувати, встановлювати найсуттєвіші зв'язки і залежність між явищами, фактами, робити висновки, загалом контролює власну діяльність. Відповідь його (її) логічна, хоч і має неточності
	9	Учень (учениця) добре володіє вивченим матеріалом, застосовує знання в стандартних ситуаціях, уміє аналізувати й систематизувати інформацію, використовує загальновідомі докази із самостійною і правильною аргументацією
IV. Високий	10	Учень (учениця) має міцні знання, здатний (а) використовувати їх у практичній діяльності, робити висновки, узагальнення, аргументувати їх
	11	Учень (учениця) на високому рівні володіє узагальненими знаннями в обсязі та в межах вимог навчальних програм, аргументовано використовує їх у різних ситуаціях, уміє знаходити інформацію та аналізувати її, ставити і розв'язувати проблеми
	12	Учень (учениця) має системні глибокі знання в обсязі та в межах вимог навчальних програм, усвідомлено використовує їх у стандартних та нестандартних ситуаціях. Уміє самостійно аналізувати, оцінювати, узагальнювати опанований матеріал, самостійно користуватися джерелами інформації, приймати рішення

Четвертий рівень - високий. Знання учня є глибокими, міцними, системними; учень вміє застосовувати їх для виконання творчих завдань,

Його навчальна діяльність позначена вмінням самостійно оцінювати різноманітні ситуації, явища, факти, виявляти і відстоювати особисту позицію. Водночас, визначення високого рівня навчальних досягнень, зокрема оцінки 12 балів, передбачає знання та уміння в межах навчальної програми і не передбачає участі школярів у олімпіадах, творчих конкурсах тощо.

Кожний наступний рівень вимог вбирає в себе вимоги до попереднього, а також додає нові характеристики. Критерії оцінювання навчальних досягнень реалізуються в нормах оцінок, які встановлюють чітке співвідношення між вимогами до знань, умінь і навичок, які оцінюються, та показником оцінки в балах.

1.2.3. Визначення навчальних досягнень учнів з математики

Загальні критерії оцінювання покладено в основу визначення навчальних досягнень учнів з конкретних навчальних предметів. При відповіді на дане питання доцільно визначити відповідні досягнення з конкретного навчального предмету. Як приклад, наведемо критерії оцінювання навчальних досягнень учнів з математики (Затверджені наказом МОН України №371 від 5.05.2008 року).

До навчальних досягнень учнів з математики, які підлягають оцінюванню, належать:

- теоретичні знання, що стосуються математичних понять, тверджень, теорем, властивостей, ознак, методів та ідей математики;

- знання, що стосуються способів діяльності, які можна подати у вигляді системи дій (правила, алгоритми);

- здатність безпосередньо здійснювати уже відомі способи діяльності відповідно до засвоєних правил, алгоритмів (наприклад, виконувати певне тотожне перетворення виразу, розв'язувати рівняння певного виду, виконувати геометричні побудови, досліджувати функцію на монотонність, розв'язувати текстові задачі розглянутих типів тощо);

- здатність застосовувати набуті знання і вміння для розв'язання навчальних і практичних задач, коли шлях, спосіб такого розв'язання потрібно попередньо визначити (знайти) самому.

Відповідно до ступеня оволодіння зазначеними знаннями і способами діяльності виокремлюються такі рівні навчальних досягнень школярів з математики:

Початковий рівень - учень (учениця) називає математичний об'єкт (вираз, формули, геометричну фігуру, символ), але тільки в тому випадку, коли цей об'єкт (його зображення, опис, характеристика) запропоновано йому (їй) безпосередньо; за допомогою вчителя виконує елементарні завдання.

Середній рівень - учень (учениця) повторює інформацію, операції, дії, засвоєні ним (нею) у процесі навчання, здатний(а) розв'язувати завдання за зразком.

Достатній рівень - учень (учениця) самостійно застосовує знання в стандартних ситуаціях, вміє виконувати математичні операції, загальні методи і послідовність (алгоритм) яких йому(їй) знайомі, але зміст та умови виконання змінені.

Високий рівень - учень (учениця) здатний(а) самостійно орієнтуватися в нових для нього(неї) ситуаціях, складати план дій і виконувати його; пропонувати нові, невідомі йому(їй) раніше розв'язання, тобто його(її) діяльність має дослідницький характер.

Оцінювання якості математичної підготовки учнів з математики здійснюється у двох аспектах: рівень оволодіння теоретичними знаннями та якість практичних умінь і навичок, здатність застосовувати вивчений матеріал під час розв'язування задач і вправ.

I. Початковий:

1 - Учень (учениця) розпізнає один із кількох запропонованих математичних об'єктів (символів, виразів, геометричних фігур тощо), виділивши його серед інших; читає і записує числа, переписує даний математичний вираз, формулу; зображує найпростіші геометричні фігури (малює ескіз);

2 - Учень (учениця) виконує однокрокові дії з числами, найпростішими математичними виразами; впізнає окремі математичні об'єкти і пояснює свій вибір;

3 - Учень (учениця) порівнює дані або словесно описані математичні об'єкти за їх суттєвими властивостями; за допомогою вчителя виконує елементарні завдання.

II. Середній:

4 - Учень (учениця) відтворює означення математичних понять і формулювання тверджень; називає елементи математичних об'єктів;

формулює деякі властивості математичних об'єктів; виконує за зразком завдання обов'язкового рівня;

5 - Учень (учениця) ілюструє означення математичних понять, формулювання теорем і правила виконання математичних дій прикладами із пояснень вчителя або підручника; розв'язує завдання обов'язкового рівня за відомими алгоритмами з частковим поясненням;

6 - Учень (учениця) ілюструє означення математичних понять, формулювання теорем і правила виконання математичних дій власними прикладами; самостійно розв'язує завдання обов'язкового рівня з достатнім поясненням; записує математичний вираз, формулу за словесним формулюванням і навпаки.

III. Достатній:

7 - Учень (учениця) застосовує означення математичних понять та їх властивостей для розв'язання завдань у знайомих ситуаціях; знає залежності між елементами математичних об'єктів; самостійно виправляє вказані йому (їй) помилки; розв'язує завдання, передбачені програмою, без достатніх пояснень;

8 - Учень (учениця) володіє визначеним програмою навчальним матеріалом; розв'язує завдання, передбачені програмою, з частковим поясненням; частково аргументує математичні міркування й розв'язування завдань

9 - Учень (учениця): вільно володіє визначеним програмою навчальним матеріалом; самостійно виконує завдання в знайомих ситуаціях з достатнім поясненням; виправляє допущені помилки; повністю аргументує обґрунтування математичних тверджень; розв'язує завдання з достатнім поясненням

IV. Високий:

10 - Знання, вміння й навички учня (учениці) повністю відповідають вимогам програми, зокрема: учень (учениця) усвідомлює нові для нього (неї) математичні факти, ідеї, вміє доводити передбачені програмою математичні твердження з достатнім обґрунтуванням; під керівництвом учителя знаходить джерела інформації та самостійно використовує їх; розв'язує завдання з повним поясненням і обґрунтуванням;

11 - Учень (учениця) вільно і правильно висловлює відповідні математичні міркування, переконливо аргументує їх; самостійно знаходить

джерела інформації та працює з ними; використовує набуті знання і вміння в незнайомих для нього (неї) ситуаціях; знає передбачені програмою основні методи розв'язання завдання і вміє їх застосовувати з необхідним обґрунтуванням;

12 - Учень (учениця) виявляє варіативність мислення і раціональність у виборі способу розв'язання математичної проблеми; вміє узагальнювати та систематизувати набуті знання; здатний(а) до розв'язування нестандартних задач і вправ.

1.3. Історія, сучасний стан та тенденції розвитку систем оцінювання в освіті

1.3.1. Системи оцінювання в освіті: історичний аспект

Розвиток освіти, як сфери діяльності, об'єктивно пов'язаний з еволюцією системи оцінювання. Найдавнішою є словесна оцінка. В часи античності вона знаходила відображення через холістичне оцінювання – це тип оцінювання, що спирається на загальне враження особи, яка визнається експертом у певній сфері, про рівень виконаної роботи.

Точно визначити, коли виникла бальна система оцінювання достатньо складно, але найчастіше науковці пов'язують її появу з навчальними закладами, які організовували члени ордену єзуїтів. Вважається, що в XVI – XVII ст. в єзуїтських школах учні ділилися за розрядами, котрі позначалися цифрами, підвищуючи розряд, учень збільшував кількість особистих привілеїв. Приблизно в цей же період, в Німеччині, виникає трьохбальна система оцінювання. Її специфіка полягала у тому, що кожен учень отримував бали, що позначали розряд і визначали його особисту успішність серед учнів класу (1-й – кращий, 2-й – середній, 3-й – гірший). Пізніше середній розряд, до якого належала найбільша кількість учнів, поділили на класи і так виникла п'ятибальна шкала.

У XVIII ст. в Німеччині, що була тоді "інкубатором" педагогічних інновацій і впливала на розвиток педагогічної думки в Європі, зародився прообраз 12-бальної системи оцінювання. У школі Philanthropinum, відкритій І.Б.Базедовим у 1774 р. у м. Дессау (Німеччина), застосовувалася своєрідна система заохочення вихованців: на особливій дошці проти прізвища учнів ставили крапки, по числу яких визначалися успіхи й відповідні їм пільги.

Вихованця, який одержав певну кількість крапок, нагороджували яким-небудь знаком відмінності або ласим блюдом. Число крапок обмежили дванадцятьма.

У дореволюційній Росії єдина 12-бальна система використалася у всіх військово-навчальних закладах, а в гімназіях функціонувала 5-ти бальна система, яка співіснувала з 3-х, 8-ми, 10-ти, 12-ти бальними системами оцінювання знань в інших типах навчальних закладів. 1837 року Міністерство народної освіти офіційно ввело 5-ти бальну систему та перехідні іспити. За цією системою, яка по суті була 6-ти бальною, тому що починалася не з «1», а з «0», балам відповідали словесні оцінки: «1» – слабкі успіхи; «2» – посередні; «3» – достатні; «4» – добрі; «5» – відмінні. Використовували цю систему непослідовно: оцінки за старанність за бальною системою не виставлялися; у контрольних та екзаменаційних відомостях рівень знань учнів визначався словесно («вельми похвально», «похвально» «добре», «вельми посередньо», «переконливо», «скромно», «дуже скромно», «погано», «кепсько», «дуже погано»).

У травні 1918 року були відмінені бальні оцінки знань і поведінки учнів, а також перехід учнів до наступного класу на основі оцінок, заборонялися всі види іспитів. Свідоцтва про закінчення школи видавалися на основі відгуків педагогічної ради. Відмінялася усна перевірка знань учнів на уроці; фронтальне опитування, письмові роботи залікового характеру допускалися лише як крайні заходи. Рекомендувалося проводити бесіди з учнями по темі, що вивчається, усні та письмові доповіді, звіти учнів про виконану роботу, ведення робочих щоденників. Широкого поширення набули тестові завдання, які використовувалися як форма самоперевірки. Навчання без оцінок призвело до зниження якості знань, рівня навченості та дисципліни, що зумовило поновлення систематичності обліку знань учнів (1932 р). У вересні 1935 року в школі було встановлено п'ять словесних оцінок: «дуже погано», «погано», «посередньо», «добре», «відмінно»; у січні 1944 року було прийнято рішення про заміну словесної системи оцінки успішності і поведінки цифровою п'ятибальною системою. Поступово оцінка «1» стала використовуватися все менше, а починаючи з 50-х років менше почала використовуватися й оцінка «2». П'ятибальна система оцінок фактично перетворилася у трибальну, а для більшості учнів, які не можуть навчатися на «4» і «5», — двобальну. Така оціночна система слабо стимулювала навчальну

діяльність учнів і зумовила введена з вересня 2000 року 12-бальну систему оцінювання навчальних досягнень учнів, побудовану за принципом урахування особистих досягнень учнів в опануванні ними змісту загальної середньої освіти.

1.3.2. Тенденції розвитку систем оцінювання в освіті

За кордоном у визначенні системи оцінювання спостерігається велика різноманітність, як у самих принципах, так і в конкретному виборі способів виставлення оцінок. У зарубіжних школах практикуються різні системи оцінювання знань, умінь, навичок, прийняті різні системи оцінок, включаючи стобальну (Афганістан, Бангладеш, Бахрейн, Бурунді, Гаїті, Гондурас, Єгипет, Замбія, Ізраїль, Індія, Йорданія, Ірак, Іран, Китай, Ліберія, Мексика, Нікарагуа, ОАЕ, Пакистан, Панама, Південна Корея, Тайвань, Ефіопія, ЮАР, Японія); тридцятибальну (Італія); двадцятибальну (Бельгія, Венесуела, Ліван, Португалія, Франція та 14 країн Африки); п'ятнадцятибальну (Алжир, Марокко, Туніс); дванадцятибальну (1-12 Україна, 0-12 Уругвай); десяти-, двобальну та ін. Так, при оцінці знань в середніх навчальних закладах Німеччини поширена шестирівнева шкала оцінок. Кожній з них відповідає сума балів з урахуванням тенденції оцінки: відмінно (1) = 15-14-13 балів; добре (2) = 12-11-10 балів; задовільно (3) = 9-8-7 балів; нижче задовільного (4) = 6-5-4 бали; слабо (5) = 3-2-1 бал; незадовільно (0) = 0 балів. У Франції система оцінювання знань коливається в залежності від сфери їх використання. Наприклад, початкова школа використовує 5-бальну шкалу, і це розумно, бо діти в перші роки навчання вивчають прості речі. В неповній середній школі (коледжі) використовується 10-бальна система: обсяг навчального матеріалу більший, тому є необхідність враховувати нюанси сприйняття дітьми інформації, яка їм подається. Повна та вища школи працюють в 20-бальному діапазоні, всередині якого функціонує ще й коефіцієнт значимості, різний для кожного навчального закладу (наприклад, філософія в рамках філологічної спеціалізації має коефіцієнт 5, а в рамках спеціалізації природничої може мати коефіцієнт 2). Якщо система освіти Франції з багатовіковими традиціями в першу чергу прагне до «шліфування» системи оцінювання знань на кожному етапі освіти, то африканські та східні країни, багато з яких лише нещодавно взяли за ліквідацію неписьменності, головне завдання вбачають в іншому. Організатори навчання в цих країнах концентрують увагу

в основному на середній освіті і відслідковують бодай малі рухи учнів до знань. Але особливу увагу в таких школах (у тих же Ефіопії, Танзанії) приділяють ставленню дітей не тільки до навчання, а й одне до одного, а також до справ, які мають громадське значення, що дуже важливо в тих країнах, котрі ще тільки будують свою державність.

Отже, аналіз історичного процесу розвитку системи оцінювання навчальних досягнень учнів дозволяє зробити висновок, що основною системою оцінювання є бальна, яка має особливості в різних країнах, щодо критеріїв, форм і методів оцінювання, з урахуванням національних та історичних особливостей конкретної країни. Бальна система оцінювання рівня навчальних досягнень учнів має певні недоліки, основними з яких є наступні: а) не існує еталона за допомогою якого можна було б виміряти й об'єктивно оцінити знання учнів; б) оцінювання у балах псує відносини між вчителем і учнями, створює основу для обоюдної недовіри; в) вчитель з джерела знань перетворюється на контролера, який може помилятися; г) необхідність систематичного оцінювання відволікає вчителя від його основних обов'язків і перетворює урок на постійне опитування, тестування. Не зважаючи на певні недоліки цієї системи, вона залишається найбільш простим і доступним засобом оцінювання в освіті. Для мінімізації негативних аспектів оцінювання в балах, вчитель повинен дотримуватися таких педагогічних вимог при оцінюванні навчальних досягнень учнів:

1. *Всебічність оцінки*, тобто максимальне врахування усіх проявів навчальної діяльності учнів, встановлення зв'язку з усіма видами роботи учня на уроці, знань теоретичних положень з практичними вміннями і навичками.

2. *Індивідуальний характер оцінки*. Передбачає оцінювання успішності кожного учня окремо; за результатами його власної навчальної діяльності, враховуючи темперамент, характер, здібності, нахили, інтереси, потреби, мотиви, особливості мислення, пам'яті, уваги, волі кожного окремого учня. Не можна допускати підміни результатів учіння окремих учнів підсумками роботи класу чи групи школярів.

3. *Систематичність оцінювання*, що означає регулярність проведення контролю успішності учнів протягом усього процесу навчання, поєднання його іншими сторонами навчальної роботи і позитивний вплив на увесь хід навчання школярів.

4. *Тематичність* передбачає оцінку навчальної діяльності учнів не тільки за її кінцевим результатом за чверть чи навчальний рік, а й за кожний урок, кожному темі і розділ. Цьому і слугує тематичний облік успішності, який полягає у тому, що, проводячи поурочне індивідуальне і фронтальне опитування, контрольні, практичні, лабораторні та самостійні роботи вчитель визначає, яку тему, який розділ програми учень засвоїв краще, а який - гірше, як спрямувати свою подальшу роботу.

5. *Об'єктивність*. Оцінка повинна відображати дійсний рівень знання навчального матеріалу, передбаченого програмами. Недопустимим є як завищення, так і заниження оцінок. Завищення оцінок заводять в оману як батьків, так і учнів щодо дійсного стану справ. Цей недолік особливо виразно проявляється тоді, коли учень переходить в іншу школу, або до більш вимогливого вчителя, або вступає у вищі чи середні спеціальні навчальні заклади. Шкідливим є і заниження оцінок, зайва вимогливість, адже у дітей втрачається стимул, мотивація до навчання.

6. *Диференційований характер оцінки* вимагає від учителя враховувати індивідуальні особливості учнів, специфічні особливості навчального предмета і навчального матеріалу. Тому, у відповідності з цими особливостями, необхідно застосовувати різні методики обліку успішності і використовувати різні види оцінок.

7. *Гласність і зрозумілість оцінки*. Вчитель повинен довести до всіх учнів вимоги щодо оцінювання знань, умінь і навичок, а учні повинні розуміти за що їм поставлена та чи інша оцінка. Оцінку у вигляді цифрового балу необхідно супроводжувати оціночним судженням, пояснювати позитивне і негативне у відповідях, накреслювати шляхи подальшої роботи.

8. *Єдність вимог вчителів*. Передбачає врахування усіма вчителями єдиних вимог до оцінювання успішності, які передбачаються навчальними програмами. Порушення єдності вимог, різнобій в оцінках виключають можливість об'єктивного і адекватного порівняльного відображення якості навчання у різних школах.

9. *Оцінка повинна відображати певний обсяг знань, умінь і навичок з кожного навчального предмета, знання конкретного матеріалу, а не виступати у ролі оцінки за поведінку учня на уроках чи на перерві.*

1.4. Зовнішнє незалежне оцінювання (ЗНО): зміст, процедура та організаційно-технологічне забезпечення

1.4.1. Зміст та процедура ЗНО

Зовнішнє незалежне оцінювання (ЗНО) – процес (комплекс організаційних процедур) визначення рівня навчальних досягнень випускників загальноосвітніх навчальних закладів (ЗНЗ). Головна мета ЗНО – рівний доступ громадян України до вищої освіти, а головне завдання – справедливе оцінювання знань випускників як важлива складова піднесення якості української освіти до європейського та світового рівнів.

Дане оцінювання є:

- ✓ **зовнішнім**, оскільки процедура відбувається на базі спеціально підготовлених пунктів тестування і виключає можливість оцінювання навчальних досягнень учнів у своєму навчальному закладі;
- ✓ **незалежним**, бо здійснюється організацією, яка не залежить і не підпорядкована ЗНЗ і ВНЗ, хоч тісно співпрацює з ними. В Україні такими установами є Український центр оцінювання якості освіти (УЦОЯО) та дев'ять регіональних центрів;
- ✓ **об'єктивним**, оскільки ставить однакові вимоги та забезпечує рівні умови всім учасникам в Україні;
- ✓ **прозорим**, бо забезпечує можливість спостереження з боку громадськості за дотриманням усієї процедури;
- ✓ **шансом** для учнів при вступі до вищого навчального закладу.

Зовнішнє незалежне оцінювання навчальних досягнень як обов'язкова процедура для всіх бажаючих вступити до вищих навчальних закладів в освітній системі України впроваджується з 2008 року. Форма ЗНО – тести. Призначення тесту – оцінити ступінь підготовленості випускників загальноосвітніх навчальних закладів з метою конкурсного відбору для навчання у вищих навчальних закладах. При розробці змістового наповнення тестів враховується необхідність перевірки навчальних досягнень учасників зовнішнього незалежного оцінювання за такими рівнями класифікації когнітивної сфери (таксономія Бенджаміна Блума): знання; розуміння; застосування; аналіз; синтез; оцінювання.

Українська історія запровадження тестових технологій на вступних випробуваннях бере свій початок із 1991-1992 навчального року, коли

Львівський національний університет імені І. Франка, а згодом і Національний університет “Києво-Могилянська академія” вперше провели вступні випробування лише у формі тестів. На базі цього досвіду з ініціативи та підтримки Міжнародного фонду “Відродження” в 1999 році було запроваджено експериментальний проект з “клаптиковою” участю вже кількох університетів. Після Указу Президента України від 4 липня 2005 року “Про невідкладні заходи щодо функціонування та розвитку освіти України” зовнішнє незалежне оцінювання навчальних досягнень почало здійснюватися як загальнодержавна програма з таких предметів: українська мова, математика, історія, основи економіки. У 2006-2007 рр. тестування відбувалося в експериментальному варіанті для бажаючих узяти в ньому участь з української мови, математики, історії (історії України та всесвітньої історії); ЗНО з хімії, біології, фізики проводилося лише для випускників шкіл Харківської області. У 2008 році було проведено зовнішнє оцінювання навчальних досягнень усіх абітурієнтів. Окрім того, 2008 року скоротили до восьми кількість предметів, з яких проходило тестування: зарубіжну літературу, основи економіки, основи правознавства, всесвітню історію виключили. Натомість, випускники змогли скласти тести з іноземних мов: англійської, німецької, іспанської або французької. Випускники змогли обрати для здавання не 3, а до п’яти предметів ЗНО. Тестування з української мови й літератури надалі лишилося обов’язковим для всіх випускників. Починаючи з цього року подання заяв до ВНЗ можна відслідковувати он-лайн. У 2011 році зовнішнє оцінювання для бажаючих стати студентами вищих навчальних закладів проводилося з таких предметів: українська мова та література, історія України, математика, біологія, географія, фізика, хімія, російська мова, одна з іноземних мов (за вибором): англійська, німецька, французька, іспанська. З цих же предметів буде проводитися ЗНО у 2012 році.

1.4.2. Організаційно-технологічне забезпечення ЗНО

Щорічно чинною нормативною базою регулюється проведення ЗНО навчальних досягнень осіб, які виявили бажання вступати до вищих навчальних закладів України в поточному навчальному році. На сайтах Українського та регіональних центрів оцінювання якості освіти розміщується інформація щодо умов реєстрації, особливостей проходження (процедури), організаційно-технологічного забезпечення ЗНО в поточному навчальному

році. На вище зазначених сайтах розміщено програми зовнішнього незалежного оцінювання, затверджені наказом Міністерства освіти і науки, молоді та спорту України, у розділі «Підготовка до ЗНО». У регіональних центрах оцінювання якості освіти реєструються особи, які виявили бажання взяти участь у пробному зовнішньому незалежному оцінюванні 2012 року. Реєстрація завершиться 30 листопада 2011 року (пробне ЗНО відбудеться 24 та 31 березня 2012 року).

Положенням про основні засади організаційно-технологічного забезпечення зовнішнього незалежного оцінювання у поточному році визначається:

- хто може бути учасником ЗНО, їхні права і обов'язки;
- суб'єкти адміністрування ЗНО та їхні повноваження;
- процедура реєстрації учасників ЗНО;
- кадрове забезпечення ЗНО та організація фахової підготовки осіб, залучених для його проведення;
- порядок розробки, тиражування, комплектації аудиторних комплектів тестів, організації їх доставки до пунктів тестування та охорони;
- створення та організації роботи пунктів тестування;
- процедура обробка результатів ЗНО (конфіденційна інформація);
- представлення результатів ЗНО;
- процедура подання та розгляду апеляції на результати тестування;
- організація додаткової сесії тестування;
- права та сфери громадського контролю;
- фінансування ЗНО.

Питання для самоконтролю

1. Охарактеризуйте сутність, завдання і форми поточного оцінювання навчальних досягнень учнів.
2. На конкретному прикладі доведіть, що поточне оцінювання виконує заохочувальну, стимулюючу та діагностичну функції.
3. Поясніть, як за допомогою тематичного оцінювання можна забезпечити систематизацію й узагальнення навчального матеріалу.
4. Визначте особливості рейтингової системи оцінювання.

5. Проаналізуйте як рейтингова система оцінювання сприяє підвищенню навчальної мотивації й відповідальності; сприяє об'єктивності оцінювання.
6. Покажіть залежність вибору методів оцінювання від мети і форми оцінювання навчальних досягнень учнів.
7. Визначте загальнопедагогічні вимоги, що становлять основу оцінювання навчальних досягнень учнів.
8. Проаналізуйте характеристики відповіді учня (осмисленість, цілісність, логічність, обґрунтованість, правильність) та якості знань (повнота, глибина, гнучкість, дієвість, системність, узагальненість, міцність).
9. Співставте поняття «знання», «уміння» і «навички».
10. Охарактеризуйте рівні навчальних досягнень учнів.
11. Визначте загальні критерії оцінювання навчальних досягнень учнів, як вони реалізуються в нормах оцінок.
12. Розгляньте критерії і норми оцінювання з конкретного навчального предмету.
13. Визначте сутність холістичного оцінювання. Наведіть приклади його використання на сучасному етапі.
14. Охарактеризуйте бальну систему оцінювання в єзуїтських школах та школах Німеччини.
15. Яка система оцінювання функціонувала у дореволюційній Росії?
16. До яких негативних наслідків призвела відміна бальної системи оцінювання в період з 1918 по 1935 рр.?
17. У чому виявляються переваги 12-бальної системи оцінювання над 5-бальною.
18. Окресліть основні ідеї критиків і прихильників бальної системи оцінювання.
19. Назвіть основні педагогічні вимоги до оцінювання навчальних досягнень учнів.
20. Яка мета зовнішнього незалежного оцінювання? За рахунок чого забезпечується об'єктивність і прозорість ЗНО?
21. Зовнішнє незалежне оцінювання навчальних досягнень це обов'язкова процедура для отримання атестату про середню освіту чи воно необхідно для всіх бажаних вступити до вищих навчальних закладів в освітній системі України?
22. Визначте основні етапи запровадження тестових технологій на вступних випробуваннях.
23. Які предмети виносяться на ЗНО у 2012р.?
24. Що регулюється Положенням про основні засади організаційно-технологічного забезпечення зовнішнього незалежного оцінювання

Список джерел, використаних у розділі I

1. Вілмут Дж. Використання портфоліо для навчання та оцінювання: Навч.посіб. / За ред. І.Є.Булах, М.Р.Мруга. – К.: Майстер-клас, 2007. – 48 с.
2. Вілмут Дж. Оцінювання для навчання: Навч.посіб. / За ред. І.Є.Булах, М.Р.Мруга. – К.: Майстер-клас, 2007. – 170 с.
3. Вілмут Дж. Оцінювання письмових робіт учнів: Навч.посіб. / За ред. І.Є.Булах, М.Р.Мруга. – К.: Майстер-клас, 2007. – 56 с.
4. Гронлунд Норман Е. Оцінювання студентської успішності: Практ. посібник. – К.: Навчально-методичний центр «Консорціум із удосконалення менеджмент-освіти в Україні», 2005. – 312 с.
5. История педагогики и образования. От зарождения воспитания в первобытном обществе до конца XX века.: Учебное пособие для педагогических учебных заведений /Под ред. А.И.Пискунова. – 2-е изд., испр. и дополн. – М.: ТЦ Сфера, 2005. – 512 с.
6. Наказ Міністерства освіти і науки, молоді та спорту України 13.04.2011 №329 Критерії оцінювання навчальних досягнень учнів (вихованців) у системі загальної середньої освіти [Електронний ресурс]. – режим доступу: http://osvita.ua/legislation/Ser_osv/18438
7. Основи педагогічного оцінювання: Ч. I. Теорія: Навчально-методичні та інформаційно-довідкові матеріали для педагогічних працівників / За заг. ред. Ірини Булах. — К.: Майстер-клас, 2005. — 96 с.
8. Педагогічне оцінювання і тестування. Правила, стандарти, відповідальність. Наукове видання / Я. Я. Болюбаш, І. Є. Булах, М. Р. Мруга, І. В. Філончук. – К.: Майстер-клас, 2007.–272 с.
9. Gronlund N.E. Measurement and evaluation in teaching.- Macmillan Pub. Co., 1965 - 420 pp.
10. Gronlund N.E., Linn R.L. Measurement and evaluation in teaching, Sixth edition. New York: Macmillan, 1990.
11. Linn R.L., Miller M.D. Measurement and assessment in teaching, 9th Edition. Upper Saddle River, NJ: Prentice Hall, 2005.
12. Miller M.D., Linn R.L., Gronlund N.E. Measurement and assessment in Teaching, 10th Edition. Upper Saddle River, NJ: Pearson Education, Inc., 2009.
13. <http://osvita.ua/school/materials/estimation/2419>

II. ВСТУП до ОСВІТНІХ ВИМІРЮВАНЬ: ПОНЯТІЙНИЙ АПАРАТ та СТАТИСТИЧНІ МЕТОДИ

У цьому розділі представлені загальні відомості про вимірювання в освіті, яке передбачає вимірювання досягнень учнів у навчальних закладах різного рівня та студентів вищих навчальних закладів, а також вимірювання знань, умінь, компетенцій кожної особи, яка бере участь у тестуванні при складанні шкільного іспиту, на іспиті в університеті, при участі у вступному тестовому випробуванні, при прийомі на роботу, при моніторинговому дослідженні тощо. При цьому більше уваги приділено більш вузькому поняттю - педагогічному вимірюванню.

Представлені необхідні відомості про типи шкал для відображення результатів вимірювання. Приділена увага статистичним методам оцінки результатів вибіркового спостереження у педагогіці та психології, наведено загальні методи вивчення зв'язку соціально-педагогічних явищ і процесів.

Описані нормативно-орієнтовані та критеріально-орієнтовані вимірювання в освіті, а також відповідні типи тестів. Описані основні характеристики нормо-орієнтованого та критеріально-орієнтованого підходів в освіті.

Висвітлено такі важливі педагогічні проблеми, як класифікація навчальних цілей та таксономія Блума.

2.1. Основні складові вимірювання в освіті

2.1.1. Поняття вимірювання в освіті та латентні змінні

З розвитком педагогіки, психології та соціології виникла потреба у введенні не тільки якісних, але і кількісних оцінок для величин, що відрізняються за ступенем прояву тієї або іншої властивості. Якісні оцінки є менш точними у порівнянні з кількісними, що пояснюється використанням різних способів та інструментів вимірювання. Наприклад, поняття «обізнаний», «підготовлений» тощо, які диференціюють учнів за рівнем знань, визначаються суб'єктивно вчителем або групою вчителів. Часто якісні оцінки виражають за допомогою чисел, які вибирають на основі експертних

угод. Ці числа можуть трактуватися у різний спосіб. Так, у традиційному педагогічному контролі у кожного вчителя є свої уявлення про те, за що потрібно ставити оцінки високого, достатнього, середнього та початкового рівнів.

Неоднозначність вимірювання в освіті посилюється латентним (прихованим, таким, що виключає можливість безпосереднього вимірювання) характером вимірюваних змінних.

Латентною називається змінна, значення якої в ході спостереження не доступні для безпосереднього вимірювання, а лише можуть бути оцінені у відповідності із висунутою гіпотезою за значеннями явних змінних.

Через латентність оцінюванню піддаються не самі характеристики навченості, а їх емпіричні референти – спостережувані ознаки вимірюваних характеристик. Вибір останніх відбувається інтуїтивно, тому їх відповідність латентним характеристикам потребує доведення на основі експертного і статистичного аналізу емпіричних результатів вимірювання.

Вимірювання — це процедура надання числових значень деяким характеристикам об'єктів у відповідності з певними правилами (найбільш поширене визначення, введене у 1946 р. американським психологом С.Стівенсом).

Вимірювання – це процедура надання числових значень (які зазвичай називаються тестовими балами) певним властивостям, характеристикам тестованих, що правильно відображає розташування тестованих на шкалі, в залежності від виявлення в них вимірюваної властивості (за Лордом і Новіком).

2.1.2. Компоненти та інструменти процесу вимірювань в освіті

Процес педагогічних вимірювань включає такі **компоненти**:

- вибір предмету вимірювання (латентних характеристик об'єктів) та їх числа;
- вибір емпіричних референтів (спостережуваних характеристик об'єктів);
- вибір вимірювальних процедур;
- конструювання і використання інструментів вимірювання;
- вибір шкали (якщо вимірювана змінна одна) або шкал (якщо вимірюють більше ніж одну змінну при багатовимірних вимірюваннях);

- побудова відображення результатів вимірювання на шкалу (шкали у разі багатовимірних вимірювань) за визначеними процедурами і правилами;
- обробка, аналіз та інтерпретація результатів вимірювання.

Інструмент вимірювання включає два компоненти.

Перший компонент — сам вимірювальний пристрій, роль якого в педагогічних вимірюваннях найчастіше, але не завжди, виконує тест.

У самому узагальненому вигляді під **тестом** можна розуміти сукупність контрольних завдань у стандартизованій формі, що мають необхідні системоутворюючі статистичні характеристики і забезпечують обґрунтовані оцінки концептуально виділеній змінній (змінним) вимірювання з високою об'єктивністю.

Другий компонент вимірювального інструменту — заздалегідь підготовлена **шкала**, яка призначена для фіксації результатів вимірювання і на якій відкладаються значення (кількісні або якісні) вимірюваної змінної.

У процесі впорядкування оцінок кожному елементу сукупності спостережуваних емпіричних даних ставиться у відповідність певний бал, що встановлює положення спостережуваного елемента на шкалі, де можна розміщувати сирі (первинні) бали (сума балів за окремі завдання тесту) або похідні бали, що отримуються в результаті перетворення первинних балів для підвищення зіставності та зручності інтерпретації результатів тестованих.

Шкала з відкладеними оцінками змінної є метою вимірювання. При вимірюваннях з високою надійністю і валідністю вона адекватно відображає вимірювані характеристики і представляє їх без істотних спотворень.

2.2. Класифікація шкал: основні кількісні та якісні шкали

З математичної точки зору, процес вимірювання є відображення стану вимірюваного об'єкту на деяку множину дійсних чисел (або на деяку множину точок числової осі), яка називається **шкалою**. Найважливішою характеристикою шкали є набір тих співвідношень між її елементами, які мають зміст та розумне тлумачення в рамках шкали.

Шкали бувають **кількісні**, за якими відмінності у результатах вимірювань визначаються кількісною величиною, та **якісні**, за якими відмінності у результатах вимірювань визначаються якісною величиною. Нижче розглянемо основні види шкал: три кількісні шкали (порядкову, метричну та нормовану) та одну якісну (номінальну).

1. **Порядкова шкала** - це така кількісна шкала, на якій результати вимірювань можна тільки порівнювати між собою, тобто для будь-яких відміток b_1 і b_2 змістовий сенс мають тільки співвідношення типу

$$b_1 < b_2 \text{ або } b_1 = b_2.$$

Прикладом може служити система шкільних оцінок, що виставляються учням залежно від їх успіхів в навчанні, оскільки із співвідношення відміток $b_1 < b_2$ для учнів A_1 і A_2 можна лише сказати, що A_1 вчиться гірше A_2 . Якщо ж, наприклад, $b_2 - b_1 = 1$, то твердження "успіхи A_2 на 1 вище, ніж успіхи A_1 " не пояснює, яка відмінність між учнями і, по суті, позбавлена значення.

Прикладами порядкової шкали є шкали первинних балів (в абсолютному або відносному виразі) як для учасників тестування, так і для тестових завдань. Максимум, що можна зробити в рамках цих шкал, це упорядкувати учасників тестування або тестові завдання в порядку зростання (або зменшення) оцінок відповідних латентних параметрів.

Основними статистиками порядкових шкал є медіана, квантіли і рангова кореляція.

2. **Інтервальна (метрична) шкала** – така кількісна шкала, для будь-якої пари відміток b_1 і b_2 змістовний сенс мають не тільки співвідношення типу $b_1 < b_2$ або $b_1 = b_2$, але і різниця $b_1 - b_2$, при цьому величина $|b_1 - b_2|$ називається **метрикою** і трактується як відстань між двома елементами метричної множини, виражена в певних одиницях і має осмислене тлумачення.

У метричній шкалі відсутня нульова відмітка, тобто відсутній початок координат, в ній неможливо визначити місцезнаходження об'єкту в деякій єдиній системі координат. Метричні шкали застосовують для фіксації взаємного положення вимірюваних об'єктів один відносно одного.

З математичної точки зору вказана ситуація означає, що на множині визначено поняття метрика (відстань), але не визначено поняття норма (довжина).

Прикладом метричної шкали є шкала єдиного параметру логістичної моделі Раша $\theta - \beta$, який вимірюється в логітах (див. Розділ V). Тут розглядається різниця рівня підготовки учасника тестування та рівня складності тесту, тобто відстань між цими величинами, а не власне величини цих параметрів.

3. **Нормованою шкалою** називається така метрична (інтервальна) шкала, в якій визначений початок відліку. В такій шкалі визначено не тільки поняття метрики, але і поняття норми, що дозволяє вимірювати "довжини" (тобто певне місцеположення відносно нуля, початку відліку).

У нормованій шкалі можна визначати не тільки різниці латентних параметрів $\theta - \beta$, але кожен величину θ або β окремо (див. Розділ V). Однією із задач сучасної теорії параметризації та моделювання тестів є побудова такої шкали, що дозволили би подолати основний недолік класичної теорії - залежність оцінок одного учасника тестування від використаного тесту і контингенту всіх учасників тестування або певної групи учасників

4. **Номінальна шкала** – це якісна шкала, заснована на якісних змінних, непіддатливих кількісному вимірюванню.

Прикладом може бути стать учасників тестування, належність певній країні, тощо. Очевидно, що навіть якщо у відповідність до кожної якості поставлено число, то ніяких змістовних співвідношень окрім $a = b$ або $a \neq b$, між такими числами немає. При цьому вибір чисел замість реальних імен або інших способів ідентифікації не є обов'язковим.

Якщо ознак, що розрізняють об'єкти або події, тільки два, то номінальна шкала називається *дихотомною*. Прикладом можуть виступати елементи матриці відповідей учасників тестування на завдання тесту: вірне виконання завдання ("так") позначається одиницею, помилкове ("ні") - нулем.

Для характеристики номінальних даних часто використовуються такі (дескриптивні) статистики, як пропорція і процентне відношення.

Використовування тієї або іншої шкали з перерахованих чотирьох накладає відбиток і на застосовність тих або інших методів математичної

обробки, якій звичайно піддаються початкові дані. Змістова інтерпретація результатів математичної обробки даних тестування може бути дана лише в тому випадку, якщо методи цієї обробки адекватні тим шкалам, до яких віднесена початкова інформація.

2.3. Два типи вимірювання та два типи тестів: нормативно-орієнтовані та критеріально-орієнтовані підходи в освіті

2.3.1. Класифікація типів вимірювання в освіті за метою

Мета **нормативно-орієнтованого вимірювання** – диференціація тестованих за результатами виконання тесту. Для коректної інтерпретації результатів тестування бал кожного тестованого необхідно порівнювати з нормами виконання тесту.

Норми – це сукупність показників, що відображають результати виконання тесту певною вибіркою випробовуваних – релевантною нормативною групою, що репрезентативно представляє генеральну сукупність тестованих.

Тестових норм, за допомогою яких можна інтерпретувати результати всіх учнів за будь-якими тестами, не існує. Область застосування будь-якої норми обмежується даним тестом і конкретною сукупністю тестованих, тому норми не є абсолютними і постійними.

До норм пред'являють наступні вимоги:

- норми повинні бути диференційованими. Наприклад тести для загальноосвітніх і профільних шкіл необхідно стандартизувати на різних вибірках;
- норми повинні відображати реальний контингент і актуальні вимоги до якості навчальних досягнень, ґрунтуючись на сучасній ситуації в освіті;
- норми повинні бути репрезентативними, тому вони завжди встановлюються емпірично відповідно до результатів тестування вибірки стандартизації (державною — для ЗНО, обласною — для атестації шкіл, внутрішньошкільною — для атестації учнів в школі).

При нормативно-орієнтованому вимірюванні змістовим компонентом, що оцінюється (доменом оцінювання), виступає не лише зміст тієї чи іншої програми з дисципліни, а й зміст, яким забезпечуються певні кваліфікаційні вимоги або вимоги щодо спроможності набуття певного освітнього чи кваліфікаційного рівня, наприклад, здатності навчатися у вищому навчальному закладі або отримати певну професійну кваліфікацію. Домен оцінювання великий, вимірюється в “ширину”.

Під час нормативно-орієнтованого вимірювання тест, як правило, складається з тестових завдань середньої складності.

Результат тестування – відносний (порівняння із norm-group).

Сфера використання: відбір, ранжирування, порівняння тестованих між собою, оцінювання програм, порівняння освітніх інституцій.

Мета **критеріально-орієнтованого вимірювання** – визначити рівень досягнень особи, що тестується, незалежно від результатів інших. Отриманий кожним тестованим бал порівнюється із стандартами виконання – критеріями, встановленими експертним шляхом.

При критеріально-орієнтованому підході за наслідками тестування можна:

- виявити засвоєні і не засвоєні знання, уміння і навички та побудувати індивідуальну освітню траєкторію кожного учня;
- ранжувати тестованих за відсотком виконання і побудувати рейтингові шкали;
- розбити тестованих на дві групи за допомогою одного критеріального балу або на декілька груп за допомогою декількох критеріальних балів.

При критеріально-орієнтованому вимірюванні змістовий компонент вибирається не широкий, а з певним ступенем деталізації програмного матеріалу, що оцінюється. Домен оцінювання вузький, вимірюється в “глибину”.

Складність тестових завдань відповідає складності оцінюваного домену.

Результат – абсолютний, порівнюється з критерієм, стандартом.

Сфера використання: сертифікація, ліцензування, оцінка мінімальної компетентності.

У таблиці представлено основні відомості про нормативно-орієнтовані та критеріально-орієнтовані підходи до вимірювання в освіті

Характеристики	Нормативно-орієнтований підхід	Критеріально-орієнтований підхід
Середнє число тестованих, що правильно виконали майже всі завдання тесту	5-10%	80-90%
Область для порівняння результатів тестованих	Результати інших тестованих	Змістова область або сукупність видів навчальної діяльності
Діапазон охоплення цілей перевірки	Широкий, охоплює багато цілей і видів навчальної діяльності	Вузкий, охоплює декілька цілей контролю
Репрезентативність охоплення змісту предмету	Помірна, фрагментарна (включаються не всі розділи)	Велика (включається весь матеріал, який можна операціоналізувати)
Варіативність балів	Висока	Низька
Добір завдань за складністю	Основна частина завдань має складність 40-60%	Основна частина завдань має складність 80-90%

При педагогічних вимірюваннях інтерпретація балів тестованих може мати різний характер у залежності від того, в який спосіб порівнюються оцінки. У відповідності з підходом до інтерпретації результатів тестування виділяють нормативно-орієнтовані та критеріально-орієнтовані тести.

2.3.2. Класифікація тестів за метою вимірювання: нормативно-орієнтовані та критеріально-орієнтовані тести

У вітчизняній та зарубіжній літературі немає єдиного підходу до класифікації видів педагогічних тестів. Визначилися два підходи в тестології до класифікації тестів – тести, орієнтовані на критерій (критеріально-орієнтовані) і тести, орієнтовані на норму (нормативно-орієнтовані).

Нормативно-орієнтований тест (norm-referenced test) дозволяє ранжувати випробовуваних за рівнем знань. Такий тест дозволяє порівнювати учбові досягнення випробовуваних один з одним. Метою нормативно-орієнтованого тесту є упорядкування випробовуваних за рівнем їх підготовленості.

Критеріально-орієнтований тест (criterion-referenced test) дозволяє виявити ступінь засвоєння випробовуваним певного розділу в заданій предметній області. Ці тести появилися в 60-их роках минулого століття, тобто значно пізніше нормативно-орієнтованих. Критеріально-орієнтовані

тести в свою чергу діляться на domain-referenced test (орієнтовані на предметну область) і mastery-tests (кваліфікаційні тести). Метою критеріально-орієнтованого тесту є визначення – чи знає випробовуваний стандартний учбовий матеріал. Критеріально-орієнтований педагогічний тест є системою завдань, яка дозволяє виміряти рівень учбових досягнень відносно повного обсягу знань, умінь і навичок, що повинні бути засвоєні.

Зовнішньо обидва типи тестів мають багато спільного, однак вони мають наступні відмінності.

1. *Мета створення тесту.* Нормативно-орієнтовані тести створюються спеціально для того, щоб порівняти випробовуваних в тій області змісту, для якої призначений тест. Ці тести можна використовувати, наприклад, для відбору абітурієнтів у вузи, тому що абітурієнтів потрібно ранжувати, щоб вибрати найкращих.
2. *Рівень деталізації області змісту.* Розробка тесту починається зі створення його специфікації і ці специфікації для обох типів тестів дуже відрізняються. Специфікації критеріально-орієнтованих тестів значно детальніше описують елементи області змісту, тому що це дозволить адекватно інтерпретувати результати тестування. Для нормативно-орієнтованих тестів рівень деталізації значно нижчий, бо значно важливіше отримати варіативні тестові завдання.
3. *Статистична обробка результатів тестування.* Шкальовані бали нормативно-орієнтованого тестування базуються на тестових нормах, отриманих на «вибірках стандартизації». При критеріально-орієнтованому тестуванні тестові бали не пов'язані з якоюсь нормативною групою випробовуваних.
4. *Аналіз і відбір тестових завдань.* Для нормативно-орієнтованих тестів велике значення мають статистичні характеристики – рівень складності завдання, його диференціююча здатність. Якщо завдання має середній рівень складності та високу диференціюючу здатність, то воно вважається хорошим для нормативно-орієнтованого тесту. Ці статистичні характеристики не мають великого значення для критеріально-орієнтованого тесту. Головним критерієм для включення завдання в тест є відповідність специфіці та елементу області змісту.
5. *Розподіл випробовуваних за індивідуальними балами* має різний характер для обох видів тестів. Для нормативно-орієнтованого тесту

крива розподілу симетрична і близька до гаусової кривої. У випадку критеріально-орієнтованого тесту ця крива не симетрична і зазвичай зсунута в область високих індивідуальних балів.

6. Інтерпретація результатів тестування. Результати нормативно-орієнтованого тестування інтерпретуються на основі статистично обґрунтованих тестових норм. При цьому є можливість визначити положення випробовуваного відносно нормативної групи. Результати критеріально-орієнтованого тестування інтерпретуються з точки зору повноти засвоєння області змісту, згідно детальної специфікації тесту.

Критеріально-орієнтовані тести потрібні для атестації випробовуваних в певній області змісту. Такі тести використовуються у підсумковому тестуванні, наприклад, по закінченню навчання в середній школі. Тут важливо в'яснити – чи засвоєна на належному рівні шкільна програма. Питання ранжування тут не мають значення.

Якщо критеріально-орієнтовані тести використати як нормативно-орієнтовані, то з причини малої дисперсії тестових результатів ці результати будуть мати низьку надійність. Вірно і обернене – застосування нормативно-орієнтованих тестів у критеріально-орієнтованому тестуванні також дасть малонадійні результати. Це пов'язано з сильною варіацією тестових завдань за трудністю в нормативно-орієнтованому тесті.

2.3.3. Інші типи класифікації тестів.

Крім наведеного підходу тести можна класифікувати за іншими ознаками:

1. *За процедурою створення:* стандартизовані та не стандартизовані.
2. *По засобах пред'явлення:* бланкові, предметні, апаратурні, практичні, комп'ютерні.
3. *За направленістю:* тести інтелекту, особистісні тести, тести досягнень.
4. *За характером дії:* вербальні, невербальні.
5. *За головною орієнтацією:* тести швидкості, тести потужності або результативності, змішані тести.
6. *За ступенем однорідності задач:* гомогенні, гетерогенні.
7. *За цілями використання (тільки для тестів у системі освіти):* попередній визначаючий тест, формуючий тест, діагностичний тест, підсумковий тест.

2.4. Цілі навчання та таксономія Блума

2.4.1. Класифікація цілей навчання

Чітке формулювання цілей, які виражені через результати діяльності, піддається надійній і об'єктивній оцінці. Чіткому визначенню цілей навчання сприяє їх класифікація (таксономія). Поняття "таксономія" взято з біології (у перекладі з грецької означає "розміщення у порядку", "закон"). Американські педагоги та психологи створювали таксономію навчальних цілей впродовж двадцяти років, починаючи з кінця 50-х рр. минулого століття. Системи класифікації цілей навчання вимагали розробки цілей у трьох сферах (доменах): пізнавальній, емоційній та психомоторній.

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

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

Цілі навчання *психомоторної сфери* становлять третю групу. Вони включають ті чи інші види моторної (рухливої) маніпулятивної діяльності нервово-м'язової координації. Це навички письма, мовні навички, фізичні та трудові навички.

2.4.2. Таксономія Блума

Цілі навчання у когнітивній сфері можуть бути виражені через такі елементи засвоєння (їх ще називають елементами таксономії Блума): знання, розуміння, застосування, аналіз, синтез і оцінка.

1. **Знання.** Ця категорія означає ~~застосування~~ застосування та відтворення матеріалу, який вивчається. Йдеться про ~~застосування~~ застосування і відтворення

термінів, конкретних фактів, методів і процедур, основних понять, правил, принципів, цілісних теорій.

Інтелектуальні вміння: впізнавання фактів, термінів, умов, понять, визначень, принципів.

Ключові слова для використання в основі тестових завдань: визначте, перерахуйте, ідентифікуйте, назвіть. Хто? Коли? Де? Який?

2. Розуміння. Показником розуміння значення вивченого може бути здатність людини встановлювати зв'язок одного матеріалу з іншим, перетворювати його із одної форми вираження в іншу, переводити його з однієї "мови" на іншу (наприклад, зі словесної у графічну, математичну і навпаки). Показником розуміння може також бути інтерпретація матеріалу учнем (пояснення, короткий виклад), прогнозування майбутніх наслідків, що впливають із наявних даних.

Інтелектуальні вміння: пояснення, інтерпретація вивченого навчального матеріалу.

Ключові слова для використання в основі тестових завдань: поясніть, передбачте, інтерпретуйте, виведіть, додайте, перетворіть, перекладіть, наведіть приклад, обчисліть.

3. Застосування. Цей елемент засвоєння означає вміння використовувати вивчений матеріал у конкретних умовах і нових ситуаціях. Сюди входить застосування правил, методів, вміння розбивати матеріал на складові поняття, законів, принципів, теорій.

Інтелектуальні вміння: використання понять або принципів, щоб розв'язати проблему (задачу) в типових і конкретних ситуаціях.

Ключові слова для використання в основі тестових завдань: застосуйте, розв'яжіть, покажіть, використайте, змініть, продемонструйте, обчисліть.

4. Аналіз. До цієї категорії належить виділення частин цілого, виявлення взаємозв'язку між ними, осмислення принципів організації цілого. Навчальні результати характеризуються осмисленням не тільки змісту навчального матеріалу, а і його внутрішньої структури.

Інтелектуальні вміння: деформування системи на складові для виявлення відношень та ієрархії, організація зв'язків між частинами.

Ключові слова для використання в основі тестових завдань: диференціюйте, порівняйте, відрізнити, поєднайте.

Учень, який добре оволодів цією категорією навчальних цілей, бачить помилки й огріхи в логіці міркувань, бачить різницю між фактами і наслідками, оцінює значимість даних.

5. **Синтез.** Ця категорія означає вміння комбінувати елементи, щоб одержати ціле з новою системною властивістю. Таким новим продуктом може бути повідомлення, план дій, нова схема тощо.

Інтелектуальні вміння: створення нового зі складових (елементів).

Ключові слова для використання в основі тестових завдань: спроектуйте, сконструюйте, розробіть, сформулюйте, уявіть, створіть, змініть так, щоб... , напишіть есе.

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

Інтелектуальні вміння: формування судження, яке ґрунтується на заздалегідь визначених критеріях.

Ключові слова для використання в основі тестових завдань: Спроектуйте, сконструюйте, розробіть, сформулюйте, уявіть, створіть, змініть так, щоб.... Що було б краще...?

2.5. Випадкові величини та їх числові характеристики

Процес вимірювання в освіті передбачає вміння аналізувати випадкові чинники, оцінювати шанси, прогнозувати результати, висувати гіпотези, приймати рішення в умовах, які мають імовірнісний характер. Закономірності, властиві кожному явищу, проявляють себе через сукупність випадковостей. Для багатьох явищ вплив випадку є настільки суттєвим, що дослідження їх неможливе без вивчення і кількісної оцінки такого впливу. Розглянемо деякі математичні моделі явищ, які враховують вплив випадку на прикладі випадкових величин та законів їх розподілу.

2.5.1 Випадкова величина. Закон розподілу випадкової величини

Випадковою величиною називається числова функція, визначена у просторі елементарних подій. Позначають випадкові величини великими

латинськими або малими грецькими літерами $X=X(u)$, $Y=Y(u)$, $Z=Z(u)$, $\xi(\omega)$, $\eta(\omega)$, $\zeta(\omega)$ а їхні значення – відповідними малими літерами. Факт, що випадкова величина X набуває значення x , є випадковою подією, позначається ($X=x$).

Дискретною називають випадкову величину, можливими значеннями якої є окремі ізольовані числа, які ця величина приймає з певними ймовірностями. Іншими словами, можливі значення дискретної випадкової величини можна перенумерувати. Число можливих значень дискретної випадкової величини може бути скінченним або нескінченним (але зліченим).

Якщо значення, які може приймати дана випадкова величина X , заповнюють скінчений чи нескінчений проміжок (a, b) числової осі Ox , то випадкова величина називається **неперервною**.

Співвідношення, яке дає можливість встановити тим чи іншим способом зв'язок між можливими значеннями випадкової величини і їх відповідними ймовірностями, називають **законом розподілу** випадкової величини.

Закон розподілу дискретної випадкової величини X може бути заданий **рядом розподілу**: у вигляді таблиці, перша строчка якої містить можливі значення x_i , а друга – ймовірності p_i :

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

причому $p_1+p_2+\dots+p_n=1$.

Якщо множина можливих значень X нескінчена (злічена), то ряд $p_1+p_2+\dots+p_n+\dots$ збігається і його сума дорівнює одиниці.

Закон розподілу дискретної випадкової величини X можна зобразити графічно, для чого в прямокутній системі координат будують точки $M_1(x_1; p_1)$, $M_2(x_2; p_2)$, ..., $M_n(x_n; p_n)$, де x_i – можливі значення X , p_i – відповідні ймовірності, і з'єднують їх відрізками прямих. Отриману фігуру (ламану) називають **многокутником розподілу (полігоном)**.

Закон розподілу неперервної випадкової величини X зручно задавати **аналітично** за допомогою **функції щільності ймовірності** $p(x)$. Графік функції $p(x)$ називається **кривою розподілу**.

Функція $F(x)=P(X<x)$ називається **функцією розподілу ймовірності** випадкової величини X . Функція $F(x)$ існує як для дискретних, так і для неперервних випадкових величин. Якщо $p(x)$ – функція щільності розподілу ймовірності неперервної випадкової величини X , то

$$F(x) = \int_{-\infty}^x p(x)dx.$$

З цієї рівності випливає, що $p(x)=F'(x)$.

Використовуючи поняття функції розподілу, можна дати інше означення неперервної випадкової величини. Випадкова величина називається **неперервною**, якщо її інтегральна функція розподілу $F(x)$ неперервна.

2.5.2. Диференціальна та інтегральна функції розподілу випадкової величини та їх властивості

Функцію щільності імовірності $p(x)$ називають **диференціальною функцією розподілу ймовірності**, а функцію розподілу $F(x)$ імовірності випадкової величини X – **інтегральною функцією розподілу ймовірності**.

Функція щільності імовірності $p(x)$ має такі властивості:

1. $p(x) \geq 0$;

2. $\int_{-\infty}^{\infty} p(x)dx = 1$ (умова нормування щільності);

3. Ймовірність $P(a < X < b)$ того, що значення, яке приймає випадкова величина X , попадає в проміжок (a, b) , визначається рівністю

$$P(a < X < b) = \int_a^b p(x)dx.$$

Геометрично ймовірність попадання випадкової величини в проміжок (a, b) дорівнює площі відповідної криволінійної трапеції, обмеженої кривою розподілу, віссю Ox і прямими $x=a$, $x=b$.

Функція розподілу імовірності $F(x)$ має такі властивості:

1. $F(x)$ – неспадна функція;

2. $F(-\infty)=0$, $F(+\infty)=1$;

3. Ймовірність $P(a < X < b)$ того, що значення, яке приймає випадкова величина X , попадає в проміжок (a, b) , визначається рівністю

$$P(a < X < b) = F(b) - F(a).$$

2.5.3. Числові характеристики центральної тенденції для випадкових величин

Для дослідження випадкових величин використовують їх числові характеристики, зокрема, для визначення центральної тенденції обчислюють математичне сподівання, моду і медіану.

Під **математичним сподіванням** випадкової величини розуміють її середнє значення. Для дискретної випадкової величини математичне сподівання визначається як сума добутків усіх можливих її значень на ймовірності цих значень. Якщо випадкова величина X характеризується наведеним вище рядом розподілу, то математичне сподівання $M(X)$ обчислюється за формулою

$$M(X) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n \quad \text{або} \quad M(X) = \sum_{i=1}^n x_i p_i.$$

Якщо $n = \infty$, то $M(X) = \sum_{i=1}^{\infty} p_i x_i$ (за умови, що ряд абсолютно збігається).

Нехай $p(x)$ – щільність ймовірності неперервної випадкової величини X .

Тоді її математичне сподівання визначається рівністю
$$M(X) = \int_{-\infty}^{+\infty} x p(x) dx$$

(за умови, що інтеграл збігається абсолютно).

Точка осі Ox , що має абсцису, рівну $M(X)$, називається **центром розподілу** цієї випадкової величини. В околі цієї точки «розсіяні» всі її значення x_1, x_2, \dots, x_n .

Імовірнісний зміст математичного сподівання: математичне сподівання випадкової величини X наближено дорівнює середньому арифметичному зваженому її спостережуваних значень.

Властивості математичного сподівання.

1. Якщо C – стала, то $M(C) = C$.

Сталу C можна розглядати, як випадкову величину, яка набуває лише одного значення C з ймовірністю одиниця. Тому $M(C) = C \cdot 1 = C$.

2. Якщо C – стала, то $M(CX) = CMX$.

3. Математичне сподівання суми випадкових величин X і Y дорівнює сумі математичних сподівань цих величин

$$M(X+Y)=MX+MY.$$

Означення. Випадкові величини X і Y , які набувають відповідно значень

$$\{x_1, x_2, \dots, x_k\}, \{y_1, y_2, \dots, y_m\},$$

називаються незалежними, якщо

$$P\{X=x_i, Y=y_j\} = P\{X=x_i\} \cdot P\{Y=y_j\}, \quad i=1, \dots, k, \quad j=1, \dots, m.$$

4. Математичне сподівання добутку незалежних випадкових величин X і Y дорівнює добутку математичних сподівань цих величин:

$$M(X \cdot Y) = MX \cdot MY.$$

Модю дискретної випадкової величини X називають її найбільш імовірнісне значення. **Модю** неперервної випадкової величини X називається те її значення, при якому щільність розподілу максимальна. Позначають символом M_0 .

Медіаною неперервної випадкової величини X називається таке її значення μ , для якого однаково ймовірно виявиться випадкова величина менше чи більше μ , тобто $P(X < \mu) = P(X > \mu) = 0,5$. Позначають $M_e = \mu$.

Геометрично мода є абсцисою точки кривої (полігону) розподілу, ордината якої максимальна. Вертикальна пряма, проведена в точці з абсцисою $x = \mu$, ділить навпіл площу, обмежену кривою розподілу.

Якщо пряма $x = a$ є віссю симетрії кривої розподілу $y = f(x)$, то

$$M_0 = M_e = M(X) = a.$$

2.5.4. Числові характеристики варіації для випадкових величин

Дисперсією випадкової величини X називають математичне сподівання квадрата відхилення випадкової величини від її математичного сподівання

$$D(X) = M[X - M(X)]^2.$$

Якщо ввести позначення $M(X) = m$, то формули для обчислення дисперсії дискретної випадкової величини X запишуться у вигляді

$$\left[\begin{array}{l} D(X) = \sum_{i=1}^n p_i (x_i - m)^2, \\ D(X) = \sum_{i=1}^{\infty} p_i (x_i - m)^2, \quad (n \rightarrow \infty) \end{array} \right.$$

а для неперервної випадкової величини X дисперсія обчислюється так

$$D(X) = \int_{-\infty}^{+\infty} (x - m)^2 f(x) dx.$$

Середнім квадратичним відхиленням випадкової величини X називається величина, яка обчислюється за формулою

$$\sigma_x = \sqrt{D(X)}.$$

Середнє квадратичне відхилення – це міра розкиду значень випадкової величини навколо її математичного сподівання. Вона має ту ж розмірність, що й випадкова величина X .

Властивості дисперсії.

1. $D(X) = MX^2 - (MX)^2$.
2. Якщо C – стала, то $D(C) = 0$.
3. Якщо C – стала, то $D(CX) = C^2 \cdot DX$.
4. Дисперсія суми двох незалежних випадкових величин X і Y дорівнює сумі їх дисперсій $D(X+Y) = DX + DY$.

Наслідки:

1. $D(X+C) = DX$.
2. $D(X-Y) = DX + DY$.

Початковим моментом s -го порядку дискретної випадкової величини X , заданої рядом розподілу (таблицею) називається сума ряду

$$\alpha_s = x_1^s \cdot p_1 + x_2^s \cdot p_2 + \dots + x_n^s \cdot p_n + \dots$$

Для неперервної випадкової величини X зі щільністю розподілу $p(x)$ початковим моментом s -го порядку називається інтеграл

$$\alpha_s = \int_{-\infty}^{+\infty} x^s p(x) dx.$$

Початковий момент l -го порядку випадкової величини X дорівнює математичному сподіванню цієї випадкової величини.

Центральним моментом s -го порядку називається математичне сподівання від $(X - M(X))^s$

$$\mu_s = M(X - M(X))^s \quad (s = 1, 2, 3, \dots).$$

Для дискретної випадкової величини $\mu_s = \sum_{i=1}^n (x_i - M(X))^s p_i$;

для неперервної випадкової величини $\mu_s = \int_{-\infty}^{\infty} (x - M(X))^s p(x) dx$.

При $s = 1$, $\mu_1 = 0$, при $s=2$, $\mu_2 = D(X)$.

Центральні і початкові моменти I, II, III, і IV порядків пов'язані співвідношеннями

$$\begin{aligned} \mu_1 &= 0, \mu_2 = \alpha_2 - \alpha_1^2, \\ \mu_3 &= \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3, \mu_4 = \alpha_4 - 4\alpha_1\alpha_2 + 6\alpha_1^2\alpha_2 - 3\alpha_1^4. \end{aligned}$$

Якщо розподіл симетричний відносно математичного сподівання, то всі центральні моменти непарного порядку рівні нулю, тобто $\mu_1 = \mu_3 = \dots = 0$.

2.5.5. Числові характеристики форми для випадкових величин

Третій центральний момент характеризує асиметрію закону розподілу випадкової величини. Якщо $\mu_3 = 0$, то випадкова величина X симетрично розподілена відносно $M(X)$. Оскільки μ_3 має розмірність випадкової величини в кубі, то вводять безрозмірну величину – коефіцієнт асиметрії

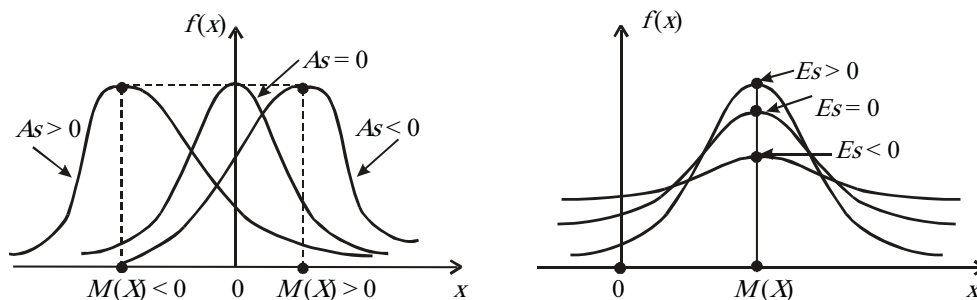
$$As = \frac{\mu_3}{\sigma^3}.$$

Центральний момент четвертого порядку використовується для визначення ексцесу, що характеризує вершину розподілу. Ексцес обчислюється за формулою

$$Es = \frac{\mu_4}{\sigma^4} - 3.$$

Для нормального закону розподілу $E_s = 0$ (крива Гауса), якщо $E_s > 0$, то розподіл гостровершинний, Якщо $E_s < 0$, то розподіл плосковершинний.

При різних значеннях A_s , E_s графіки $f(x)$ мають вигляд.



2.6. Основні розподіли дискретних випадкових величин

2.6.1. Рівномірний закон розподілу ймовірностей

Дискретна випадкова величина X має рівномірний закон розподілу, якщо вона набуває значень $1, 2, \dots, n$ з однаковими ймовірностями.

Ряд розподілу має вигляд:

$X = x_k = k$	1	2	3	...	N
$P(X = k)$	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$...	$\frac{1}{n}$

Числові характеристики, пов'язані з параметром розподілу n ,

$$M(X) = \frac{n+1}{2}, \quad D(X) = \frac{n^2-1}{12}.$$

2.6.2. Гіпергеометричний закон розподілу ймовірностей

Дискретна випадкова величина X має гіпергеометричний закон розподілу, якщо вона набуває значень $0, 1, \dots, k$ з ймовірностями

$$P(X = m) = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n}, \quad m = \overline{0, k}.$$

Параметри розподілу N, M, n – натуральні числа, причому $M < N$, $n < N$, $k = \min(M, n)$.

Гіпергеометричний закон розподілу ймовірностей описує такі випадки: нехай задано деяку множину однотипних елементів, число яких дорівнює N ; з них M елементів мають, наприклад, ознаку A , а решта $N - M$

елементів – ознаку В; коли із цієї множини навмання беруть n елементів, то число елементів k з ознакою А (або В), що трапляється серед n навмання взятих елементів, буде дискретною випадковою величиною з гіпергеометричним законом розподілу.

Числові характеристики цього закону обчислюються за формулами:

$$M(X) = n \cdot \frac{N}{M}, \quad D(X) = \frac{nM(N-M)(N-n)}{N^2(N-1)}.$$

2.6.3. Геометричний закон розподілу ймовірностей

Дискретна випадкова величина X розподілена за геометричним законом з параметром p , якщо вона набуває значень з множини натуральних чисел з ймовірностями

$$P(X = m) = p(1 - p)^{m-1} = pq^{m-1}, \quad m = 1, 2, \dots, n, \dots$$

Ряд розподілу має вигляд:

X	1	2	3	4	...	n	...
P	p	pq	pq^2	pq^3	...	pq^{n-1}	...

За геометричним законом розподілена, наприклад, випадкова величина, яка дорівнює числу проведених випробувань Бернуллі до настання першого «успіху».

Числові характеристики розподілу $M(X) = \frac{1}{p}, \quad D(X) = \frac{q}{p^2}.$

2.6.4. Біноміальний закон розподілу ймовірностей

Дискретна випадкова величина X має біноміальний закон розподілу з параметрами n та $0 < p < 1$, якщо ймовірність її можливих значень обчислюється за формулою Бернуллі

$$P(X = k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, 2, 3, \dots, n.$$

Ряд розподілу має вигляд:

$X = k$	0	1	2	3	...	n
$P(X = k)$	q^n	npq^{n-1}	$C_n^2 p^2 q^{n-2}$	$C_n^3 p^3 q^{n-3}$...	p^n

За біноміальним законом розподілена, наприклад, випадкова величина, яка дорівнює числу «успіхів» при n випробуваннях Бернуллі.

Математичне сподівання і дисперсія відповідно дорівнюють

$$M(X) = np, D(X) = npq.$$

2.6.5. Пуассонівський закон розподілу ймовірностей

Дискретна випадкова величина X розподілена за законом Пуассона, якщо вона набуває значень $0, 1, \dots, n, \dots$ з ймовірностями, які обчислюються за формулою Пуассона

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, 3, \dots, n, \dots$$

λ – параметр розподілу. Якщо число випробувань скінченне, то $\lambda = np$.

Ряд розподілу має вигляд

$X = k$	0	1	2	...	n	...
$P(X = k)$	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{1}{2} \lambda^2 e^{-\lambda}$		$\frac{\lambda^n e^{-\lambda}}{n!}$...

Числові характеристики розподілу $M(X) = D(X) = \lambda$.

Таким чином, теоретико-ймовірнісний зміст параметра λ полягає в тому, що параметр λ в законі Пуассона дорівнює математичному сподіванню і дисперсії випадкової величини X .

2.7. Основні розподіли неперервних випадкових величин

2.7.1 Рівномірний закон розподілу

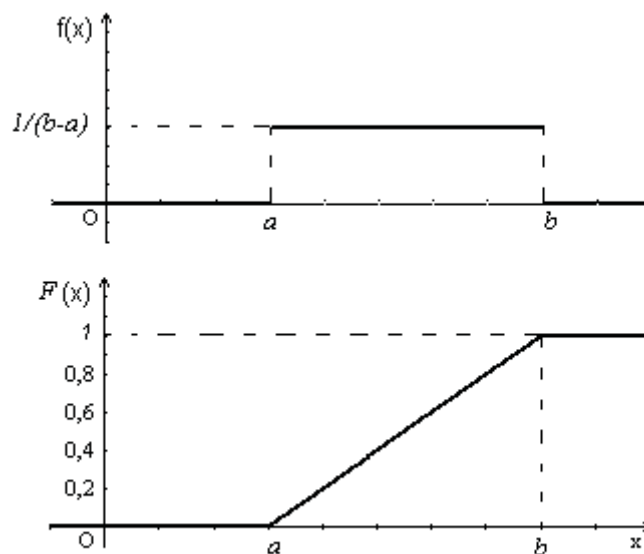
Неперервна випадкова величина X , що визначена на проміжку $[a, b]$, має рівномірний закон розподілу, якщо її щільність розподілу ймовірностей визначається формулою

$$p(x) = \begin{cases} 0, & x < a, \\ \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & x > b. \end{cases}$$

Функція розподілу ймовірностей

$$F(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x \leq b, \\ 0, & x > b. \end{cases}$$

Графіки щільності розподілу і функції розподілу ймовірностей мають вигляд



Ймовірність того, що випадкова величина X належить проміжку $[\alpha; \beta] \subset [a; b]$ обчислюється за формулою

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} \frac{dx}{b-a} = \frac{\beta - \alpha}{b-a}.$$

Числові характеристики розподілу:

$$M(X) = \frac{b+a}{2}, \quad D(X) = \frac{(b-a)^2}{12}, \quad \sigma(X) = \frac{b-a}{2\sqrt{3}}, \quad Me = M(X) = \frac{b+a}{2}.$$

2.7.2. Показниковий (експоненціальний) закон розподілу

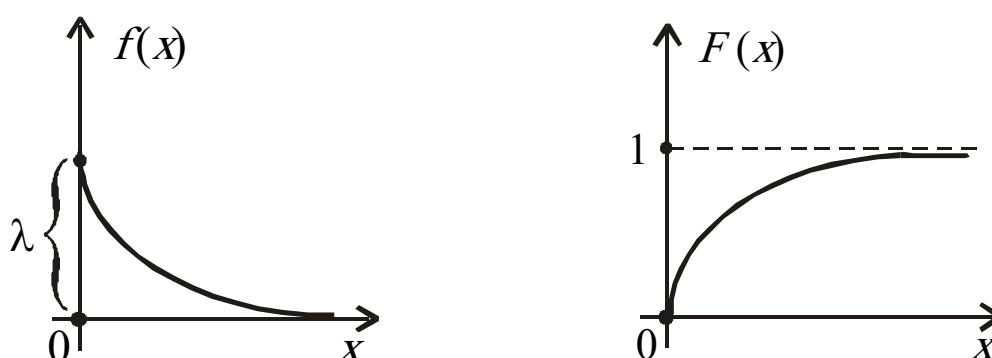
Неперервна випадкова величина X , називається розподіленою за показниковим (експоненціальним) законом, якщо її щільність розподілу ймовірностей визначається формулою:

$$f(x) = \begin{cases} 0, & x < 0; \\ \lambda e^{-\lambda x}, & x \geq 0. \end{cases}$$

Функція розподілу ймовірностей

$$F(x) = \begin{cases} 0, & x < 0; \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

Графіки щільності розподілу і функції розподілу ймовірностей мають вигляд



Числові характеристики розподілу

$$M(X) = \frac{1}{\lambda}, \quad D(X) = \frac{1}{\lambda^2}, \quad \sigma(X) = \frac{1}{\lambda}.$$

Серед усіх законів неперервних випадкових величин лише експоненціальному притаманна властивість – відсутність післядії, а саме: якщо пов'язати випадкову величину із часом, то для цього закону минуле не впливає на передбачення подій у майбутньому. Цю властивість експоненціального закону використовують у марківських випадкових процесах, теорії надійності тощо.

2.7.3. Нормальний закон розподілу

Випадкова величина X називається нормально розподіленою, якщо її щільність розподілу ймовірностей визначається формулою

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < x < \infty,$$

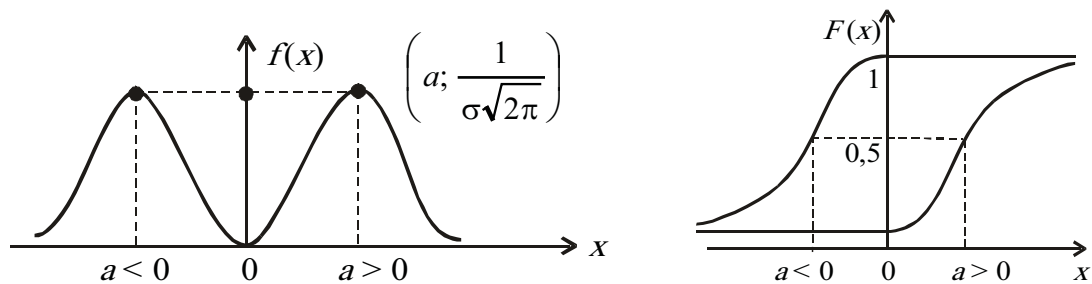
де $a=M(X)$, $\sigma=\sigma(X)$. Отже, нормальний закон визначається параметрами a і σ . Для нормального закону $A_s = E_s = 0$.

Графік щільності розподілу ймовірностей симетричний відносно прямої $x=a$ і має вісь Ox асимптотою (крива Гауса). Точки $x=a\pm\sigma$ є точками перегину.

Функція розподілу ймовірностей

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Графіки $p(x)$, $F(x)$ для загального нормального закону залежно від параметрів a і σ мають вигляд



Мода розподілу рівна медіані й середньому значенню $M_o=M_e=M_X=a$.

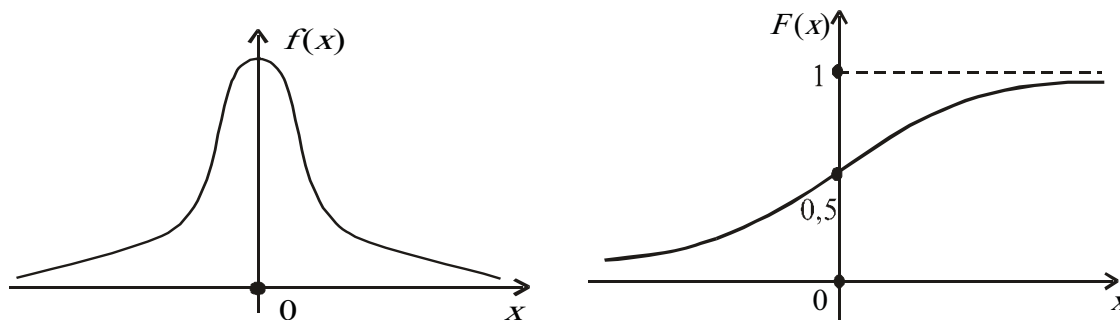
Якщо $a = 0$ і $\sigma = 1$, то нормальний закон називають нормованим. Нормований нормальний закон позначають $N(0; 1)$. У цьому разі

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty,$$

тобто $p(x) = \varphi(x)$ є функцією Гаусса,

$$F(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{x^2}{2}} dx - \text{функція Лапласа.}$$

Для нормованого нормального закону графіки функцій $p(x)$, $F(x)$ мають вигляд



Ймовірність того, що нормально розподілена випадкова величина набуває значень з деякого проміжку $[\alpha; \beta]$

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} p(x) dx = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

Правило трьох сигм для нормального закону

Якщо $\alpha = a - 3\sigma$, $\beta = a + 3\sigma$ то, враховуючи, що функція Лапласа непарна, згідно з вище вказаною формулою маємо

$$\begin{aligned} P(a - 3\sigma < X < a + 3\sigma) &= P(|x - a| < 3\sigma) = \\ &= 2\Phi\left(\frac{3\sigma}{\sigma}\right) = 2\Phi(3) = 2 \cdot 0,49865 = 0,9973 \end{aligned}$$

Фактично ця подія при одному експерименті здійсниться, а тому її вважають практично вірогідною. Звідси

$$P(|x - a| > 3\sigma) = 1 - P(|x - a| < 3\sigma) = 1 - 0,9973 = 0,0027.$$

Тобто ймовірність того, що внаслідок проведення експерименту випадкова величина X , яка має закон розподілу $N(a; \sigma)$, не потрапить у проміжок $[a - 3\sigma; a + 3\sigma]$, дорівнює $0,0027 = 0,27\%$, тобто практично вважається, що ця подія внаслідок проведення одного експерименту не здійсниться.

2.8. Закон великих чисел, його застосування в теорії оцінювання

Лема 1. Нехай X невід'ємна випадкова величина. Тоді для довільного $\varepsilon > 0$

$$P\{X \geq \varepsilon\} \leq \frac{MX}{\varepsilon} \text{ (нерівність Маркова).}$$

Наслідок. Для довільного $\varepsilon > 0$ справедлива нерівність:

$$P\{X < \varepsilon\} \geq 1 - \frac{MX}{\varepsilon} \text{ (нерівність Маркова).}$$

Лема 2. Нехай X невід'ємна випадкова величина. Тоді для довільного $\varepsilon > 0$

$$P\{|X - MX| \geq \varepsilon\} \leq \frac{DX}{\varepsilon^2} \text{ (нерівність Чебишова).}$$

Наслідок. Для довільного $\varepsilon > 0$ справедлива нерівність:

$$P\{|X - MX| < \varepsilon\} \geq 1 - \frac{DX}{\varepsilon^2} \text{ (нерівність Чебишова).}$$

Теорема Чебишова (закон великих чисел). Нехай $X_1, X_2, \dots, X_n, \dots$ – послідовність попарно незалежних випадкових величин з однаковими математичними сподіваннями $MX_i = a$ ($i=1, 2, \dots, n$) і обмеженими дисперсіями $DX_i \leq c$ ($i=1, 2, \dots, n$), c – стала. Розглянемо числову послідовність

$$p_n = P\left\{\left|\frac{X_1 + X_2 + \dots + X_n}{n} - a\right| < \varepsilon\right\}, \quad n=1, 2, \dots$$

Тоді $\lim_{n \rightarrow \infty} p_n = 1$.

Доведення.

Позначимо через η_n випадкові величини $\frac{X_1 + X_2 + \dots + X_n}{n}$. Згідно з властивостями математичного сподівання

$$M(\eta_n) = M\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n}(MX_1 + MX_2 + \dots + MX_n) = \frac{na}{n} = a.$$

Використовуючи властивості дисперсії та умову теореми, оцінимо зверху дисперсію випадкової величини η_n

$$D\eta_n = D\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n^2}(DX_1 + DX_2 + \dots + DX_n) \leq \frac{nc}{n^2} = \frac{c}{n}.$$

Застосувавши до випадкової величини η_n нерівність Чебишова дістанемо, що для довільного $\varepsilon > 0$

$$P\{|\eta_n - M\eta_n| < \varepsilon\} \geq 1 - \frac{D\eta_n}{\varepsilon^2},$$

а звідси $p_n = P\{|\eta_n - M\eta_n| < \varepsilon\} \geq 1 - \frac{c}{n\varepsilon^2} \rightarrow 1, n \rightarrow \infty$.

Враховуючи, що ймовірність не може бути більшою від 1, матимемо $\lim_{n \rightarrow \infty} p_n = 1$.

Формула $P\{|\eta_n - M\eta_n| < \varepsilon\} \geq 1 - \frac{c}{n\varepsilon^2}$ має багато застосувань у теорії оцінювання. Ймовірність $P\{|\eta_n - M\eta_n| < \varepsilon\}$ називають надійністю оцінювання, мале, додатне число ε – точністю, n – число спостережуваних значень випадкової величини.

Доведена теорема дає обґрунтування правила середнього арифметичного в практиці вимірювань. Нехай потрібно виміряти деяку фізичну величину a . Здійснивши n незалежних вимірювань, ми дістанемо n значень цієї величини x_1, x_2, \dots, x_n . Кожне значення x_i є значенням випадкової величини X_i , математичне сподівання якої дорівнює a (ця умова означає, що вимірювання позбавлені систематичних помилок). Крім того, вважатимемо, що виконано умову $DX_i \leq c$; це означає, що всі вимірювання здійснюються з деякою гарантованою точністю. Тоді з теореми Чебишова випливає, що

$$P\left\{\left|\frac{X_1 + X_2 + \dots + X_n}{n} - a\right| < \varepsilon\right\} \rightarrow 1,$$

тобто, при достатньо великому числі вимірювань з імовірністю, як завгодно близькою до одиниці, середнє арифметичне результатів вимірювань буде як завгодно мало відрізнятись від вимірюваної величини.

Теорема Бернуллі. *Нехай $\xi(n)$ число появ події A при n послідовних незалежних випробуваннях, в кожному з яких імовірність настання події A дорівнює p . Тоді $P\left\{\left|\frac{\xi(n)}{n} - p\right| < \varepsilon\right\} \geq 1 - \frac{1}{4n\varepsilon^2}$.*

Наслідок (найпростіша форма закону великих чисел).

$$\lim_{n \rightarrow \infty} P\{|v_n - p| < \varepsilon\} = 1$$

Справді, для випадкових величин $\xi_1, \xi_2, \dots, \xi_n, \dots$, виконуються умови теореми Чебишова і тому

$$\lim_{n \rightarrow \infty} P\{|v_n - p| < \varepsilon\} = \lim_{n \rightarrow \infty} P\left\{\left|\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} - p\right| < \varepsilon\right\} = 1.$$

Смисл цього наслідку в тому, що з імовірністю, як завгодно близькою до одиниці, можна стверджувати, що при досить великій кількості незалежних випробувань частота події як завгодно мало відрізняється від її ймовірності при окремому випробуванні.

2.9. Центральна гранична теорема та її застосування в освітніх вимірюваннях

Зміст закону великих чисел полягає в тому, що при усередненні великого числа випадкових доданків усе менше відчувається характерне для випадкових величин неконтрольоване розкидання їхніх значень, і в границі при $n \rightarrow \infty$ це розкидання зникає зовсім, тобто випадкова величина вироджується в не випадкову.

Однак при скінченному числі доданків випадкове розкидання залишається і виникає питання дослідження характеру цього розкидання. Фундаментальний результат у цьому напрямку, відомий як центральна гранична теорема, полягає в тому, що для широкого класу незалежних випадкових величин граничний закон розподілу їх нормованої суми, незалежно від типу розподілу доданків, наближається до нормального закону розподілу. Існує кілька формулювань центральної граничної теореми, розглянемо деякі.

Теорема Ляпунова. Якщо X_1, X_2, \dots, X_n – незалежні випадкові величини, у кожній з яких існує математичне сподівання $M(X_i) = a_i$, дисперсія $D(X_i) = \sigma_i^2$, абсолютний центральний момент третього порядку $M(|X_i - a_i|^3) = m_i$

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n m_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}} = 0,$$

то закон розподілу суми $Y_n = X_1 + X_2 + \dots + X_n$ при $n \rightarrow \infty$ необмежено наближається до нормального з математичним сподіванням $\sum_{i=1}^n a_i$ і дисперсією $\sum_{i=1}^n \sigma_i^2$.

Необмежене наближення закону розподілу суми $Y_n = X_1 + X_2 + \dots + X_n$ до нормального закону при $n \rightarrow \infty$ у відповідності до властивостей нормального закону означає, що

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{Y_n - \sum_{i=1}^n a_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \right| \leq z \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt = 0,5 + \Phi(z),$$

де $\Phi(z)$ – функція Лапласа.

Зміст умови теореми Ляпунова полягає в тому, щоб в сумі $Y_n = X_1 + X_2 + \dots + X_n$ не було доданків, вплив яких на розсіювання Y_n дуже великий у порівнянні з впливом всіх інших, а також не повинно бути великої кількості випадкових доданків, вплив яких дуже малий у порівнянні з сумарним впливом інших. Таким чином, питома вага кожного окремого доданку повинна прямувати до нуля при збільшенні кількості доданків.

Центральна гранична теорема залишається справедливою і у випадку, коли випадкові доданки не є однаково розподіленими, що широко використовується в практиці вимірювань.

2.10. Емпіричні дані вимірювань в освіті та методи визначення законів їх розподілу

Основою для висунення гіпотези про закон розподілу ознаки генеральної сукупності на основі отриманих даних результатів вимірювань в освіті може бути наявність теоретичних передумов про характер зміни ознаки. До них, зокрема, відносять виконання умов, що є підґрунтям теореми Ляпунова.

У деяких випадках підставою для висунення гіпотези про закон розподілу ознаки генеральної сукупності можуть бути певні формальні

властивості здобутого статистичного розподілу, а саме: рівність нулю A_s і E_s для нормального розподілу, рівність вибіркової середньої і вибіркового середнього квадратичного відхилення для експоненціального розподілу.

Інколи підґрунтям для висновків про характер гіпотетичного розподілу можуть бути форми полігону, гістограми.

Оскільки всі припущення про закон розподілу ознаки генеральної сукупності мають риси гіпотез, а не категоричних тверджень, то вони мають бути перевірені з допомогою критерію згоди. Критерії згоди дозволяють відповісти на питання про те, чи розбіжність між емпіричними (статистичними) і теоретичними розподілами є настільки незначною, що вона може бути приписана впливу випадковості, чи ні.

Емпіричними називаються частоти, які спостерігаються при реалізації вибірки, а *теоретичними* — які обчислюються за формулами.

Дискретний закон розподілу. Теоретичні частоти для дискретної випадкової величини обчислюємо за формулою

$$n'_i = nP_i,$$

де n — обсяг вибірки; P_i — імовірність спостережуваного значення $X = x_i$, яка обчислюється за умови, що ознака X має взятий за припущенням закон розподілу ймовірностей.

Неперервний закон розподілу. Якщо ознака X генеральної сукупності має неперервний розподіл ймовірностей, то теоретичні частоти обчислюються за

формулою
$$n'_i = n P_i,$$

де n — обсяг вибірки,

P_i — імовірність того, що випадкова величина X потрапить в i -й частковий інтервал. Вона обчислюється за формулами того закону розподілу, який припускаємо на основі обробки статистичного розподілу вибірки.

Так, наприклад, якщо є підстави для припущення, що ознака генеральної сукупності X має нормальний закон розподілу, то теоретичні частоти в цьому разі можна обчислювати за формулами

$$n'_i = \frac{nh}{\sigma_B} \cdot \varphi(u_i) = \frac{nh}{\sigma_B} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \bar{x}_B)^2}{2\sigma_B^2}},$$

де n — обсяг вибірки; h — довжина часткового інтервалу;

\bar{x}_B — вибіркова середня величина; σ_B — вибіркове середнє квадратичне відхилення;

$\varphi(u_i)$ — щільність імовірностей для загального нормального закону розподілу

або

$$n'_i = n \cdot \left(\Phi \left(\frac{x_{i+1} - \bar{x}_B}{\sigma_B} \right) - \Phi \left(\frac{x_i - \bar{x}_B}{\sigma_B} \right) \right),$$

де $\Phi \left(\frac{x_{i+1} - \bar{x}_B}{\sigma_B} \right)$, $\Phi \left(\frac{x_i - \bar{x}_B}{\sigma_B} \right)$ — значення функції Лапласа.

Критерій згоди Пірсона. Критерій згоди Пірсона є випадковою величиною, що має розподіл χ^2 з $k = q - m - 1$ ступенями вільності і визначається за формулою

$$\chi^2 = \sum_{i=1}^q \frac{(n_i - np_i)^2}{np_i},$$

де q — число часткових інтервалів інтервального статистичного розподілу вибірки; m — число параметрів, якими визначається закон розподілу ймовірностей генеральної сукупності згідно з нульовою гіпотезою, обчислених за вибіркою.

Так, наприклад, для закону Пуассона, який характеризується одним параметром λ число $m = 1$, для нормального закону число $m = 2$, оскільки цей закон визначається двома параметрами $\mu = M(X)$ і σ .

Якщо $n_i = np_i$ (усі емпіричні частоти збігаються з теоретичними), то $\chi^2 = 0$, у протилежному випадку $\chi^2 > 0$. Визначивши при заданому рівні значущості α і числу ступенів вільності критичну точку $\chi_{кр}^2(\alpha; k = q - m - 1)$, за таблицею «Критичні точки розподілу χ^2 » будується правобічна критична область. Якщо виявиться, що спостережуване значення критерію $\chi_{сп}^2 > \chi_{кр}^2$, то основну гіпотезу про закон розподілу ознаки генеральної сукупності

відхиляють, а у протилежному випадку ($\chi_{\text{сп}}^2 < \chi_{\text{кр}}^2$) основну гіпотезу приймають.

2.11. Вибірковий метод та оцінка результатів вибіркового спостереження у педагогіці й психології

2.11.1. Вибірковий метод

Вибірковий метод полягає у здійсненні вибіркового спостереження – виду несучільного спостереження, при якому отримують характеристику всієї сукупності одиниць на основі дослідження деякої її частини. Цей метод відрізняється двома ознаками: 1) визначається кількість одиниць генеральної сукупності, яку належить вибірково обстежити, 2) визначається порядок відбору одиниць, за якого вибіркова сукупність достатньою мірою буде репрезентувати генеральну.

Основними завданнями вибіркового спостереження є: 1) вивчення середнього розміру досліджуваної ознаки; 2) визначення частки досліджуваної ознаки в певній сукупності.

Вибіркове спостереження одержало широке практичне застосування завдяки своїм перевагам перед суцільним: 1) економія часу та коштів внаслідок скорочення обсягів роботи; 2) зведення до мінімуму псування та втрат досліджуваних об'єктів; 3) необхідність детального обстеження кожної одиниці при неможливості обстежити всі одиниці; 4) досягнення більшої точності результатів завдяки скороченню помилок реєстрації.

Вибірковий метод ґрунтується на випадковому відборі одиниць, які спостерігаються. Точність результатів, які одержують за допомогою вибіркового методу, залежить від способу відбору одиниць, ступеня коливання ознаки у сукупності та від чисельності (обсягу) вибірки.

2.11.2. Оцінка результатів вибіркового спостереження

За способом відбору одиниць для спостереження розрізняють такі види вибірок: 1) власне випадкова, 2) механічна, 3) типова (районована), 4) серійна (гніздова).

Результатом проведення вибіркового спостереження є визначення генеральних показників – середнього значення ознаки (\overline{x}_G) та частки ознаки в генеральній сукупності (p) за вибірковими даними. У педагогіці та психології розрізняють два способи поширення одержаних даних на генеральну сукупність (екстраполювання): 1) спосіб прямого перерахунку; 2) спосіб поправочних коефіцієнтів.

Суть способу прямого перерахунку полягає в тому, що вибіркoву середню досліджуваної ознаки множать на чисельність одиниць генеральної сукупності з урахуванням значення граничної помилки репрезентативності вибіркової середньої (або частки).

Спосіб поправочних коефіцієнтів застосовують під час вибіркового спостереження для перевірки й уточнення даних суцільного спостереження. Зіставляючи дані вибіркового спостереження з даними суцільного спостереження, обчислюють поправочний коефіцієнт, яким користуються для внесення поправок у матеріали суцільних спостережень.

2.11.3. Обробка малих вибірок

Малою вибіркою називається вибіркова сукупність, яка складається з порівняно невеликої кількості одиниць (20-30).

На відміну від великої вибірки, для малих вибірок середнє квадратичне відхилення обчислюють за формулою:

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}},$$

де $n - 1$ – кількість ступенів вільності варіації.

Під ступенем вільності варіації розуміють число варіантів, які можуть приймати довільні значення, не змінюючи їх загальної характеристики (середньої).

Середня помилка малої вибірки розраховується за формулою:

$$\mu = \frac{S}{\sqrt{n - 1}}$$

Оцінити результати малої вибірки за нормальним розподілом неможливо, оскільки при малій кількості спостережень розподіл ймовірності для середньої величини значною мірою залежить від характеру розподілу

індивідуальних величин. Англійський математик-статистик Вільям Госсет (Ст'юдент) знайшов закон розподілу вибірових середніх від генеральної середньої для малих вибірок і склав спеціальні таблиці, в яких наводяться значення t при невеликому обсязі вибірки та заданому рівні ймовірності P .

Знаючи середню помилку малої вибірки μ та коефіцієнт t , можна визначити граничну помилку малої вибірки:

$$\Delta = t \cdot \mu$$

Таким чином, при вибірках невеликого обсягу методи оцінювання результатів вибіркового спостереження відрізняються від тих, які застосовують в теорії великих вибірок. Це пояснюється тим, що за малих обсягів, досліджувані показники більше залежать від випадковостей вибірки.

2.12. Загальні методи вивчення зв'язку соціально-педагогічних явищ і процесів

2.12.1. Типи зв'язків між величинами

Вивчення соціально-педагогічних явищ і процесів показує, що варіація кожної досліджуваної ознаки знаходиться в тісному взаємозв'язку з варіацією інших ознак, що характеризують певну сукупність одиниць. За ступенем залежності між явищами розрізняють два види зв'язку: функціональний і стохастичний, різновидом якого є кореляційний зв'язок. Усі ці зв'язки причинно-наслідкові. Вивчаючи закономірності зв'язку, причини та умови об'єднують в одне поняття — фактор. Відповідно до цього ознаки, які характеризують причини чи умови зміни певних явищ, називаються факторними, а ті, що характеризують наслідки — результативними. Як правило, значення факторної ознаки позначають через x , а результативної — через y .

Функціональні зв'язки характеризуються повною відповідністю між причиною і наслідком, тобто між факторною і результативною ознаками. При функціональному зв'язку кожному значенню x можна поставити y відповідність чітко визначене значення y . Такий зв'язок можна записати так: $x_1 \rightarrow y_1, x_2 \rightarrow y_2, \dots, x_n \rightarrow y_n$.

Стохастичні зв'язки, на відміну від функціональних, неоднозначні. Вони виявляються як узгодженість варіації двох чи більше ознак. Для них характерно те, що кожному значенню ознаки x відповідає певна множина

значень результативної ознаки y , які утворюють умовний розподіл результативної ознаки. Стохастичний зв'язок, характеризуючи множинність взаємозалежностей причин і наслідків, виявляється в зміні умовних розподілів.

Якщо умовні розподіли результативної ознаки замінити її середнім значенням \bar{y} , то утвориться різновид стохастичного зв'язку, який називається **кореляційним зв'язком**. Кореляційний зв'язок, на відміну від функціонального, виявляється не в кожному окремому випадку, а в середньому, в цілому для багатьох випадків: $x_1 \rightarrow \bar{y}_1, x_2 \rightarrow \bar{y}_2, \dots, x_n \rightarrow \bar{y}_n$.

У найпростішому випадку застосування кореляційного зв'язку величина результативної ознаки розглядається як наслідок зміни лише одного фактора.

При функціональному зв'язку для кожного значення факторної ознаки можна визначити відповідне значення результативної. При кореляційному зв'язку встановлюється лише тенденція зміни результативної ознаки при зміні величини факторної.

Статистичні методи дослідження зв'язків можуть показати наявність зв'язку між ознаками, особливості його прояву в певних умовах, кількісно охарактеризувати таку залежність, а пояснення існуючим змінам повинен дати змістовний аналіз, який ґрунтується на теорії освітніх вимірювань (класичній та сучасній).

2.12.2. Достовірність дослідження кореляційного зв'язку

При дослідженні кореляційних залежностей між ознаками розв'язуються такі завдання:

- попередній аналіз властивостей досліджуваних сукупностей одиниць;
- встановлення факту наявності зв'язку між ознаками, визначення його напрямку та форми;
- вимірювання ступеня щільності зв'язку між ознаками;
- побудова регресійної моделі зв'язку, тобто знаходження аналітичного вираження зв'язку (коефіцієнти, або невідомі параметри регресії шукають методом найменших квадратів);
- оцінювання адекватності моделі, її інтерпретація та рекомендації щодо практичного використання.

Для того щоб результати кореляційного аналізу знайшли практичне використання та були достовірними, досліджувані сукупності одиниць повинні відповідати таким вимогам: 1) бути *однорідними* (одним з методів дослідження однорідності є розрахунок відносних показників варіації); 2) повинна виконуватися вимога *достатньої кількості спостережень*; 3) фактори, які підлягають дослідженню, повинні бути *незалежними* один від одного (наявність тісного зв'язку між ними може бути свідченням того, що вони характеризують одні і ті самі сторони досліджуваного явища і значною мірою дублюють один одного); 4) оскільки всі основні положення теорії кореляції розроблялися стосовно припущення про нормальний характер розподілу, досліджувані сукупності повинні *відповідати умовам нормального розподілу*; 5) при побудові кореляційних моделей фактори повинні мати *кількісне вираження*, інакше побудувати модель кореляційної залежності неможливо.

Для відповіді на питання про наявність чи відсутність кореляційного зв'язку між ознаками використовується ряд спеціальних методів, серед яких: дисперсійний аналіз та елементарні прийоми (паралельне порівняння рядів значень результативної та факторної ознаки; графічне зображення фактичних даних за допомогою поля кореляції; побудова групової та кореляційних таблиць) та інші методи.

Одним з найпростіших методів аналізу взаємозв'язків є метод порівняння паралельних рядів, суть якого полягає в тому, що отримані в результаті групування і лічильного оброблення матеріали статистичного спостереження розташовуються паралельними рядами. При цьому значення факторної ознаки впорядковують в порядку зростання чи спадання і паралельно записують відповідні значення результативної ознаки. Це дає можливість, порівнюючи їх, простежити співвідношення, виявити наявність зв'язку і його напрямок.

2.12.3. Лінійна кореляція

Досліджуючи аналітичну залежність між випадковими величинами X і Y , обмежуються вивченням залежності між однією з них і умовним математичним сподіванням іншої, тобто залежністю:

$$\overline{y_x} = f^*(x) \text{ — вибіркоче рівняння регресії } Y \text{ на } X, \quad \overline{y_x} = M(Y | X = x);$$

$$\overline{x_y} = g^*(y) \text{ — вибіркоче рівняння регресії } X \text{ на } Y, \quad \overline{x_y} = M(X | Y = y);$$

У наведених вибіркових рівняннях регресії $\overline{y_x}$ і $\overline{x_y}$ — вибіркові умовні математичні сподівання, відповідно, Y на X і X на Y , а $f^*(x)$ і $g^*(y)$ — вибіркові функції регресії. Аналітичні вирази для функцій $f^*(x)$ і $g^*(y)$ будують на підставі проведеної вибірки $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Характер відповідної регресійної моделі допомагає вибрати діаграма розсіювання точок (x_i, y_i) на площині.

Припустимо, що розподіл результативної ознаки Y у генеральній сукупності є нормальним; дисперсія результативної ознаки Y не залежить від факторної ознаки X ; між результативною та факторною ознаками існує лінійний зв'язок. Ці обмеження приводять до дослідження найпростішої регресійної моделі — лінійної регресії, коли вибіркоче рівняння регресії записують у вигляді:

$$\overline{y_x} = ax + b.$$

Основним методом отримання точкових оцінок (що виражаються одним числом) для параметрів a і b рівняння регресії є *метод найменших квадратів*.

Нехай вибірка $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ обсягу n є незгрупованою.

Ідея методу найменших квадратів полягає в тому, що за точкові оцінки a і b параметрів $a^\#$ і $b^\#$ вибирають такі числа, для яких пряма $\overline{y_x} = ax + b$ є «найближчою» до точок $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Мірою відхилення шуканої прямої від точок (x_i, y_i) є величина

$$S(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2,$$

тобто сума квадратів різниць між теоретичними ординатами прямої та емпіричними ординатами точок (x_i, y_i) для одних і тих же значень $x = x_i$.

Якщо числа a і b є такими, що функція $S(a, b)$ набуває найменшого значення (мінімум), то пряма $\overline{y_x} = ax + b$ найменше відхиляється від точок (x_i, y_i) , тобто сумарно є наближеною до точок, які зображають результати вибірки.

Оскільки точкові оцінки a і b для невідомих параметрів $a^\#$ і $b^\#$ повинні реалізувати мінімум функції $S(a, b)$, то для їх знаходження маємо систему рівнянь

$$\begin{cases} \frac{\partial S(a,b)}{\partial a} = 2 \sum_{i=1}^n [y_i - (ax_i + b)] \cdot (-x_i) = 0, \\ \frac{\partial S(a,b)}{\partial b} = (-2) \sum_{i=1}^n [y_i - (ax_i + b)] = 0, \end{cases}$$

яку елементарними перетвореннями зводимо до вигляду

$$\begin{cases} a \left(\sum_{i=1}^n x_i^2 \right) + b \left(\sum_{i=1}^n x_i \right) = \sum_{i=1}^n x_i y_i, \\ a \left(\sum_{i=1}^n x_i \right) + b \cdot n = \sum_{i=1}^n y_i. \end{cases}$$

Зауваження. У випадку згрупованої вибірки для визначення невідомих параметрів $a^\#$ і $b^\#$ матимемо систему двох лінійних рівнянь

$$\begin{cases} a \left(\sum_{i=1}^m n_{x_i} x_i^2 \right) + b \left(\sum_{i=1}^m n_{x_i} x_i \right) = \sum_{i=1}^n \sum_{j=1}^k n_{x_i y_j} x_i y_j, \\ a \left(\sum_{i=1}^m x_i \right) + b \cdot n = \sum_{j=1}^k n_{y_j} y_j. \end{cases}$$

Припускаючи, що ознака X не є сталою, тобто серед вибіркових значень x_1, \dots, x_n обов'язково є різні числа, отримаємо, що визначник системи

$$\begin{vmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{vmatrix} = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 > 0.$$

Тому система рівнянь має єдиний розв'язок

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{\sigma_x^2}}, \quad b = \bar{y} - \bar{a} \cdot \bar{x},$$

де

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \overline{\sigma_x^2} = \overline{D_x} = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2, \quad \overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i.$$

Підставивши у рівняння лінійної регресії значення a і b дістанемо шукане рівняння лінійної регресії

Коефіцієнт a називають *коефіцієнтом регресії*. Він характеризує відношення величини приросту результативної ознаки $\Delta \bar{y}_x$ до величини приросту Δx факторної ознаки.

Лінійне рівняння регресії можна подати також через точкову оцінку коефіцієнта кореляції

$$\bar{y}_x - \bar{y} = r_B \frac{\sigma_y}{\sigma_x} (x - \bar{x}), \quad r_B = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{n \cdot \sigma_x \cdot \sigma_y}.$$

Зауваження. У разі, коли припущення про лінійність зв'язку між результативною ознакою Y і факторною ознакою X порушується (про це можна зробити висновок з діаграми розсіювання вибірки), то будують нелінійні регресійні моделі. Ці моделі можуть виражатися, наприклад, рівняннями

$$\begin{aligned} \bar{y}_x &= ax^2 + bx + c, \\ \bar{y}_x &= ax^3 + bx^2 + cx + d, \\ \bar{y}_x &= \frac{a}{x} + b. \end{aligned}$$

Точкові оцінки параметрів у наведених нелінійних моделях також можна знайти методом найменших квадратів.

2.12.4. Рангова кореляція

Нехай вибірка складається з n незалежних одиниць, які мають ознаки A і B . Під якісною ознакою розуміють ознаку, яку не можливо точно виміряти, але вона дозволяє порівнювати одиниці вибірки між собою, тобто, розташовувати їх у порядку погіршення, або покращення якості.

Розташуємо одиниці вибірки в порядку погіршення їх якості і присвоїмо кожній з них порядковий номер, який назвемо *рангом*. Кореляція між рангами відображає у певній мірі залежність між ознаками.

Нехай за ознакою A елементи мають ранги X_1, X_2, \dots, X_n , а за B — Y_1, Y_2, \dots, Y_n , де всі X і Y є перестановками перших чисел натурального ряду. Утворимо різниці $d_k = X_k - Y_k$, $k = 1; \dots, n$. Значення d_k утворюють міру сили між ознаками A і B , якщо всі $d_k = 0$, то відповідність між ознаками повна. Припустимо, що серед d_k є відмінні від нуля. Тоді на практиці для обчислення щільності

зв'язку користуються **вибірковим коефіцієнтом рангової кореляції Спірмена**:

$$r_c = 1 - \frac{6}{n^3 - n} \sum_{k=1}^n d_k^2, \quad k = \overline{1, n}.$$

Крім вибіркового коефіцієнта рангової кореляції Спірмена існують інші показники зокрема рангової кореляції Кендалла.

Нехай за ознакою А елементи мають ранги X_1, X_2, \dots, X_n , а за В – Y_1, Y_2, \dots, Y_n , де праворуч від $Y_1 \in R_1$ рангів більших за Y_1 , праворуч від $Y_2 \in R_2$ рангів більших за Y_2 і т.д., праворуч від $Y_{p-1} \in R_{p-1}$ рангів більших за Y_{p-1} . Позначимо через $R = R_1 + R_2 + \dots + R_{p-1}$.

Тоді **вибірковий коефіцієнт рангової кореляції Кендалла** матиме вигляд:

$$r_K = \frac{4 \cdot R}{n \cdot (n-1)} - 1.$$

Не можна сказати який із вибірових коефіцієнтів, Спірмена чи Кендалла, дає точніший результат зв'язку між рангами. На даний період користуються обома вибіровими коефіцієнтами.

Для всіх вибірових коефіцієнтів кореляції виконується нерівність $|r| \leq 1$, якщо $r=0$, то зв'язку немає, якщо $r=\pm 1$, то зв'язок функціональний.

ПРАКТИЧНІ ЗАВДАННЯ

Приклад 1. Студент прийшов на екзамен, вивчивши тільки 20 із 25 питань програми. У екзаменаційному білеті 3 питання. Випадкова величина X – кількість питань, на які студент зможе відповісти. Скласти закон розподілу випадкової величини X . Знайти математичне сподівання і середнє квадратичне відхилення.

Розв'язання:

Так як питання не повторюються, то випадкова величина X – кількість питань, на які студент зможе відповісти, розподілена за гіпергеометричним законом. Оскільки студент, прийшовши на екзамен, не вивчив 5 питань програми, а в екзаменаційному білеті 3 питання, то випадкова величина X може приймати значення 0, 1, 2, 3 з ймовірностями, які обчислюються за формулами:

$$P(X = m) = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n}, \text{ де } n, M, N \text{ – параметри розподілу.}$$

За умовою задачі $n=3, M=20, N=25$. Тоді

$$P(0) = \frac{C_{20}^0 \cdot C_5^3}{C_{25}^3} = \frac{1}{230} = 0,004, \quad P(1) = \frac{C_{20}^1 \cdot C_5^2}{C_{25}^3} = \frac{20}{230} = 0,087,$$

$$P(2) = \frac{C_{20}^2 \cdot C_5^1}{C_{25}^3} = \frac{95}{230} = 0,413, \quad P(3) = \frac{C_{20}^3 \cdot C_5^0}{C_{25}^3} = \frac{114}{230} = 0,496.$$

Запишемо закон розподілу

X	0	1	2	3
P	0,004	0,087	0,413	0,496

Контроль: $\sum_{i=1}^4 p_i = 0,004 + 0,087 + 0,413 + 0,496 = 1$, отже, ймовірності обчислено правильно.

Математичне сподівання і середнє квадратичне відхилення обчислимо відповідно за формулами:

$$MX = n \frac{M}{N}, \quad \sigma x = \sqrt{DX} = \sqrt{\frac{n \cdot M \cdot (N - M) \cdot (N - n)}{N^2 \cdot (N - 1)}}.$$

Таким чином, математичне сподівання $MX = n \frac{M}{N} = 3 \cdot \frac{20}{25} = 2,4$, дисперсія

$$DX = \frac{3 \cdot 20(25 - 20)(25 - 3)}{25^2(25 - 1)} = 0,44, \quad \text{а середнє квадратичне відхилення}$$

$$\sigma = \sqrt{0,44} = 0,663.$$

Приклад 2. Результати тестування абітурієнтів вищих закладів освіти деякого міста наближено можна вважати розподіленими за нормальним законом з середнім значенням 500 балів і середнім квадратичним відхиленням 100 балів. Знайти: а) який відсоток усіх результатів охоплюють оцінки від 300 до 700 балів? б) яка найнижча оцінка серед 10% найвищих оцінок? в) яка кількість абітурієнтів з 10000, можна сподіватися, матимуть оцінки, не нижчі від 675 балів?

Розв'язання:

За умовою задачі середнє значення нормального розподілу $a=MX=500$; а середнє квадратичне відхилення $\sigma=100$.

а) Ймовірність того, що випадкова величина X (результати тестування абітурієнтів вищих закладів освіти) міститься в межах від 300 до 700 балів знайдемо за формулою

$$P(\alpha < x < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right).$$

$$P(300 < x < 700) = \Phi\left(\frac{700 - 500}{100}\right) - \Phi\left(\frac{300 - 500}{100}\right) = \Phi(2) - \Phi(-2) = 2\Phi(2) = 2 \cdot 0.4772 = 0.9544$$

$0.9544 = 95.44\%$. Отже, 95,44% усіх результатів охоплюють оцінки від 300 до 700 балів;

б) За умовою $P(\alpha < x < \infty) = 0,5 - \Phi\left(\frac{\alpha - 500}{100}\right) = 10\% = 0,1$,

$$0.5 - 0.1 = \Phi\left(\frac{\alpha - 500}{100}\right); \quad \Phi\left(\frac{\alpha - 500}{100}\right) = 0,4.$$

За таблицями функції Лапласа $\frac{\alpha - 500}{100} = 1,28$; $\alpha - 500 = 128$; $\alpha = 628$.

Отже, серед 10% найвищих оцінок найнижча оцінка 628 балів.

в) Знайдемо ймовірність того, що абітурієнти вищих закладів освіти матимуть оцінки, не нижчі від 675 балів

$$P(675 \leq x) = 0,5 - \Phi\left(\frac{675 - 500}{100}\right) = 0,5 - \Phi(1,75) = 0,5 - 0,4599 = 0,0401.$$

За статистичним означенням $\frac{k}{n} = \nu(A) \approx P(A)$; $\frac{k}{10000} = 0,0401$. Звідки $k=401$ абітурієнт.

Приклад 3. Знайти вибірковий коефіцієнти кореляції Спірмена (табл. 3.1), перевірити його на значимість:

Таблиця 3.1

Ранги	Учень									
	1	2	3	4	5	6	7	8	9	10
Оцінки з математики	9	3	1	4	2	8	5	6	7	10
Оцінки з хімії	6	7	3	2	1	8	5	4	9	10

Розв'язання:

Упорядкуємо в таблиці оцінки з математики в порядку зростання (табл.3.2).

Таблиця 3.2

Ранги	Учень									
	3	5	2	4	7	8	9	6	1	10
Оцінки з математики, x_i	1	2	3	4	5	6	7	8	9	10
Оцінки з хімії, y_i	3	1	7	2	5	4	9	8	6	10

Знайдемо різниці $d_i = x_i - y_i$, запишемо ці числа та їх квадрати в таблицю 3.3:

Таблиця 3.3

i	1	2	3	4	5	6	7	8	9	10
d_i	-2	1	-4	2	0	2	-2	0	3	0
d_i^2	4	1	16	4	0	4	4	0	9	0

Тоді коефіцієнт рангової кореляції Спірмена обчислимо за формулою:

$$r_C = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n d_i^2 .$$

Отже, $r_C = 1 - \frac{6}{10 \cdot (10^2 - 1)} \cdot (4 + 1 + 16 + 4 + 0 + 4 + 4 + 0 + 9 + 0) = \frac{41}{55} \approx 0,75$, слід

вважати, що оцінки з хімії та оцінки з математики тісно зв'язані між собою.

Перевірку на значимість вибіркового коефіцієнта рангової кореляції можна провести строго. Гіпотеза $H_0: |r_s|=0$ при альтернативній гіпотезі $H_\alpha: |r_s|>0$ і обсязі вибірки $n \geq 10$ перевіряється за значенням випадкової величини

$$t = |r_C| \cdot \sqrt{\frac{n-2}{1-r_C^2}} .$$

Якщо гіпотеза H_0 вірна, то ця статистика має розподіл Стюдента з $k=n-2$ ступенями вільності. Закон розподілу вихідної двовимірної генеральної сукупності не має значення, хоча передбачається, що складові X і Y генеральної сукупності – неперервні випадкові величини.

У нашому випадку

$$t_{cn} = |0,75| \cdot \sqrt{\frac{10-2}{1-(0,75)^2}} \approx 3,21.$$

Якщо взяти рівень значущості $\alpha=0,05$, то при $k=8$ $t_{кр}=1,86$ (критична область правостороння). Так як $t_{кр} < t_{cn}$, то гіпотеза H_0 відхиляється.

Таким чином, коефіцієнт рангової кореляції Спірмена значимий.

Приклад 4. Знайти вибіркового коефіцієнти кореляції Кендела (табл. 4.1), перевірити його на значимість:

Таблиця 4.1

Ранги	Учень									
	1	2	3	4	5	6	7	8	9	10
Оцінки з математики	9	3	1	4	2	8	5	6	7	10
Оцінки з хімії	6	7	3	2	1	8	5	4	9	10

Розв'язання:

Упорядкуємо в таблиці оцінки з математики в порядку зростання (табл.4.2).

Таблиця 4.2

Ранги	Учень									
	3	5	2	4	7	8	9	6	1	10
Оцінки з математики, x_i	1	2	3	4	5	6	7	8	9	10
Оцінки з хімії, y_i	3	1	7	2	5	4	9	8	6	10

Знайдемо кількість рангів, більших за y_i , позначимо ці числа відповідно $R_1=7$,

$$R_2=8, R_3=3, R_4=6, R_5=4, R_6=4, R_7=1, R_8=1, R_9=1,$$

їх сума

$$R = R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8 + R_9 = 7 + 8 + 3 + 6 + 4 + 4 + 1 + 1 + 1 = 35.$$

Тоді коефіцієнт рангової кореляції Кендела обчислимо за формулою:

$$r_K = \frac{4R}{n \cdot (n-1)} - 1.$$

$$\text{Отже, } r_K = \frac{4 \cdot 35}{10 \cdot (10-1)} - 1 = \frac{5}{9} \approx 0,56.$$

Значимість коефіцієнта рангової кореляції Кендела (H_0 : в генеральній сукупності $r_K=0$) при обсязі вибірки $n \geq 10$ перевіряється за допомогою статистики:

$$\lambda = \frac{r_K}{\sqrt{\frac{2 \cdot (2n+5)}{9n \cdot (n-1)}}},$$

яку можна вважати приблизно нормально розподіленою випадковою величиною з середнім $M(\lambda)=0$ і дисперсією $D(\lambda)=1$ (якщо вірна нульова гіпотеза). Для нашого прикладу при альтернативній гіпотезі $H_a: r_K > 0$ і рівні значущості $\alpha=0,05$, за таблицею функції Лапласа знаходимо, що $\lambda_{кр}=1,65$. Між тим

$$\lambda_{cn} = \frac{0,56}{\sqrt{\frac{2 \cdot (2 \cdot 10 + 5)}{9 \cdot 10 \cdot (10-1)}}} \approx 2,25.$$

Так як $t_{кр} < t_{cn}$, то гіпотеза H_0 відхиляється (критична область правостороння).

Таким чином, коефіцієнт рангової кореляції Кендела значимий.

Приклад 5. Групі школярів молодших класів був даний стандартний тест на перевірку швидкості читання. Потім зі школярами був проведений спеціальний курс занять, після чого дітям було запропоновано другий тест. Швидкість читання оцінювалася в балах, результати представлені в таблиці 5.1.

Таблиця 5.1

№ за списком	1	2	3	4	5	6	7	8	9	10
Результати I тестування	7	5	6	8	3	8	7	8	2	4
Результати II тестування	6	7	9	8	6	7	7	9	4	7

Чи можна сказати, що заняття зі збільшення швидкості читання ефективні, тобто результати істотно покращилися? Передбачається, що обидва тести еквівалентні за складністю, а оцінки можна вважати нормально розподіленими.

Розв'язання:

Розглянемо вибірку, варіанти якої – різниця балів, отриманих кожним школярем до і після занять. Віднімемо від балів, отриманих під час другого тестування, бали, які були отримані за перший тест. Отримаємо таблицю 5.2.

Таблиця 5.2

№	1	2	3	4	5	6	7	8	9	10
X	-1	2	3	0	3	-1	0	1	2	3

Припустимо, що вправи не впливають на швидкість читання. Тоді отриману вибірку слід вважати вибіркою з нормально розподіленої генеральної сукупності з середнім значенням $\mu=0$, отже, сформулюємо основну гіпотезу так: $H_0: \mu=0$. Якщо ж під час спеціального курсу занять вправи збільшують швидкість читання (альтернативна гіпотеза), то математичне сподівання μ повинно бути більше нуля, тобто $H_a: \mu>0$.

Для перевірки нульової гіпотези потрібно використати статистику:

$$t = \frac{\bar{x} - a}{\frac{S}{\sqrt{n}}}, \text{ яка має розподіл Стюдента з числом ступенів вільності } k=n-1.$$

Розрахуємо середнє значення та виправлене середнє квадратичне відхилення отриманого результату за формулами:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{10} x_i, \quad S = \sqrt{\frac{n}{n-1} \cdot \left(\frac{1}{n} \cdot \sum_{i=1}^{10} x_i^2 - (\bar{x})^2 \right)}.$$

Отримаємо:

$$\bar{x} = \frac{-1+2+3+0+3+(-1)+0+1+2+3}{10} = 1,2;$$

$$S = \sqrt{\frac{10}{9} \left(\frac{(-1)^2 + 2^2 + 3^2 + 0^2 + 3^2 + (-1)^2 + 0^2 + 1^2 + 2^2 + 3^2}{10} - (1,2)^2 \right)} = 1,62.$$

Використовуючи таблиці Стюдента знайдемо критичне значення критерію t за рівнем значущості $\alpha=0,05$ і числом ступенів вільності $k=10-1=9$:

$$t_{кр}(\alpha=0,05; k=9)=1,83.$$

Обчислимо спостережуване значення критерію:

$$t_{сн} = \frac{1,2 - 0}{\frac{1,62}{\sqrt{10}}} = 2,34.$$

Так як $t_{сн} > t_{кр}$, то основна гіпотеза відхиляється і приймається альтернативна.

Отже, варто визнати, що заняття пройшли успішно і середня швидкість читання у дітей покращилася.

Приклад 6. Було проведено два тестування в одній і тій же групі. Перед другим тестуванням група пройшла підготовку. Чи можна за результатами, зображеними в таблиці, зробити висновок, що підготовка була ефективною (тобто результати другого тестування істотно покращилися)? Використайте T-критерій Вілкоксона.

Таблиця 6.1

№з/п	1	2	3	4	5	6	7	8	9	10	11
Тест 1	5,05	6,48	5,16	7,3	4,7	7,25	5,85	6,62	5,15	4,83	6,2
Тест 2	7,2	7,43	5,58	7,46	7,05	12,95	5,55	9,85	7,5	6,38	14,35

Розв'язання:

Знайдемо різниці результатів другого й першого тестування $t_{№2} - t_{№1}$, та їхні значення за абсолютною величиною $|t_{№2} - t_{№1}|$. Результати тестування та розрахункові значення запишемо в таблицю 6.2. Як бачимо, більшість зміщень – більше нуля. Їх і будемо вважати типовими.

Сформулюємо гіпотези:

основна H_0 : Інтенсивність зміщень в типовому напрямку переважає;

альтернативна H_α : Інтенсивність зміщень в типовому напрямку не переважає.

Таблиця 6.2

№ тестованого	Тест №1	Тест №2	$t_{№2} - t_{№1}$	$ t_{№2} - t_{№1} $	Ранг
1	5,05	7,2	2,15	2,15	6
2	6,48	7,43	0,95	0,95	4
3	5,16	5,58	0,42	0,42	3
4	7,3	7,46	0,16	0,16	1
5	4,7	7,05	2,35	2,35	7,5
6	7,25	12,95	5,7	5,7	10
7	5,85	5,55	-0,3	0,3	2
8	6,62	9,85	3,23	3,23	9
9	5,15	7,5	2,35	2,35	7,5
10	4,83	6,38	1,55	1,55	5
11	6,2	14,35	8,15	8,15	11

Критичне значення при $n=11$ та при рівні значущості $\alpha=0,05$ $T_{кр}(11; 0,05)=13$.

Обчислимо суми рангів окремо для від'ємних та додатних відхилень.

$R_+ = 6+4+3+1+7,5+10+9+7,5+5+11=64$; $R_- = 2$

За емпіричне значення береться мінімальне з R_+ і R_- . Тому, $T_{емп}=2$.

Оскільки емпіричне значення менше за критичне, то підтверджується гіпотеза H_0 .

Отже, дійсно підготовка була ефективною, так як результати другого тестування істотно покращилися порівняно з результатами першого тестування.

Питання для самоконтролю:

1. Які змінні називаються латентними?
2. Як трактується поняття «вимірювання в освіті»?
3. Назвіть компоненти процесу педагогічних вимірювань.
4. Що називають інструментом вимірювання?
5. Назвіть чим відрізняються кількісні шкали від якісних.
6. Які типи кількісних шкал Ви знаєте?
7. До якого типу шкал відноситься 12-ти бальна шкала шкільних оцінок? Шкала балів ЗНО?
8. Чи обґрунтоване додавання балів порядкової шкали?
9. Що називають метрикою інтервальної шкали?
10. Чим відрізняються метрична та нормована шкали?
11. За якою шкалою можна розрізнити дихотомічні об'єкти?
12. Що називають нормами?
13. Чи є норми абсолютними і постійними?
14. У чому полягає мета нормо-орієнтованого вимірювання?
15. Які вимоги пред'являють до норм?
16. Назвіть основні характеристики нормо-орієнтованого вимірювання.
17. У чому полягає мета критеріально-орієнтованого вимірювання?
18. Назвіть основні характеристики критеріально-орієнтованого вимірювання.
19. Назвіть відмінності між нормо-орієнтованими і критеріально-орієнтованими тестами.
20. Яке призначення нормативно-орієнтованого тесту?
21. Яке призначення критеріально-орієнтованого тесту?
22. Що спільного між нормативно-орієнтованим та критеріально-орієнтованим тестами?
23. У чому відмінності між нормативно-орієнтованим та критеріально-орієнтованим тестами?
24. Вкажіть ознаки за якими можна класифікувати тести.
25. Дайте визначення гомогенного та гетерогенного тестів.
26. Як можна класифікувати тести за цілями використання? Наведіть визначення відповідних видів тестів?
27. Наведіть класифікацію навчальних цілей.
28. Що означає термін «таксономія»? У чому полягає суть таксономії Блума?

29. Що означає категорія «знання»?
30. Що є показником розуміння навчального матеріалу?
31. Що означає категорія «застосування»?
32. Які інтелектуальні вміння характеризують категорію «аналіз»?
33. Що означає категорія «синтез»?
34. Які ключові слова використовуються при створенні тестових завдань рівня «оцінювання»?
35. Наведіть приклади завдань у тестовій формі різних рівней таксономії Блума.
36. Дайте означення дискретної випадкової величини, що має рівномірний закон розподілу на множині $\{1, 2, \dots, n\}$. Поясніть теоретико-ймовірнісне значення параметра розподілу.
37. Охарактеризуйте дискретну випадкову величину, що має гіпергеометричний розподіл. Скільки параметрів має розподіл? Яке їх теоретико-ймовірнісне значення?
38. Дайте означення дискретної випадкової величини, що розподілена за біноміальним законом. Скільки параметрів має розподіл? Яке їх теоретико-ймовірнісне значення? Наведіть приклад застосування у педагогічних вимірюваннях.
39. Охарактеризуйте дискретну випадкову величину, що має геометричний закон розподілу. Поясніть теоретико-ймовірнісне значення параметра розподілу.
40. Як обчислити теоретичні частоти спостережуваного значення $X = x_i$ за умови, що ознака X має закон розподілу Пуассона? Як визначити параметр розподілу?
41. Запишіть функцію щільності для неперервної випадкової величини, розподіленої рівномірно на проміжку $[a; b]$. Поясніть теоретико-ймовірнісне значення параметрів розподілу.
42. Охарактеризуйте випадкову величину, що має нормальний закон розподілу. Скільки параметрів має розподіл? Яке їх теоретико-ймовірнісне значення? Наведіть приклад застосування у педагогічних вимірюваннях.
43. Дайте означення неперервної випадкової величини, що має показниковий (експоненціальний) закон розподілу. Поясніть теоретико-ймовірнісне значення параметра розподілу.
44. Запишіть функцію щільності для випадкової величини, розподіленої за стандартним законом $N(0; 1)$. Поясніть правило двох сигм; трьох сигм.
45. Які критерії використовують для перевірки статистичних гіпотез про закон розподілу ознаки генеральної сукупності? Сформулюйте деякі з них.

46. Яке спостереження називають вибіркоvim? Назвіть його суть і завдання.
47. У чому переваги вибіркового спостереження порівняно з суцільним?
48. Яких умов необхідно дотримуватися у разі відбору одиниць при вибіркоvimу спостереженні?
49. Чому при вибіркоvimу спостереженні завжди виникають помилки (похибки), як їх класифікують і що вони характеризують?
50. Які Ви знаєте види і способи відбору у вибіркоvu сукупність? Як здійснюють власне-випадковий відбір, механічний, типовий?
51. На що вказує коефіцієнт довіри і як його позначають?
52. Чи впливає обсяг вибірки на її точність і якою мірою?
53. Від чого залежить обсяг вибірки?
54. Як поширюються дані вибіркового спостереження на генеральну сукупність?
55. Як визначають необхідну чисельність вибірки для середньої і частки в разі повторного і неповторного відбору?
56. В чому полягають основні завдання статистичного вимірювання взаємозв'язків між явищами?
57. Який зв'язок називають функціональним і як він проявляється? Який зв'язок називають стохастичним і в чому полягає його суть?
58. Які Ви знаєте форми кореляційної залежності і як вони співвідносяться із стохастичним зв'язком?
59. У чому полягає суть лінії регресії і які основні моделі кореляційного зв'язку?
60. За допомогою якого методу оцінюють параметри регресійної моделі?
61. Як розраховують кореляційне відношення і що воно відображає?
62. Як перевіряють істотність зв'язку в аналітичному групуванні?
63. Які функції найчастіше застосовують для побудови регресійних моделей?
64. Які показники використовують для вимірювання щільності зв'язку в регресійній моделі і як їх розраховують?
65. Чим відрізняються параметричні та непараметричні методи вимірювання зв'язків? В яких випадках застосовують коефіцієнт асоціації?

Список джерел, використаних у розділі II

1. Аванесов В.С. Теория и методика педагогических измерений [Материалы публикаций в открытых источниках и Интернет]. – ЦТ и МКО УГТУ-УПИ, 2005. – 98 с.
2. Булах І.Є., Мруга М.Р. Створюємо якісний тест, навчально-методичний посібник, 2-е видання. – К.: Майстер-клас, 2009. – 176 с.
3. Звонников В.И., Челышкова М.Б. Контроль качества обучения при аттестации. – М.: Логос, 2011.– 272с.
4. Ким В.С. Тестирование учебных достижений. Монография.- Уссурийск: Изд-во УГПИ, 2007. -214 с.
5. Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов.- М.: Логос, 2010.- 668 с.
6. Майоров А. Н. Теория и практика создания тестов для системы образования. (Как выбирать, создавать и использовать тесты для целей образования) / А. Н. Майоров. – М.: Интеллект-центр, 2001. – 296 с.
7. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов.- М.: Прометей, 2000.- 168 с.
8. Челышкова Н. Б. Теория и практика конструирования педагогических тестов : учебное пособие / Н. Б. Челышкова. – М. : Логос, 2002. – 432 с.
9. Allen M.J., Yen W.M. Introduction to Measurement Theory. Belmont, CA: Wadsworth, 1979.
10. Andrich D. Advanced Social and Educational Measurement. Perth: Murdoch University, 2001.- 128 pp.
11. Brennan R. Educational Measurement.- Westport, CT: Praeger, 2006.- 796 pp.
12. Engelhard G.Jr., Wilson M. Objective measurement: Theory into practice, Vol. 3. Norwood, New Jersey: Ablex, 1996.
13. Gronlund N.E. Measurement and evaluation in teaching.- Macmillan Pub. Co., 1965 - 420 pp.
14. Gronlund N.E., Linn R.L. Measurement and evaluation in teaching, Sixth edition. New York: Macmillan, 1990.
15. Linn R.L., Miller M.D. Measurement and assessment in teaching, 9th Edition. Upper Saddle River, NJ: Prentice Hall, 2005.
16. Miller M.D., Linn R.L., Gronlund N.E. Measurement and assessment in Teaching, 10th Edition. Upper Saddle River, NJ: Pearson Education, Inc., 2009.

III. КОНСТРУЮВАННЯ ТЕСТІВ

На початку розділу введені основні поняття - тестового завдання та тесту, наведені вимоги до завдань у тестовій формі. Охарактеризовані різні форми, принципи формулювання та оцінювання тестових завдань: з вибором однієї та декількох правильних відповідей, на встановлення відповідності, на встановлення правильної послідовності, відкритої форми з короткою та розгорнутою відповіддю. Далі охарактеризовані основні етапи конструювання педагогічного тесту, описаний процес планування змісту тесту та загальні принципи його відбору. Приділено увагу принципам та підходам до експертизи якості змісту тесту, а також процесу розробки супровідної документації тесту та створення інструкцій для учня і для викладача.

Висвітлене питання створення моделі тестування тобто процесу створенню сукупності тестових завдань та опрацюванню результатів тестування, порядку пред'явлення завдань учасникам тестування, а також методу визначення його рівня знань за результатами тестування.

У заключному пункті наведені відомості про сучасний метод вимірювання в освіті - адаптивне тестування.

3.1. Тестові завдання та тест

3.1.1. Поняття завдання у тестовій формі, тестового завдання, тесту

Поняття завдання у тестовій формі, тестове завдання, тест є ієрархічно залежними і відповідають трьом основним етапам розробки педагогічних тестів:

1. Створення завдання в тестовій формі.
2. Створення тестових завдань.
3. Конструювання тесту.

Завдання в тестовій формі – це одиниця контрольного матеріалу, що відповідає низці формальних вимог і характеризується певним рівнем складності.

‘При розробці тесту враховується не тільки коректність змісту завдання, але й статистичні вимоги до нього. Не всі завдання в тестовій формі можуть стати тестовими завданнями. **Тестове завдання** – складова одиниця тесту,

яка відповідає не тільки формальним вимогам до завдань у тестовій формі, але й певним статистичним вимогам. Кожне завдання є невідомою частиною цілого – тесту, – який являю собою систему завдань. Видалення хоча б одного системного завдання неминуче призведе до виникнення прогалини на континуумі знань, що перевіряються, відповідно до погіршення якості вимірювання.

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

Вимоги до завдань у тестовій формі:

- мета;
- стилість;
- технологічність;
- логічна форма вислову;
- визначеність місця для відповідей;
- однаковість правил оцінки відповідей;
- правильність розташування елементів завдання;
- однаковість інструкції для всіх екзаменованих;
- адекватність інструкції формі і змісту завдання.

3.1.2. Порівняльний аналіз різних форм тестових завдань

Форми тестових завдань:

- завдання з вибором правильної відповіді (MCQ – Multiple Choice Question);
- завдання на встановлення послідовності (Arranging);
- завдання на встановлення відповідності (Correspondence);
- завдання з короткою відповіддю (Short Answer);
- завдання з розгорнутою формою відповіді (Essay);
- фасетні завдання.

Таблиця 3.1. Переваги та недоліки різних форм тестових завдань

Форма тестових завдань	Переваги	Недоліки
Завдання з вибором правильної відповіді (завдання закритої форми)	<p>Оцінювання результатів швидке, легке, об'єктивне, надійне.</p> <p>Завдання добре структуровані й чіткі.</p> <p>Дають можливість вимірювати результати навчання від простих до складних.</p> <p>Неправильні варіанти відповідей дають діагностичну інформацію про рівень сформованості знань і вмінь.</p>	<p>Складання запитань потребує багато часу.</p> <p>Часто складно знайти правдоподібні дистрактори.</p> <p>Завдання неефективні для оцінювання вміння розв'язувати проблеми.</p> <p>Рівень сформованості навичок читання може впливати на результати оцінювання.</p>
Завдання на встановлення послідовності	<p>Перевіряють сформованість алгоритмічного мислення.</p> <p>Читання варіантів відповідей потребує мало часу.</p> <p>Завдання добре структуроване й чітке.</p> <p>Результати менше залежать від угадування, ніж у завданнях з однією правильною відповіддю.</p>	<p>Обмежуються переважно оцінюванням простих результатів навчання.</p> <p>Такі завдання обмежують творчу діяльність тестованих.</p>
Завдання на встановлення відповідності	<p>Оцінювання результатів легке, об'єктивне й надійне.</p> <p>Запитання мають компакту форму.</p> <p>Щоб прочитати завдання та вибрати варіанти відповідей, тестований витрачає небагато часу.</p> <p>Завдання легко складати, якщо об'єднати кілька завдань, що мають однаковий набір варіантів відповідей.</p>	<p>Переважно оцінюють прості результати навчання на рівні знань, що базуються на асоціаціях.</p> <p>Важко складати завдання, які містять достатню кількість однорідних відповідей.</p> <p>Значна чутливість до невідповідних ключів.</p>
Завдання з короткою відповіддю	<p>Відповідь на запитання формулюється коротко й однозначно.</p> <p>Можуть оцінювати різноманітні результати навчання на рівні знань.</p>	<p>Складно формалізувати правильну відповідь.</p> <p>Перевірка завдань є досить трудомісткою й до певної міри суб'єктивною.</p> <p>Не можна застосовувати для</p>

Форма тестових завдань	Переваги	Недоліки
	<p>Завдання такої форми легко складати.</p> <p>Не треба шукати кілька варіантів відповідей.</p> <p>Угадування відповіді менш імовірно, ніж у завданнях на вибір відповіді.</p> <p>Необхідність відтворення відповіді по пам'яті.</p> <p>Простота перевірки.</p> <p>Використовують для розв'язання задач на обчислення, оцінювання знання термінів, фактів, назв подій тощо.</p>	<p>оцінювання складних результатів навчання.</p> <p>На оцінювання результатів може впливати рівень грамотності тестованого, якщо відповідь треба написати словами.</p>
<p>Завдання з розгорнутою формою відповіді</p>	<p>Оцінюють результати навчання найвищого рівня (уміння аналізувати, синтезувати, оцінювати).</p> <p>Підготовка завдань потребує менше часу порівняно із завданнями закритої форми</p> <p>Перевіряють уміння тестованих інтегрувати й застосовувати ідеї, висловлювати власне ставлення щодо запропонованої теми.</p>	<p>Оцінювання результатів є трудомістким і потребує багато часу.</p> <p>Оцінювання таких завдань є найбільш суб'єктивним порівняно з іншими формами завдань і часто знижує надійність тесту.</p> <p>Складно розрахувати час, необхідний для виконання завдання.</p> <p>Складно передбачити результати тестованих через високий ступінь свободи щодо структури твору (есе), стилю викладу та глибини розкриття теми.</p> <p>Каліграфічний почерк, хороше естетичне оформлення роботи й фальсифікації підчас проведення тестування (списування, порушення процедур тощо) підвищують оцінки, а поганий почерк, орфографічні й пунктуаційні помилки їх знижують.</p>

3.2. Завдання закритої форми

3.2.1. Завдання з вибором однієї правильної відповіді

Принципи формулювання та добору варіантів відповідей завдань з однією правильною відповіддю:

- **Суперечливість.** Характеризується використанням частки заперечення “не”, а також слів заперечень. При цьому друга відповідь заперечує не смисл самого завдання, а зміст першої відповіді. Використовується логічна схема “А або не А”.

Наприклад: *Трикутник зі сторонами 3, 7 і 9 см:*

А існує

Б не існує

Функція $y = 2x^2$ обернена

А має

Б не має

- **Протилежність.** Близький за змістом з попереднім принципом, але певною мірою відрізняється логічними властивостями та мовними тонкощами. Протилежні відповіді допускають можливість потенційного існування іншого перехідного стану.

Наприклад: *Значення екстремуму функції $y = -x^2 + 1$*

А додатне

Б від'ємне

Графік непарної функції симетричний відносно:

А вісі абсцис

Б вісі ординат

В початку координат

- **Однорідність.** Добір відповідей здійснюється за ознакою єдиного роду, виду, чи відображають основні сторони, грані явища.

Наприклад: *Тіло, що падає на землю у Північній півкулі, відхиляється на:*

А північ

Б південь

В захід

Г схід

- **Кумуляція.** Означає, що зміст другої відповіді включає в себе (акумулює) зміст першої, зміст третьої – другої і т. д. відповідей.

Наприклад: *Ступінь родючості ґрунтів обумовлює:*

А сукупність елементів

Б структура та сукупність елементів

В мікроорганізми, структура та сукупність елементів

- **Сполучення.** Використовується сполучення слів (знаків) по два чи по три в кожній відповіді. Сполучаються або однорідні пари відповідей, або одне слово з декількома іншими в усіх відповідях, або за принципом ланцюжка, коли останнє слово в першій відповіді стає першим в другій, останнє в другій – першим в третій і т. д.

Наприклад: *Ядро атому складається з:*

А нейтронів і протонів

Б протонів і електронів

В електронів і нейтронів

Принципи оцінювання завдань з однією правильною відповіддю: за правильну відповідь в кожному завданні загальноприйнято ставити один бал, за неправильну – нуль.

Формула корекції балів, отриманих у результаті виконання завдань з k відповідями, з яких один правильний:

$$X_{ci} = X_i - \frac{W_i}{k-1}$$

де X_{ci} - зкорегований на здогадку тестовий бал; X_i - тестовий бал випробовуваного і до корекції; W_i - число помилкових відповідей у цього випробовуваного; k - число відповідей в кожному завданні тесту.

3.2.2. Завдання з вибором декількох правильних відповідей

Змістову основу завдань цієї форми складають, головним чином, класифікаційні знання. Відповіді до завдання обов'язково повинні відноситися до одного роду або виду термінів. Кількість відповідей – від 5 до 12. При цьому співвідношення кількості правильних і неправильних відповідей серед усіх завдань повинно бути різним – від одного до усіх.

Особливість цієї форми завдань – в них доводиться визначати не тільки правильні відповіді, але й самостійно оцінювати повноту власної відповіді.

Наприклад: *Ссавцями є:*

- 1 *єхидна*
- 2 *кенгуру*
- 3 *качкодзьоб*
- 4 *черепашка*
- 5 *кіт*
- 6 *змія*
- 7 *дельфін*
- 8 *жаба*

Принципи оцінювання завдань з вибором декількох правильних відповідей:

1. за правильну відповідь (вибір усіх, без винятку, правильних відповідей) – один бал. За неправильну відповідь (навіть у випадку хоча б одного помилкового вибору) – нуль балів.
2. за кожну правильну відповідь особа отримує один бал і по мірі виконання завдання набирає певну кількість балів за вказані правильні відповіді.
3. за повну правильну відповідь – три бали, за одну допущену помилку зняти один бал, за другу помилку зняти ще один бал, за третю – зняти останній бал.

3.3. Завдання на встановлення відповідності та послідовності

3.3.1. Завдання на встановлення відповідності (логічні пари)

Завдання цієї форми дають змогу перевірити так звані асоціативні знання: взаємозв'язок фактів, подій, явищ, форм та змісту, співвідношення між різними властивостями, формулами, законами, датами тощо.

Назва форми відповідає виду діяльності учасників тестування – необхідно встановити відповідність елементів одного стовпчика з елементами другого стовпчика. Формальна вимога – різна кількість елементів у правому та лівому стовпцях. Правдоподібні елементи

розташовуються у правому стовпчику та виконують роль дистракторів. Їхня кількість має бути більшою за число елементів ліворуч. Літери та цифри використовуються як ідентифікатори елементів двох стовпчиків.

У кожне завдання на встановлення відповідності необхідно включати тільки однорідний матеріал. Бажано розташовувати відповіді в алфавітному або числовому порядку.

Наприклад: *Установіть відповідність між функціями та похідними*

A $y = \sin 2x \cos x$

B $y = \cos 2x \sin x$

B $y = 2(\sin^3 x - \sin x)$

Г $y = 2(\cos^2 x - \cos x)$

1 $y = \cos^2 x - 5\sin^2 x \cos x$

2 $y = 4\cos x - 6\cos^3 x$

3 $y = 2\cos^3 x - 4\sin^2 x \cos x$

4 $y = 2\sin x - 6\cos^2 x \sin x$

5 $y = 6\sin^2 x \cos x - 2\cos x$

6 $y = 6\sin^3 x + 2\sin x$

Принципи оцінювання завдання на встановлення відповідності:

1. За правильну відповідь (встановлення усіх, без винятку, правильних відповідностей) – один бал. За неправильну відповідь (навіть у випадку хоча б одного помилкового вибору) – нуль балів.
2. За кожну правильно встановлену відповідність особа отримує один бал і по мірі виконання завдання набирає певну кількість балів за вказані правильні відповіді.
3. За усі правильні відповіді давати два бали, за одну помилку знижувати оцінку до одного балу, за дві та більше – нуль балів.

3.3.2. Завдання на встановлення правильної послідовності

Варіанти завдань даної форми:

- Послідовність історичних подій.
- Послідовність дій і операцій.
- Послідовність різних процесів (історичних, літературних, художніх, лінгвістичних тощо).
- Ланцюжок розумових дій, який утворює систему знань, умінь, уявлень.
- Перевірка знань певних понять і термінів різних дисциплін.

Наприклад: *Установіть заповнення орбіталей електронами*

А 4s

Б 3p

В 4f

Г 3d

Д 5p

Назвемо принципи оцінювання завдання на встановлення правильної послідовності. Оцінка може варіюватися в залежності від важливості, значимості та складності конкретного завдання. Найчастіше використовується дихотомічна оцінка 0–1 бал, при якій один бал дається за правильну розстановку усіх рангів у завданні, нуль – при помилці у відповіді. Характерно, якщо помилка зроблена на початку, то і наступні ранги визначаються неправильно.

Безпомилкове рішення оцінюється, наприклад, у три бали. Помилка в кінці завдання знижує оцінку на один бал, помилка всередині на два, а помилка на початку – на три бали.

3.4. Завдання відкритої форми

3.4.1. Завдання відкритої форми з короткою відповіддю

Специфічність цієї форми завдань полягає в тому, що в них не пропонується перелік відповідей. Особа повинна сама дописати відповідь, яка свідчить про наявність або відсутність необхідних знань.

Завдання відкритої форми використовуються там, де необхідно повністю виключити отримання правильної відповіді шляхом вгадування і тим самим підвищити якість педагогічного вимірювання.

Етапи створення завдання відкритої форми з короткою відповіддю

1. Формулювання короткого та точного запитання.
2. Формулювання короткої та точної відповіді.
3. Вилучення з відповіді ключового слова, яке особа повинна доповнити.

Наприклад:

Оператор _____ пов'язує файл на диску з файловою змінною, а оператор _____ відкриває файл для читання.

Принципи оцінювання: Використовується дихотомічна оцінка 0–1 бал, при якій один бал дається за правильну відповідь на завдання, нуль – при помилці у відповіді.

3.4.2. Завдання відкритої форми з розгорнутою відповіддю

Тестований повинен написати твір (есе) або розв'язати задачу.

Етапи розробки критеріїв оцінювання відповідей на відкриті запитання:

- Розробити схему оцінювання ще до проведення тесту паралельно зі складанням переліку вимог до завдання в основній частині тесту.
- Написати повну відповідь на запитання.
- Перелічити складові правильної відповіді та вказати, скільки балів тестований отримає за кожну складову, навести приклади.

Схема оцінювання:

- Указати максимальний бал
- Навести приклади повністю правильних відповідей
- Надати опис елементів, які мають бути присутні для отримання повного балу, із зазначенням того, скільки балів тестований отримає за кожний елемент
- Навести приклади частково правильних відповідей
- Надати опис елементів, які можуть бути частково зараховані, та кількість балів, які тестований за них отримає.

Принципи оцінювання завдань відкритої форми з розгорнутою відповіддю: діапазон оцінювання може змінюватися: 0-1, 0-2, 0-3, 0-4 бали тощо (переважно оцінювання проводить людина).

3.5. Основні етапи конструювання педагогічного тесту

3.5.1. Алгоритм створення тесту

Процес створення тесту, його наукового обґрунтування, переробки і покращення Челишкова Н.Б. пропонує розбити на ряд етапів.

1. Визначення мети тестування, вибір виду тесту і підходу до його створення.
2. Аналіз змісту учбової дисципліни.

3. Визначення структури тесту і стратегії розташування завдань.
4. Розробка специфікації тесту, апіорний вибір довжини тесту і часу його виконання.
5. Створення тестових завдань.
6. Відбір завдань в тест і їх ранжування за обраною стратегією пред'явлення на основі апіорних авторських оцінок труднощі завдань.
7. Експертиза змісту тестових завдань і тесту.
8. Експертиза форми тестових завдань.
9. Переробка змісту і форми завдань за результатами експертизи.
10. Розробка методики апробаційного тестування.
11. Розробка інструкцій для учнів і для викладачів, які проводять апробацію тесту.
12. Проведення апробаційного тестування.
13. Збір емпіричних результатів.
14. Статистична обробка результатів виконання тесту.
15. Інтерпретація результатів обробки з метою покращення якості тесту. Перевірка відповідності характеристик тесту науково обґрунтованим критеріям якості.
16. Корекція змісту і форми завдань на основі даних попереднього етапу. Чистка тесту і додання нових завдань для оптимізації діапазону значень параметра труднощі та покращення системотвірних властивостей завдань тесту. Оптимізація довжини тесту і часу його виконання на основі апостеріорних оцінок характеристик тесту. Оптимізація порядку розташування завдань в тесті.
17. Повторення етапу апробації для виконання чергових кроків для підвищення якості тесту.
18. Інтерпретація даних обробки, встановлення норм тесту і створення шкали для оцінки результатів випробовуваних.

Із алгоритму створення тесту нескладно бачити, що виникає своєрідний цикл, тому що після чистки тесту розробнику приходиться повертатися до етапу збору емпіричних даних, причому, зазвичай, не один раз, а два, три і більше. У певному сенсі цей цикл нескінченний, але не тому, що всі завдання погані та розробник не має достатнього досвіду створення тесту. Просто процес конструювання тесту досить важкий, тому що необхідний

усесторонній підхід до оцінки якості тесту і характеристик тестових завдань, до перевірки їх системотвірних властивостей.

Взагалі кажучи, можна вважати, що задача оптимального підбору складу тесту не має єдиного рішення, тому що не все тут визначається якістю тестового матеріалу, а багато ще залежить від рівня підготовки групи учнів. Завдання, які добре працюють на одній вибірці учнів, можуть виявитися непридатними на іншій, тому що вони будуть або дуже легкими, або дуже важкими і їх не виконає правильно ні один учень групи.

Успіх створення тесту багато в чому залежить від високої якості навчального тестового матеріалу, яка забезпечується правильним відбором змісту, що буде перевірятися, і умінням розробника коректно відобразити його в завданнях тесту. Дуже важливим є також етап обробки емпіричних результатів тестування, для виконання якого необхідні спеціальні програмні засоби для професійної розробки тестів.

3.5.2. Модифіковані алгоритми створення тестів різних типів

Майоров О.М. наводить трохи інший але схожий перелік етапів створення педагогічних тестів. Він також наводить приклад двох переліків етапів, які рекомендовані Голландським інститутом СІТО та екзаменаційним синдикатом Кембриджського університету.

Для нормативно-орієнтованих та критеріально-орієнтованих тестів на етапах конструювання вирішуються різні задачі, які подаються у таблиці 3.2.

Таблиця 3.2. Задачі тестів

Етапи	Нормативно-орієнтовані Тести	Критеріально-орієнтовані тести
Визначення мети тестування	Оцінка учня відносно інших учнів	Опис засвоєного або не засвоєного учбового матеріалу. Оцінка учня відносно набору характеристик достатніх для досягнення результату
Необхідність в запрошенні фахівців	Орієнтація на статистичні методи оцінювання. Спеціалісти необхідні для статистичної обробки результатів	Орієнтація на експертне оцінювання, проведення робіт по визначенню переліку і обсягу учбових задач, критерію досягнення необхідного рівня підготовки. Спеціалісти для експертизи учбового матеріалу.

Етапи	Нормативно-орієнтовані Тести	Критеріально-орієнтовані тести
Відбір змісту учбового матеріалу	Може бути відібраний широкий спектр матеріалу.	Відбирається матеріал, який стосується обмеженої тематики
Апробація тестових завдань	У тому числі для визначення труднощі і дискриміна-тивності	Для відпрацювання змісту завдань
Корекція змісту і форми завдань	Відбираються завдання з максимальною дискриміна-тивністю. Дуже прості і дуже складні завдання відкидаються	Відбираються завдання, які максимально покривають область, що відповідає критерію. Складні і прості завдання включаються в тест
Нормування тесту	Статистичні норми, що розраховуються для нормативної вибірки, або визначення місця учня відносно інших учнів, які брали участь в тестуванні	Опис критерію досягнення результату або опис тих учбових задач, які учень повинен виконати.

3.6. Планування змісту тесту та загальні принципи його відбору

Після визначення цілей тестування і їх конкретизації необхідно *розробити план і специфікацію тесту*.

3.6.1. Розробка плану тесту

При розробці плану виконується приблизна розкладка процентного співвідношення змісту розділів і визначається необхідне число завдань з кожного розділу дисципліни виходячи з важливості розділу і кількості годин, відведених на його вивчення в програмі.

Розкладка починається з підрахунку планованого числа завдань в тесті, яке потім неодноразово може змінюватися. Зазвичай граничне число не перевищує 60 - 80 завдань і це залежить від часу тестування та врахування середнього часу виконання завдання тесту. Загалом план задається у вигляді таблиці, у якій міститься інформація: назва розділу (теми або змісту), який буде контролюватися; кількість завдань з цього розділу, номери завдань. Також таблиця може містити інші необхідні дані.

3.6.2. Два підходи до розробки короткої специфікації тесту

Після виконання першого кроку по плануванню змісту розробляється специфікація тесту, в якій фіксується структура, зміст перевірки і процентне

співвідношення завдань в тесті. Інколи специфікацію роблять у розгорнутій формі, яка містить вказівки на тип завдань, що будуть використовуватися для оцінки досягнень учнів у відповідності з наміченими цілями тесту, час виконання тесту, число завдань, особливості проведення тестування, які можуть вплинути на характеристики тесту і т.д.

Перший підхід до створення короткої специфікації базується на спряженні системи знань і вмінь з процентним співвідношенням завдань по різних розділах або по різних змістових лініях дисципліни. Приклад реалізації подібного спряження наведений Чельшковою Н.Б. (таблиця 3.3.). Запропонований в ній перелік знань і умінь має досить загальний характер. До нього включені: А – знання понять, означень, термінів; В – знання законів і формул; С – уміння застосовувати закони і формули до розв'язання задач; D – уміння інтерпретувати результати на графіках і схемах; Е – уміння проводити оціночні судження.

Таблиця 3.3. Гіпотетична специфікація тесту

№ п/п	Заплановані до перевірки знання і уміння	Змістові лінії (розділи) дисципліни				Сумарне число по кожному пункту
		I (20%)	II (10%)	III (30%)	IV (40%)	
1	A(10%)	1	1	2	2	6
2	B(20%)	2	1	4	5	12
3	C(30%)	4	2	5	7	18
4	D(30%)	4	2	5	7	18
5	E(10%)	1	-	2	3	6
Разом		12	6	18	24	60

У процесі роботи над тестом початкова розкладка завдань буде змінюватися. Це пояснюється тим, що не всі завдання виявляться вдалим і доцільними в тій мірі, як це вважається на етапі планування тесту. Тому якусь одну частину завдань розробник вилучає після експертизи завдань, а другу – в процесі поглибленого аналізу результату статистичної обробки емпіричних даних виконання тесту. У цьому зв'язку після корекції тесту необхідна доробка специфікації для приведення її у відповідність з остаточними пропорціями змісту тесту.

При другому підході специфікація містить процентне співвідношення завдань, яке планується відповідно до пропорцій розділів і видів передбачуваної діяльності випробовуваного в процесі виконання тесту. Приклад подібної специфікації наведений Чельшковою Н. Б. (таблиця 3.4.).

Таблиця 3.4. Коротка специфікація підсумкового тесту з фізики

Зміст предмету	Передбачувана діяльність випробовуваного		
	Репродуктивний рівень		Продуктивний рівень
	Знання визначень, фактологічного матеріалу і т.п.	Застосування знань основних законів в знайомій ситуації за зразком, на основі узагальнюючого алгоритму	Застосування знань в незнайомій ситуації
Загальноосвітня частина тесту (40 завдань) <i>Із них:</i>	8 завдань (20%)	24 завдання (60%)	8 завдань (20%)
Механіка (10 завдань, 25%)	2 завдання	6 завдань	2 завдання
Молекулярна фізика і термодинаміка (8 завдань, 20%)	1 завдання	5 завдань	2 завдання
Електродинаміка, СТО (16 завдань, 40%)	3 завдання	10 завдань	3 завдання
Геометрична оптика (2 завдання, 5%)	1 завдання	1 завдання	-
Квантова фізика (4 завдання, 10%)	1 завдання	2 завдання	1 завдання
Поглиблена частина тесту (20 завдань)	-	4 завдання (20%)	16 завдань (80%)

3.6.3. Загальні принципи правильного відбору змісту тестів

Існують загальні принципи, які сприяють у певній мірі правильному відбору змісту тестів.

Принцип репрезентативності регламентує не тільки повноту відображення, але і значимість змістовних елементів тесту. Зміст завдань повинен бути таким, щоб по відповідях на них можна було зробити висновок про знання або незнання всієї програми розділу або курсу, що перевіряється.

Принцип системності передбачає підбір змістових елементів, що відповідають вимогам системності і зв'язаних між собою загальною структурою знань. За умови дотримання принципу системності тест можна використати для виявлення не тільки обсягу знань, але і для оцінки якості структури знань учнів.

3.7. Експертиза якості змісту тесту

Успіх створення тесту багато в чому залежить від якості навчального тестового матеріалу, яка забезпечується правильним плануванням змісту і

специфіки тесту та вмінням розробника коректно реалізувати цей план при розробці завдань тесту.

Оцінка якості змісту тесту зазвичай проводиться за певною методикою незалежними експертами, які не брали участі у розробці тесту. Зазвичай, число експертів становить не менше трьох чоловік по кожному тесту.

Методика експертизи якості тесту зазвичай включає три розділи, які вибудовуються відповідно трьом напрямкам роботи експертів. Перед початком роботи кожний експерт повинен познайомитися зі специфікацією репрезентованого тесту, яка містить пояснення по його структурі та планованому до перевірки змісту.

3.7.1. Експертиза окремих завдань тесту

Робота експерта по першому напрямку полягає в аналізі змісту окремих завдань тесту. Експерт повинен заповнити таблицю, яка, наприклад, розроблена для атестаційного тесту випускників школи.

1.	№ завдання	1	2	3	...
2.	№ правильної відповіді				
3.	№ вимоги (одної або декількох)				
4.	Рівень базовості (Б/П/С)				
5.	Значимість змісту завдання (0,1,2,3)				
6.	Очікуваний процент виконання випробовуваними із задовільною підготовкою				
7.	Очікуваний процент виконання випробовуваними вибірки				
8.	Очікуваний час виконання завдання, хв.				
9.	Невдалі завдання				
10.	Коментарі про якість завдання				

У другому рядку наводяться номери вибраних експертом правильних відповідей до закритих завдань тесту. Відповіді до відкритих завдань наводяться експертом поза таблицю на окремому листку. Для відкритих завдань з короткою відповіддю в таблицю заноситься ця відповідь.

У третьому рядку розташовуються результати аналізу змісту завдання в термінах вимог до рівня підготовки. Експерт зіставляє зміст кожного завдання тесту з переліком занумерованих вимог, які входять в комплект матеріалів для експертизи. Якщо окреме завдання контролює декілька вимог, то

експерт наводить тільки одну вимогу, яка в основному перевіряється цим завданням.

У четвертому рядку таблиці проставляється рівень базовості. Диференціація завдань по рівнях дозволяє розділити їх на три групи: Б – завдання перевіряє ступінь досягнень на мінімальному рівні, П – завдання дозволяє виставити випробовуваному добру або відмінну оцінку, С – завдання виходить за межі мінімуму змісту і рівня вимог до підготовки випробовуваних.

Оцінку значимості змісту завдання експерти виставляють за чотирибальною шкалою: 0 – завдання не виправдано включено в тест, 1 – зміст завдання є незначимим, 2 – значимим, 3 – найважливішим опорним елементом.

Очікуваний процент виконання випробовуваними із задовільною підготовкою – це орієнтований процент від гіпотетичної репрезентативної вибірки випробовуваних із задовільною підготовкою, які здатні виконати кожне завдання тесту.

У сьомому рядку експерт також наводить орієнтовний процент випробовуваних, які правильно виконали кожне завдання тесту, але орієнтуючись в своїх гіпотетичних оцінках не на групу «трійочників», а на вибірку, де рівномірно представлені всі категорії відміток від двох до п'яти.

При виборі очікуваного часу виконання завдання експерт повинен орієнтуватися на учня з середньою підготовкою.

У дев'ятому рядку таблиці експерт проставляє зірочки на місцях, що відповідають номерам невдалих завдань тесту.

Всі зауваження до завдань відмічаються в останньому рядку.

Майоров А.Н. наводить найпростіший варіант експертизи завдань, де експертам пропонується за трьох- або п'ятибальною шкалою оцінити завдання з точки зору

- відповідності цілям тестування (відповідає – частково відповідає – не відповідає),
- однозначності формулювань (однозначна – не зовсім однозначна – неоднозначна),
- придатності варіантів відповідей (придатні – частково придатні – не придатні).

Також запропонований якісний варіант змістовної експертизи тестових завдань. Західні тестологи пропонують використовувати для експертизи

тестових завдань скрінінг, розуміючи під скрінінгом систематичну перевірку всіх аспектів питання. Оцінка завдань в рамках скрінінга проводиться за 6 розділами. Експертам видається оціночний лист, основний зміст якого складається з таких розділів: доречність, застосування мови, інформація, питання і відповідь, варіанти відповіді (для закритих завдань), застосування контексту.

3.7.2. Експертиза якості змісту паралельних варіантів тесту

Другий напрямок роботи експерта пов'язаний з аналізом якості змісту всього тесту, що має декілька паралельних варіантів. Аналіз виконується для кожного варіанту окремо, а потім вся інформація об'єднується по групі варіантів в тих таблицях, які пропонуються в методиці експерту. При виконанні робіт за другим напрямком треба мати на увазі, що зміст, наприклад, атестаційного тесту визначається як оптимальне відображення вимог до рівня підготовки випускників в системі завдань тесту. Вимога оптимальності виділяє певні критерії якості відображення.

Перший критерій – повнота охоплення вимог до рівня підготовки випускників кожним варіантом тесту і всіма варіантами разом.

Другий критерій якості – правильність пропорцій змісту тесту.

Третій критерій – перевірка відповідності змісту рецензованої системи завдань специфікації тесту.

3.7.3. Заключний етап експертизи: висновки та рекомендації

Третій напрямок роботи експерта розрахований на підготовку узагальнюючих висновків і рекомендацій на покращення змісту тесту. У третьому розділі рецензії експерт наводить своє загальне враження про зміст тесту. Тут можуть бути висловлені всі сумніви і побажання експерта, його рекомендації на покращення змісту. Можлива оцінка співвідношення завдань, які перевіряють знання теорії предмету і його практики. Бажано виявити завдання, які призначені для перевірки системи понять, а також завдання інтегративного характеру, призначені для перевірки вмінь учнів узагальнювати знання по різним розділах предмету.

3.8. Розробка супровідної документації тесту та створення інструкцій для учня і для викладача

3.8.1. Методичне оснащення тесту

Створення тесту передбачає розробку для його проведення необхідної супровідної документації, яка складає методичне оснащення тесту. Методичне оснащення повинно вирішувати одну із основних задач тестування – задачу об'єктивності отримуваних результатів – воно повинно забезпечувати однаковість умов для всіх випробовуваних.

Методичне оснащення – складова частина тесту, яка складається із комплексу відомостей, інструкцій і рекомендацій, які в комплексі забезпечують всім випробовуваним рівні умови на всіх етапах використання тесту. Методичне оснащення включає в себе декілька частин:

- відомості, які необхідно знати користувачу тесту про інструмент;
- правила і вимоги пред'явлення тесту випробовуваним;
- правила обробки результатів;
- рекомендації по інтерпретації отриманих результатів.

Підбір оснащення залежить від призначення та виду тестів і від кількості учасників.

Основним видом тестів шкільних досягнень є групові тести. Розробка методичного оснащення буде відрізнятися залежно від того, який спосіб пред'явлення тесту вибраний – індивідуальні тести, тести з використання тестових зошитів або бланків, тести виконувани на комп'ютері.

У випадку використання тестових зошитів, всі помітки, обчислення, відповіді робляться випробовуваними тільки в тестовому зошиті, який містить необхідні інструкції для учня і всю необхідну інформацію по тестових завданнях. Проведення тестування з використанням тестових зошитів найбільш надійний якісний і об'єктивний спосіб заповнення тестового інструменту. Всі міжнародні дослідження, вся атестація учнів в розвинутих країнах проходить тільки з використанням тестових зошитів.

3.8.2. Пояснювальні та інструктивні матеріали до тесту

Керівництво до тесту пишеться для його користувачів, у ньому подаються основні відомості про тест:

- 1) призначення і педагогічний або психолого-педагогічний зміст;

- 2) обмеження і показники для застосування;
- 3) склад тесту;
- 4) інформація про апробацію тесту (цілі апробації, об'єм і склад вибірки, основні статистичні характеристики);
- 5) інструкція для того, хто проводить тест (вказівки до проведення тесту);
- 6) ключі;
- 7) дані про трудність і дискримінативність завдань;
- 8) дані про надійність, валідність і труднощі всього тесту;
- 9) інші статистичні матеріали;
- 10) правила обробки даних;
- 11) будова шкал;
- 12) правила і особливості інтерпретації результатів.

Інструкції для того, хто проводить тест (зазвичай це педагог, який керує процесом застосування тесту), і випробовуваного (зазвичай це учень, який виконує тест) входять до складу всіх стандартизованих тестів і визначають його об'єктивний характер. Інструкція для педагога наводиться, зазвичай, в керівництві до тесту, а інструкція для учня – в тестовому зошиті.

Склад інструкції для того, хто проводить тест:

- 1) умови проведення тесту;
- 2) матеріали і прилади, що вимагаються при проведенні (олівці, резинки, прилади, чернетки, бланки для відповідей, відношення до використання калькуляторів та ін.); порядок їх розміщення і пред'явлення;
- 3) часові обмеження;
- 4) повний текст інструкції для випробовуваного;
- 5) поведінка експериментатора під час проведення тесту (що він зобов'язаний, що може робити і що йому заборонено);
- 6) опис відповідей на можливі типові питання, у тому числі варіант відповіді на ті питання, на які йому відповідати заборонено;
- 7) вказівка про вирішення питання про вгадування;
- 8) необхідна кваліфікація для осіб, які проводять тестування, перевіряють правильність виконання, аналізують результати;
- 9) питання конфіденційності і доступу до отримуваної в результаті тестування інформації;

- 10) інструкція для перевірки результатів і занесення результатів на магнітний носій.

Склад інструкції для випробовуваного:

- 1) опис призначення тесту (якщо це не суперечить умовам тестування);
- 2) правила заповнення бланків для відповідей (тестових зошитів);
- 3) зразки розв'язання задач-зразків, що найменше, по одному на кожний тип завдань, поданих в тесті;
- 4) зразки виправлення невірно виконаного завдання і зразки виправлення виправленого.

Викладення інструкцій повинно бути ясным, доступним і детальним. Від цього залежить розуміння їх випробовуваними і відсутність лишніх питань.

3.9. Моделі педагогічного тестування

У процесі тестування значна увага приділяється створенню сукупності тестових завдань та опрацюванню результатів тестування. Крім того, важливе значення має порядок пред'явлення завдань учню/студенту і метод визначення його рівня знань за результатами тестування, тобто *модель тестування*. Виокремлення поняття моделі пов'язано з тим, що класифікація стратегій тестування тісно пов'язана з методикою інтерпретації тестових результатів.

Модель педагогічного тестування – це схема пред'явлення тестових завдань і оцінювання результатів тестування. Існують різні класифікації моделей і їх реалізація фактично можлива з використанням комп'ютерних технологій. Розглянемо основні моделі.

3.9.1. Класична модель

Дана модель є найпершою і найпростішою.

Є n завдань з певної області знань, з декількох областей знань або частини області знань (розділу, теми і т. п.). Із цієї множини завдань випадковим чином вибирається k завдань ($k < n$), які пропонуються випробовуваному. Випробовуваний вибирає або вводить правильну, на його думку, відповідь. Результат відповіді за кожне завдання оцінюється як

«правильно» або «неправильно». Результатом тестування є процент правильних відповідей випробовуваного.

Переваги: простота реалізації.

Недоліки:

1. із-за випадкової вибірки не можна наперед визначити, які завдання за складністю попадуться випробовуваному – одному випробовуваному можуть дістатися легкі завдання, іншому – складні;
2. оцінка залежить тільки від кількості правильних відповідей і не враховується складність завдань.

Класична модель із-за своїх недоліків має найнижчу надійність, тому що відсутність обліку параметрів завдань часто не дозволяє об'єктивно оцінити знання випробовуваного.

3.9.2. Класична модель з урахуванням складності завдань

Ця модель є узагальненням першої моделі. Кожне завдання має певний рівень складності T_i , $i=1\dots n$. Порядок вибірки завдань і їх оцінки такий самий як у першій моделі. При підрахунку результату тестування враховується складність завдання, на яке випробовуваний відповів правильно. Чим вища складність завдання, тим вищий результат тестування. Для завдань, на які була дана неправильна відповідь, складність не враховується.

Недолік моделі – це недолік 1 першої моделі.

3.9.3. Модель за зростанням складності

Є n завдань з певної області знань, з декількох областей знань або частини області знань (розділу, теми і т. п.). Кожне завдання має певний рівень складності T_i , $i=1\dots n$. Є m рівнів складності. У тесті повинні бути завдання всіх рівнів складності. Із цієї множини завдань випадковим чином вибирається k завдань ($k < n$). Вибрані завдання сортуються за зростанням складності, після чого пропонуються випробовуваному. Кількість завдань з кожного рівня складності повинно бути однаковим, або розподіл завдань за рівнями складності повинен підпорядковуватися нормальному закону.

Результати тестування обраховуються аналогічно моделі з урахуванням складності. Дана модель забезпечує паралельність тестів за складністю, тобто надійність результатів ще вища, ніж в попередніх моделях.

3.9.4. Модель з розділенням завдань за рівнями засвоєння

Розрізняють п'ять рівнів засвоєння учбового матеріалу:

1. Нульовий рівень (Розуміння).
2. Перший рівень (Впізнавання).
3. Другий рівень (Відтворення).
4. Третій рівень (Застосування).
5. Четвертий рівень (Творча діяльність).

Рівень позначимо через α і він може змінюватися від 0 до 4. Завдання створюються для кожного із п'яти рівнів. Спочатку проводиться тестування з використанням завдань по рівню 0, потім по рівню 1, 2, і т. д. Перед переходом з рівня на рівень обчислюється степінь володіння учбовим матеріалом на даному рівні і визначається можливість переходу на наступний рівень. Для вимірювання степені володіння учбовим матеріалом на кожному рівні використовується коефіцієнт $K_\alpha = P_1/P_2$, де P_1 – кількість правильно виконаних суттєвих операцій в процесі тестування, P_2 – загальна кількість суттєвих операцій. Під суттєвими розуміють ті операції, які виконуються на рівні α , який перевіряється. Операції, що належать до більш низьких рівнів, в число суттєвих не входять. Тому $0 \leq K_\alpha \leq 1$.

Таким чином рівень засвоєння учбового матеріалу може бути використаний для оцінки якості знань у випробовуваного і виставлення оцінки. Рекомендуються наступні критерії для виставлення оцінки:

- $K_\alpha < 0,7$ – незадовільно;
- $0,7 \leq K_\alpha < 0,8$ – задовільно;
- $0,8 \leq K_\alpha < 0,9$ – добре;
- $0,9 \leq K_\alpha$ – відмінно.

При $K_\alpha < 0,7$ потрібно продовжити процес навчання на тому ж рівні.

3.9.5. Модель з урахуванням часу відповіді на завдання

У даній моделі при визначенні результату тестування враховується час відповіді на кожне завдання. Врахування часу відповіді може здійснюватися за формулам:

$$R_i = \begin{cases} (t_{\max} - t_{\text{відп}}) / t_{\max}, & \text{якщо відповідь правильна,} \\ 0, & \text{якщо відповідь неправильна.} \end{cases}$$

Якщо $R_i < 0$ при правильній відповіді, то покладаємо $R_i = 0$.

У цих формулах R_i - результат відповіді на i -те завдання; t_{\max} - максимальний час, який відводиться на виконання i -того завдання; $t_{\text{відп}}$ - час затрачений випробовуваним на виконання завдання, тому $0 \leq R_i < 1$. Параметр t_{\max} може задаватися як константа для всіх завдань тесту або обчислюватися для кожного окремо взятого завдання в залежності від його складності, тобто $t_{\max} = f(T_i)$, тому що логічно припустити, що для відповіді на складне завдання потрібно більше часу, ніж на просте завдання. Інша можлива залежність параметра t_{\max} - від індивідуальних здібностей випробовуваного, які повинні бути визначені раніше.

$$\text{Результат тестування } R = \frac{\sum_{i=1}^k R_i}{k}, \quad R \in [0,1).$$

Модель з урахуванням часу відповіді на завдання також дозволяє підвищити надійність результатів тестування, особливо разом з моделлю з урахуванням складності завдань.

3.9.6. Модель з обмеженням часу на тест

Є n завдань з певної області знань, з декількох областей знань або частини області знань (розділу, теми і т. п.). Із цієї множини завдань випадковим чином вибирається k завдань ($k < n$), які пропонуються випробовуваному, і вказується максимальний час на проходження тесту. Для оцінки результатів тестування беруться тільки ті завдання, на які встиг відповісти випробовуваний за даний час. Сам тест може бути побудований за моделями 1-4.

У деяких роботах рекомендується обов'язково сортувати завдання за зростанням складності та встановлювати такий час тестування, за який на всі завдання не зможе відповісти ні один навіть найсильніший випробовуваний. Такий підхід пропонується застосовувати при тестуванні на бланках, коли випробовуваний бачить перед собою зразу всі завдання. Суть його в тому, що коли випробовуваний відповість на завдання, а час у нього ще залишається, він може почати перевіряти свої відповіді, сумніватися, а в підсумку може виправити правильні відповіді на неправильні. Тому рекомендується або обмежувати час на тест або забирати бланк зразу після відповіді на всі завдання.

Окрім розглянутих найпростіших моделей існують також інші складніші моделі, які комбінуються з використанням розглянутих шести моделей.

Одну із перших класифікацій моделей педагогічного тестування запропонував Г.Давідян: класична модель, класична модель з урахуванням складності, модель зі зростаючою складністю, модель з розподілом завдань за рівнями засвоєння, модель з урахуванням часу відповіді на завдання, модель з обмеженням часу на тест.

О.Матвієвська пропонує таку критеріально-рівневу типологію моделей: класична, класична з урахуванням складності, за зростанням складності, з розділенням знань за рівнями засвоєння, з урахуванням часу відповіді, з обмеженням часу, адаптивна, на нечіткій математиці.

Л.Зайцева, Н.Прокофєва виділяють такі методи організації контролю знань:

- неадаптивні (строга послідовність, випадкова вибірка, комбінований метод, в основі якого випадкова вибірка доповнена строгою послідовністю);
- частково-адаптивні (випадкова вибірка з урахуванням окремих параметрів моделі студента – кожному студенту генерується набір завдань, який відповідає його рівню підготовленості; контроль на основі відповіді студента – за заздалегідь розробленим сценарієм вказують наступне питання залежно від правильності відповіді; контроль на основі навчального матеріалу – послідовність видачі завдань аналогічна послідовності вивчення навчального матеріалу; модульно-рейтинговий метод);
- адаптивні (контроль за моделлю студента, контроль за моделлю студента і навчального матеріалу).

Існують також інші класифікації. Сьогодні за допомогою комп'ютерних технологій починають впроваджуватися адаптивні моделі тестування.

3.10. Адаптивне тестування - сучасний метод вимірювання в освіті

3.10.1. Поняття адаптивного тестування та методики його організації

Адаптивне тестування – це широкий клас методик тестування, що передбачають зміну послідовності завдань в процесі тестування з врахуванням відповідей тестованого на попередні завдання. В процесі

проходження тесту створюється модель тестованого, яка використовується для вибору наступного завдання в залежності від цього рівня підготовленості.

Один з можливих підходів отримання експертної оцінки знань – використання технології штучного інтелекту та експертних систем.

Математична модель адаптивного контролю знань визначає рівень навченості тестованих залежно від труднощі завдань. Теоретичною основою адаптивного контролю є теорія IRT у поєднанні з дидактичним принципом індивідуалізації навчання. Інший підхід до створення педагогічних тестів і до інтерпретації результатів їх виконання представлений в так званій сучасній теорії педагогічних вимірів Item Response Theory (IRT), що отримала широкий розвиток в 60-і - 80-і роки у ряді західних країн (див. розділ V).

До найбільш значущих переваг IRT відносять вимір значень параметрів тестованих і завдань тесту в одній і тій же шкалі, що дозволяє співвіднести рівень знань будь-якого тестованого з мірою труднощі кожного завдання тесту. Саме на цій властивості оцінок параметрів тестованих і завдань заснована організація сучасного адаптивного контролю знань. Але неможливо точно виміряти знання тестованих різного рівня підготовки за допомогою одного і того ж тесту. Це одна з причин того, що в практиці прагнули зазвичай створювати тести, розраховані на вимір знань тестованих найчисленнішого, середнього рівня підготовленості. Природно, що при такій орієнтації тесту знання у сильних і слабких тестованих вимірювалися з меншою точністю.

Обробка результатів тестування по IRT: визначається групова адаптивність, на відповідність середнього логіка складності, завдань тесту $V_{\text{ср.}}$ і середнього логіка навченої випробовуваних $Q_{\text{ср.}}$:

$$A_{\text{гр}} = 1 - [Q_{\text{ср.}} - V_{\text{ср.}}]$$

Групова адаптивність $A_{\text{гр.}} = 1$ при ідеальній відповідності $V_{\text{ср.}}$ і $Q_{\text{ср.}}$. Результати попереднього тестування групи тестованих мають значення $A_{\text{гр.}}$ далеко не рівними одиниці, тому наступним кроком є зміна значення групової адаптивності шляхом виключення з тесту "непрацюючих" завдань в цій групі тестованих з $V=0$ і визначення рівня підготовленості кожного тестованого, а також отримання в цьому випадку індивідуальних характеристичних кривих тестованих. Надалі визначається істинний бал як сума усієї вірогідності відповідей кожного тестованого на кожне завдання тесту і оцінюється рівень знань.

Таким чином, з'являється можливість для цього тестованого вибрати відповідний його рівню набір тестових завдань і тестовані можуть бути протестовані завданнями підібраними індивідуально для них. Для групи тестованих створюються адаптивні тести, що мають різну довжину і час виконання для сильних, слабких і середніх учнів цієї групи. Такий процес вимагає комп'ютерної технології створення, зберігання тестів, проведення тестування і обробки результатів тестування.

Застосування адаптованих тестів дозволить підвищити об'єктивність оцінки рівня знань випробовуваних з використанням меншої кількості тестових завдань в порівнянні із звичайним тестуванням за рахунок зниження погрішності виміру.

3.10.2. Алгоритми формування тесту при адаптивному тестуванні

Тестування методом ланцюжків питань. Автоматизована система контролю знань пропонує викладачеві використовувати систему ланцюжка питань, коли декілька питань об'єднуються у фіксовану послідовність (ланцюжок) за деякою смисловою ознакою, визначеною викладачем, а кожному питанню в ланцюжку привласнюється деякий коефіцієнт важливості цього питання в цьому ланцюжку. Цей коефіцієнт в діапазоні від 0 до 1 і сума коефіцієнтів питань в ланцюжку повинна бути рівною 1. Сенс коефіцієнта розкривається при обробці результатів тестування: оцінка за відповіді на питання, об'єднані в ланцюжок, виставляється залежно від важливості питань, на які були дані правильні відповіді. Ланцюжок може містити необмежене число питань, об'єднаних за семантичною ознакою усередині вибраної теми тестування. Виродженим випадком структури ланцюжка є наявність в ланцюжку усього лише одного питання. В цьому випадку коефіцієнт його важливості, очевидно, встановлюється рівним одиниці.

Автоматизований контроль знань за методикою уточнюючих питань. Концепція базується на автоматизації методики уточнюючих питань, що широко використовується в педагогічній практиці для виявлення глибини знань учнів. Відносна важливість питань, що ставляться, визначається їх ваговими коефіцієнтами, що враховуються при обчисленні результатів тестування.

Кожен ланцюжок є послідовністю близьких з тематики питань, що формулюються для уточнення знань тестованого. Чергове питання в ланцюжку задається тільки після отримання відповіді на попереднє питання. Залежно від стратегії тестування, що обирається організатором контролю знань, чергове питання в ланцюжку може пред'являтися до першої помилки ("строгий" викладач), або тестованому надається можливість демонструвати максимум знань, відповідаючи на усі питання цієї тематичної послідовності.

У ланцюжок може об'єднуватися необмежена кількість тематично близьких питань. У виродженому випадку ланцюжок складається з єдиного питання, коефіцієнт важливості якого встановлюється рівним 1.

Процедура кількісного оцінювання знань, виявлених в ході тестування, складається з трьох етапів. На першому розраховуються бали, набрані за правильні відповіді у рамках кожної окремої тематичної послідовності. На другому етапі розраховується сумарний бал S_{Σ} за відповіді на усі питання тесту з урахуванням кількості ланцюжків питань, на які тестований встиг відповісти за відведений час. На третьому етапі визначається підсумкова оцінка знань тестованого. Для цього набраний ним сумарний бал S_{Σ} проектується на оцінну шкалу, що має вигляд $[0; l_1; l_2; l_3; 1]$, де $0 < l_1 < l_2 < l_3 < 1$ - межі інтервальних діапазонів оцінок, що задаються викладачем при організації тестування.

Метод адаптивного автоматизованого тестування знань. Адаптивним тестуванням знань називатимемо спосіб екзаменаційного контролю рівня підготовки учня, при якому процедура вибору і пред'явлення йому чергового тестового завдання на $(t+1)$ -му кроці тестування визначається його відповідями на попередніх t кроках тесту. Математичну основу такого обліку складає модель об'єднання тестових завдань в тематичні послідовності із зваженим ранжуванням як окремих завдань, так і цілих послідовностей і виведенням підсумкової оцінки за тест з урахуванням нормованої суми балів, що накопичується за вибрані тестованим варіанти відповідей.

ПРАКТИЧНІ ЗАВДАННЯ

Завдання 1. Удосконалити тестове завдання за формою

Організації вивчають спеціалізацію клітин у процесі розвитку:

- A** на молекулярному рівні
- B** на клітинному рівні
- C** на популяційно-видовому рівні
- D** все з вищезазначеного

Розв'язання: У варіантах відповідей А-С повторюються слова *на ... рівні*, їх можна перемістити у формулювання питання. Дистрактор *все з вищезазначеного* є небажаними для використання. Кращий варіант формулювання завдання:

Організації вивчають спеціалізацію клітин у процесі розвитку на рівні:

- A** молекулярному
- B** клітинному
- C** популяційно-видовому
- D** організмівому

Завдання 2. Удосконалити тестове завдання за формою

Вказати, яка температура поверхні характерна для червоних зірок (°C):

- A** 2 000 – 3 000
- B** 6 000 – 7 000
- C** 12 000
- D** 20 000

Розв'язання: У варіантах відповідей А-В указано діапазон значень, у варіантах відповідей С-Д указано конкретне значення. Кращий варіант формулювання завдання:

Вказати, яка температура поверхні характерна для червоних зірок (°C):

- A** 2 000 – 3 000
- B** 6 000 – 7 000
- C** 12 000 – 13 000
- D** 20 000 – 25 000

Завдання 3. Удосконалити тестові завдання за формою

Організми, що вимерли, досліджуються:

- A** в науці палеонтології
- B** в науці систематиці
- C** в науці археології
- D** нічого з вищезазначеного

Розв'язання: У варіантах відповідей А-С повторюються слова *в науці*, їх можна перемістити у формулювання питання. Дистрактор *нічого з вищезазначеного* є небажаними для використання. Кращій варіант формулювання завдання:

Організми, що вимерли, досліджуються в науці

- A** палеонтології
- B** систематиці
- C** археології
- D** в еволюційному вченні

Завдання 4. Удосконалити тестові завдання за формою

Вирізнити причини більшої суворості клімату Антарктиди порівнянно з Арктикою:

- A** в Арктиці, на відміну від Антарктиди, клімат пом'якшує Північний Льодовитий океан
- B** абсолютні висоти території в Арктиці більші, ніж в Антарктиді
- C** у європейському секторі Арктики пом'якшує клімат Північноатлантична течія
- D** в Антарктиді, на відміну від Арктики, дують стокові вітри, які формують клімат цієї південної області

Розв'язання: При формулюванні варіантів відповідей порушено принцип стислості. Кращій варіант формулювання завдання:

Вирізнити причини більшої суворості клімату Антарктиди порівнянно з Арктикою:

- A** Відсутність Північного Льодовитого океану
- B** Менші абсолютні висоти території
- C** Відсутність Північноатлантичної течії
- D** Наявність стокових вітрів

Завдання 5. Дібрати чотири дистрактори до тестових завдань

1. Спростіть вираз $\frac{3x+12}{x^2-16}$

Розв'язання: Правильна відповідь: $\frac{3}{x-4}$, тому можливі варіанти

дистракторів:

$$\frac{3}{4-x};$$

$$\frac{3}{x+4};$$

$$-\frac{3}{x+4};$$

$$\frac{1}{x-4}.$$

Завдання 6. Дібрати чотири дистрактори до тестових завдань

Про якого великого київського князя йдеться в уривку з історичного джерела: «Отець бо його ... землю ... хрещенням просвітив, а сей великий князь ... засіяв книжними словами серця віруючих людей...»?

Розв'язання: Правильна відповідь: *Ярослава Мудрого*,

тому можливі варіанти дистракторів:

Святослава Ігоревича
Володимира Великого
Володимира Мономаха
Игоря Рюриковича

Завдання 7. Дібрати чотири дистрактори до тестових завдань

У туриста є 10 однакових за розмірами консервних банок, серед яких 4 банки – з тушко-ваним м'ясом, 6 банок – з рибою. Під час зливи етикетки відклеїлися. Турист навмання взяв одну банку. Яка ймовірність того, що вона буде з рибою?

Розв'язання: Оскільки правильна відповідь $3/5$, то можливі варіанти дистракторів:

$$1/10;$$

$$2/3;$$

$$1/6;$$

$$2/5.$$

Завдання 8. Дібрати чотири дистрактори до тестових завдань

Основу економічного розвитку Київської Русі становили:

Розв'язання: Оскільки правильна відповідь:

землеробство та чорна металургія,

то можливі варіанти дистракторів:

скотарство та землеробство

чорна металургія та торгівля

торгівля та промисли

промисли та чорна металургія

Завдання 5. Тест містить 60 питань з п'ятьма відповідями. 1-й учень правильно відповів на 37 питань, 2-й – на 56 питань. Знайти для кожного учня скорегований на здогадку тестовий бал. Порівняти результати. Зробити висновки.

Розв'язання: Корекція розраховується за формулою:

$$X_{ci} = X_i - \frac{W_i}{k-1}$$

де X_{ci} – зкорегований на здогадку тестовий бал;

X_i – тестовий бал i -го випробовуваного до корекції;

W_i – число помилкових відповідей у цього випробовуваного;

k – число відповідей в кожному завданні тесту.

$$X_{c1} = 37 - \frac{23}{5-1} = 31,25 \approx 31,$$

$$X_{c2} = 56 - \frac{4}{5-1} = 55.$$

Висновок: із збільшенням кількості правильних відповідей число балів, що віднімаються на здогадку, зменшується. Таким чином, у добре підготовлених учнів корекція балів на здогадку не знімає багато балів, тоді як у слабких і середніх така корекція знижує результати.

Питання для самоконтролю

1. Що називають «завданням у тестовій формі», «тестовим завданням», «тестом»?
2. У чому полягають принципові відмінності між цими поняттями?
3. Які вимоги до завдань у тестовій формі?
4. У чому полягає суть вимоги «технологічність»?
5. У чому полягає суть вимоги «логічна форма вислову»?
6. Які існують форми тестових завдань?
7. Назвіть переваги та недоліки різних форм тестових завдань.
8. За якими принципами формулюються завдання з вибором однієї правильної відповіді? Наведіть приклади.
9. Які принципи оцінювання завдань з вибором однієї правильної відповіді?
10. Для перевірки якого виду знань призначені завдання з вибором декількох правильних відповідей? Наведіть приклади.
11. Чи використовуються завдання з вибором декількох правильних відповідей на випробуваннях високих ставок? Чому?
12. Які принципи оцінювання завдань з вибором декількох правильних відповідей?
13. Наведіть формулу корекції на здогадку первинних тестових балів. Проаналізуйте її.
14. Для перевірки якого виду знань призначені завдання на встановлення відповідності? Наведіть приклади.
15. Які принципи композиції завдань на встановлення відповідності?
16. Які принципи оцінювання завдань на встановлення відповідності?
17. Для перевірки якого виду знань призначені завдання на встановлення правильної послідовності? Наведіть приклади.
18. Які принципи композиції завдань на встановлення правильної послідовності?
19. Які принципи оцінювання завдань на встановлення правильної послідовності?
20. У чому полягає специфічність завдань відкритої форми з короткою відповіддю?
21. Назвіть етапи створення завдання відкритої форми з короткою відповіддю. Наведіть приклади.

22. У чому полягає специфічність завдань відкритої форми з розгорнутою відповіддю?
23. Які існують вимоги до завдань у тестовій формі?
24. Назвіть переваги та недоліки різних форм тестових завдань.
25. Назвіть етапи конструювання педагогічного тесту.
26. Розкрийте суть першого етапу конструювання педагогічного тесту і в чому його необхідність?
27. Розкрийте суть дев'ятого етапу конструювання педагогічного тесту і в чому його необхідність?
28. Розкрийте суть шістнадцятого етапу конструювання педагогічного тесту і в чому його необхідність?
29. Які задачі вирішуються для нормативно-орієнтованих та критеріально-орієнтованих тестів на етапах конструювання?
30. У чому відмінності для нормативно-орієнтованих та критеріально-орієнтованих тестів на етапі апробації тестових завдань?
31. Назвіть перелік етапів створення педагогічних тестів, який наводить Майоров О.М.
32. У чому полягає розробка плану змісту тесту?
33. Яка таблиця заповнюється при розробці плану змісту тесту?
34. У чому полягає розробка специфікації тесту?
35. Що повинна містити специфікація тесту в розгорнутій формі?
36. Охарактеризуйте один із найпоширеніших підходів до створення короткої специфікації тесту.
37. Охарактеризуйте таблицю гіпотетичної специфікації тесту.
38. Охарактеризуйте таблицю короткої специфікації підсумкового тесту з фізики.
39. Назвіть та охарактеризуйте загальні принципи, які сприяють в певній степені правильному відбору змісту тестів.
40. Які розділи включає методика експертизи якості тесту?
41. Яке завдання повинен виконати експерт по першому напрямку роботи?
42. Охарактеризуйте таблицю для атестаційного тесту випускників школи, яку повинен заповнити експерт.
43. У чому суть найпростішого варіанту експертизи завдань?
44. Що таке скрінінг і в чому його суть?
45. Яке завдання повинен виконати експерт по другому напрямку роботи?
46. Яке завдання повинен виконати експерт по третьому напрямку роботи?
47. Наведіть означення методичного оснащення і для чого воно потрібне?
48. З яких частин складається методичне оснащення?
49. Від чого залежить зміст методичного оснащення?
50. Яка інформація повинна міститися в керівництві до тесту?

51. Яка інформація повинна міститися в інструкції для особи, яка проводить тест?
52. Яка інформація повинна міститися в інструкції для випробовуваного?
53. Що називають моделлю педагогічного тестування?
54. Охарактеризуйте класичну модель. Її переваги і недоліки.
55. Охарактеризуйте класичну модель з урахуванням складності завдань. Її переваги і недоліки.
56. Охарактеризуйте модель за зростанням складності. Її переваги і недоліки.
57. Охарактеризуйте модель з розділенням завдань за рівнями засвоєння. Її переваги і недоліки.
58. Охарактеризуйте модель з урахуванням часу відповіді на завдання. Її переваги і недоліки.
59. Охарактеризуйте модель з обмеженням часу на тест. Її переваги і недоліки.
60. Запропонуйте свій варіант моделі адаптивного тестування.
61. Що є теоретичною основою адаптивного тестування?
62. Для тестованих якого рівня підготовленості добре підходить класичне тестування?
63. Що таке групова адаптивність і як вона визначається?
64. В чому полягає суть тестування методом ланцюжків питань?
65. В чому відмінність між методом ланцюжків питань та методом уточнюючих питань?
66. Дайте визначення методу адаптивного автоматизованого тестування знань.
67. Як визначається підсумкова оцінка знань при використанні методу уточнюючих питань?

Список джерел, використаних у розділі III

1. Аванесов В.С. Теория и методика педагогических измерений [Материалы публикаций в открытых источниках и Интернет]. – ЦТ и МКО УГТУ-УПИ, 2005. – 98 с.
2. АДАПТИВНОЕ ТЕСТИРОВАНИЕ Учебно-методическое пособие. [Электронный ресурс].- Режим доступа: http://www.gumer.info/bibliotek_Buks/Pedagog/testing/index.php .- Назва з екрану
3. Булах І.Є., Мруга М.Р. Створюємо якісний тест, навчально-методичний посібник, 2-е видання. – К.: Майстер-клас, 2009. – 176 с.
4. Давидян Г.Ю. Модели педагогического тестирования [Электронный ресурс] / Третья областная научно-практическая конференция «Информационные технологии в образовании» «ИТО-Ростов 2003», Ростов-на-Дону, 9-10 октября 2003 / Донской институт информатизации образования. - <http://www.doniinfo.aaanet.ru/ito/2003/>.
5. Зайцева Л.В., Прокофьева Н.О. Модели и методы адаптивного контроля знаний [Электронный ресурс] //Образовательные технологии и общество.–2004.– Т.7.- №7.- http://ifets.ieee.org/russian/depository/v7_i4/pdf/1.pdf
6. Звонников В.И., Челышкова М.Б. Контроль качества обучения при аттестации. – М.: Логос, 2011.– 272с.
7. Ким В.С. Тестирование учебных достижений. Монография.- Уссурийск: Изд-во УГПИ, 2007. -214 с.
8. Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов.- М.: Логос, 2010.- 668 с.
9. Майоров А. Н. Теория и практика создания тестов для системы образования. (Как выбирать, создавать и использовать тесты для целей образования) / А. Н. Майоров. – М.: Интеллект-центр, 2001. – 296 с.
10. Матвиевская Е.Г. Педагогические измерения. Опыт и проблемы исследования [Электронный ресурс] // CREDO NEW теоретический журнал. – <http://credonew.ru/content/view/769/33/>.
11. Челышкова Н. Б. Теория и практика конструирования педагогических тестов : учебное пособие / Н. Б. Челышкова. – М. : Логос, 2002. – 432 с.
12. Andrich D. Advanced Social and Educational Measurement. Perth: Murdoch University, 2001.- 128 pp.
13. Brennan R. Educational Measurement.- Westport, CT: Praeger, 2006.- 796 pp.
14. Gronlund N.E. Measurement and evaluation in teaching.- Macmillan Pub. Co., 1965 - 420 pp.
15. Gronlund N.E., Linn R.L. Measurement and evaluation in teaching, Sixth edition. New York: Macmillan, 1990.
16. Linn R.L., Miller M.D. Measurement and assessment in teaching, 9th Edition. Upper Saddle River, NJ: Prentice Hall, 2005.

IV. КЛАСИЧНА ТЕОРІЯ ТЕСТУВАННЯ

У цьому розділі наведені основні положення класичної теорії тестування, підходи до обробки результатів тестування, методи статистичного аналізу та інтерпретація його результатів. Класична теорія тестів є у значній мірі актуальною у наш час, незважаючи на появу більш сучасних теорій. У ній достатньо уваги приділяється вирішенню однієї із центральних задач теорії тестів – надійності тесту. При цьому вводяться такі важливі характеристики окремих тестових завдань та усього тесту як валідність, розподільна здатність тощо.

Основоположником класичної теорії тестів (Classical Theory of mental tests) є відомий британський психолог, автор факторного аналізу, Чарльз Едвард Спірман (Charles Edward Spearman, 1863-1945). Значний вклад в розвиток цієї теорії вніс Льюїс Гуттман (Louis Guttman, 1916-1983). Всебічно і повно класична теорія тестів вперше була викладена у фундаментальній праці Гарольда Гулліксена (Gulliksen H., 1950 р.), а сучасне її викладення наведене у книзі Л. Крокер і Дж. Алгіни (1986 р.).

4.1. Основні положення класичної теорії тестування

Класична теорія тестів ґрунтується на наступних п'яти основних положеннях.

1. *Емпірично отриманий результат вимірювання є сумою істинної оцінки та похибки вимірювання:*

$$X = T + E,$$

де X – тестова оцінка, що спостерігається, тобто отримується емпіричним шляхом, T – істинна оцінка індивідуума, E – оцінка похибки вимірювання. Цю формулу запропонував Чарльз Спірман, який логічно та математично довів, що тестові оцінки характеристик людей завжди містять помилкові компоненти вимірювання. Цю формулу вважають класичною моделлю істинної оцінки.

2. *Істинний результат вимірювання (істинну оцінку) можна виразити як математичне сподівання $M(X)$:*

$$T = M(X).$$

Обґрунтуємо це положення. *Змінна* – це величина, яка може приймати будь-яке значення із набору допустимих значень. *Випадкова змінна* може бути визначена як змінна, що *приймає свої значення згідно ряду розподілу ймовірностей*.

Кожного разу, коли випробуваний виконує тест, його оцінку за цей тест можна вважати значенням випадкової величини. Дійсно, припустимо, що тест має 50 завдань. Оцінки випробуваних можуть знаходитись в межах від 0 до 50. Після виконання тесту випробуваний отримає якусь оцінку в цих межах. На цю оцінку можуть впливати систематичні та випадкові помилки. Якщо уявити ідеальну ситуацію, коли випробуваний виконує цей тест скільки завгодно разів (він не втомлюється і повністю забуває завдання після попереднього тестування), то отримувані оцінки появлятимуться з певною частотою, яка може використовуватися для підрахунку ймовірності появи оцінки за тест. У цьому смислі тестову оцінку X можна розглядати як значення випадкової величини. Якщо випадкова величина X задана розподілом

X	x_1	x_2	x_n
P	p_1	p_2	p_n

то її математичне сподівання обчислюється за формулою

$$M(X) = \sum_{j=1}^n x_j \cdot p_j,$$

де x_j – j -те значення X (одна із оцінок за тест), p_j – ймовірність отримання оцінки x_j . Для i -того учасника тестування отримуємо

$$T_i = M(X_i) = \sum_{j=1}^n x_{ij} \cdot p_{ij}.$$

Останню формулу можна трактувати наступним чином. *Для кожного випробуваного математичне сподівання можна розглядати як середнє всіх тестових оцінок, які він міг би отримати при багатократному виконанні*

тесту. Тоді його істинну оцінку можна інтерпретувати як середнє оцінок, отриманих по нескінченному числу тестувань при використанні одного і того ж тесту. Істинна оцінка є статистичним поняттям, що базується на очікуваному значенні, отриманому в даному процесі вимірювання.

3. Середнє значення похибок оцінок для генеральної сукупності випробуваних дорівнює нулю:

$$M(E) = 0.$$

4. Кореляція між істинною оцінкою та її похибкою для генеральної сукупності випробуваних дорівнює нулю:

$$\rho_{TE} = 0.$$

5. Кореляція між похибками оцінок двох будь-яких тестів дорівнює нулю:

$$\rho_{E1 E2} = 0.$$

Серед розглянутих п'яти положень перші два є означеннями. Решту три можна отримати логічними міркуваннями із означень. Положення 3-5 ще називають припущеннями або аксіомами моделі. Вони описують ті основні властивості істинних оцінок та їх похибок, які дозволяють застосувати класичну модель істинної оцінки до дослідження надійності тестових балів.

У класичній теорії використовуються два означення паралельних і еквівалентних тестів.

Паралельні тести повинні відповідати вимогам 1-5, істинні компоненти одного тесту T_1 повинні бути рівними істинним компонентам другого тесту T_2 для кожної вибірки випробуваних, що відповідають на обидва тести. Припускається, що $T_1 = T_2$ і, крім того, рівні дисперсії $S_1^2 = S_2^2$.

Еквівалентні тести повинні відповідати всім вимогам паралельних тестів за виключенням одного: істинні компоненти одного тесту не обов'язково повинні дорівнювати істинним компонентам другого паралельного тесту, але відрізнятися вони повинні на одну і ту ж константу c . Умова еквівалентності двох тестів записується у вигляді:

$$T_1 = T_2 + c_{12},$$

де c_{12} – константа відмінності результатів першого і другого тестів.

4.2. Первинний аналіз результатів тестування

4.2.1. Матриця результатів тестування та її редукування

Статистична обробка результатів тестування починається з формування матриці тестових результатів.

Матриця результатів тестування – це матриця розмірності $N \times M$, яка містить числові градації індикатора досліджуваної латентної змінної, де M – число індикаторів, N – число випробовуваних.

Ця матриця є прямокутною таблицею, рядки якої відповідають випробовуваним, а стовпці – індикаторам змінних. Для випадку тестування навчальних досягнень індикаторами змінних є тестові завдання. У такої матриці рядки відповідають учасникам тестування, а стовпці – тестовим завданням (номерам завдань). На перетині рядків і стовпців знаходиться число, яке зазвичай є оцінкою даного випробовуваного за відповідь на дане тестове завдання.

Відповідь випробовуваного може оцінюватися числами в деякому діапазоні, наприклад від 0 до 9. Можна припустити, що 0 відповідає відсутності знань, а 9 – наявності повних знань для даного тестового завдання. Проміжні варіанти оцінюються числами в діапазоні від 0 до 9. Тоді матриця результатів тестування називається *політомічною* і вона зображена таблицею 4.1.

Таблиця 4.1. Політомічна матриця результатів тестування

№	Випробовувані	Номери завдань		
		1	2	3
1	Антоненко	2	3	0
2	Дмитренко	5	9	4
3	Іванчук	8	7	4
4	Шульга	7	4	3
5	Ясулович	3	5	1

Зауважимо, що числа, розташовані в комірках таблиці, відліковуються по порядковій шкалі. Це означає, що з цими числами не можна виконувати

арифметичних операцій (додавати, віднімати і т.д.), бо формально це є символи для градації знань випробовуваних.

На практиці найчастіше використовується дихотомічне оцінювання, тобто коли за невірну відповідь на завдання тесту виставляється 0, а за вірну – 1. Тоді матриця результатів тестування складається із нулів та одиниць і вона називається *бінарною* або *дихотомічною*. Приклад дихотомічної матриці наведений в таблиці 4.2.

Таблиця 4.2. Дихотомічна матриця (11x9)

	Вихідні номери	1	2	3	4	5	6	7	8	9
Анісімова	1	1	0	0	0	0	0	0	1	0
Бугай	2	0	1	0	0	0	0	0	1	1
Воропайчук	3	1	1	0	0	1	1	0	1	0
Дем'яненко	4	1	0	0	0	0	0	0	1	1
Заїка	5	1	1	0	0	0	1	1	1	0
Ломака	6	1	1	1	1	0	0	1	1	1
Монсур	7	1	1	1	1	1	1	1	1	1
Оффенбах	8	1	1	1	0	1	1	1	1	1
Розенфельд	9	1	1	0	0	0	0	0	1	1
Чорнецький	10	1	1	1	0	0	1	0	1	1
Ясулович	11	1	1	1	0	0	1	0	1	1

За даною матрицею обчислюються величини: індивідуальний бал випробовуваних, кількість правильних відповідей випробовуваних на кожне завдання тесту. Позначимо матрицю результатів через $A = \{a_{ij}\}$, тобто вона формально складається з елементів a_{ij} . Індекси елемента a_{ij} інформують, що він розташований на перетині i -го рядка та j -го стовця.

Індивідуальний бал X_i i -го випробовуваного – це кількість правильних відповідей на тест. Цей бал обчислюється за формулою

$$X_i = \sum_{j=1}^M a_{ij},$$

де M – кількість завдань у тесті ($M=9$ у таблиці 4.2), наприклад, індивідуальний бал другого учасника тестування дорівнює

$$X_2 = \sum_{j=1}^9 a_{2j} = 3.$$

Кількість правильних відповідей R_j на j -те завдання обчислюється за формулою $R_j = \sum_{i=1}^N a_{ij}$, де N – кількість учасників тестування ($N=11$ у

таблиці 4.2). Для 3-го завдання $R_3 = \sum_{i=1}^N a_{i3} = 5$.

Також можна обчислювати величини:

- кількість неправильних відповідей на j -те завдання - $W_j = N - R_j$;
- долю правильних відповідей на j -те завдання - $p_j = \frac{R_j}{N}$;
- долю не правильних відповідей на j -те завдання - $q_j = 1 - p_j$.

В таблиці 4.3 наведені значення величин X_i та R_j . Крайній правий стовпець містить значення X_i індивідуальних балів випробовуваних, а найнижчий рядок містить значення R_j - кількість правильних відповідей на завдання тесту.

Таблиця 4.3. Бінарна матриця з величинами X_i та R_j

Вихідні номери	1	2	3	4	5	6	7	8	9	X_i
1	1	0	0	0	0	0	0	1	0	2
2	0	1	0	0	0	0	0	1	1	3
3	1	1	0	0	1	1	0	1	0	5
4	1	0	0	0	0	0	0	1	1	3
5	1	1	0	0	0	1	1	1	0	5
6	1	1	1	1	0	0	1	1	1	7
7	1	1	1	1	1	1	1	1	1	9
8	1	1	1	0	1	1	1	1	1	8
9	1	1	0	0	0	0	0	1	1	4
10	1	1	1	0	0	1	0	1	1	6
11	1	1	1	0	0	1	0	1	1	6
R_j	10	9	5	2	3	6	4	11	8	58

Спочатку виконується упорядкування рядків за спаданням X_i , а потім упорядковуються стовпці за зростанням R_j (іноді упорядкування стовпців виконують за спаданням балів). Після упорядкування стовпців та рядків таблиця 4.3 набуде вигляду таблиці 4.4.

Таблиця 4.4. Упорядкована бінарна таблиця

Вихідні номери	4	5	7	3	6	9	2	1	8	X_i
7	1	1	1	1	1	1	1	1	1	9
8	0	1	1	1	1	1	1	1	1	8
6	1	0	1	1	0	1	1	1	1	7
11	0	0	0	1	1	1	1	1	1	6
10	0	0	0	1	1	1	1	1	1	6
3	0	1	0	0	1	0	1	1	1	5
5	0	0	1	0	1	0	1	1	1	5
9	0	0	0	0	0	1	1	1	1	4
2	0	0	0	0	0	1	1	0	1	3
4	0	0	0	0	0	1	0	1	1	3
1	0	0	0	0	0	0	0	1	1	2
R_j	2	3	4	5	6	8	9	10	11	58

Профіль випробовуваного (учасника тестування) – це послідовність значень індикатора в упорядкованій матриці тестових результатів. Для таблиці 4.4 профіль випробовуваного 7 є послідовністю із дев'яти одиниць.

Після упорядкування таблиці порушилась нумерація рядків і стовпців, але їх не потрібно перенумеровувати, щоб не втратити початкову інформацію про номери завдань та номери учасників тестування.

Із таблиці 4.4 видно, що завдання №8 виконали всі 11 випробовуваних. Це завдання не дозволяє диференціювати випробовуваних, тому його потрібно вилучити із тесту, якщо тест є нормативно-орієнтованим. Відповідно, якби на якесь завдання не відповів жодний випробовуваний ($R_j = 0$), то його теж потрібно було вилучити. Аналогічно вилучаються рядки, що складаються тільки з одиниць або нулів, бо тест не дає ніякої інформації про випробовуваного. Після вилучення завдання №8 і тестованого №7 отримується редукована бінарна матриця (таблиця 4.5).

Таблиця 4.5. Редукована бінарна матриця

	Вихідні номери	4	5	7	3	6	9	2	1	
Вихідні номери	Нові номери	1	2	3	4	5	6	7	8	X_i
8	1	0	1	1	1	1	1	1	1	7
6	2	1	0	1	1	0	1	1	1	6
11	3	0	0	0	1	1	1	1	1	5
10	4	0	0	0	1	1	1	1	1	5
3	5	0	1	0	0	1	0	1	1	4
5	6	0	0	1	0	1	0	1	1	4
9	7	0	0	0	0	0	1	1	1	3
2	8	0	0	0	0	0	1	1	0	2
4	9	0	0	0	0	0	1	0	1	2
1	10	0	0	0	0	0	0	0	1	1
	R_j	1	2	3	4	5	7	8	9	39

При цьому учасникам тестування та тестовим завданням присвоєні нові номери, які далі всюди використовуються при аналізі результатів тестування.

4.2.2. Побудова рядів результатів тестування та графічна інтерпретація

Дані, представлені у редукованій бінарній матриці, використовуються для графічної інтерпретації результатів тестування. Для цього упорядковують дані у вигляді різних рядів. Незгрупований та ранжований ряди представлені таблицею 4.6.

Таблиця 4.6. Незгрупований та ранжований ряди

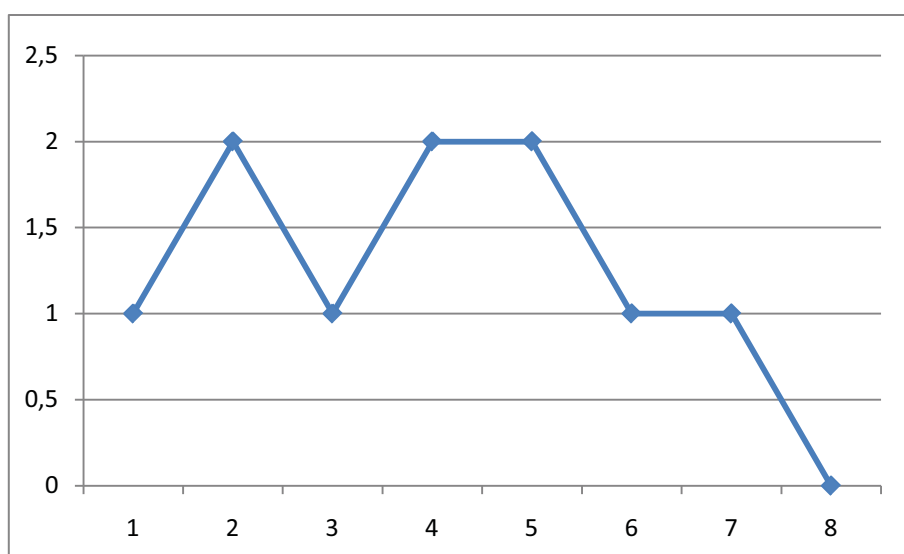
Незгрупований ряд		Ранжований ряд			
Вихідний номер	Бал X_i	Вихідний номер	Новий номер	Ранг	Бал X_i
1	1	8	1	1	7
2	2	6	2	2	6
3	4	11	3	3,5	5
4	2	10	4	3,5	5
5	4	3	5	5,5	4
6	6	5	6	5,5	4
8	7	9	7	7	3
9	3	2	8	8,5	2
10	5	4	9	8,5	2
11	5	1	10	10	1

Частота тестового балу – це кількість випробовуваних, які мають даний бал. Частотний ряд представлений таблицею 4.7.

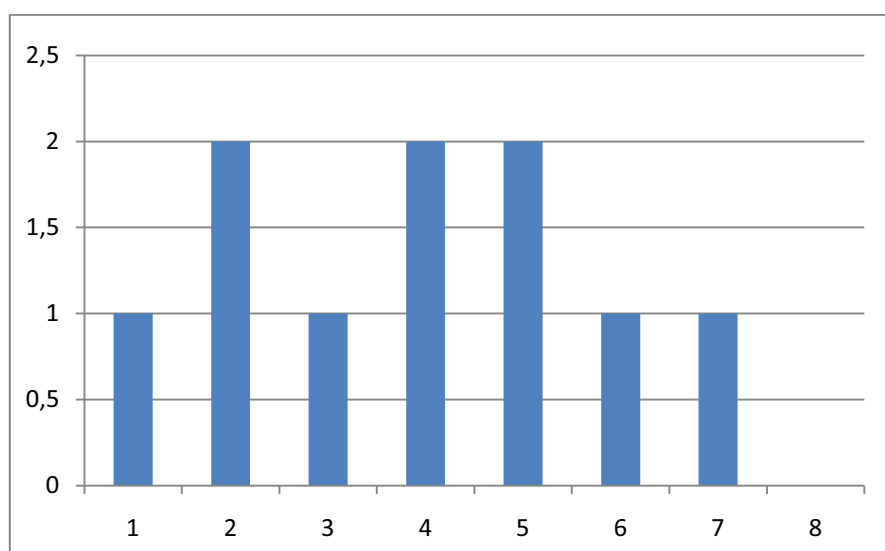
Таблиця 4.7. Частотний ряд

Бал X_i	1	2	3	4	5	6	7	8
Частота	1	2	1	2	2	1	1	0

На основі таблиці частот будуються різні діаграми – полігон частот, гістограми і т. п. Перевагу віддають гістограмі, тому що вона зручна для візуального порівняння емпіричного розподілу з теоретичним нормальним.



Полігон частот



Гістограма частот

4.3. Основні статистичні характеристики результатів тестування

4.3.1. Міри центральної тенденції тестових балів

Тестові бали випробовуваних зазвичай групуються поблизу деяких найбільш ймовірних значень, які можна охарактеризувати трьома мірами центральної тенденції – модою, медіаною і середнім. Міри центральної тенденції призначені для виявлення «центрального положення», біля якого в основному групується множина значень заданого розподілу даних.

Мода – це таке значення, яке у вибірці зустрічається найчастіше.

- Якщо всі значення X_i зустрічаються з однаковою частотою, то мода відсутня.
- Коли два сусідні значення X_i мають однакові частоти, які більші за частоту будь-якого іншого значення, то мода визначається як середнє сусідніх значень X_i .
- Якщо два несуміжні значення X_i мають однакові частоти, які більші за частоту будь-якого іншого значення, то існує дві моди. У цьому випадку кажуть, що група оцінок має бімодальний розподіл.

Для наведеного вище прикладу (табл.4.6), єдина мода відсутня (зустрічаються три значення – 2, 4 та 5, які повторюються двічі).

Нагадаємо, що *медіана* – це значення X_i , яке ділить упорядкований набір даних навпіл так, що одна половина значень менша за медіану, а друга – більша. Із ранжованого ряду, видно, що медіана балів учасників тестування рівна пів-сумі балів п'ятого та шостого учасників: $(4+4)/2=4$.

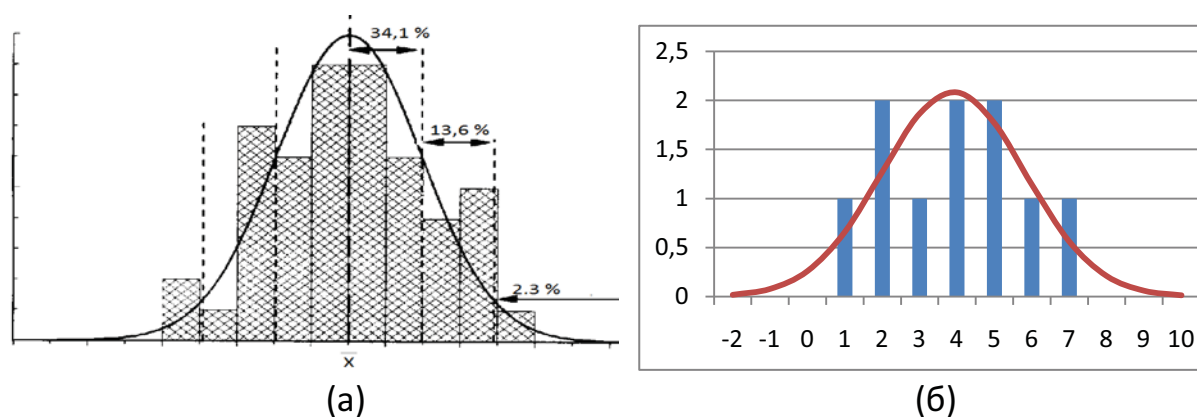
Середнє вибіркове (середнє арифметичне) обчислюється за формулою

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_N}{N} \text{ або } \bar{X} = \frac{\sum_{i=1}^N X_i}{N},$$

де N – кількість учасників тестування. На відміну від моди на величину середнього впливають значення всіх результатів. Нескладно обчислити, що для наведеного вище прикладу $\bar{X} = 3,9$ (перевірте це самостійно!).

Міри центральної тенденції в певній мірі допомагають при оцінці якості тесту в тому випадку, коли вона проводиться за результатами апробації на репрезентативній вибірці учнів. Зазвичай вважають, що хороший нормативно-орієнтований тест забезпечує нормальний розподіл

індивідуальних балів репрезентативної вибірки учнів, якщо середнє значення балів знаходиться в центрі розподілу, а решта значень концентруються навколо середнього за нормальним законом, тобто приблизно 70% значень в центрі, а інші наближаються до нуля на краях розподілу. На рисунку зображено емпіричний розподіл частот балів та відповідний нормальний розподіл за результатами реального тестування (а) та за результатами нашого прикладу (б).



Якщо тест забезпечує близький до нормального розподіл балів, то це означає, що на його основі можна визначити стійке середнє значення балів, яке приймається за одну із репрезентативних норм виконання тесту. Обернений висновок, загалом, невірний: стійкість тестових норм зовсім не передбачає обов'язкового нормального розподілу емпіричних результатів виконання тесту.

Нормальний розподіл унімодальний і симетричний, тобто половина результатів, яка розташована нижче моди, точно співпадає з другою половиною, розташованою вище, а мода і середнє значення рівні.

4.3.2. Міри мінливості тестових балів

Для характеристики ступеня розсіювання окремих значень навколо середнього використовуються різні міри: розмах, дисперсія, стандартне відхилення.

Розмах вимірює на шкалі відстань, в межах якої змінюються всі значення показника в розподілі. Наприклад, для розподілу індивідуальних балів в таблиці 4.6 розмах дорівнює $7-1=6$.

Варіаційний розмах легко обчислюється, але використовується дуже рідко при характеристиці розподілу балів по тесту. По-перше, розмах є дуже приблизним показником, тому що не залежить від степеня змінюваності проміжних значень, розташованих між крайніми значеннями в розподілі балів по тесту. По-друге, крайні значення індивідуальних балів, зазвичай, ненадійні, тому що містять в собі значну похибку вимірювання.

Дисперсія. Підрахунок дисперсії ґрунтується на обчисленні відхилення кожного значення показника від середнього арифметичного в розподілі. Для індивідуальних балів значення відхилень

$$X_i - \bar{X} \quad (i = 1, 2, \dots, N)$$

несуть інформацію про варіацію сукупності значень балів N учнів, тобто відображають міру неоднорідності результатів по тесту. Якщо всі індивідуальні бали однакові, то відхилення рівні 0 . Якщо ж індивідуальні бали не співпадають, то відхилення можуть бути додатними і від'ємними. Сума всіх відхилень буде рівна 0 . Тому для характеристики варіації тестових балів використовують квадрати відхилень. Дисперсія обчислюється так

$$S_X^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}.$$

Для обчислень краще використовувати наступну формулу дисперсії

$$S_X^2 = \frac{N \sum_{i=1}^N X_i^2 - (\sum_{i=1}^N X_i)^2}{N(N - 1)},$$

яка може бути отримана із першої. Дисперсія пропорційна не $1/N$, а $1/(N - 1)$. Це зроблено для того, щоб для невеликих N отримати незміщену оцінку генеральної дисперсії. Емпірична дисперсія, яка відповідає розподілу із наведеного вище прикладу, дорівнює $3,66$ (перевірте це самостійно!).

Стандартне відхилення дорівнює кореню квадратному із дисперсії

$$S_X = \sqrt{S_X^2}.$$

Оцінка стандартного відхилення для вище розглянутого прикладу дорівнює $1,91$ (перевірте це самостійно!).

Дисперсія відіграє важливу роль в оцінці якості нормативно-орієнтованих тестів. Низька дисперсія індивідуальних балів свідчить про слабку диференціацію випробовуваних в групі, а надто висока дисперсія

характерна для випадку, коли всі учні відрізняються за числом виконаних завдань. У обох випадках треба переробляти тест. При переробці тесту слід керуватися простим правилом: якщо перевірка узгодженості емпіричного розподілу з нормальним дає позитивні результати, а дисперсія зростає, то це означає, що здійснюється підвищення диференціуючої здатності тесту і процес покращення тесту.

4.3.3. Перевірка гіпотези про нормальний закон розподілу результатів тестування

Результати нормативно-орієнтованого тестування при великих вибірках зазвичай мають розподіл, близький до нормального.

Неперервна випадкова величина X має нормальний закон розподілу (закон Гауса) з параметрами μ і σ^2 , якщо її щільність ймовірності має вигляд

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}},$$

де σ^2 - дисперсія, μ - константа, яка задає зсув розподілу по осі X .

Стандартне відхилення є корисною мірою варіації для випадку нормального розподілу балів випробовуваних.

Існують наближена методи перевірки близькості емпіричного розподілу до нормального. Якщо виконується приблизне співвідношення $\bar{X} \approx 3 \cdot S_X$, то можна вважати дисперсію оптимальною, а розподіл тестових балів близький до нормального. Однак це вірно не завжди.

За грубу оцінку нормальності розподілу можна прийняти співвідношення

$$\bar{X} - 3S_X \leq X \leq \bar{X} + 3S_X,$$

що ще раз підтверджує інформацію із пункту 4.3.1. Якщо майже всі значення тестових балів X містяться у цьому інтервалі, то у першому наближенні можна вважати емпіричний розподіл нормальним.

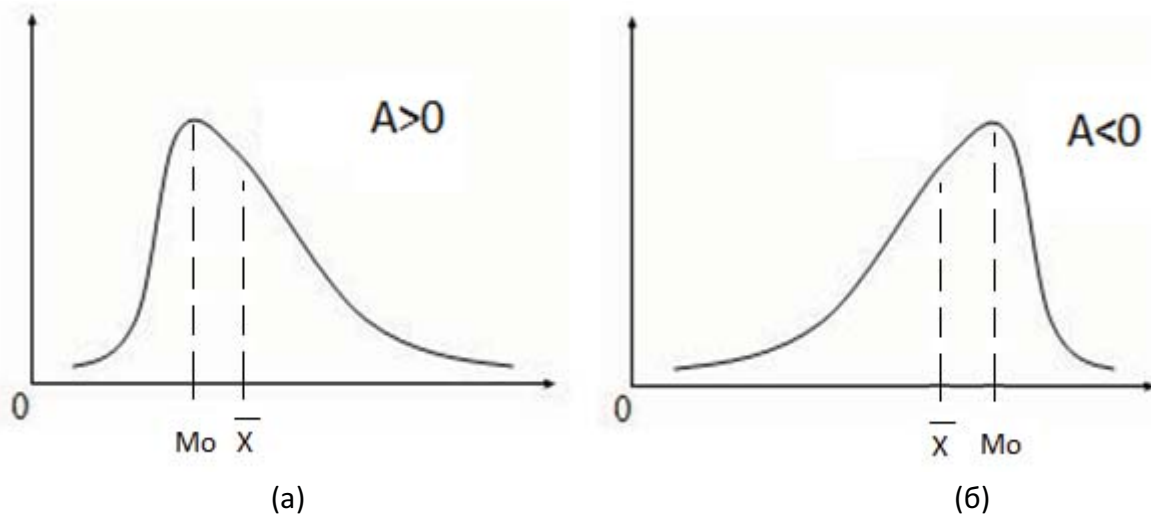
Для перевірки близькості розподілу емпіричних даних до нормального потрібно перевірити гіпотезу про нормальний розподіл генеральної сукупності за критерієм Пірсона.

4.3.4. Асиметрія та ексцес, їх інтерпретація

Степінь відхилення емпіричного розподілу від симетричного, характерного для нормальної кривої, оцінюється за допомогою *асиметрії*, яка обчислюється за формулою

$$A = \frac{\sum_{i=1}^N (X_i - \bar{X})^3}{S_X^3 \cdot N}.$$

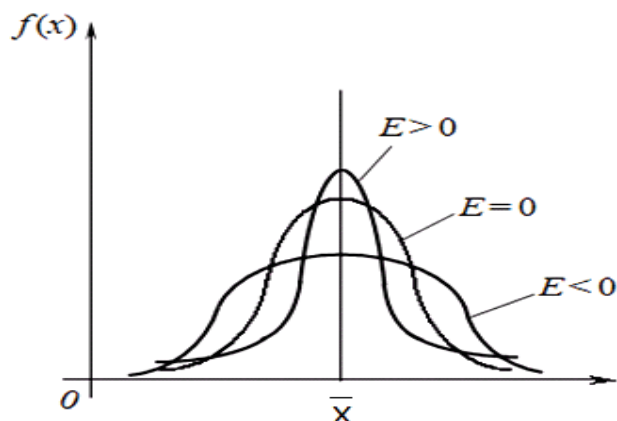
Нижче зображені два типи асиметричних розподілів. Асиметрія розподілу від'ємна, якщо основна частина значень індивідуальних балів лежить справа від середнього значення, що зазвичай характерно для досить легких тестів (а). Асиметрія розподілу балів додатна, якщо більшість учнів отримали оцінки нижчі середнього балу. Цей ефект спостерігається для досить трудних тестів (б). Для нормального розподілу характерна нульова асиметрія, тобто тест добре збалансований за трудністю.



Асиметрія у нашому прикладі є додатною величиною $0,06 > 0$ (радимо перевірити самостійно), значення асиметрії додатне при цьому близьке до нуля, що говорить про його достатньо добре підібраний за трудністю тест.

За допомогою ексцесу можна виявити форму вершини розподілу. Гостровершинна крива має додатній ексцес, середньовершинна крива – нульовий, плосковершинна крива – від'ємний. При нормальному розподілі

ексцес дорівнює нулю. Зауважимо, що поняття «ексцес» можна застосовувати лише до унімодальних розподілів. На рисунку зображені гостровершинну, середньовершинну та плосковершинну криві.



Додатний ексцес означає менший за нормальний розкид результатів тестування відносно середнього значення в околі цього значення. Від'ємний ексцес означає, що результати тестування є сильно розкиданими в околі їх середнього значення.

Нагадаємо, що ексцес обчислюють за формулою
$$E = \frac{\sum_{i=1}^N (X_i - \bar{X})^4}{S_X^4 \cdot N} - 3.$$

Ексцес у нашому прикладі є від'ємною величиною $-0,87 < 0$ (радимо перевірити самостійно), що говорить про значну скупченість білів учасників у вузькому околі навколо середнього балу.

4.4. Кореляційний аналіз результатів тестування

4.4.1. Коефіцієнт кореляції Пірсона

Між завданнями тесту існують зв'язки, які можна виявити в процесі тестування за допомогою кореляції. Для виявлення питання про наявність зв'язку між двома величинами X і Y розглядають суму

$$\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y}).$$

Ця сума буде великою і додатною, коли X і Y сильно зв'язані прямою залежністю, близькою до нуля у випадку відсутності зв'язку і від'ємна, коли X

і Y сильно зв'язані оберненою залежністю. Для того, щоб ця сума не залежала від кількості значень X і Y , її ділять на $N-1$. Отримана величина S_{XY} називається *коваріацією* X і Y та є мірою їх зв'язку:

$$S_{XY} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{N-1}.$$

Для виключення впливу стандартних відхилень на величину зв'язку, коваріацію S_{XY} ділять на стандартні відхилення S_X і S_Y :

$$r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} \quad \text{або} \quad r_{XY} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}}.$$

Отримана міра зв'язку між X і Y називається *коефіцієнтом кореляції Пірсона*. Для обчислень більш зручною є наступна формула коефіцієнта

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

Коефіцієнт кореляції Пірсона r_{XY} змінюється в межах від -1 до 1. Чим ближче значення коефіцієнта Пірсона до 1, тим тісніший зв'язок між результатами тестувань X і Y . Якщо коефіцієнт Пірсона рівний нулю, то зв'язок відсутній. Якщо він менше нуля, то зв'язок між X і Y є взаємно оберненим, тобто спостерігаються протилежні результати тестування.

4.4.2. Φ -коефіцієнт кореляції. Кореляційна матриця тестових завдань

Для оцінки зв'язку між двома завданнями тесту коефіцієнт кореляції Пірсона r_{XY} треба перетворити, тому що результати виконання завдань

представлені в дихотомічній шкалі. Коефіцієнт кореляції Пірсона для дихотомічних даних називається *коефіцієнтом «фі»* та обчислюється за формулою

$$\varphi_{mk} = \frac{p_{mk} - p_m p_k}{\sqrt{p_m q_m \cdot p_k q_k}},$$

де p_{mk} - доля випробовуваних, які виконали правильно обидва завдання тесту (m -те і k -те), p_m - доля випробовуваних, які правильно виконали m -те завдання, $q_m = 1 - p_m$, p_k - доля випробовуваних, які правильно виконали k -те завдання, $q_k = 1 - p_k$.

За коефіцієнтом φ_{mk} на основі стовпів редукованої матриці результатів тестування розраховується кореляційна матриця, яка є квадратною матрицею розмірності $M \times M$, де M – кількість завдань, симетричною відносно головної одиничної діагоналі. В останньому рядку таблиці можна розмістити суми кореляційних коефіцієнтів, величини яких свідчать про тісноту зв'язку окремого завдання з усім тестом. Для розглянутого вище прикладу маємо кореляційну таблицю 8×8 (таблиця 4.8).

Таблиця 4.8. Кореляційна матриця тестових завдань

$j \setminus j$	1	2	3	4	5	6	7	8
1	1	-0,167	0,5092	0,4082	-0,333	0,2182	0,1667	0,1111
2	-0,167	1	0,2182	0,1021	0,5	-0,218	0,25	0,1667
3	0,5092	0,2182	1	0,3563	0,2182	-0,048	0,3273	0,2182
4	0,4082	0,1021	0,3563	1	0,4082	0,5345	0,4082	0,2722
5	-0,333	0,5	0,2182	0,4082	1	-0,218	0,5	0,3333
6	0,2182	-0,218	-0,048	0,5345	-0,218	1	0,2182	-0,218
7	0,1667	0,25	0,3273	0,4082	0,5	0,2182	1	-0,167
8	0,1111	0,1667	0,2182	0,2722	0,3333	-0,218	-0,167	1
$\sum \varphi_{mk}$	1,9134	1,8521	2,7999	3,4898	2,4082	1,2687	2,7038	1,7166

Кореляція завдань одне з одним не повинна бути дуже високою ($r_{XY} \leq 0,3$), інакше завдання будуть дублювати одне одного. Якщо кореляція між двома завданнями близька до одиниці, то одне із них лишнє. Від'ємна кореляція завдання з іншими завданнями небажана. Якщо завдання від'ємно корелює з більшістю інших завдань, то це означає, що відповіді на нього

протилежні результатам по інших завданнях. Це завдання може мати грубі помилки або перевіряти знання з іншої предметної області. Взагалі, для тематичних тестів характерна висока кореляція між завданнями, що виправдане призначенням тестів.

Із кореляційної таблиці видно, що завдання 6 має чотири від'ємних кореляційних коефіцієнтів з завданнями 2, 3, 5 та 8. Крім того у окремих пар тестових завдань також зустрічаються від'ємні або додатні близькі до нуля кореляційні коефіцієнти, що свідчить про певну суперечливість системи тестових гавань. Такі дані є сигналом для розробників тесту до експертної оцінки якості завдань, оцінки тематичної відповідності їх до всієї системи, а також їх рівня трудності. Найменшу суму кореляційних коефіцієнтів для 6-го завдання (див. в останньому рядку) підтверджують попередню інформацію про незначну корельованість цього завдання з усім тестом.

4.4.3. Коефіцієнт бісеріальної кореляції

Коефіцієнт бісеріальної кореляції використовується для оцінювання валідності окремих завдань тесту. Він використовується в тому випадку, коли один набір значень розподілу задається в дихотомічній шкалі, а другий – в інтервальній. Під цю ситуацію попадає підрахунок кореляції між результатами виконання кожного завдання (дихотомічна шкала) і сумою балів випробовуваних (інтервальна шкала) по завданнях тесту. Формула коефіцієнта бісеріальної кореляції має вигляд

$$(r_{bis})_j = \frac{(\bar{X}_1)_j - (\bar{X}_0)_j}{S_X} \cdot \frac{(N_1)_j - (N_0)_j}{uN\sqrt{N^2 - N}},$$

де $(\bar{X}_1)_j$ - середнє значення індивідуальних балів випробовуваних, які виконали вірно j -те завдання тесту; $(\bar{X}_0)_j$ - середнє значення індивідуальних балів випробовуваних, які виконали невірно j -те завдання тесту; S_X - стандартне відхилення по множині значень індивідуальних балів; $(N_1)_j$ - число випробовуваних, які виконали вірно j -те завдання тесту; $(N_0)_j$ - число випробовуваних, які виконали невірно j -те завдання тесту; N – загальне число випробовуваних, $N = (N_1)_j + (N_0)_j$; u – квантиль нормованого нормального розподілу порядку N_1 / N .

Обчислення $(r_{\text{pbis}})_j$ за вказаною формулою вимагає спеціальних таблиць для знаходження ординати стандартної нормальної кривої. Тому нерідко використовують інший коефіцієнт кореляції, який називається точково-бісеріальним коефіцієнтом - r_{pbis} . Він обчислюється простіше і не вимагає виконання гіпотезі про нормальний характер розподілу. Формула для обчислення r_{pbis} має вигляд

$$(r_{\text{pbis}})_j = \frac{(\bar{X}_1)_j - (\bar{X}_0)_j}{S_x} \cdot \sqrt{\frac{(N_1)_j \cdot (N_0)_j}{N(N-1)}}.$$

Загалом завдання можна вважати валідним (про «валідність» докладніше див. розділ IV, пункт 4.9.), коли значення $(r_{\text{pbis}})_j \approx 0,5$. Оцінка валідності завдання дозволяє судити про те, наскільки завдання придатне для роботи у відповідності з загальною метою створення тесту. Якщо ця мета – диференціація учнів за рівнем підготовки, то валідні завдання повинні чітко відділяти добре підготовлених від слабо підготовлених учнів.

Головну роль в оцінці валідності завдання відіграє різниця $(\bar{X}_1)_j - (\bar{X}_0)_j$, що знаходиться в чисельнику формули. Чим більше значення цієї різниці, тим краще завдання диференціює випробовуваних. Якщо ця різниця близька до нуля, то завдання має низьку диференціацію. При від'ємній різниці потрібно завдання вилучити із тесту, бо в ньому перемагають слабкі учні, а сильні або відповідають неправильно або пропускають завдання.

Табл.4.9. Точково-бісеріальна кореляція тестових завдань з усім тестом

<i>J</i>	1	2	3	4	5	6	7	8
$(r_{\text{pbis}})_j$	0,3859	0,4411	0,6376	0,8328	0,6064	0,3248	0,6616	0,3492

Значення точково-бісеріальної кореляції тестових завдань з усім тестом (таблиця 4.9.) свідчать про невисоку (але прийнятну) корельованість завдань 6, 8 та 1 з усім тестом, їх структуру та зміст слід проаналізувати експертам. Завдання 2, 3, 5 та 7 мають кореляцію у прийнятному інтервалі. Завдання 4 має високий степінь кореляції, тому і його зміст та форму слід проаналізувати. Такі завдання слід залишати тільки у критерійно-орієнтованих вузько спеціальних тестах.

4.5. Основні характеристики тестових завдань

4.5.1. Трудність тестових завдань

Оцінка трудності завдань проводиться по-різному залежно від вибраної теорії методів обробки емпіричних результатів. У класичній теорії в дихотомічному випадку оцінка трудності j -го завдання обчислюється за формулою

$$p_j = \frac{R_j}{N},$$

де p_j - доля правильних відповідей на j -те завдання, R_j - кількість учнів, які правильно виконали j -те завдання, N – кількість учнів в групі випробовуваних. Трудність часто обчислюється в процентах за формулою

$$P_j = \frac{R_j}{N} \cdot 100\%.$$

Зауважимо, що у рамках класичної теорії трудність завдань тим більше, чим більше учасників тестування його розв'язали, що протирічить загальноприйнятому тлумаченню поняття «трудності». У сучасній теорії тестування (див. розділ V) ця некоректність виправлена.

У добре збалансованому тесті за трудністю повинно бути декілька найважчих завдань з p_j близькою до 0, декілька легеньких завдань з p_j близькою до 1. Решта завдань повинні мати p_j в проміжку від 0,3 до 0,7. Основна маса завдань по трудності повинні наближатися до середини нормального розподілу.

Дисперсія для кожного завдання тесту обчислюється за формулою

$$\sigma_j^2 = p_j q_j \quad (j = 1, 2, \dots, N),$$

де $q_j = 1 - p_j$. Оскільки добуток $p_j q_j$ досягає максимального значення при $p_j = q_j = 0,5$, то для нормативно-орієнтованих тестів найвдалішими вважаються завдання з середньою трудністю 0,5. Ці завдання забезпечують максимальний вклад у загальну дисперсію тесту.

Середній рівень трудності завдань тесту обчислюється за формулою

$$\mu_p = \frac{\sum_{i=1}^k p_i}{k},$$

де k – число завдань в тесті. При виконанні групою випробовуваних декількох тестів різної довжини нерідко перевагу віддають середній трудності завдань, а не сирих балам.

Для нашого прикладу трудність тестових завдань у процентах матиме значення, наведені у таблиці 4.10.

Таблиця 4.10. Трудність у процентах

№ завдання	1	2	3	4	5	6	7	8
Трудність завдань у процентах - P_j	10%	20%	30%	40%	50%	70%	80%	90%

Дисперсія для кожного завдання тесту наведена у таблиці 4.11.

Таблиця 4.11. Дисперсія завдань

J	1	2	3	4	5	6	7	8
p_j	0,1	0,2	0,3	0,4	0,5	0,7	0,8	0,9
q_j	0,9	0,8	0,7	0,6	0,5	0,3	0,2	0,1
σ_j^2	0,09	0,16	0,21	0,24	0,25	0,21	0,16	0,09

Максимальний внесок у загальну дисперсію тест робить завдання 5, а також завдання 4, 3 та 6. Ці чотири завдання знаходяться у центральній частині ряду, саме вони є найвдалішими для нормативно-орієнтованих тестів. Дисперсія за результатами крайніх завдань 1 та 2, а також 7 та 8 є невисокою. Такі завдання рекомендується включати у невеликій кількості у збалансований тест.

Середній рівень трудності завдань нашого тесту $\mu_p = 0,49$ (перевірте самостійно), значення близьке до 0,5 – це є достатньо вдалим результатом.

4.5.2. Правдоподібність дистракторів

Тестові завдання закритої форми з вибором одної або декількох правильних відповідей крім правильних відповідей містять також правдоподібні неправильні відповіді, які називаються дистракторами. Оцінка правдоподібності дистракторів ґрунтується на підрахунку долі випробовуваних, які обрали неправильну відповідь.

Розглянемо приклад, коли група із 100 випробовуваних виконували завдання із п'ятьма варіантами відповідей, із яких третя відповідь правильна, а решта є дистракторами. Якщо правильно виконали завдання 60 випробовуваних, то його трудність $p_3=0,6$. Решта 40 випробовуваних виконали завдання неправильно, вибравши за правильну відповідь один із дистракторів. В ідеалі, вибір одного із чотирьох дистракторів серед 40 випробовуваних повинен бути рівномірним, тобто один той самий дистрактор повинні бути обрати 10 випробовуваних, тому що дистрактори повинні бути однаково привабливі. Доля відповідей на кожний дистрактор повинна бути $10/100=0,1$ або 10%.

Реальні результати на практиці рідко співпадають з описаною гіпотетичною ситуацією. У добре створених завданнях з вибором відповідей учні, які неправильно відповіли, повинні розподілятися майже рівномірно між всіма запропонованими дистракторами. Дистрактори, які вибирають менше 5% невірно відповівших випробовуваних, повинні бути вилучені з тесту та замінені на інші.

Поглиблений варіант дистракторного аналізу побудований на підрахунку значення точково-бісеріального коефіцієнта кореляції для кожного дистрактора в завданнях тесту. Якщо коефіцієнт бісеріальної кореляції дистрактора від'ємний і менший за $-0,2$, то сильні випробовувані не будуть вибирати даний дистрактор за правильну відповідь. Додатні або близькі до нуля значення коефіцієнта для дистракторів вказують на необхідність їх вилучення для вдосконалення неправильних відповідей. Правильним відповідям повинні відповідати додатні коефіцієнти, які перевищують $0,5$.

Припустимо, що в прикладі використовувались завдання закритої форми з п'ятьма варіантами відповідей, серед яких одна вірна, тобто за умови вдало обраних дистракторів слід очікувати рівномірний розподіл невірних відповідей – по 25%.

Розглянемо структуру відповідей на окремі тестові завдання (1, 2 та 3), представлені у наведених нижче таблицях 4.12, 4.13 та 4.14.

У завданні №1 (табл. 4.12.) вірна відповідь на третьому місці. Перший та другий дистрактори обрали по одному учаснику, що становить 11,11% від усіх дев'яти невірних відповідей. Третій дистрактор обрали троє (33,33%), а четвертий дистрактор – четверо учасників (44,45%). Такий нерівномірний розподіл говорить про невдалий підбір дистракторів завдання №1.

Таблиця 4.12. Відповіді на завдання №1

		Номери дистракторів k					Бали
		1	2	вірно	3	4	
Учасник №	1	1	0	0	0	0	7
	2	0	0	1	0	0	6
	3	0	0	0	1	0	5
	4	0	0	0	0	1	5
	5	0	0	0	1	0	4
	6	0	0	0	0	1	4
	7	0	0	0	0	1	3
	8	0	0	0	1	0	2
	9	0	0	0	0	1	2
	10	0	1	0	0	0	1
		11,11%	11,11%		33,33%	44,45%	
$(r_{pbis})_k$		0,57	-0,53	0,39	-0,08	-0,18	

Поглиблений аналіз кореляційної залежності результатів обрання дистракторів завдання №1 з загальними балами учасників демонструє деякі цікаві залежності. При тому, що число вибору дистракторів 1 та 2 однакове (по одному разу) коефіцієнт кореляції у першого дистрактора достатньо велике додатне значення 0,57, а у другого від'ємне число -0,53. Пояснюється цей факт тим, що перший дистрактор обрав добре підготовлений найсильніший учасник, що скоріше за все могло статися з причини його неуважності, а не за рахунок оригінального формулювання дистрактора. Взагалі кажучи, вважається, що дистрактор №1 є неякісним і його слід замінити, однак існує протилежна думка про те що такі варіанти відповіді потрібні у невеликій кількості для перевірки уважності сильних учасників тестування. Другий дистрактор обрав учасник, який показав найслабший результат, тобто цей дистрактор виконує призначену йому роль заплутування та відволікання уваги. Дистрактори 3 та 4 дають близьку до нуля від'ємну кореляцію, оскільки вони були обраними учасниками як із високими так і з низькими балами, ці дистрактори також слід проаналізувати та корегувати.

У завданні №2 (табл. 4.13.) вірна відповідь була розміщена на п'ятому місці. Всі чотири дистрактори обрали по два учасника, що становить 25,00% від усіх восьми невірних відповідей, тобто у завдання №2 підбір дистракторів є достатньо вдалим.

Таблиця 4.13. Відповіді на завдання №2

		Номери дистракторів k					Вірно	Бали
		1	2	3	4			
Учасник №	1	0	0	0	0	1	7	
	2	0	0	1	0	0	6	
	3	0	0	0	1	0	5	
	4	1	0	0	0	0	5	
	5	0	0	0	0	1	4	
	6	0	0	1	0	0	4	
	7	1	0	0	0	0	3	
	8	0	0	0	1	0	2	
	9	0	1	0	0	0	2	
	10	0	1	0	0	0	1	
		25,00%	25,00%	25,00%	25,00%			
	$(r_{pbis})_k$	0,028	-0,66	0,303	-0,11	0,441		

Проведено поглиблений аналіз кореляційної залежності результатів обрання дистракторів завдання №2 з загальними балами учасників. Виявлено, що тільки один коефіцієнт кореляції (у другого дистрактора) є прийнятним -0,66. Четвертий дистрактор дав близьку до нуля від'ємну кореляцію, а дистрактори 1 та 3 мають додатний коефіцієнт кореляції з загальними балами учасників, причиною чого є обрання цих дистракторів деякими сильними учасниками тестування. Отже, незважаючи на рівномірний розподіл кількості обрань дистракторами, три з них потребують додаткового аналізу.

У завданні №3 вірна відповідь була розміщена на першому місці. Перший, другий та третій дистрактори обрали по два учаснику, що становить 28,57% від усіх семи невірних відповідей. Четвертий дистрактор обрав один учасник тестування (14,29%), отже розподіл обрань дистракторів завдання №3 є достатньо рівномірним (він є найкращим із можливих для семи учасників, які дали невірну відповідь)

Таблиця 4.14. Відповіді на завдання №3

		Номери дистракторів k					Бали
		вірно	1	2	3	4	
Учасник №	1	1	0	0	0	0	7
	2	1	0	0	0	0	6
	3	0	0	0	1	0	5
	4	0	1	0	0	0	5
	5	0	0	1	0	0	4
	6	1	0	0	0	0	4
	7	0	0	1	0	0	3
	8	0	0	0	1	0	2
	9	0	1	0	0	0	2
	10	0	0	0	0	1	1
			28,57%	28,57%	28,57%	14,29%	
$(r_{pbis})_k$		0,638	-0,11	-0,11	-0,11	-0,53	

Аналогічно до двох попередніх завдань проведено поглиблений аналіз кореляційної залежності результатів обрання дистракторів завдання №3 з загальними балами учасників. Виявлено, що всі коефіцієнти кореляції є від'ємними, найкращий коефіцієнт має четвертий дистрактор.

Зауважимо, що у нашому прикладі розглянуто випадок невеликої кількості учасників тестування. У цьому випадку оцінювання якості дистракторів є неточним. Особливо складно робити аналіз дистракторів більш простих завдань, які не змогли розв'язати мала кількість тестованих. Такі дослідження потрібно проводити для великих репрезентативних вибірок учасників.

4.5.3. Дискримінативність тестового завдання

Дискримінативністю називається здатність завдання диференціювати учнів на кращих і гірших. Висока дискримінативність – важлива характеристика вдалого тестового завдання. У нормативно-орієнтованих тестах завдання повинні мати високу дискримінативну здатність. В якості показника дискримінативності завдань використовують п'ять параметрів.

Один із них (показник розпізнавальної здатності) базується на виборі порогових оцінок в певних точках розподілу критеріальної оцінки, решта чотири є різними видами коефіцієнтів кореляції: точкова-бісеріальна, бісеріальна, коефіцієнт фі, тетрахорична.

Показник розпізнавальної здатності застосовується тільки для дихотомічного оцінювання завдань. Він розраховується за простою формулою

$$D_j = (p_1)_j - (p_0)_j,$$

де D_j - індекс дискримінативності для j -го завдання тесту; $(p_1)_j$ - доля учнів, які правильно виконали j -те завдання у підгрупі з 27% кращих учнів за результатами виконання тесту; $(p_0)_j$ - доля учнів, які правильно виконали j -те завдання у підгрупі з 27% гірших учнів за результатами виконання тесту.

Значення індексу D_j належать інтервалу $[-1;1]$. Максимальне значення $D_j=1$ досягається в тому випадку, коли всі учні із підгрупи кращих вірно виконали j -те завдання тесту, а з підгрупи гірших це завдання не виконав вірно ні один учень. У цьому випадку завдання має максимальний диференціюючий ефект. Завдання з $D_j=0$ або $D_j<0$ потрібно з тесту вилучити.

Базуючись на практичному досвіді Ебель запропонував наступні принципи для інтерпретації D_j :

1. Якщо $D_j \geq 0,40$, то завдання функціонує досить задовільно.
2. Якщо $0,30 \leq D_j \leq 0,39$, то потрібна невелика корекція завдання або в ньому немає необхідності.
3. Якщо $0,20 \leq D_j \leq 0,29$, то завдання потрібно переглянути.
4. Якщо $D_j \leq 0,19$, то завдання потрібно вилучити із тесту або повністю переробити.

Розглянемо приклад, наведений у попередніх пунктах. Із десятих осіб виділимо трьох учасників тестування (учасники 1, 2 та 3), які показали сильний результат, та трьох учасників (учасники 8,9 та 10), які показали гірший результат (табл. 4.15.).

Таблиця 4.15. Групи сильних та слабих учасників

	№	1	2	3	4	5	6	7	8	
Високі результ.	1	0	1	1	1	1	1	1	1	7
	2	1	0	1	1	0	1	1	1	6
	3	0	0	0	1	1	1	1	1	5
	4	0	0	0	1	1	1	1	1	5
	5	0	1	0	0	1	0	1	1	4
	6	0	0	1	0	1	0	1	1	4
	7	0	0	0	0	0	1	1	1	3
Низькі результ.	1	0	0	0	0	0	1	1	0	2
	9	0	0	0	0	0	1	0	1	2
	10	0	0	0	0	0	0	0	1	1

Обчислимо $(p_1)_j$ - долі учнів, які правильно виконали j -те завдання у підгрупі з 27% кращих учнів за результатами виконання тесту; $(p_0)_j$ - долі учнів, які правильно виконали j -те завдання у підгрупі з 27% гірших учнів за результатами виконання тесту

j	1	2	3	4	5	6	7	8
$(p_1)_j$	0,3333	0,3333	0,6667	1	0,6667	1	1	1
$(p_0)_j$	0	0	0	0	0	0,6667	0,3333	0,6667

Отже, для нашого прикладу отримаємо такі значення показника розпізнавальної здатності $D_j = (p_1)_j - (p_0)_j$:

№ тест. Завдання	1	2	3	4	5	6	7	8
D_j	0,3333	0,3333	0,6667	1	0,6667	0,3333	0,6667	0,3333

Робимо висновок, що маємо одне завдання з номером 4 з найкращою роздільною здатністю; три завдання з номерами 3,5,7 функціонують задовільно; чотири завдання з номерами 1,2,6,8 потребують корекції.

Кореляційні показники дискримінативності завдань. Коефіцієнт точково-бісеріальної кореляції можна обчислювати за формулами

$$(r_{pbis})_j = \frac{(\bar{X}_1)_j - \bar{X}}{S_X} \sqrt{\frac{p_j}{q_j}} \quad \text{або} \quad (r_{pbis})_j = \frac{(\bar{X}_1)_j - (\bar{X}_0)_j}{S_X} \sqrt{p_j q_j},$$

де $(r_{pbis})_j$ - коефіцієнт точково-бісеріальної кореляції для j -го завдання; $(\bar{X}_1)_j$ - середнє значення індивідуальних балів учнів, які виконали вірно j -те завдання; $(\bar{X}_0)_j$ - середнє значення індивідуальних балів учнів, які виконали невірно j -те завдання; \bar{X} - середнє значення балів по всій вибірці учнів; S_X - стандартне відхилення по множині індивідуальних балів.

Чим вища кореляція, тим краще завдання тесту. Завдання з близькими до нуля або від'ємними значеннями $(r_{pbis})_j$ повинні бути вилучені з тесту. На думку багатьох фахівців, завдання з $(r_{pbis})_j < 0,2$ повинні бути вилучені з тесту.

Замість $(r_{pbis})_j$ можна використовувати $(r_{bis})_j$ при умові, що значення латентної змінної результатів виконання тесту розподілені нормально. Оскільки значення $(r_{bis})_j$ трохи вищі значень $(r_{pbis})_j$, то всі завдання з $(r_{bis})_j < 0,3$ повинні бути вилучені з тесту.

Коли оцінки по завданнях, що оцінюються дихотомічно, повинні бути скорельовані з оцінками по дихотомічному критерію (наприклад, успіх або провал, стать випробовуваного і т.п.), можна використовувати ϕ -коефіцієнт кореляції. Також можна використовувати коефіцієнт тетрахомічної кореляції при умові, що завдання і оцінки критерію отримуються шляхом дихотомізації даних за двома нормально розподіленими змінними. Цей коефіцієнт використовується рідко із-за обчислювальної складності, але він рекомендується в тих випадках, коли кореляції будуть використовуватися в подальшому факторному аналізі.

Для нашого прикладу коефіцієнт точково-бісеріальної кореляції матиме такі значення (тут для наочності наведені також значення бісеріальної кореляції):

j	1	2	3	4	5	6	7	8
$(r_{pbis})_j$	0,3661	0,4184	0,6049	0,79	0,5753	0,3082	0,6276	0,3312
$(r_{bis})_j$	0,3859	0,4411	0,6376	0,8328	0,6064	0,3248	0,6616	0,3492

Нескладно бачити, що у даному прикладі маємо досить високий кореляційний показник дискримінативності завдань, з чого робимо висновок, що завдання добре диференціюють учнів і можна залишити всі завдання даного тесту.

4.5.4 Надійність завдань і показники валідності

Показник надійності завдання обчислюється за формулою $\sigma_i \rho_{iX}$, де σ_i - стандартне відхилення i -го завдання, ρ_{iX} - кореляція між i -тим завданням та загальною тестовою оцінкою. Для дихотомічного оцінювання завдань формула набуває вигляду $\sqrt{p_i q_i} \rho_{iX}$, де ρ_{iX} - точково-бісеріальна кореляція між i -тим завданням и загальною тестовою оцінкою. Такий показник розглядають як показник надійності, так і показник валідності завдання, що залежить від трактування числового значення точково-бісеріальної кореляції. Його пропонують використовувати у тих випадках, коли метою відбору завдань є підвищення надійності тестових оцінок на основі вибору завдань, які добре диференціюють за відповідним критерієм.

Рівність $\sigma_X^2 = (\sum \sigma_i \rho_{iX})^2$ може бути корисною при аналізі завдань в ситуації, коли розробник тесту встановив для дисперсії загальної тестової оцінки бажану мінімальну величину. Як тільки починається процес відбору завдань тесту, сума показників надійності завдань по мірі добавлення кожного нового завдання буде збільшуватися і так до тих пір, поки не буде досягнутий бажаний мінімальний рівень дисперсії загальної тестової оцінки.

Якщо розробник тесту встановив мінімальну величину для коефіцієнта внутрішньої обґрунтованості, виміряного за допомогою коефіцієнта альфа, то величину цього коефіцієнта також можна переоцінити по мірі добавлення кожного нового завдання, використовуючи тільки дані на рівні завдань за формулою

$$\rho_\alpha = \frac{k}{k-1} \left[1 - \frac{\sum \sigma_i^2}{(\sum \sigma_i \rho_{iX})^2} \right],$$

де k – число завдань, відібраних до даного моменту.

Якщо розробник тесту хоче визначити мінімальну величину коефіцієнта валідності як кореляцію між вибраними завданнями і деяким зовнішнім

критерієм, то коефіцієнт валідності, по мірі того як добавляється кожне нове завдання, може бути наближено оцінений із даних на рівні завдань за формулою

$$\rho_{XY} = \frac{\sum \sigma_i \rho_{iY}}{\sum \sigma_i \rho_{iX}}.$$

4.6. Надійність тесту та її числові характеристики

4.6.1. Надійність тесту в класичній теорії.

Надійністю (reliability) називається характеристика тесту, яка відображає точність тестових вимірювань, а також стійкість тестових результатів до дії випадкових факторів. Отже, термін «надійність» має два значення. По-перше, тест вважається надійним, якщо він забезпечує високу точність вимірювань. З цієї точки зору, надійність є *мірою тісноти зв'язку істинних оцінок випробуваних з їхніми оцінками, що спостерігаються*. По-друге, тест вважається надійним, якщо він дає при повторному виконанні близькі результати при умові, що підготовка учнів не змінилася за час до повторного виконання тесту. Тут надійність виступає *мірою тісноти зв'язку між оцінками, що спостерігаються по двох паралельних тестах*.

У вимозі перевірки тесту на надійність реалізується важлива ідея методологічного характеру, яка пов'язана з необхідністю обґрунтування якості тестових вимірювань. Надійність обчислюють за допомогою коефіцієнта кореляції.

4.6.2. Показник надійності.

Розглянемо формулу для обчислення коефіцієнта кореляції між істинними оцінками (T) і оцінками (X), що спостерігаються. Зв'язок між цими оцінками задається рівністю

$$X_i = T_i + E_i$$

для i -того випробуваного. Для генеральної сукупності випробуваних цю рівність у відхиленнях можна переписати так

$$X_i - \bar{X} = (T_i - \bar{T}) + (E_i - \bar{E}),$$

де \bar{X} , \bar{T} , \bar{E} - відповідні середні величини. Коефіцієнт кореляції між величинами T та X матиме вигляд

$$\rho_{XT} = \frac{i \sum (X_i - \bar{X})(T_i - \bar{T})}{N \cdot \sigma_X \cdot \sigma_T},$$

де $\frac{i \sum (X_i - \bar{X})(T_i - \bar{T})}{N}$ - коваріація, σ_X , σ_T - стандартні відхилення величин

T та X . Виконаємо перетворення

$$\begin{aligned} \rho_{XT} &= \frac{i \sum (X_i - \bar{X})(T_i - \bar{T})}{N \cdot \sigma_X \cdot \sigma_T} = \frac{i \sum ((T_i - \bar{T}) + (E_i - \bar{E}))(T_i - \bar{T})}{N \cdot \sigma_X \cdot \sigma_T} = \\ &= \frac{i \sum (T_i - \bar{T})^2}{N \cdot \sigma_X \cdot \sigma_T} + \frac{i \sum (E_i - \bar{E})(T_i - \bar{T})}{N \cdot \sigma_X \cdot \sigma_T}. \end{aligned}$$

Оскільки $\frac{i \sum (T_i - \bar{T})^2}{N} = \sigma_T^2$ - дисперсія величини T , то перший доданок

набуває вигляду $\frac{i \sum (T_i - \bar{T})^2}{N \cdot \sigma_X \cdot \sigma_T} = \frac{\sigma_T^2}{\sigma_X \cdot \sigma_T} = \frac{\sigma_T}{\sigma_X}$. Перетворимо другий доданок

$$\frac{i \sum (E_i - \bar{E})(T_i - \bar{T})}{N \cdot \sigma_X \cdot \sigma_T} = \frac{i \sum (E_i - \bar{E})(T_i - \bar{T})}{N \cdot \sigma_X \cdot \sigma_T} \cdot \frac{\sigma_T \cdot \sigma_E}{\sigma_T \cdot \sigma_E} = \left(\frac{i \sum (E_i - \bar{E})(T_i - \bar{T})}{N \cdot \sigma_E \cdot \sigma_T} \right) \cdot \left(\frac{\sigma_T \cdot \sigma_E}{\sigma_X \cdot \sigma_E} \right)$$

Перший вираз в дужках є кореляцією ρ_{TE} між істинною оцінкою та її похибкою, яка дорівнює 0 за четвертим положенням класичної теорії, тобто весь другий доданок рівний нулю. Остаточно отримуємо рівність

$$\rho_{XT} = \frac{\sigma_T}{\sigma_X}.$$

Величина ρ_{XT} називається *показником надійності*, який рівний відношенню стандартного відхилення істинних оцінок до стандартного відхилення оцінок, що спостерігаються. Зауважимо, що це – кореляція між істинними оцінками та всіма оцінками, що спостерігаються по результатах множини тестувань, які повторюються. Даний вираз не має значного практичного застосування, тому що істинні оцінки безпосередньо не

спостерігаються і ми не можемо отримати всі можливі оцінки, які спостерігаються для кожного із випробуваних. Однак ймовірно провести двічі тестування групи випробуваних за допомогою одного і того ж тесту або двох його форм. Якщо два тести відповідають вимогам паралельності, то можна встановити математичний зв'язок між ρ_{XT} - кореляцією між істинними оцінками і оцінками, що спостерігаються, та $\rho_{XX'}$ - кореляцією між оцінками, що спостерігаються, по двох паралельних тестах.

Згідно класичної теорії істинної оцінки, два тести є паралельними, якщо:

1. кожний випробуваний має одну і ту ж істинну оцінку по обох формах;
2. дисперсії похибок для двох форм рівні.

Такі тести будуть, як наслідок, мати рівні середні значення і рівні дисперсії. Також розумно припустити, що паралельні тести будуть відповідати один одному за змістом.

4.6.3. Коефіцієнт надійності

Коефіцієнт надійності можна означити як кореляція між оцінками по паралельних тестових формах.

Позначимо відхилення оцінок випробуваних по двох паралельних тестах як $x_1 = X_{i1} - \bar{X}_1$ і $x_2 = X_{i2} - \bar{X}_2$. Згідно моделі, ці оцінки у відхиленнях можуть бути записані у вигляді

$$x_1 = t_1 + e_1, \quad x_2 = t_2 + e_2.$$

Формула для кореляції між тестовими оцінками X_1 і X_2 , що спостерігаються, через їх відхилення має вигляд

$$\rho_{X_1 X_2} = \frac{\sum x_1 \cdot x_2}{N \cdot \sigma_{X_1} \cdot \sigma_{X_2}}.$$

Її можна записати так

$$\rho_{X_1 X_2} = \frac{\sum (t_1 + e_1) \cdot (t_2 + e_2)}{N \cdot \sigma_{X_1} \cdot \sigma_{X_2}}.$$

Якщо виконати алгебричні перетворення, то вираз набуде вигляду

$$\rho_{X_1 X_2} = \frac{\sum t_1 \cdot t_2}{N \sigma_{X_1} \sigma_{X_2}} + \frac{\sum t_1 \cdot e_2}{N \sigma_{X_1} \sigma_{X_2}} + \frac{\sum t_2 \cdot e_1}{N \sigma_{X_1} \sigma_{X_2}} + \frac{\sum e_1 \cdot e_2}{N \sigma_{X_1} \sigma_{X_2}}.$$

Користуючись положеннями класичної моделі тестової оцінки, можна показати, що кожний із трьох останніх доданків дорівнює нулю (у другому і третьому можна отримати кореляцію виду $\rho_{TE} = 0$, а в четвертому - $\rho_{E_1 E_2} = 0$). Оскільки значення істинних оцінок для будь-якого випробуваного вважаються рівними по всіх випадках тестування, то $t_1 = t_2$ і $\sigma_1 = \sigma_2$ за означенням паралельних тестів. Таким чином

$$\rho_{X_1 X_2} = \frac{\sum t_1 \cdot t_2}{N \sigma_{X_1} \sigma_{X_2}} = \frac{\sum t_1^2}{N \sigma_{X_1}^2}$$

або

$$\rho_{X_1 X_2} = \frac{\sigma_T^2}{\sigma_X^2}.$$

Таким чином, *коефіцієнт надійності* може бути математично отриманий як *відношення дисперсії істинної оцінки до дисперсії оцінки, яка спостерігається* (або є *квадратом показника надійності*).

При інтерпретації інформації про надійність важливо враховувати відмінності між:

1. $\rho_{X_1 X_2}$ - долею дисперсії оцінки, що спостерігається, яка може бути пояснена варіацією істинних оцінок випробуваних;
2. $(\rho_{X_1 X_2})^2$ - долею дисперсії оцінки, що спостерігається по одному із двох паралельних тестів, яка могла би бути передбачена із дисперсії оцінки, що спостерігається по другому із двох паралельних тестів;
3. ρ_{XT} - кореляцією між істинними оцінками і оцінками, що спостерігаються.

Важливо зауважити, що розглянутий коефіцієнт надійності для ряду тестових оцінок є просто теоретичним поняттям. Його можна було б отримати, якби насправді існували паралельні тести.

4.6.4. Надійність складених тестових оцінок

Існують два методи визначення надійності складеної оцінки в термінах статистичних властивостей її компонентів.

Прогностична **формула Спірмена-Брауна**

$$\rho_{CC'} = \frac{k\rho_{ii'}}{1 + (k-1)\rho_{ii'}}$$

обчислює надійність складеної оцінки C , як функцію від окремих компонентів. У цій формулі величина C визначає складену загальну оцінку за k строго паралельних тестових форм, $\rho_{ii'}$ - коефіцієнт надійності одної із паралельних форм (i -тої).

У реальних ситуаціях тестування не можна бути впевненим, що всі форми строго паралельні. Тоді можна отримати нижню оцінку для коефіцієнта надійності складеної оцінки

$$\rho_{CC'} \geq \frac{k}{k-1} \left(\frac{\sum_{i \neq j} \sigma_{ij}}{\sigma_C^2} \right) \quad \text{або} \quad \rho_{CC'} \geq \frac{k}{k-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_C^2} \right).$$

Вираз у правій частині називається **коефіцієнтом альфа Кронбаха**. Якщо тест компонується із непаралельних субтестів, то можна оцінити нижню межу коефіцієнта надійності за допомогою коефіцієнта альфа. Його обчислення вимагає знання числа k субтестів, дисперсії σ_C^2 складених оцінок і суми $\sum_{i \neq j} \sigma_{ij}$ всіх коваріацій між всіма субтестами.

4.7. Методи оцінки надійності та її прогнозування

4.7.1. Методи двократного тестування

Оцінка надійності нормативно-орієнтованих тестів проводиться різними методами, які за способом виконання можна умовно розділити на дві групи. Перша група методів базується на двократному тестуванні, яке проводиться за допомогою одного і того ж тесту або за допомогою двох паралельних форм тесту. Друга група передбачає однократне тестування при оцінці надійності тесту.

Ретестовий метод оцінки надійності (test-retest reliability) ґрунтується на підрахунку кореляції індивідуальних балів випробовуваних, отриманих в результаті двократного виконання одного і того ж тесту. Зазвичай повторне тестування проводиться через 2-3 тижні. Низька кореляція буде наслідком ненадійності тесту. Для підрахунку коефіцієнта надійності використовується формула Пірсона

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

де $(r_{XY})_{\text{рет}}$ - коефіцієнт надійності тесту за ретестовим методом; X_i - індивідуальний бал i -го випробуваного в першому тестуванні; Y_i - індивідуальний бал i -го випробуваного в другому тестуванні. Цей коефіцієнт кореляції називають *коефіцієнтом стійкості (стабільності)*.

Метод паралельних форм (parallel-form reliability) найкраще підходить для оцінки надійності в рамках класичної теорії тестів, але з практичної точки зору це найменш ефективний метод, бо потрібно доводити паралельність форм. Цей метод аналогічний попередньому з використанням двох паралельних форм тесту. Коефіцієнт надійності обчислюється теж за формулою Пірсона. Цей коефіцієнт кореляції називають *коефіцієнтом еквівалентності*. Чим вищий коефіцієнт еквівалентності, тим з більшою впевненістю можна розраховувати на те, що оцінки по різних формах можуть використовуватися як взаємно замінювані.

4.7.2. Методи однократного тестування.

Метод розщеплення тесту на дві частини (split-half method) найпоширеніший із-за своєї зручності. Він дозволяє обчислити коефіцієнт надійності при однократному виконанні учнями тесту. Для оцінки надійності результати тестування ділять на дві частини: в одну включають дані випробовуваних з парними, а в другу – з непарними номерами завдань тесту. Зауважимо, що способів поділу на дві частини існує багато. Коефіцієнт

надійності $(r_{XY})_{\text{розщ}}$ обчислюють, маючи на увазі, що величина X відповідає першій частині, а Y – другій. Для оцінки надійності всього тесту використовується формула Спірмена-Брауна при $k=2$ (дві форми)

$$r_H = \frac{2(r_{XY})_{\text{розщ}}}{1 + (r_{XY})_{\text{розщ}}}.$$

Розглянутий метод розщеплення ґрунтується на припущенні паралельності двох половинок тесту, що не завжди вірно. Кореляція двох половинок зростає по мірі росту гомогенності тесту. У зв'язку з цим метод розщеплення нерідко називають методом оцінки *внутрішньої узгодженості* тесту.

Методи, що ґрунтуються на коваріаціях завдань. Існують три методи, які приводять до ідентичних результатів: коефіцієнт альфа Кронбаха, формула Кьюдера-Річардсона, дисперсійний аналіз Хойта.

Коефіцієнт альфа. Кронбах запропонував обчислювати коефіцієнт внутрішньої узгодженості тесту за формулою

$$\tilde{\alpha} = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_X^2} \right),$$

k - число завдань в тесті; S_i^2 - дисперсія по i -му завданню; S_X^2 - загальна дисперсія по тесту. Коефіцієнт альфа може застосовуватися до результатів представлених дихотомічною шкалою або завдань оцінювання за допомогою широкого діапазону вагових коефіцієнтів.

Формула Кьюдера-Річардсона застосовується тільки до завдань, які оцінюються дихотомічно. Вона має вигляд

$$KR_{20} = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k p_i q_i}{S_X^2} \right)$$

та ідентична коефіцієнту альфа для дихотомічного випадку. Остаточні результати оцінювання надійності представлені в таблиці 4.16.

Таблиця 4.16. Оцінювання надійності

Головне джерело помилки	Коефіцієнт надійності	Процедури отримання даних	Статистична трактовка даних
Зміни учасників з часом	Коефіцієнт стабільності	Тестування, проміжок чекання, повторне тестування	Обчислення коефіцієнта кореляції Пірсона $(r_{XY})_{\text{пер}}$
Вибіркова перевірка змісту від форми до форми	Коефіцієнт еквівалентності	Пред'явлення форм 1 і 2	Обчислення коефіцієнта кореляції Пірсона $(r_{XY})_{\text{пер}}$
Вибіркова перевірка змісту або виявлення некоректних завдань	Коефіцієнт внутрішньої узгодженості	Пред'явлення одної форми один раз	<p>А) Ділення тесту на дві половини, підрахунок кореляції оцінок $(r_{XY})_{\text{розщ}}$ по половиних тесту, обчислення надійності за формулою Спірмена-Брауна</p> $r_H = \frac{2(r_{XY})_{\text{розщ}}}{1 + (r_{XY})_{\text{розщ}}}$ <p>Б) Обчислення дисперсій по завданнях, а також коефіцієнта альфа за формулою</p> $\tilde{\alpha} = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_X^2} \right)$

Зауважимо також, що прогнозувати оцінки істинних балів випробовуваних по розподілу спостережуваних балів і коефіцієнту надійності тесту дозволяють методи регресійного аналізу, відповідне регресійне рівняння має вигляд $T_i = \bar{X} + r_H(X_i - \bar{X})$.

Перевіримо надійність на нашому прикладі. Розділимо початковий тест на два субтести (парні і непарні номери завдань), отримаємо:

Завдання					бал за субтест
№	1	3	5	7	
1	0	1	1	1	3
2	1	1	0	1	3
3	0	0	1	1	2
4	0	0	1	1	2
5	0	0	1	1	2
6	0	1	1	1	3
7	0	0	0	1	1
8	0	0	0	1	1
9	0	0	0	0	0
10	0	0	0	0	0
	1	3	5	8	

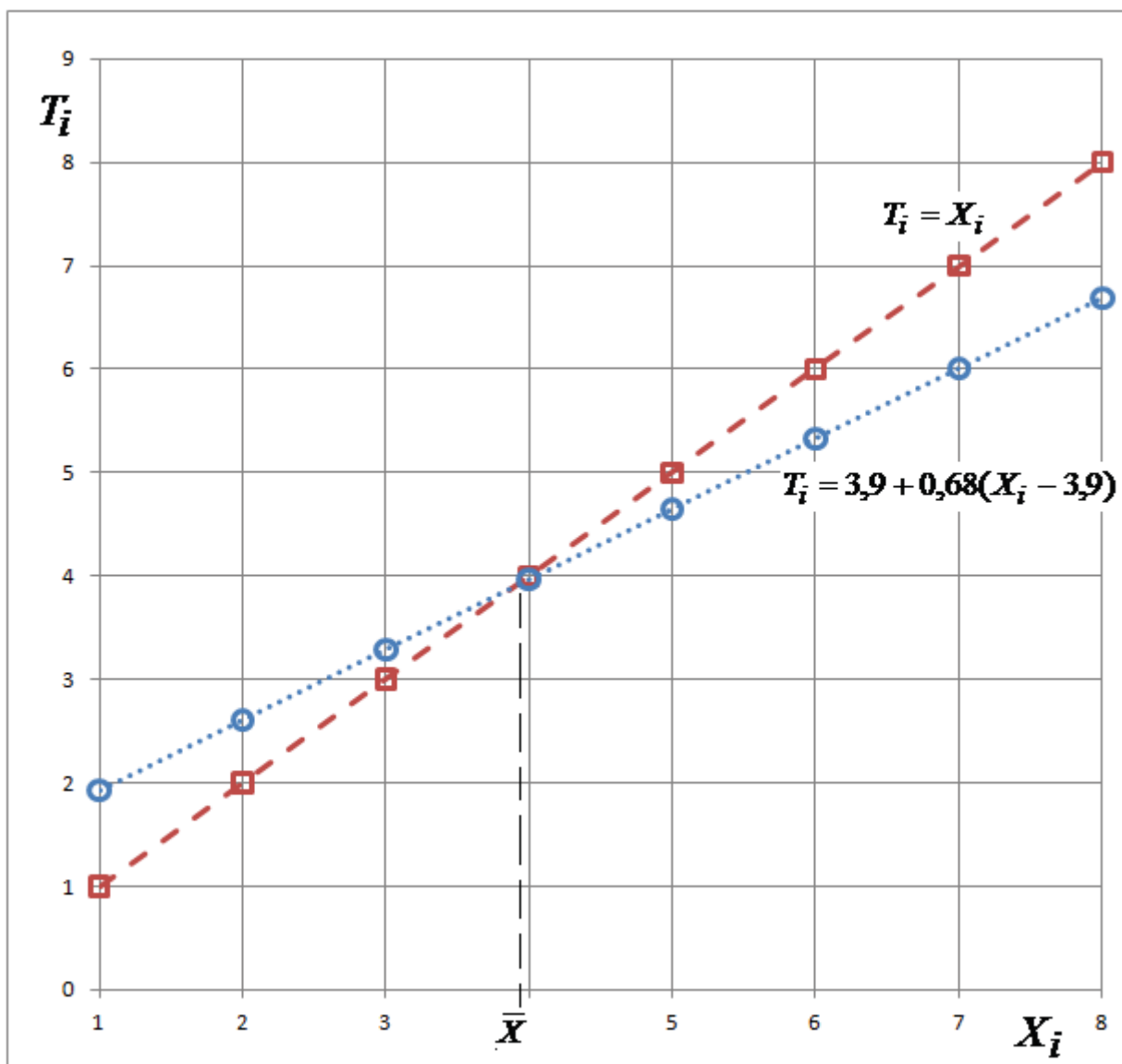
Завдання					бал за субтест
№	2	4	6	8	
1	1	1	1	1	4
2	0	1	1	1	3
3	0	1	1	1	3
4	0	1	1	1	3
5	1	0	0	1	2
6	0	0	0	1	1
7	0	0	1	1	2
8	0	0	1	0	1
9	0	0	1	1	2
10	0	0	0	1	1
	2	4	7	9	

Підрахуємо кореляції оцінок $(r_{XY})_{розщ} = 0,52$ по половиних тесту і обчислимо надійність за формулою Спірмена-Брауна $r_n = 0,68$.

Використавши регресійне рівняння зробимо прогноз оцінок істинних балів за лінійною функцією регресії $T_i = 3,9 + 0,68(X_i - 3,9)$

отримані бали	8	7	6	5	5	4	4	3	2	2	1
спрогнозовані бали	6,69	6,01	5,33	4,65	4,65	3,97	3,97	3,29	2,61	2,61	1,93

Представимо результат графічно. На рисунку пунктирна пряма лінія з квадратними маркерами є графіком отриманих при тестуванні балів, тобто $T_i = X_i$. Точкова лінія з круглими маркерами є лінією прогнозу істинних балів $T_i = 3,9 + 0,68(X_i - 3,9)$.



Нескладно бачити, що учасники тестування, які отримали кількість балів, близьку до середнього значення, фактично отримують бали, достатньо близькі до реальних. При цьому ті учасники тестування, які отримали високі бали при тестуванні, мають за прогнозом нижчі істинні бали. І навпаки, учасники тестування, які отримали низькі бали при тестуванні, мають за прогнозом вищі істинні бали. Отже, за класичною теорією тестування тест завищує реальну оцінку сильних учасників та занижує її для слабих учасників. Очевидно, чим вище надійність, тим ближче результати тестування до істинних рівнів підготовки учасників.

4.8. Фактори впливу на надійність та похибки її обчислення

4.8.1. Фактори, що впливають на коефіцієнти надійності

Гомогенність групи. Величина коефіцієнта надійності залежить від варіації як істинних оцінок, так і оцінок по індивідуумах. Таким чином гомогенність групи випробовуваних є важливим фактором при розробці тесту і його виборі. Розробники тесту повинні враховувати склад і варіативність групи випробовуваних, для яких розробляється тест.

Часові обмеження. При фіксованому часі виконання тесту одні випробовувані встигають його завершити, а інші – ні. Це означає, що дисперсія швидкості роботи випробовуваних є частиною дисперсії істинних оцінок. Існують тести на швидкість виконання і в них час є важливою характеристикою. Однак, якщо швидкість відповіді є неважливою, тоді час виконання тесту повинен бути таким, щоб майже всі випробовувані могли завершити роботу. Найкращим методом оцінювання надійності швидкісного тесту буде ретестовий метод або метод еквівалентних форм.

Довжина тесту. При конструюванні тесту корисно вміти визначити необхідну його довжину, щоб досягти запланованої надійності тесту. Для цього можна скористатися формулою Спірмена-Брауна

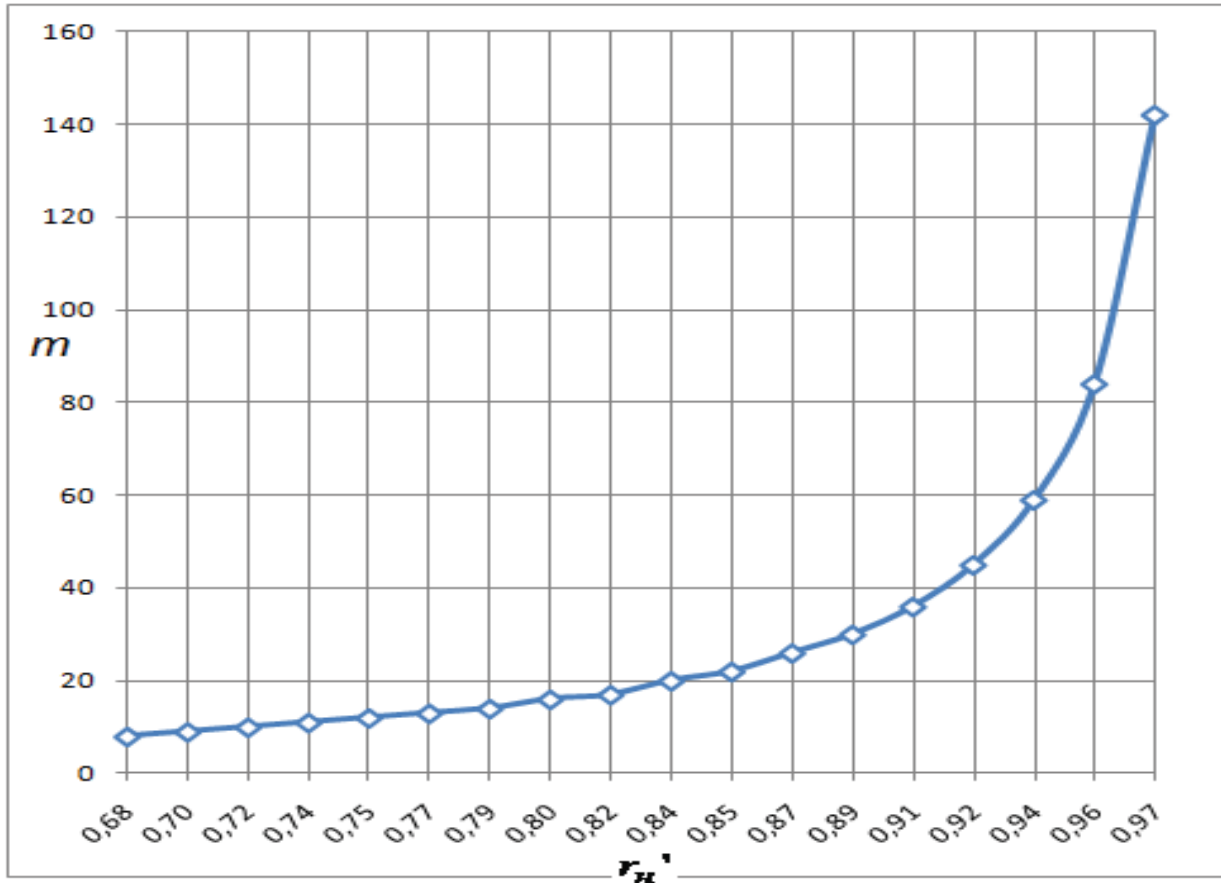
$$r'_H = \frac{kr_H}{1 + (k-1)r_H},$$

де k – число разів, на яке збільшується довжина тесту; r_H – коефіцієнт надійності тесту до збільшення його довжини; r'_H – коефіцієнт надійності після збільшення його довжини. Оскільки формула Спірмена-Брауна отримана для паралельних форм тесту, то збільшення довжини тесту повинно здійснюватися за рахунок завдань, які не змінюють змістовної області та завдання повинні мати характеристики близькі до характеристик початкових завдань. Цю формулу можна використовувати по різному. З одної сторони, маючи початкову надійність r_H та задавши число k , можна обчислити очікувану надійність r'_H . З іншої сторони, маючи показники r_H та r'_H , можна обчислити число k .

На нашому прикладі перевіримо у скільки раз потрібно подовжити тест із 8 завдань для того, щоб надійність ($r_H = 0,68$) збільшилась на 2,5%, 5%, 5,5% і т.д. Використовуючи формулу Спірмена-Брауна для тесту з довжиною 8 завдань отримаємо

Збільшення надійності r_H на %	Надійність нового тесту r_H'	Кратність збільшення довжини k	Число m завдань у новому тесті
0,00%	0,68	1	8
2,50%	0,70	1,083	9
5,00%	0,72	1,177	10
7,50%	0,74	1,283	11
10,00%	0,75	1,4036	12
12,50%	0,77	1,541	13
15,00%	0,79	1,702	14
17,50%	0,80	1,890	16
20,00%	0,82	2,115	17
22,50%	0,84	2,386	20
25,00%	0,85	2,722	22
27,50%	0,87	3,147	26
30,00%	0,89	3,703	30
32,50%	0,91	4,461	36
35,00%	0,92	5,557	45
37,50%	0,94	7,281	59
40,00%	0,96	10,389	84
42,50%	0,97	17,662	142
45,00%	0,99	54,501	437

Для обчислення цілих значень довжини нового більш надійного тесту виконуємо округлення результату з надлишком, щоб забезпечити нову надійність. На рисунку наведено графік залежності від бажаної надійності r_H' числа завдань m у новому тесті, за умови того, що вихідний тест містив 8 завдань та мав надійність $r_H = 0,68$.



Нескладно бачити, що при наближенні коефіцієнта надійності до одиниці кількість завдань для її забезпечення починає різко зростати. При цьому слід врахувати, що кількість завдань не рекомендується давати більше 50-70. Отже, найкращий результат, який теоретично можна досягти за даних умов за рахунок збільшення довжини - це коефіцієнт надійності нового тесту на рівні 0,95.

4.8.2. Похибка вимірювання надійності

У класичній теорії тестів виконується одна із важливих рівностей

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2.$$

З неї легко отримати співвідношення $\frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}$ або

$$\rho_{XX'} = 1 - \frac{\sigma_E^2}{\sigma_X^2}.$$

Ця формула є іншим виразом для коефіцієнта надійності. Переписавши її в статистичних характеристиках $r_H = 1 - \frac{S_E^2}{S_X^2}$ і виконавши перетворення,

отримаємо

$$S_E = S_X \sqrt{1 - r_H}.$$

Отримана формула використовується для обчислення S_E , яке трактується як стандартне відхилення результатів випробовуваного від його істинного балу, отримане при виконанні ним великого числа паралельних форм тесту. В ідеальній ситуації результати тестування i -того випробовуваного, який виконував багато разів один і той самий тест, утворюють нормальний розподіл навколо істинного балу T_i зі стандартним відхиленням S_E .

На практиці S_E розглядається як статистична величина, яка відображає степінь точності окремих вимірювань, тому величину S_E використовують для визначення меж довірчого інтервалу, всередині якого повинен знаходитися істинний бал випробовуваного.

Загальноприйнятим є підхід, коли довірчий інтервал будується навколо спостережуваного балу випробовуваного, хоч це не зовсім вірно, тому що мова повинна йти про окіл навколо істинного балу. Однак в прикладних дослідженнях зазвичай цей факт ігнорується і довірчий інтервал, наприклад, при заданому ризику допустити помилку $\alpha = 0,05$, тобто в п'яти випадках із ста, приймається рівним $(X_i - 1,96S_E; X_i + 1,96S_E)$, де X_i – спостережуваний бал i -го випробовуваного. Істинний бал цього випробовуваного може знаходитися в будь-якій точці цього інтервалу. Таким чином, стандартна помилка вимірювання є стандартною похибкою оцінки істинних балів на основі спостережуваних результатів тестових вимірювань. З ростом S_E межі довірчого інтервалу будуть розширюватися, а разом з цим будуть збільшуватися можливі межі відхилення істинного балу від спостережуваних результатів вимірювання.

Помилка вимірів надійності для нашого прикладу $S_E = 1,1$. Побудуємо інтервал надійності навколо спостережуваної оцінки для кожного учасника при $\alpha = 0,05$

Випробуваний	Спостережуваний бал	95% довірчий інтервал	95% інтервал істинного балу
1	8	(5,8; 10,2)	(5,8; 8)
2	7	(4,8; 9,2)	(4,8; 8)
3	6	(3,8; 8,2)	(3,8; 8)
4	5	(2,8; 7,2)	(2,8; 7,2)
5	5	(2,8; 7,2)	(2,8; 7,2)
6	4	(1,8; 6,2)	(1,8; 6,2)
7	4	(1,8; 6,2)	(1,8; 6,2)
8	3	(0,8; 5,2)	(0,8; 5,2)
9	2	(-0,2; 4,2)	(0; 4,2)
10	2	(-0,2; 4,2)	(0; 4,2)
11	1	(-1,2; 3,2)	(0; 3,2)

Інтерпретуємо отриманий результат для 5-го випробуваного: у 95% випадків ми можемо бути впевнені, що істина оцінка даного випробуваного лежить в інтервалі від 1,8 до 6,2. Для всіх учасників інтервал 95% попадання істинного балу не може перевищувати 8 балів (див. 95% інтервал істинного балу учасників 1, 2 та 3), т.я. у тесті після вилучення одного завдання було 8 завдань, та не може бути меншим 0 балів (див. 95% інтервал істинного балу учасників 9, 10 та 11).

Використавши прогнозовані істинні оцінки отримані за допомогою регресійного аналізу в пункті 2.7.2 побудуємо інтервал надійності навколо істинної оцінки для кожного учасника при точності $\alpha = 0,05$ та помилці вимірів $S_E = 0,9$.

Випробуваний	Прогнозований бал	95% довірчий інтервал	95% довірчий інтервал істинного балу
1	6,69	(4,89; 8,49)	(4,89; 8)
2	6,01	(4,21; 7,81)	(4,21; 7,81)
3	5,33	(3,53; 7,13)	(3,53; 7,13)
4	4,65	(2,85; 6,45)	(2,85; 6,45)
5	4,65	(2,85; 6,45)	(2,85; 6,45)
6	3,97	(2,17; 5,77)	(2,17; 5,77)
7	3,97	(2,17; 5,77)	(2,17; 5,77)
8	3,29	(1,49; 5,09)	(1,49; 5,09)
9	2,61	(0,81; 4,41)	(0,81; 4,41)
10	2,61	(0,81; 4,41)	(0,81; 4,41)
11	1,93	(0,13; 3,73)	(0,13; 3,73)

Розглянемо знову результат 5-го випробуваного: в даному випадку з 95-ти відсотковою впевненістю можна стверджувати, що його істинний бал знаходиться у межах від 2,85 до 6,45

При використанні прогнозованих оцінок спостерігаємо звуження довірчого інтервалу на 0,8.

4.9. Валідність тесту у класичній теорії

4.9.1. Валідність для критеріально-орієнтованих тестів

Валідність – це характеристика здатності тесту служити поставленій меті вимірювання. Кронбах описує валідизацію як процес, за допомогою якого розробник або користувач тесту збирає доведення для обґрунтування тих висновків, які повинні бути зроблені з тестових оцінок. Існує три основних типи досліджень валідності:

1. Змістовна валідність для ситуацій, в яких користувач тесту хоче з тестових оцінок випробовуваних отримати висновок для більшої області завдань, які подібні тим, що безпосередньо використовуються в тесті.
2. Валідність, зв'язана з критерієм, для ситуацій, в яких користувач тесту хоче із тестової оцінки випробовуваного отримати висновок відносно виконання індивідуумом діяльності по деякій поведінковій змінній практичного значення.
3. Конструктна валідність для ситуацій, в яких не проходить змістова і критерійна валідність, але користувач тесту хоче отримати висновок із тестових оцінок результатів виконання тесту, які можуть бути згруповані за ознакою вимірювання певного психологічного конструкта.

Загалом при дослідженні критеріально-орієнтованої валідності треба виконати наступні кроки:

1. встановити придатний критерій поведінки і вибрати метод для його вимірювання;

2. сформувати необхідну вибірку випробовуваних, які репрезентативно представляють тих, для кого буде в кінцевому рахунку використовуватися тест;
3. пред'явити тест і вести облік оцінок кожного випробовуваного;
4. коли дані за критерієм будуть доступні, отримати міру виконання критерію для кожного випробовуваного;
5. визначити силу зв'язку між виконанням діяльності за критерієм і тестовими оцінками.

Розрізняють два типи критеріально-орієнтованої валідності: прогностичну і конкурентну. Прогностична валідність характеризує степінь, в якій тестові оцінки передбачають результати вимірювання за критерієм в деякий момент в майбутньому. Конкурентна валідність визначає зв'язок між тестовими оцінками і вимірюваннями критерію, зробленими в той час, коли тест був виданий випробовуваним.

Дані дослідження критеріально-орієнтованої валідності можуть бути проаналізовані декількома способами. Якщо вимірювання за критерієм є розподілені неперервно, то можна обчислити коефіцієнт кореляції Пірсона між тестовою оцінкою і мірою критерію. Таку статистичну величину називають *коефіцієнтом валідності*. Якщо змінна критерію приймає дискретні значення, то коефіцієнт валідності обчислюється як коефіцієнт кореляції за відповідною формулою.

Припустимо, що в нашому тесті критерієм виступає отримання заліку з дисципліни (даний вид валідності є критеріально-прогностичним з припущення, що залік отримують ті випробувані, які отримали не менше 4-х балів). За таких умов критеріальні дані будуть мати такий вигляд:

Учасник №	спостережуваний бал по тесту	зараховано-незараховано
1	7	1
2	6	1
3	5	1
4	5	1
5	4	1
6	4	1
7	3	0
8	2	0
9	2	0
10	1	0

Нескладно обчислити коефіцієнт валідності (коефіцієнт кореляції Пірсона між тестовою оцінкою і мірою критерію), який буде рівним 0,86, що свідчить про достатньо високу критеріально-прогностичну валідність даного тесту.

Розглянемо тепер приклад конкурентної валідності. Припустимо, що потрібно визначити наскільки близькі оцінки групи учасників тестування, наведені вище, та результати цих же людей, отримані ними за результатами екзамену за стобальною шкалою.

Учасник №	Спостережуваний бал по тесту	Оцінка за екзамен
1	7	45
2	6	80
3	5	60
4	5	94
5	4	45
6	4	23
7	3	42
8	2	40
9	2	50
10	1	20

Нескладно бачити, що коефіцієнт конкурентної валідності буде рівним 0,55, що значно нижче за коефіцієнт критеріально-прогностичної валідності 0,82. Це пояснюється тим, що деякі із тестованих, які показали високі результати за тестом, отримали посередні бали на екзамені за 100-бальною шкалою (наприклад, учасник №1). При цьому деякі учасники тестування, які отримали низькі бали за тестом, мають далеко не найслабші результати на екзамені (наприклад, учасник № 9).

4.9.2. Інтерпретація результатів валідизації

Додаткова інформація, яка може допомогти при інтерпретації результатів дослідження валідності включає коефіцієнт детермінації, стандартну помилку наближеної оцінки і таблицю прогнозу.

Квадрат коефіцієнта кореляції часто використовують при інтерпретації значимості зв'язків між двома змінними. Квадрат коефіцієнта валідності називається *коефіцієнтом детермінації*. Якщо, наприклад, кореляція між

тестовою оцінкою і деякою мірою реальної діяльності рівна 0,60, то коефіцієнт детермінації буде рівним 0,36 і 36% дисперсії в результатах реальної діяльності зв'язано з дисперсією діяльності по предиктору тесту.

Для передбачення оцінки i -го індивідуума по критеріальній змінній Y формула має вигляд

$$Y_i = r_{XY}(S_Y / S_X)(X_i - \bar{X}) + \bar{Y},$$

де r_{XY} - коефіцієнт валідності, S_X - стандартне відхилення по предиктору тесту, S_Y - стандартне відхилення по критеріальних оцінках. Для побудови довірчого інтервалу навколо передбачуваної оцінки Y_i використовують стандартну похибку наближеної оцінки для вибірки валідизації $S_{YX} = S_Y \sqrt{1 - r_{XY}^2}$. Користувач тесту, який визначив передбачувану величину Y_i для випробовуваного, може бути на 68% впевненим у тому, що істинна оцінка Y_i попаде в наближений інтервал $Y_i \pm 1S_{YX}$, і на 95% - в наближений інтервал $Y_i \pm 2S_{YX}$.

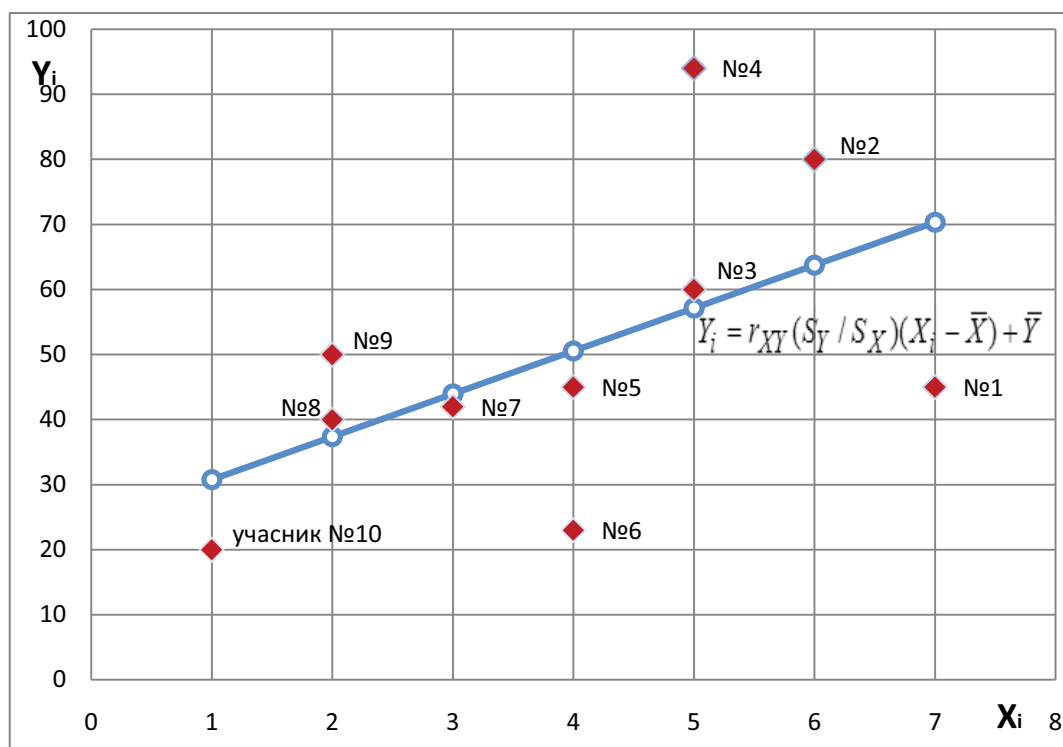
Психологічний конструкт не спостерігається безпосередньо. Спочатку його треба визначити операціонально (семантично) шляхом задання процедур, які використовуються для вимірювання конструкта. Крім того, психологічний конструкт вимагає синтаксичного означення шляхом постулювання конкретних зв'язків між мірами конструкта, по-перше, з мірами других конструктів в теоретичній системі та, по-друге, з мірами певних критеріїв реального світу.

Оцінка конструктної валідності вимагає отримання загального висновку на основі свідств багатьох типів з різних джерел інформації. Для побудови процесу валідизації широко використовуються чотири підходи: кореляції між мірою конструкта і мірою раніше визначеного конструкта; метод контрастних груп; факторний аналіз; матриця «багатовимірні характеристики - множинні методи».

Використовуючи вище наведений приклад про екзаменаційні оцінки отримуємо прогноз по критеріальній змінній за формулою $Y_i = 0,55 \cdot (23,0/1,9)(X_i - 3,9) + 49,9$, що має вигляд:

Учасник № i	Тест X_i	Екзамен	Прогноз екзамену Y_i
1	7	45	70,34
2	6	80	63,74
3	5	60	57,15
4	5	94	57,15
5	4	45	50,56
6	4	23	50,56
7	3	42	43,97
8	2	40	37,37
9	2	50	37,37
10	1	20	30,78

Представимо результат графічно. На рисунку пунктирна пряма лінія з круглими маркерами є графіком прогнозованих оцінок i -го індивідуума по критеріальній змінній Y , а точки є оцінками студента за екзамен.



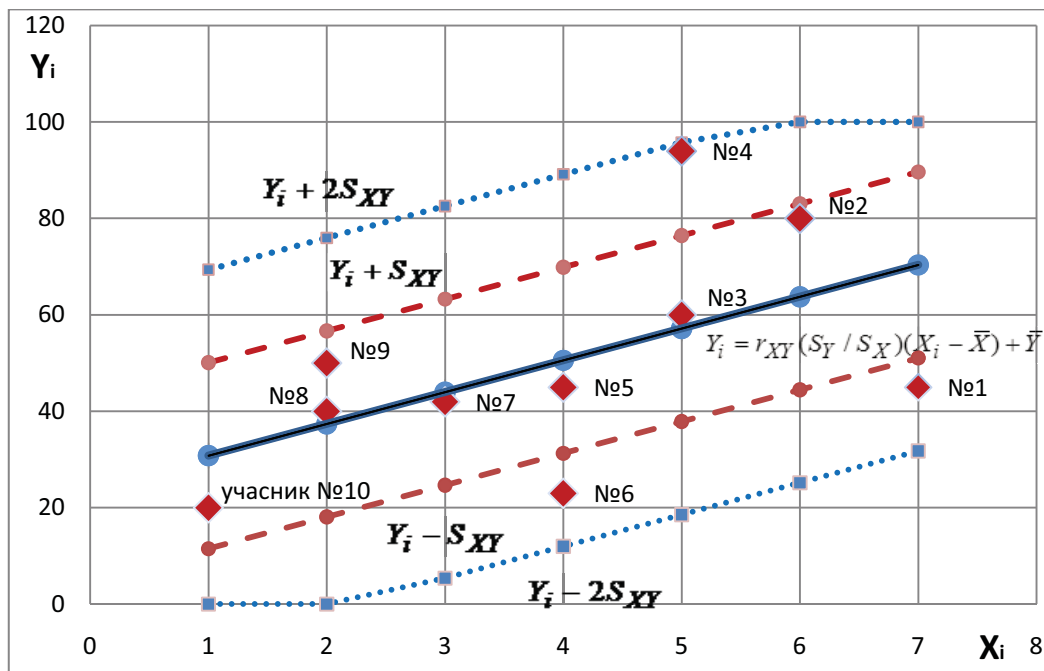
Із графічного зображення прогнозованих оцінок та оцінок на екзамені видно, що обидва учасники тестування №8 та №9, які отримали за тест по два бали, за прогнозом мали б скласти екзамен на дещо нижчий бал за ті бали, що вони отримали на екзамені. Пара учасників №5 та №6, які набрали по чотири бали під час тестування, виконали екзаменаційні завдання гірше за

прогнозовані результати, причому учасник №6 отримав результат значно гірший за прогнозований, а учасник №5 показав результат близький до прогнозованого. Учасники №3 та №4, які мали під час тестування по п'ять балів, продемонстрували різні результати: результат екзамена учасника № 3 нижче за прогнозований, а учасника №4 – вище. Відмітимо також, що результати учасників 10, 6, 4 та 1 істотно відрізняються від прогнозованих.

Стандартна похибка наближеної оцінки для вибірки валідизації $S_{YX} = 19,3$. Побудуємо довірчі інтервали критеріальної оцінки

Учасник № i	68% довірчий інтервал ($Y_i - S_{YX}; Y_i + S_{YX}$)	95% довірчий інтервал ($Y_i - 2S_{YX}; Y_i + 2S_{YX}$)
1	(51,05; 89,63)	(31,76; 100)
2	(44,45; 83,03)	(25,16; 100)
3	(37,86; 76,44)	(18,57; 95,73)
4	(37,86; 76,44)	(18,57; 95,73)
5	(31,27; 69,85)	(11,98; 89,14)
6	(31,27; 69,85)	(11,98; 89,14)
7	(24,68; 63,26)	(5,39; 82,55)
8	(18,08; 56,66)	(0; 75,95)
9	(18,08; 56,66)	(0; 75,95)
10	(11,49; 50,07)	(0; 69,36)

Можна зробити висновок, що із 68% критеріальна оцінка для, наприклад, 5-го випробуваного буде належати інтервалу від 31,27 до 69,85. Для того ж випробуваного із 95% впевненістю можна сказати, що критеріальна оцінка буде належати інтервалу від 11,98 до 89,14. Взагалі кажучи, верхня межа довірчих інтервалів для учасників 1 та 2 виходить за 100 балі, а нижня межа для учасників 8,9 та 10 є від'ємною, що неможливо, тому були внесені відповідні корективи. На рисунку представлені межі довірчих інтервалів: пунктирними лініями помічені межі інтервалу ($Y_i - S_{YX}; Y_i + S_{YX}$), точковими лініями межі ($Y_i - 2S_{YX}; Y_i + 2S_{YX}$).



При цьому нескладно бачити, що розподіл екзаменаційних балів близький до нормального, оскільки всі значення містяться у 95% інтервалі та сім значень (учасники №2, 3, 5, 7, 8, 9, 10) містяться у 68% інтервалі.

4.9.3. Метод контрастних груп

Якщо критерій носить неперервний характер, то для оцінювання валідності окремого завдання тесту учасників тестування, які вірно виконали дане завдання, ділять на дві групи - з високим показником критерію і з низьким показником. Так само діляться на дві групи і ті учасники, які дане завдання виконали невірно. В результаті будується таблиця зв'язаних ознак 2x2, за якою оцінка валідності завдань тесту виконується за допомогою коефіцієнта асоціації ϕ .

Для підвищення контрастності груп рекомендується в кожному з них включати по 27% від загального складу, відповідно, з високим показником критерію і з низьким показником, оскільки зменшення вказаної кількості відсотків приводить до зменшення надійності подальших викладень через нечисленність вибірок; збільшення цього числа приводить до падіння контрастності.

ПРАКТИЧНІ ЗАВДАННЯ

Приклад 1. Для кожної наведеної нижче ситуації потрібно вказати як впливає дана подія (випадково чи систематично) на помилку вимірів тестової оцінки тестованого.

- а) Викладач, що оцінює результати есе студентів, має тенденцію до завищення оцінок дівчат у порівнянні з хлопцями.

Розв'язання. Дана помилка вимірів є систематичною, так як викладач постійно використовує такий метод оцінки.

- б) Студент, що отримав математичний тест, відволікся на шум в сусідній аудиторії, в результаті чого зробив помилку у підрахунках.

Розв'язання. Дана помилка вимірів є випадковою, однак якщо вона є результатом постійної неувважності чи розсіяності студента, то перетворюється на систематичну.

- в) Учень так сильно нервує під час складання тесту, що часто залишає багато завдань невиконаними.

Розв'язання. Дана помилка вимірів є систематичною, так як учень постійно нервує під час складання іспитів чи тестів.

- г) Викладач, що читає твори учнів, проглядає текст через речення, що призводить до неоднозначних оцінок учнів.

Розв'язання. Дана помилка вимірів є систематичною, так як викладач постійно неувважно вчитується в текст.

- д) Студент на лекції робить помилку у твердженні, в результаті чого невірно відповідає на завдання в тесті, що ґрунтується на твердженні.

Розв'язання. Дана помилка вимірів є випадковою, однак це означає, що для підготовки до тесту студент не використовував окрім конспекту інших джерел, що може призводити до систематичних помилок.

Приклад 2. Тест із завданнями видається чотирьом тестованим із генеральної сукупності. Таблиця показує долі розподілу тестових оцінок для кожного тестованого при багаторазовому тестуванні. Приклад є гіпотетичним, так як самі ймовірності на практиці ніколи не розглядаються.

Тестований	Оцінка			
	0	1	2	3
A	0,5	0,5	0	0
B	0,25	0,25	0,25	0,25
C	0	0	0,5	0,5
D	0	0,5	0,5	0

a) Яка оцінка є істинною для третього тестованого?

Розв'язання. За класичною теорією істинною оцінкою вважається математичне сподівання спостережуваних оцінок. Використавши дане визначення отримаємо істину оцінку для третього тестованого: $M(X) = 2,5$

b) Яка оцінка є істинною для другого тестованого?

Розв'язання. Для другого тестованого істиною буде оцінка 1,5.

c) Яка дисперсія істинної оцінки по даному тесту?

Розв'язання. Використовуючи формулу дисперсії $S_X^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$, отримаємо дисперсію істинної оцінки по даному тесту, рівну 0,67.

d) Яка дисперсія помилкового компонента оцінки для третього тестованого?

Розв'язання. $S_E^2 = S_X^2(1 - \rho) \Rightarrow S_E^2 = 0,5625 * (1 - 0,73) = 0,15$

e) Яка дисперсія помилкового компонента оцінки для другого тестованого?

Розв'язання. $S_E^2 = S_X^2(1 - \rho) \Rightarrow S_E^2 = 0,104 * (1 - 0,73) = 0,03$

f) Який коефіцієнт надійності для даного тесту?

Розв'язання. За класичною теорією коефіцієнт надійності визначається як відношення дисперсії істинної оцінки до дисперсії спостережуваної оцінки:

$\rho = \frac{S_T^2}{S_X^2}$. Прийmemo за спостережувані оцінки (враховуючи ймовірності з таблиці): для А – 1, для В – 2, для С – 3, для D – 1; дисперсія для спостережуваних оцінок: $S_X^2 = 0,92$; дисперсія для істиних оцінок: $S_T^2 = 0,67$; звідси $\rho = 0,73$.

г) Якщо ви дали цей тест два рази, наведіть приклади двох можливих наборів оцінок, що могли бути отримані цими чотирма тестованими.

Розв'язання. Враховуючи ймовірності з таблиці отримаємо 2 набори можливих спостережуваних оцінок:

Тестований	1-й набір оцінок	2-й набір оцінок
A	1	0
B	2	1
C	3	2
D	1	2

h) В реальних ситуаціях тестування чи можливо визначити надійність тестової оцінки методом, що використовується в пункті f? Чому?

Розв'язання. Даний метод не підходить для оцінки надійності в реальних ситуаціях, тому що на практиці не розглядаються ймовірності виникнення спостережуваних оцінок.

Приклад 3. Психолог дослідник пред'явив три паралельні форми стандартизованого тесту групі тестованих. Бланки відповідей перевірялись з допомогою оптичного сканера. Пізніше досліднику повідомили, що сканер працював з перебоями, вносячи помилки в оцінювання випадковим чином в день підрахунку результатів по тестах. Оскільки дослідник використовував тільки середню оцінку для кожного тестованого, обраховану за трьома тестами, то він не побачив необхідності в тому, щоб заново оцінювати бланки відповідей, так як ефекти цих випадкових помилок повинні були усереднюватись, що б призвело до їх обнуління за трьома тестами. Чи є

вірним припущення про те, що середня помилка вимірів кожної оцінки повинна дорівнювати 0?

Розв'язання. Середнє значення помилок оцінок генеральної сукупності тестованих дорівнює нулю, однак, ми маємо вибірку із трьох тестових форм всіх можливих варіантів тестових форм із генеральної сукупності, тому значення цієї вибірки помилок оцінок може бути рівним як нулю так і будь-якому іншому числу.

Приклад 4. У таблиці приведені значення кореляції між всіма парами субтестів в батареї тестів учбових досягнень, середні значення, стандартні відхилення, а також дані по приміненню альтернативних форм оцінки надійності субтестів із числа пред'явлених студентам на початку другого курсу.

Назва субтеста	Номер субтеста	1	2	3	4	5	6	7
Вокабуляр	1	1	0,66	0,63	0,68	0,69	0,65	0,69
Лексичні навиви	2		1	0,87	0,85	0,61	0,61	0,55
Розуміння прочитаного	3			1	0,83	0,57	0,59	0,53
Мова	4				1	0,62	0,62	0,55
Математичні поняття	5					1	0,72	0,64
Природничо-наукові поняття	6						1	0,61
Суспільні науки	7							1
Стандартне відхилення (σ)		6,67	8,79	8,39	10,12	4,97	6,65	4,33
Середнє значення		19,8	22	18	37,7	16,7	15,6	16
Альтернативна форма надійності		0,87	0,95	0,95	0,94	0,81	0,9	0,81

Використовуючи таблицю дати відповіді на питання:

- A. Яка кореляція між навиками студентів по лексиці і суспільним наукам?
- B. Яка коваріація між двома цими змінними?
- C. Загальна оцінка по читанню може бути отримана шляхом складання оцінок лексичних навиків і розуміння прочитаного. Яка дисперсія загальної оцінки по читанню?

- D. Яка надійність загальної оцінки по читанню?
- E. Когнітивний психолог формує змінну шляхом складання оцінок субтестів по вокабуляру (словарному запасу), мові і суспільним наукам. Яка кореляція між цією новою змінною і загальною оцінкою по читанню?
- F. Якщо тестований отримав оцінку 22 бали по субтесту природничо-наукових понять, то в границях якого інтервалу ми можемо бути впевнені на 68% в тому, що в ньому лежить його істинна оцінка?
- G. Згідно з припущеннями класичної моделі істинної оцінки, якою повинна бути діаграма розсіювання у помилкових компонентів оцінок в субтестах з математики і природничим наукам?

Розв'язання.

- A. Із кореляційної таблиці випливає, що $r_{27} = 0,55$.
- B. Залежність між коефіцієнтом кореляції і коваріацією, визначається за формулою: $r_{AB} = \frac{cov_{AB}}{\sigma_A \sigma_B}$. Використовуючи дані таблиці, отримаємо: $cov_{AB} = 0,55 \cdot 8,79 \cdot 4,33 = 20,93$.
- C. За властивістю дисперсії: $S^2(A + B) = S^2(A) + S^2(B) + 2cov_{AB}$. Використавши величини для оцінки лексичних навиків і розуміння прочитаного, отримаємо $S_{23}^2 = 8,79^2 + 8,39^2 + 2 * 64,16 = 275,98$
- D. Так як в данному випадку ми маємо не паралельні субтести (для паралельних субтестів мають бути рівні стандартні відхилення, середні і дисперсії) то скористаємось обрахунком нижньої границі коефіцієнта точності з допомогою коефіцієнта альфа за формулою

$$\rho \geq \frac{k}{k-1} \left(1 - \frac{\sum_{i \neq j}^k cov_{ij}}{S_C^2} \right),$$

де k – кількість субтестів, S_C^2 – дисперсія складених оцінок, $\sum_{i \neq j}^k cov_{ij}$ – сума всіх коваріацій між всіма субтестами. Для даного випадку $k = 2$; $S_C^2 = S_{23}^2 = 275,98$; $cov_{23} = 64,16 \Rightarrow \rho \geq 0,47$.

Приклад 5. Для кожної ситуації вкажіть тип оцінки надійності, що найбільш підходить наступним умовам.

- а) Єдина форма стандартизованого тесту навчальних досягнень охоплює області інформатики з допомогою дихотомічно оцінюваних завдань по трьом розділам: використання пакету MSOffice, адміністрування операційної системи Windows та програмування мовою Pascal. Питання, що краще: виставити єдину оцінку чи окремі оцінки по кожній області змісту субтестів?

Відповідь. Більш довші тести є надійнішими ніж коротші, так як помилки вимірів при збільшенні вибірки зменшуються, тому більш точною буде оцінка виставлена за загальний тест.

- б) Шкільний вчитель хоче перевірити на учнях свого класу надійність підсумкового тесту по біології, що містить завдання з множинним вибором.

Відповідь. В даному випадку завданням вчителя є виявлення некоректних завдань в тесті, тому оцінкою надійності виступить коефіцієнт внутрішньої узгодженості.

- с) Психолог-консультант розроблює два набори аттитюдів один з яких призначений для використання в якості попереднього тесту, а другий в якості вихідного тесту для програми, що триває один день.

Відповідь. Маємо дві форми тесту, що будуть пред'явлені групі тестованих в короткий проміжок часу. В даному випадку розробника цікавить зв'язок отриманих оцінок, тому за оцінку надійності слід взяти коефіцієнт еквівалентності.

- д) Соціальний психолог розробляє інструментарій для оцінки згоди з аттитюдами студентів, що тільки прийняті до коледжу з метою виявлення їхнього відношення до сумісного навчання і проживання в гуртожитку. Кожна відповідь буде оцінюватись у відповідності зі шкалою: відповідь "так" - 2 бала, "можливо" - 1 бал, "ні" - 0 балів.

Відповідь. В даному випадку тест буде пред'явлений одній групі один раз, тому за оцінку надійності можна взяти коефіцієнт альфа Кронбаха.

- е) Екзамен для вступу в коледж проводить восени і навесні (використовуються різні форми тесту). Дирекція коледжу хоче бути впевненою в тому, що пора року коли проводиться тестування не впливає на результати.

Відповідь. В даному випадку тести кожного разу проводитимуть у різних групах, тому найдоцільнішою оцінкою надійності виступає коефіцієнт альфа (за формулою Кронбаха або Кюдера Річардсона) для оцінки внутрішньої узгодженості завдань.

Приклад 6. Тест, що складається з 10 завдань був пред'явлений 9-ти учням. Матриця, що об'єднує результати по завданням і учням, має наступний вигляд:

Студент	Завдання								Загальний бал
	1	2	3	4	5	6	7	8	
A	1	1	1	1	1	1	0	0	6
B	1	1	1	0	0	1	0	0	4
C	1	1	0	1	0	0	0	0	3
D	0	0	0	0	1	0	1	1	3
E	1	1	1	1	1	1	1	1	8
F	1	1	1	1	1	0	0	0	5
G	0	1	1	1	1	0	0	1	5
H	1	1	1	1	1	1	1	0	7
I	1	1	0	0	0	0	0	0	2
J	1	1	1	1	1	0	1	1	7

Потрібно дати відповіді на наступні питання.

- A. Оцінити внутрішню узгодженість тесту, використовуючи метод поділу навпіл (створити два субтести з парних і непарних завдань). Обрахувати дві оцінки, що були б отримані для кожного учня по субтестам.

Розв'язання. Створимо 2 субтести і обчислимо суму балів студентів по ним

Субтест А (непарні номери завдань)

Студент	Завдання				Загальний бал
	1	3	5	7	
A	1	1	1	0	3
B	1	1	0	0	2
C	1	0	0	0	1
D	0	0	1	1	2
E	1	1	1	1	4
F	1	1	1	0	3
G	0	1	1	0	2
H	1	1	1	1	4
I	1	0	0	0	1
J	1	1	1	1	4

Субтест В (парні номери завдань)

Студент	Завдання				Бал
	2	4	6	8	
A	1	1	1	0	3
B	1	0	1	0	2
C	1	1	0	0	2
D	0	0	0	1	1
E	1	1	1	1	4
F	1	1	0	0	2
G	1	1	0	1	3
H	1	1	1	0	3
I	1	0	0	0	1
J	1	1	0	1	3

В. Знайти кореляцію між субтестами.

Розв'язання. Отримаємо кореляцію

$$r_{AB} = \frac{\sum_{i=1}^{10} (3 - 2,6)(3 - 2,4)}{\sqrt{\sum_{i=1}^{10} (3 - 2,6)^2 \sum_{i=1}^{10} (3 - 2,4)^2}} + \dots$$

$$\dots + \frac{\sum_{i=1}^{10} (4 - 2,6)(3 - 2,4)}{\sqrt{\sum_{i=1}^{10} (4 - 2,6)^2 \sum_{i=1}^{10} (3 - 2,4)^2}} = 0,74$$

Отримуємо високу кореляцію (0,74) між субтестами А і В.

С. Яка оцінка надійності для тесту повної довжини?

Розв'язання:

а) Розглянемо оцінку надійності тесту для повної довжини скоректовану за формулою Спірмана-Брауна

$$\rho = \frac{2 \cdot 0,74}{1 + 0,74} = 0,85$$

б) Розглянемо оцінку надійності тесту для повної довжини обраховану за

$$\text{формулою Рюлона: } \rho = 1 - \frac{S_D^2}{S_X^2}$$

Заг. бал по субтесту А	Заг. бал по субтесту В	Різниця оцінок D
3	3	0
2	2	0
1	2	-1
2	1	1
4	4	0
3	2	1
2	3	-1
4	3	1
1	1	0
4	3	1

Знаходимо значення дисперсій для спостережуваних оцінок $S_X^2 = 4$ і різниці оцінок $S_D^2 = 0,62 \Rightarrow \rho = 0,84$

D. Сказати, користуючись результатами візуального аналізу, який із субтестів, на вашу думку буде мати більш високу оцінку надійності, обґрунтувати відповідь.

Розв'язання. Більш високу оцінку надійності буде мати субтест А, так як результати відповідей на тестові завдання в ньому є більш однорідними, а дисперсія спостережуваних оцінок буде більшою ніж с субтесті В (це означає, що дисперсія їх істинних оцінок повинна пояснювати більший процент дисперсії спостережуваних оцінок). Такі результати вказують на те, що в формулюванні завдань, їх оформленні та проведенні самого тестування допущено мінімальну кількість помилок.

Приклад 7. Наступна таблиця містить дані по двом субтестам стандартизованої батареї тестів навчальних досягнень по використанню різних типів програмного забезпечення, кожен з яких має дві взаємозамінювані форми. Потрібно дослідити дані, наведені в таблиці, і відповісти на питання.

	Використання пакету MS Office	Використання пакету Open Office
Форма А		
Середнє	22	25
Стандартне відхилення	8	12
Число завдань	35	40
KR20	0,83	0,85
Форма В		
Середнє	22	24
Стандартне відхилення	7	12
Число завдань	35	40
KR20	0,8	0,85

- А. Які стандартні помилки обчислень для двох форм тестів? Чи завжди краще використовувати тестову форму з найменшою стандартною помилкою вимірів?

Розв'язання:

Формула для обчислення стандартної помилки має вигляд: $\sigma_E = \sigma_X \sqrt{1 - \rho}$, використавши значення з таблиці отримаємо

$$\text{для форми А: } \sigma_{E_1} = 8 \cdot \sqrt{1 - 0,83} = 3,3; \quad \sigma_{E_2} = 4,65;$$

$$\text{для форми В: } \sigma_{E_1} = 3,13; \quad \sigma_{E_2} = 4,65.$$

- В. Нехай оцінка студента X — 20 балів по використанню пакету прикладних програм MS Office за формою А, побудуйте інтервал, який імовірно містить його істинну оцінку (з 68% - ним шансом достовірності). Які були б границі інтервалу при 95%- ній впевненості в тому, що інтервал містить його істинну оцінку?

Розв'язання:

Для впевненості на 68% в тому, що інтервал містить істинну оцінку тестованого використовується стандартна помилка вимірів, що має вигляд: $X \pm \sigma_E$, для 95% - ї впевненості - $X \pm 2 \cdot \sigma_E$.

Студент X				
Стандартна помилка вимірів	68% інтервалу довіри	95% інтервалу довіри	Інтерпретація 68% інтервалу довіри для студента з спостережуваною оцінкою 20 балів	Інтерпретація 95% інтервалу довіри для студента з спостережуваною оцінкою 20 балів
3,3	20 +/- 3,3	20 +/- 6,6	В 68% випадків можемо бути впевнені, що істина оцінка студента X належить інтервалу від 16,7 до 23,3	В 95% випадків можемо бути впевнені, що істина оцінка студента X належить інтервалу від 13,4 до 26,6

C. По формі В тесту по використанню пакету Open Office студент Y отримав оцінку 10 балів, а студент Z — 29 балів. Які значення їх можливих істинних оцінок?

Розв'язання. Сирі (необроблені) оцінки за формою В загального тесту для студента Y=10 балів, а для студента Z = 29 балів. Використавши формулу для знаходження істинної оцінки: $\hat{T} = \rho(X - \bar{X}) + \bar{X}$, отримуємо $\hat{T}_Y = 0,85 \cdot (10 - 24) + 24 = 12,1$; $\hat{T}_Z = 0,85 \cdot (29 - 24) + 24 = 28,25$.

D. Розробник тесту розглядає можливість скорочення тестів по використанню пакету Open Office шляхом виключення п'яти завдань із кожної форми. Як це вплине на надійність форм?

Розв'язання. Тестові форми А і В складаються із 40 тестових завдань кожна. Розглянемо зміну надійності тесту на прикладі форми А. Використаємо формулу Спірмана-Брауна: $\rho_X = \frac{k\rho}{1+(k-1)\rho}$, де k – це число на яке домножують довжину тесту, щоб отримати нову довжину; ρ – надійність початкового тесту. $k = \frac{35}{40} = 0,875 \Rightarrow \rho_X = \frac{0,875 \cdot 0,85}{1+(0,875-1) \cdot 0,85} = 0,83$ з чого можна зробити висновок, що при скороченні початкового тесту довжиною 40 на 5 тестових завдань його надійність зменшиться на 0,02.

- Е. При дослідженні надійності не у всіх тестованих було досить часу для закінчення тесту по використанню пакету Open Office. Як це вплинуло на оцінку надійності з допомогою формули KR_{20} ?

Розв'язання. Як тільки у тестованого закінчується ліміт часу, виконання всіх завдань, що залишились з парними і непарними номерами буде повністю узгодженим, незалежно від того чи гомогенні завдання по змісту. Даний факт вказує на зменшення дисперсії по завданню, що призведе до штучного завищення надійності тесту обрахованої за формулою Кьюдера-Річардсона.

Приклад 8. Керівництво по тестуванню для індивідуального використання містить оцінки надійності, отримані методом поділу тесту з допомогою формули:

$$\rho_X = \frac{4r_{AB}\sigma_A\sigma_B}{\sigma_X^2},$$

де А і В – оцінки по половинам тесту. Довести, що дана формула еквівалентна формулі для коефіцієнта Рюлона, тобто:

$$\rho_X = \frac{4\rho_{AB}\sigma_A\sigma_B}{\sigma_X^2} = 1 - \frac{\sigma_D^2}{\sigma_X^2}.$$

Розв'язання. Оскільки $D=A-B \Rightarrow 1 - \frac{\sigma_D^2}{\sigma_X^2} = 1 - \frac{\sigma_A^2 + \sigma_B^2 - 2\sigma_A\sigma_B r_{AB}}{\sigma_X^2}$,

звівши до спільного знаменника, отримаємо:

$$\frac{4r_{AB}\sigma_A\sigma_B}{\sigma_X^2}$$

Приклад 9. На основі формули KR_{20} виведіть формулу для коефіцієнта надійності KR_{21} .

Розв'язання. Потрібно довести, що $\frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^n p_i q_i}{S_X^2} \right) = \frac{k}{k-1} \left(1 - \frac{\bar{\mu}(k-\bar{\mu})}{kS_X^2} \right)$;

Оскільки $p = \frac{X_i}{k}$; $q = 1 - \frac{X_i}{k} \Rightarrow$ підставимо р і q в формулу KR_{20} , отримаємо

$$\frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^n \frac{X_i}{k} \left(1 - \frac{X_i}{k} \right)}{S_X^2} \right)$$

знаючи, що $\mu = \frac{\sum_{i=1}^k X_i}{k}$, отримаємо $\frac{k}{k-1} \left(1 - \frac{\bar{\mu}(k-\bar{\mu})}{kS_X^2} \right)$.

Питання для самоконтролю

1. Якою рівністю визначається істинна оцінка?
2. За якою формулою можна обчислити істинну оцінку?
3. Перерахуйте основні положення класичної теорії тестів та обґрунтуйте їх.
4. Які тести називаються паралельними?
5. Які тести називаються еквівалентними?
6. Які види матриць результатів тестування ви знаєте і яка між ними різниця?
7. Яку інформацію можна отримати з матриці результатів тестування?
8. Що таке профіль учасника тестування та частота тестового балу?
9. Як обчислюється середнє вибіркове і яка його інтерпретація?
10. Як обчислюється дисперсія і яка її інтерпретація?
11. Як обчислюється стандартне відхилення і яка його інтерпретація?
12. Як обчислюється асиметрія і для чого вона потрібна?
13. Як обчислюється ексцес і для чого він потрібний?
14. Для чого потрібний коефіцієнт кореляції?
15. У яких випадках використовується коефіцієнт кореляції Пірсона, як він обчислюється і яка його інтерпретація?
16. Як створити кореляційну матрицю?
17. Яку інформацію можна отримати із кореляційної матриці?
18. У яких випадках використовується фі коефіцієнт кореляції, як він обчислюється і яка його інтерпретація?
19. У яких випадках використовується коефіцієнт бісеріальної кореляції, як він обчислюється і яка його інтерпретація?
20. У яких напрямках досліджуються параметри тестових завдань?
21. Як обчислюється трудність завдань і яка її інтерпретація?
22. З якою метою потрібно досліджувати правдоподібність дистракторів?
23. Як обчислюється коефіцієнт кореляції для кожного дистрактора в завданнях тесту?
24. Що називається дискримінативністю тестового завдання?
25. У яких випадках використовується кореляція при визначенні дискримінативності тестового завдання?
26. За якими формулами обчислюються надійність завдань і показники валідності?
27. Що розуміють під надійністю тесту?
28. Як обчислюються показник та коефіцієнт надійності, чи є зв'язок?
29. Вкажіть прогностичну формулу Спірмена-Брауна та її застосування.
30. Вкажіть коефіцієнт альфа Кронбаха та його застосування.
31. Назвіть методи двократного тестування та формули обчислення надійності.

32. Назвіть методи однократного тестування та формули обчислення надійності.
33. Які фактори впливають на коефіцієнти надійності?
34. Як знаходиться похибка вимірювання надійності?
35. За якою формулою здійснюється наближене обчислення істинних оцінок?
36. Що таке валідність тесту?
37. Вкажіть основні типи валідності.
38. Вкажіть типи критеріально-орієнтованої валідності. Що вони характеризують?
39. Що таке коефіцієнт валідності та як він обчислюється?
40. Що таке коефіцієнт детермінації?
41. Які методи використовуються для знаходження конструктивної валідності?

Список джерел, використаних у розділі IV

1. Давидян Г.Ю. Модели педагогического тестирования [Электронный ресурс] / Третья областная научно-практическая конференция «Информационные технологии в образовании» «ИТО-Ростов 2003», Ростов-на-Дону, 9-10 октября 2003 / Донской институт информатизации образования. - <http://www.doniinfo.aaanet.ru/ito/2003/>.
2. Зайцева Л.В., Прокофьева Н.О. Модели и методы адаптивного контроля знаний [Электронный ресурс] // Образовательные технологии и общество.–2004.– Т.7.– №7.– http://ifets.ieee.org/russian/depository/v7_i4/pdf/1.pdf
3. Звонников В.И., Челышкова М.Б. Контроль качества обучения при аттестации. – М.: Логос, 2011.– 272с.
4. Ким В.С. Тестирование учебных достижений. Монография.- Уссурийск: Изд-во УГПИ, 2007. -214 с.
5. Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов.- М.: Логос, 2010.- 668 с.
6. Майоров А. Н. Теория и практика создания тестов для системы образования. (Как выбирать, создавать и использовать тесты для целей образования) / А. Н. Майоров. – М.: Интеллект-центр, 2001. – 296 с.
7. Матвиевская Е.Г. Педагогические измерения. Опыт и проблемы исследования [Электронный ресурс] // CREDO NEW теоретический журнал. – <http://credonew.ru/content/view/769/33/>.
8. Челышкова Н. Б. Теория и практика конструирования педагогических тестов : учебное пособие / Н. Б. Челышкова. – М. : Логос, 2002. – 432 с.
9. Andrich D. Advanced Social and Educational Measurement. Perth: Murdoch University, 2001.- 128 pp.
10. Brennan R. Educational Measurement.- Westport, CT: Praeger, 2006.- 796 pp.
11. Gronlund N.E. Measurement and evaluation in teaching.- Macmillan Pub. Co., 1965- 420 pp.
12. Gronlund N.E., Linn R.L. Measurement and evaluation in teaching, Sixth edition. New York: Macmillan, 1990.
13. Linn R.L., Miller M.D. Measurement and assessment in teaching, 9th Edition. Upper Saddle River, NJ: Prentice Hall, 2005.

V. СТАТИСТИЧНЕ МОДЕЛЮВАННЯ ТЕСТІВ

У цьому розділі представлені основні сучасні підходи до моделювання та параметризації тестів. Наведено основні відомості про однопараметричну модель Раша, а також двох- та трьох параметричні моделі Бірнбаума.

Висвітлюються математично-статистичні методи та підходи, які розроблені з метою зробити процес педагогічного тестування універсальним інструментом вимірювання досягнень учасників тестування. Окреслені шляхи розв'язання інших важливих задач моделювання тестів, зокрема, побудови характеристичних кривих для досягнень учасників тестування та для складності тестових завдань на основі емпіричних даних та перевірки гіпотези про відповідність емпіричних даних моделі Раша. Наведено аналіз інформаційних функцій тестових завдань та всього тесту у різних моделях тестування.

Підняті також актуальні як для класичної, так і для сучасної теорії тестування питання оцінювання роздільної здатності тесту та оцінки дискримінаційної здатності тесту. Наведені відомості про надійність тесту та методи обчислення середньої квадратичної помилки вимірів та методів обчислення коефіцієнта надійності, а також класифікація та математичне тлумачення ще однієї важливої характеристики тесту – валідності.

5.1. Основні припущення та поняття сучасної теорії тестування

Введемо **функцію успіху**, яка дорівнює імовірності того, що учасник тестування вірно виконає тестове завдання $p = p(s, t)$, де s - рівень підготовленості учасника, t - рівень складності тестового завдання.

Імовірність успіху залежить не від кожного окремо взятого аргументу s та t , а лише від їх частки $\zeta = \frac{s}{t}$, тобто $p = p(s, t) = p_1(\zeta)$.

Величини s та t називають **латентними** параметрами, оскільки вони визначають деякі приховані характеристики учасників тестування і тестових завдань.

На функцію успіху накладаються такі умови:

- функція $p_1(\zeta)$ повинна бути гладкою (неперервною разом зі своєю похідною) і монотонно зростаючою на всій області свого визначення, оскільки будь-яке збільшення відношення $\zeta = \frac{s}{t}$ повинне приводити до збільшення імовірності правильного вирішення завдання;
- $\lim_{\zeta \rightarrow 0} p_1(\zeta) = p_1(0) = 0$, що робить безнадійним успіх абсолютно невідготовленого учасника тестування;
- $\lim_{\zeta \rightarrow +\infty} p_1(\zeta) = 1$, що гарантує успіх учаснику тестування, рівень підготовленості якого у багато разів перевищує трудність завдання;
- $p_1(1) = 0,5$, тобто максимальна невизначеність в прогнозі результату виконання завдання повинна бути у тому випадку, коли рівень підготовленості учасника співпадає з трудністю завдання $t = d$.

Замість вище введених латентних параметрів s та t зазвичай розглядають інші змінні $\ln s = \theta$, $\ln t = \beta$, які вимірюються однією і тією ж шкалою з одиницею вимірювання *логит* та можуть набувати довільних значень $\theta \in (-\infty, \infty)$, $\beta \in (-\infty, \infty)$.

5.2. Основні параметричні моделі в освітніх вимірюваннях

5.2.1. Однопараметрична модель Раша

Модель Раша - найпростіша модель імовірності успіху - має вигляд

$$p = p(s, t) = \frac{s}{s + t} = \frac{s/t}{1 + s/t} = \frac{\zeta}{1 + \zeta}.$$

Нескладно бачити, що модель задовольняє основним припущенням, наведеним у попередньому пункті. Після заміни змінних s та t на нові змінні θ та β , які вимірюються в логітах $\ln s = \theta$, $\ln t = \beta$ функція успіху набуде вигляду

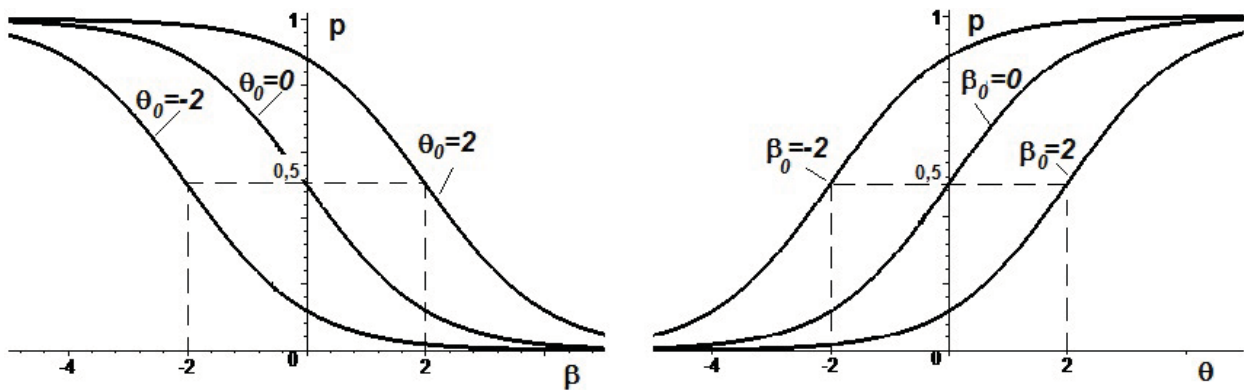
$$p = \frac{1}{1 + \exp[-(\theta - \beta)]},$$

яка називається **основною логістичною моделлю Раша**.

Імовірність успіху залежить від одного параметра $\theta - \beta$, і тому модель є **однопараметричною**. При цьому виконуються граничні випадки

$$\lim_{(\theta-\beta)\rightarrow+\infty} p = 1, \quad \lim_{(\theta-\beta)\rightarrow-\infty} p = 0, \quad \text{а також } p = 0,5, \text{ якщо } \theta = \beta$$

Імовірність p як функція β при фіксованому значенні рівня підготовленості $\theta = \theta_0$ повністю описує потенціальні можливості індивідуума з рівнем підготовленості θ_0 при виконанні завдань будь-яких складності β і тому називається **характеристичною функцією рівня підготовленості** θ_0 . Імовірність p як функція θ при фіксованому значенні $\beta = \beta_0$ характеризує можливості учасників тестування з різним рівнем підготовленості при розв'язанні завдання складності β_0 і називається **характеристикою функцією складності** β_0 . Графіки характеристичних функцій називаються відповідними **характеристичними кривими**.



Характеристичні криві рівня підготовленості учасників тестування $\theta_0 = -2, 0, 2$

Характеристичні криві рівня складності завдань $\beta_0 = -2, 0, 2$

Характеристичні криві рівня складності тесту, які відповідають різним рівням складності β_1 та β_2 , не перетинаються. Характеристичні криві рівня підготовленості учасників тестування, які відповідають різним рівням підготовленості θ_1 та θ_2 , також не перетинаються. Це свідчить про закладену у модель Раша можливість справедливого оцінювання досягнень учасників тестування шляхом їх диференціації системою тестових завдань з різними рівнями складності.

Часто у якості основної логістичної моделі Раша замість вище наведеної формули використовують таку

$$p = \frac{1}{1 + \exp[-1,7(\theta - \beta)]}$$

оскільки вона добре апроксимує функцію розподілу імовірності нормального закону з нульовим математичним сподіванням і одиничною дисперсією.

5.2.2. Двох-параметрична модель Бірнбаума

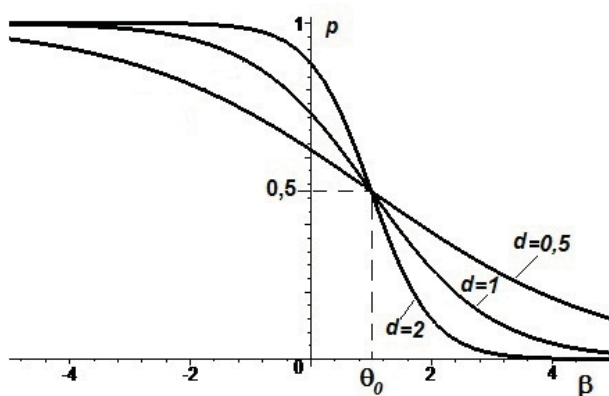
За **двохпараметричною моделлю Бірнбаума** функція успіху має вигляд

$$p = \frac{1}{1 + \exp(-d(\theta - \beta))},$$

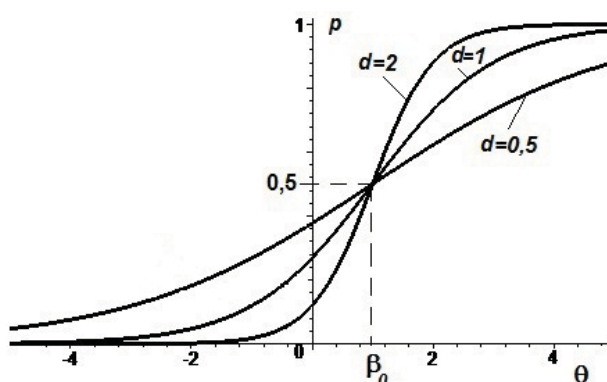
звідки видно, що вона залежить від таких параметрів

- $(\theta - \beta)$ - різниця рівня підготовки учасника тестування та рівня складності тесту, які вимірюються у логітах;
- d - **дискримінаційна здатності тестового завдання**, тобто здатності відділити краще підготовлених від слабо підготовлених учасників

Параметр d прийнято називати дискримінаційною здатністю тестового завдання, т.я. при малому d характеристична крива складності β_0 є похилою, а це означає, що учасники тестування з хорошим та поганим рівнем підготовки виконують дане завдання з приблизно рівним успіхом; якщо ж параметр d набуває великого значення, то шанси успішного виконання даного завдання учасниками з $\theta < \beta_0$ і $\theta > \beta_0$ істотно відрізняються.



Характеристичні криві рівня підготовленості учасників тестування θ_0 з дискримінаційною здатністю $d = 0.5, 1, 2$



Характеристичні криві рівня складності завдань β_0 з дискримінаційною здатністю $d = 0.5, 1, 2$

У рамках двохпараметричної моделі кожному тестовому завданню певної складності $\beta = \beta_0$ відповідає сім'я кривих, які перетинаються в єдиній точці перегину $(\theta = \beta_0; 0.5)$, у цій точці кривизна кожної кривої множини,

тобто величина тангенса кута нахилу дотичної, рівна $0,25d$. Аналогічне можна сказати про сім'ю характеристичних кривих часників тестування при протилежному напрямі нахилу дотичної.

При $d = 1$ двохпараметрична логістична модель Бірнбаума співпадає з логістичною однопараметричною моделлю Раша
$$p = \frac{1}{1 + \exp[-(\theta - \beta)]}$$
.

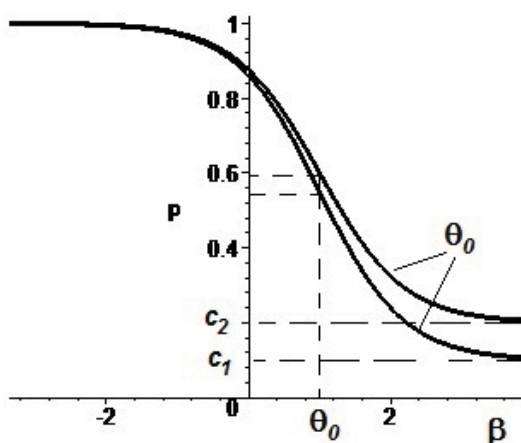
При $d = 1,7$ двохпараметрична логістична модель Бірнбаума співпадає з таким варіантом логістичної однопараметричної моделі Раша
$$p = \frac{1}{1 + \exp[-1,7(\theta - \beta)]}$$
, яка добре апроксимує функцію розподілу імовірності нормального закону з нульовим математичним сподіванням і одиничною дисперсією.

5.2.3. Трьох-параметрична модель Бірнбаума

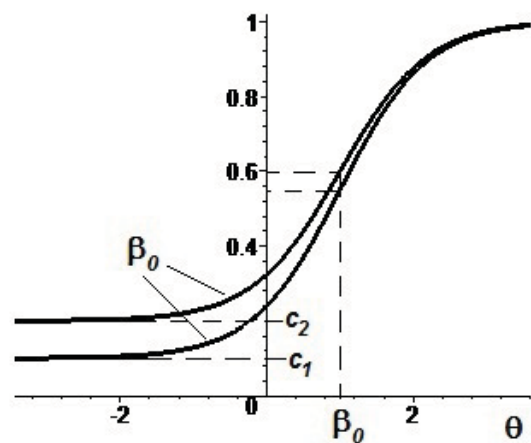
За **трьохпараметричною моделлю Бірнбаума** функція успіху

$$p = c + (1 - c) \frac{1}{1 + \exp(-d(\theta - \beta))},$$

звідки видно, що вона залежить від параметрів $(\theta - \beta)$ та d , а також від третього параметра c , який дає можливість відсікти можливість відгадування вірних відповідей.



Характеристичні криві рівня підготовленості учасників тестування θ_0 з дискримінаційною здатністю $d = 1,7$ та шансами угадування $c_1 < c_2$



Характеристичні криві рівня складності завдань β_0 з дискримінаційною здатністю $d = 1,7$ та шансами угадування $c_1 < c_2$

Граничне значення функції успіху трьох-параметричної моделі дорівнює c (а не прямує до нуля, як у двох-параметричній моделі) при фіксованому рівні складності тесту β_0 та за умови зменшення рівня підготовки $\theta \rightarrow -\infty$ і тому графік характеристичної прямої має горизонтальну асимптоту $p = c$.

Для врахування можливості відгадування вірних відповідей параметр c обирають рівним імовірності відгадування.

Функція успіху трьохпараметричної моделі не є логістичною.

Очевидно, що трьохпараметрична модель Бірнбаума перетворюється у двохпараметричну логістичну модель Бірнбаума, якщо параметр c дорівнює нулю. Якщо ж при фіксованому $c = 0$ дискримінаційна здатність $d = 1$, то дана модель перетворюється в однопараметричну модель Раша.

Зауважимо, що збільшення кількості параметрів, які ще більш повно характеризують тестові завдання і учасників тестування, призводить до зниження надійності їх практичного визначення за результатами масового тестування.

5.3. Первинні бали та їх статистичні характеристики

5.3.1. Первинні бали та їх розподіл

Нехай n кількість учасників тестування різного рівня підготовленості θ_i $i = 1, 2, \dots, n$. Тест складається з k завдань різної складності β_j $j = 1, 2, \dots, k$. Результат оцінюється за дихотомічним принципом: «1» - завдання виконано вірно, «0» - невірно. Результати тестування утворюють **матрицю відповідей** $A = \{A_{ij}\}$ розмірністю $n \times k$. Її елемент A_{ij} виражає можливий результат виконання i -го учасником j -го завдання. Елементи A_{ij} є випадковими величинами: вони дорівнюють «1» з імовірністю $p_{ij} = p(\theta_i, \beta_j)$, і дорівнюють «0» з імовірністю $p_{ij} = 1 - q_{ij}$. Тут θ_i та β_j - значення латентних параметрів рівня підготовленості i -го учасника та рівня складності j -го тестового завдання.

Реалізації випадкових величин A_{ij} позначатимемо a_{ij} , вони утворюють реалізацію матриці відповідей $a = \{a_{ij}\}$. В умовах реального тестування кількість стовпців k (число завдань в одному тесті) не перевищує 60-70, але кількість рядків n (число учасників тестування) може бути значно більше. У рамках вищеописаних параметричних моделей дані тестування дають можливість побудувати оцінки латентних параметрів. Деякі статистичні методи та відповідні алгоритми побудови ефективних та незміщених оцінок будуть наведені у наступних пунктах.

Зручними статистиками при обробці результатів масового тестування є так звані маргінальні суми елементів матриці відповідей $a = \{a_{ij}\}$ по рядку і по стовпцю.

Величина $b_i = \sum_{j=1}^k a_{ij}$, ($i = 1, 2, \dots, n$) дорівнює кількості вірно

виконаних завдань учасником з номером i називається **первинним балом i -го учасника**, відображає визначену міру успіху i -го випробовуваного при виконанні тесту. Первинний бал i -го учасника є реалізацією випадкової

величини $B_i = \sum_{j=1}^k A_{ij}$.

Величина $c_j = \sum_{i=1}^n a_{ij}$, ($j = 1, 2, \dots, k$) дорівнює кількості учасників, які

вірно виконали завдання з номером j і називається **первинним балом j -го завдання**. Різниця $n - c_j$ відображає визначену міру складності j -го завдання при виконанні цього завдання даним континентом n учасників тестування. Первинний бал j -го тестового завдання є реалізацією випадкової

величини $C_j = \sum_{i=1}^n A_{ij}$.

Нескладно бачити, що випадкові величини A_{ij} мають математичне сподівання p_{ij} та дисперсію $p_{ij}q_{ij}$, а математичні сподівання і дисперсії первинних балів

$$M\{b_i\} = \sum_{j=1}^k p_{ij} = p_i, \quad D\{b_i\} = \sum_{j=1}^k p_{ij}q_{ij}, \quad q_{ij} = 1 - p_{ij};$$

$$M\{c_j\} = \sum_{i=1}^n p_{ij} = \rho_i, \quad D\{c_j\} = \sum_{i=1}^n \rho_{ij} \gamma_{ij}, \quad \gamma_{ij} = 1 - \rho_{ij}.$$

Первинні бали мають узагальнений біноміальний розподіл.

5.3.2. Оцінки ймовірностей первинного балу учасників тестування

Імовірність $P_{ik}(b_i)$ того, що i -й учасник із k завдань вірно вирішить b_i штук, $0 \leq b_i \leq k$, дорівнює коефіцієнту при x^{b_i} у розвиненні похідної функції

$$\varphi_{ik}(x) = (p_{i1}x + q_{i1})(p_{i2}x + q_{i2}) \cdot \dots \cdot (p_{ik}x + q_{ik}).$$

Якщо всі завдання тесту мають однаковий рівень складності, тобто $\beta_j = \beta$, отже $p_{ij} = p_i$ незалежно від j , то

$$M\{b_i\} = kp_i; \quad D\{b_i\} = kp_i q_i; \quad \varphi_{ik}(x) = (p_i x + q_i)^k,$$

де $p_i = p(\theta_i, \beta)$, $q_i = 1 - p_i$.

Первинний бал B_i має біноміальний розподіл $B \in Bi(k, p)$

Для імовірності $P_{ik}(b_i)$ отримання i -м учасником тестування b_i балів із k завдань справедливі формули:

$$P_{ik}(b_i) = C_k^{b_i} \cdot p_i^{b_i} \cdot q_i^{k-b_i} \quad \text{або} \quad P_{ik}(b_i) \approx \frac{1}{\sqrt{kp_i q_i}} \varphi(u)$$

де $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$, $u = \frac{b_i - kp_i}{\sqrt{kp_i q_i}}$, $u \in N(0;1)$.

Імовірність попадання первинного бала B_i - в заданий проміжок $[m_1; m_2]$

$$P_{ik}(m_1 \leq B_i \leq m_2) \approx \Phi(x_2) - \Phi(x_1),$$

де $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{u^2}{2}} du$, $x_1 = \frac{m_1 - kp_i}{\sqrt{kp_i q_i}}$, $x_2 = \frac{m_2 - kp_i}{\sqrt{kp_i q_i}}$.

При великій кількості числа k завдань в тесті їх розподіл близький до

нормального, тому $P_{ik} \{ |B - M(B_i)| \leq \varepsilon \} = 2\Phi\left(\frac{\varepsilon}{\sqrt{D(B)}}\right)$

Більш точний результат $P_{ik} \{ m_1 < M(B_i) < m_2 \} = d$, де

$$m_1 = \frac{k}{k+l^2} \left(b_i + \frac{l^2}{2} - l \sqrt{\frac{b_i(k-b_i)}{k} + \frac{l^2}{2}} \right);$$

$$m_2 = \frac{k}{k+l^2} \left(b_i + \frac{l^2}{2} + l \sqrt{\frac{b_i(k-b_i)}{k} + \frac{l^2}{2}} \right);$$

l - значення аргументу функції Лапласа, при якому $2\Phi(l) = d$; d - рівень значущості інтервалу (m_1, m_2) .

5.3.3. Оцінки ймовірностей первинного балу тестових завдань

Всі формули для первинних балів учасників тестування справедливі також для первинних балів завдань.

Імовірність $R_{nj}(c_j)$ того, що j -те тестове завдання із n учасників вірно вирішать c_j людей, $0 \leq c_j \leq n$, дорівнює коефіцієнту при y^{c_j} у розвиненні похідної функції

$$\Psi_{nj}(y) = (\rho_{1j}y + \gamma_{1j})(\rho_{2j}y + \gamma_{2j}) \cdot \dots \cdot (\rho_{nj}y + \gamma_{nj}).$$

Якщо всі учасники мають однаковий рівень підготовки, тобто $\theta_i = \theta$, отже $\rho_{ij} = \rho_j$ незалежно від i , то

$$M\{c_j\} = n\rho_j; \quad D\{c_j\} = n\rho_j\gamma_j; \quad \Psi_{nj}(y) = (\rho_j y + \gamma_j)^n,$$

де $\rho_j = p(\theta, \beta_j)$, $\gamma_j = 1 - \rho_j$.

Первинний бал C_j має біноміальний розподіл $C_j \in Bi(n, \rho_j)$

Для імовірності $R_{nj}(c_j)$ того, що j -те тестове завдання розв'яжуть c_j осіб із n учасників тестування справедливі формули:

$$R_{nj}(c_j) = C_n^{c_j} \cdot \rho_j^{c_j} \cdot \gamma_j^{n-c_j}$$

$$\text{або } R_{nj}(c_j) \approx \frac{1}{\sqrt{n\rho_j\gamma_j}} \varphi(u),$$

де $\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$, $u = \frac{c_j - n\rho_j}{\sqrt{n\rho_j\gamma_j}}$, $u \in N(0;1)$.

Імовірність попадання первинного бала C_j - в заданий проміжок $[m_1; m_2]$

$$R_{nj}(m_1 \leq C_j \leq m_2) \approx \Phi(x_2) - \Phi(x_1),$$

$$\text{де } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{u^2}{2}} du, \quad x_1 = \frac{m_1 - n\rho_j}{\sqrt{n\rho_j\gamma_j}}, \quad x_2 = \frac{m_2 - n\rho_j}{\sqrt{n\rho_j\gamma_j}}.$$

При великій кількості n учасників тестування розподіл їх балів близький до

$$\text{нормального, тому } R_{nj}\{|C_j - M(C_j)| \leq \varepsilon\} = 2\Phi\left(\frac{\varepsilon}{\sqrt{D(C_j)}}\right).$$

Більш точний результат $R_{nj}\{m_1 < M(C_j) < m_2\} = d$, де

$$m_1 = \frac{n}{n+l^2} \left(c_j + \frac{l^2}{2} - l \sqrt{\frac{c_j(n-c_j)}{n} + \frac{l^2}{2}} \right);$$

$$m_2 = \frac{n}{n+l^2} \left(c_j + \frac{l^2}{2} + l \sqrt{\frac{c_j(n-c_j)}{n} + \frac{l^2}{2}} \right);$$

l - значення аргументу функції Лапласа, при якому $2\Phi(l) = d$; d - рівень значущості інтервалу (m_1, m_2) .

5.3.4. Важливі властивості первинних балів

Можна показати, що заміна ймовірностей p_{ij} їх осередненими значеннями p_i , призводить до того, що дисперсія первинного балу учасників тестування досягає максимуму. Це означає, що будь-які відхилення складностей завдань тесту від їх середнього рівня зменшують величину випадкових коливань первинного бала учасників тестування та, відповідно, зменшують розподіл можливих оцінок відповідних рівнів підготовленості учасників. З іншого боку, чим більш однорідний за рівнем підготовленості склад учасників тестування, тим більший розподіл оцінок рівня складності завдань слід чекати. Такий **взаємозв'язок між дисперсіями** контингенту учасників тестування і множиною тестових завдань є найважливішою властивістю тесту з незалежними завданнями, які виконують незалежні учасники. Видалення впливу цього взаємозв'язку з оцінок латентних

параметрів - одна з умов об'єктивного калібрування тестових завдань і шкалування результатів учасників тестування.

Важливою властивістю первинних балів є те, що у рамках основної моделі Раша вони є **достатніми статистиками**, тому великий масив початкових даних вдається істотно зменшити без втрати інформації заміною масиву матриці відповідей на два вектори первинних балів.

5.4. Оцінки латентних параметрів

5.4.1. Стандартні оцінки параметрів та помилки вимірювання

Після збору емпіричних даних за результатами тестування n учасників тестування системою k тестових завдань виконується статистична обробка. Перші кроки у статистичній обробці, взагалі кажучи, співпадають з алгоритмом класичної теорії, описаним у попередньому розділі.

Складається матриця тестових результатів, в якій символом x_{ij} для всіх $i = 1, \dots, n$; $j = 1, \dots, k$ позначено результат виконання i -м учасником тестування

j -го завдання тесту:

$$x_{ij} = \begin{cases} 1, & \text{якщо відповідь вірна} \\ 0, & \text{якщо відповідь невірна} \end{cases}$$

Після видалення з матриці рядків та стовпців, які містять тільки нулі або тільки одиниці, знаходять індивідуальні бали досліджуваних та кількістю правильних відповідей випробовуваних на кожне завдання тесту. Індивідуальний (первинний) бал X_i ($i = 1, \dots, n$) учасника тестування підраховується підсумовуванням всіх одиниць, отриманих ним за правильно виконані завдання тесту. Індивідуальний (первинний) бал R_j ($j = 1, \dots, k$) тестового завдання отримується підсумовуванням одиниць по стовпцях.

Далі здійснюється впорядкування матриці результатів тестування. Для цього роблять перестановку стовпців, розставляючи R_j в порядку спадання, після чого виконують перестановку рядків так, що розміщують згори вниз у порядку зростання X_i . Остаточна нумерація учасників тестування (рядків) та тестових завдань (стовпчиків) відповідає впорядкованій матриці відповідей після названих перетворень (вважаємо розмірність матриці $n \times k$ незмінною).

Виконується підрахунок доли правильних p_i і неправильних q_i відповідей кожного випробовуваного на усі завдання тесту

$$p_i = \frac{X_i}{k}, \quad q_i = 1 - p_i,$$

та рівень підготовки i -го учасника тестування у логітах $\theta_i^0 = \ln \frac{p_i}{q_i}$.

Аналогічно підраховуються долі правильних ρ_j і неправильних γ_j відповідей на кожне завдання тесту

$$\rho_j = \frac{R_j}{n}, \quad \gamma_j = 1 - \rho_j$$

та рівень складності j -го завдання у логітах $\beta_j^0 = \ln \frac{\gamma_j}{\rho_j}$.

Середні значення логітів рівня підготовки учасників та логітів складності завдань тесту визначаються формулами

$$\bar{\theta} = \frac{\sum_{i=1}^n \theta_i^0}{n}, \quad \bar{\beta} = \frac{\sum_{j=1}^k \beta_j^0}{k}.$$

Оцінки кожного з параметрів θ_i і β_j виражаються в інтервальній шкалі, але з різними значеннями середніх і різними стандартними відхиленнями. Виникає питання про перенесення їх на єдину шкалу, для чого обчислюються дисперсії оцінок латентних параметрів.

$$\text{Дисперсія по множині значень } \theta_i^0: \quad V = \frac{1}{n-1} \left[\sum_{i=1}^n (\theta_i^0)^2 - n(\bar{\theta})^2 \right].$$

$$\text{Дисперсія по множині значень } \beta_j^0: \quad U = \frac{1}{k-1} \left[\sum_{j=1}^k (\beta_j^0)^2 - k(\bar{\beta})^2 \right].$$

Уточнюючі коефіцієнти:

$$X = \sqrt{\frac{1+U/2,89}{1-UV/8,35}}, \quad Y = \sqrt{\frac{1+V/2,89}{1-UV/8,35}}.$$

Оцінки параметрів θ_i і β_j в **єдиній** інтервальній шкалі знаходяться так

$$\theta_i = \bar{\theta} + X\theta_i^0, \quad \beta_j = \bar{\beta} + Y\beta_j^0$$

Оцінка *стандартної похибки* $S_e(\theta_i)$ вимірювання рівня підготовленості θ_i

$$S_e(\theta_i) = \frac{X}{\sqrt{p_i(k - X_i)}} = \frac{X}{\sqrt{kp_i(1 - p_i)}} = \frac{X}{\sqrt{kp_i q_i}}$$

Оцінка стандартної похибки $S_e(\beta_j)$ вимірювання рівня складності β_j

$$S_e(\beta_j) = \frac{Y}{\sqrt{\rho_j(n - Y_j)}} = \frac{Y}{\sqrt{n\rho_j(1 - \rho_j)}} = \frac{Y}{\sqrt{n\rho_j\gamma_j}}.$$

5.4.2. Стійкі оцінки латентних параметрів

У рамках одно-параметричної моделі Раша за умови великої кількості учасників тестування можна обчислити уточнені стійкі значення латентних параметрів θ_i' і β_j' методом найбільшої правдоподібності Р.Фішера.

Побудуємо функцію правдоподібності

$$L(x_{ij}, \theta_i, \beta_j) = \exp \left[\sum_{i=1}^n \sum_{j=1}^k x_{ij} (\theta_i - \beta_j) \right] \left[\prod_{i=1}^n \prod_{j=1}^k (1 + \exp(\theta_i - \beta_j)) \right]^{-1}.$$

В якості ефективних оцінок найбільшої правдоподібності θ_i' і β_j' обирають такі значення θ_i і β_j , при яких функція правдоподібності досягає максимуму. Шукають максимум логарифмічної функції правдоподібності $\ln(L)$

$$\ln L(x_{ij}, \theta_i, \beta_j) = \sum_{i=1}^n \sum_{j=1}^k x_{ij} \theta_i - \sum_{i=1}^n \sum_{j=1}^k x_{ij} \beta_j - \sum_{i=1}^n \sum_{j=1}^k \ln[1 + \exp(\theta_i - \beta_j)].$$

Для знаходження максимуму логарифмічної функції правдоподібності розв'язують систему рівнянь правдоподібності

$$\frac{\partial \ln L}{\partial \theta_i} = \sum_{j=1}^k x_{ij} - \sum_{i=1}^n p_{ij} = 0,$$

$$\frac{\partial \ln L}{\partial \beta_i} = -\sum_{i=1}^n x_{ij} + \sum_{j=1}^k p_{ij} = 0,$$

яка розв'язується в ітераційному циклі шляхом послідовних підстановок оцінок параметрів θ_i і β_j з єдиної інтервальної шкали в якості початкових. Цикл переривається, якщо різниця послідовних наближень стає меншим наперед заданого значення. Останні значення θ_i' і β_j' в ітераційному циклі є уточненими стійкими значеннями латентних параметрів, різниця яких $(\theta_i' - \beta_j')$ є єдиним параметром одно-параметричної логістичної моделі Раша.

Для підбору параметрів двох- та трьох-параметричної моделі на основі емпіричних даних також існують відповідні ітераційні методи, які тут не наводяться.

5.5. Характеристичні функції та криві для емпіричних даних

5.5.1. Побудова характеристичних функцій та кривих

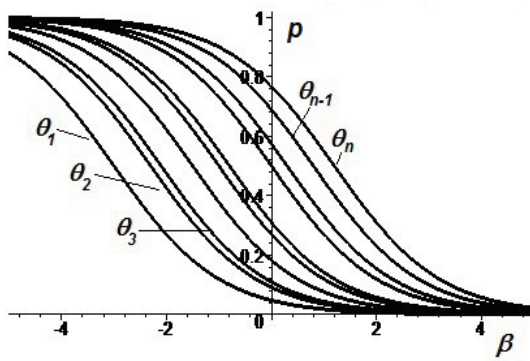
Характеристичні криві учасників тестування (PCC) та тестових завдань (ICC) будують на основі отриманих методом найбільшої правдоподібності Р.Фішера значень латентних параметрів θ_i' і β_j' та з використанням однієї з параметричних моделей тестування, наприклад, функції успіху логістичної моделі Раша у вигляді
$$p = \frac{1}{1 + \exp[-1,7(\theta - \beta)]}$$
.

Для побудови ICC підставляємо значення рівня складності для кожного тестового завдання β_j' у функцію успіху, при цьому параметр θ вважаємо змінною величиною. Отримуємо k характеристичних функцій тестових завдань
$$p_j(\theta) = \frac{1}{1 + \exp[-1,7(\theta - \beta_j')]}.$$

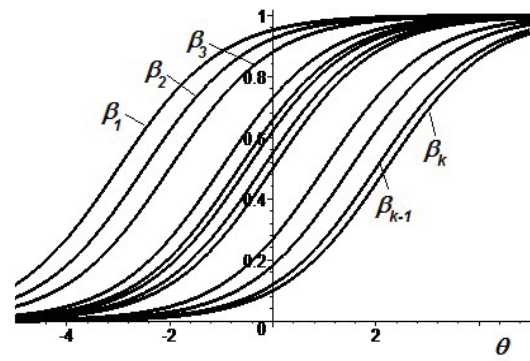
Аналогічно, для побудови PCC підставляємо значення рівня підготовленості для кожного учасника тестування θ_i' у функцію успіху, при цьому параметр β вважаємо змінною величиною. Отримуємо n характеристичних функцій учасників тестування

$$p_i(\beta) = \frac{1}{1 + \exp[-1,7(\theta_i' - \beta)]}.$$

Графіки характеристичних функцій є характеристичними кривими, приклади яких зображені на рисунку.

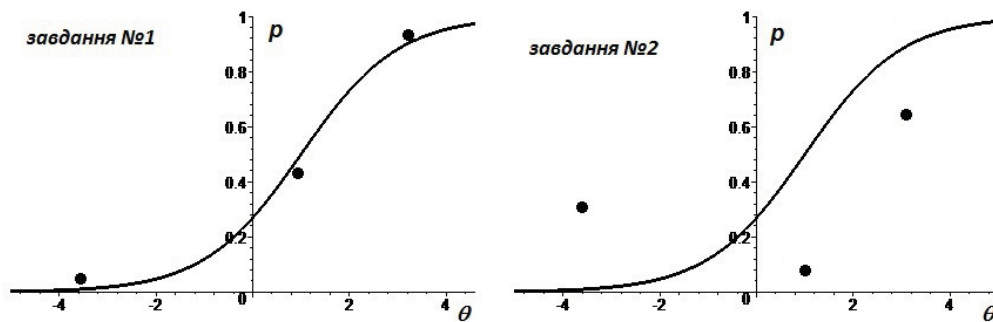


РСС - характеристичні криві учасників
тестування $p_i(\beta)$ $i = 1, \dots, n$



ІСС - характеристичні криві тестових
завдань $\rho_j(\theta)$ $j = 1, \dots, k$

Нижче на рисунках наведемо дані двох завдань, звідки нескладно бачити, що завдання №1 у значній мірі відповідає моделі, а завдання №2 не відповідає моделі Раша.



5.5.2. Два підходи теорії параметризації тестів

У сучасній теорії параметризації тестів існує два протилежні підходи:

- однопараметрична модель Раша створювалась з метою побудови «ідеальних» тестів, які б відповідали теорії;
- багатопараметричні моделі Бірнбаума створювались для покращення відповідності теорії з емпіричними даними шляхом підбору параметрів моделі.

Наведемо міркування у рамках першого підходу до моделювання тестів, за яким однопараметрична модель Раша створювалась з метою побудови «ідеальних» тестів, які б відповідали теорії. Для вирішення, чи відповідає побудована модель досліджуваному тесту, потрібно виконати наступні дії. Всіх n учасників тестування, які виконували k завдань тесту можна розподілити по шкалі θ за діапазонами рівня підготовленості. Учасники тестування розбиваються на T груп вздовж шкали θ так, щоб всі

учасники всередині даної групи мали однаковий рівень підготовленості θ_t . Всього всередині групи з номером t міститься m_t учасників тестування ($t = 0, 1, \dots, k$). В межах кожної групи r_{tj} учасників відповідають вірно на j -те завдання тесту. Для групи з рівнем підготовленості θ_t емпірична імовірність

вірної відповіді на j -те завдання тесту дорівнює
$$P(\theta_{tj}) = \frac{r_{tj}}{m_t},$$

а теоретична імовірність вірної відповіді на j -те завдання тесту для групи з рівнем підготовленості θ_t

$$p(\theta_{tj}) = \frac{1}{1 + \exp[-1,7(\theta_{tj} - \beta)]}.$$

Для перевірки основної гіпотези H_0 про відповідність отриманих емпіричних даних однопараметричній моделі IRT для завдань тесту обчислюють статистику Хі-квадрат $\chi^2 = \sum_{j=1}^k \chi(j)$, де

$$\chi^2(j) = \sum_{t=0}^T m_t \frac{(P(\theta_{tj}) - p(\theta_{tj}))^2}{p(\theta_{tj})}.$$

В умовах нульової гіпотези випадкова величина $\chi^2(j)$ має Хі-квадрат розподіл з числом степенів волі $(T - 1)$. Рекомендується нечисленні групи, де число учасників менше 6, об'єднувати з сусідніми, чим зменшити число степенів

вільності $(T - 1)$. Статистика $\chi^2 = \sum_{j=1}^k \chi(j)$ має число степенів волі Tk .

Порівнюють спостережуване значення $\chi_{\text{спост}}^2$ цієї вибіркової статистики з її критичним значенням $\chi_{\text{крит}}^2$, яке обирається з відповідної таблиці, за результатами порівняння приймається одне з двох рішень :

1) якщо $\chi_{\text{спост}}^2 > \chi_{\text{крит}}^2$ то нульова гіпотеза відкидається, тобто вважається, що гіпотетична модель Раша не узгоджується з експериментальними даними;

2) якщо $\chi_{\text{спост}}^2 \leq \chi_{\text{крит}}^2$ те вважається, що немає підстав для відхилення нульової гіпотези, тобто гіпотетична модель Раша узгоджується з експериментальними даними.

Міркування у рамках другого підходу до моделювання тестів, за яким багатопараметричні моделі Бірнбаума створювались для покращення відповідності теорії з емпіричними даними шляхом підбору параметрів моделі, істотно відрізняються. Названий підхід став можливим після побудови двох- та трьох-параметричної моделі, що дало змогу підбором другого параметру (або другого та третього параметрів) налаштувати багатопараметричну модель так, щоб вона відповідала основній гіпотезі про відповідність отриманих емпіричних даних побудованій багатопараметричній моделі IRT.

5.6. Інформаційна функція

5.6.1. Інформаційна функція тестового завдання

Кількість інформації, яка забезпечується j -м завданням тесту при фіксованому рівні підготовки учасника тестування θ , це величина, обернено пропорційна стандартній помилці вимірювання j -м завданням величини θ .

Інформаційна функція j -го тестового завдання задається формулою

$$I_j(\theta) = \frac{\left(\frac{dp_j(\theta)}{d\theta}\right)^2}{p_j(\theta)q_j(\theta)},$$

де $q_j(\theta) = 1 - p_j(\theta)$.

Для логістичної однопараметричної моделі Раша $p = \frac{1}{1 + \exp[-1,7(\theta - \beta)]}$

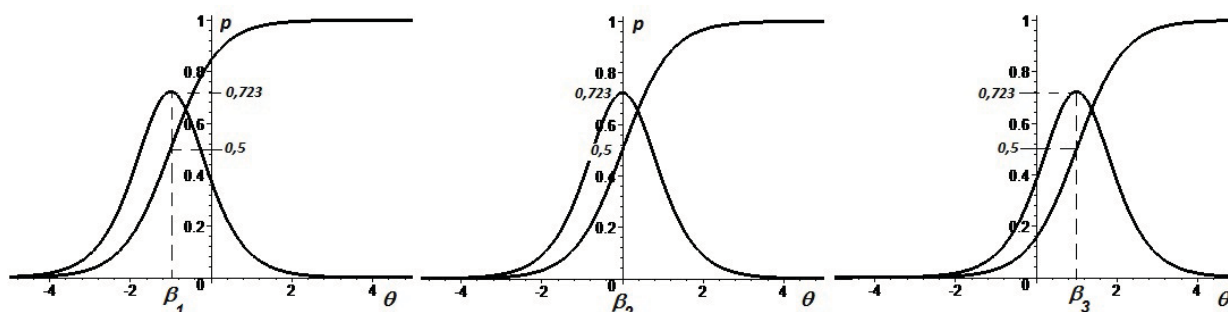
справедливе співвідношення $\frac{dp_j(\theta)}{d\theta} = 1,7 p_j(\theta)q_j(\theta)$,

тому інформаційна функція обчислюється формулою

$$I_j(\theta) = 2,89 p_j(\theta)q_j(\theta) = 2,89 \frac{\exp 1,7(\theta - \beta_j)}{1 + \exp 1,7(\theta - \beta_j)}.$$

Графіки інформаційної функції та характеристична крива для трьох тестових завдань зі складностями рівнів $\beta_1 < \beta_2 < \beta_3$ представлені на рисунку, із якого

видно, що завдання найбільш інформативні, якщо його складність співпадає з рівнем підготовленості учасника тестування.

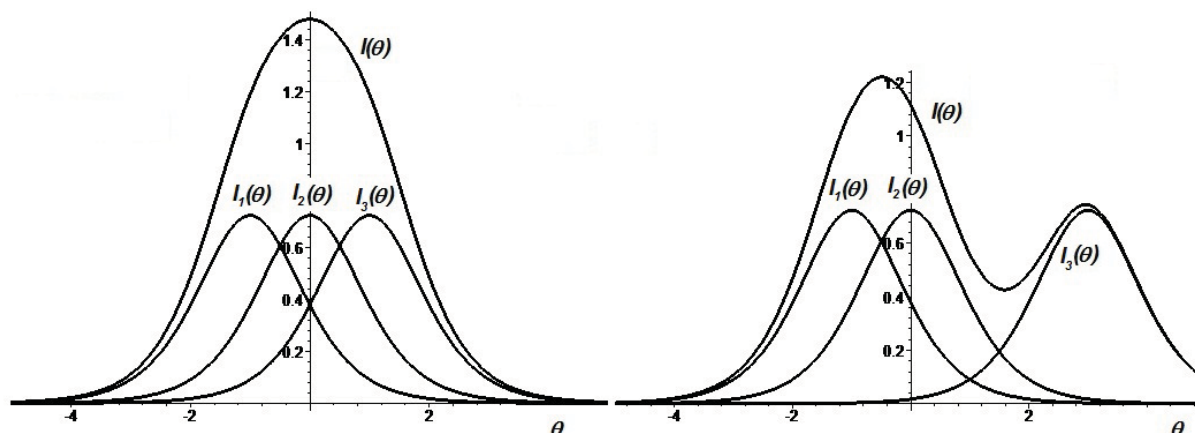


5.6.2. Інформаційна функція системи тестових завдань

Інформаційною функцією сукупності тестових завдань називають

суму інформаційних функцій кожного тестового завдання $I(\theta) = \sum_{j=1}^N I_j(\theta)$.

На рисунках наведено інформаційні функції двох різних тестових завдань. Нескладно бачити, що перший тест орієнтований на групу учасників тестування з нормально розподіленим рівнем підготовки, його інформаційна функція має один максимум. Другий тест має два екстремуму, що свідчить про необхідність його доопрацювання шляхом збільшення кількості завдань, складність яких знаходиться між екстремумами або зміною складності існуючих завдань.



Аналогічно можна побудувати інформаційні функції тестових завдань для

двох-параметричної моделі Бірнбіума $p = \frac{1}{1 + \exp[-1,7a(\theta - \beta)]}$

$$I_j(\theta) = 2,89 a_j^2 p_j(\theta) q_j(\theta) = 2,89 a_j^2 \frac{\exp 1,7 a_j (\theta - \beta_j)}{1 + \exp 1,7 a_j (\theta - \beta_j)},$$

для трьох-параметричної моделі Бірнбаума $p = c + (1 - c) \frac{1}{1 + \exp[-1,7a(\theta - \beta)]}$

$$I_j(\theta) = \frac{2,89 a_j^2 (1 - c_j)}{(c_j + \exp(1,7 a_j (\theta - \beta_j))) (1 + \exp(-1,7 a_j (\theta - \beta_j)))^2}$$

Характер інформаційних функцій для двох- та трьох-параметричної моделей в цілому зберігається, однак, існують деякі відміни від однопараметричної моделі.

У двохпараметричній моделі при однакових рівнях складності завдання та рівня підготовки учасника нахил характеристичної кривої може перевищувати величину 0,25 на відміну від однопараметричної моделі. При цьому істотно звужується робоча область завдання. Тобто збільшення дифереціюючої здатності завдання призводить до зменшення числа учасників тестування, на яких воно розраховане.

У трьох параметричній моделі параметр відгадування c_j призводить до зниження точності оцінок інших параметрів та зменшує швидкість збіжності процедури пошуку їх стійких оцінок. Інформаційна функція досягає максимуму при значення рівня підготовки $\theta_{\max} = \beta_j + \frac{1}{3,4 a_j} (1 + \sqrt{1 + 8 c_j})$.

Зі зменшенням параметру c_j відбувається зростання інформативності завдання, що нескладно пояснити тим, що можливість відгадування повинно деформувати реальну картину рівня підготовки учасників тестування.

5.7. Основні характеристики тесту за сучасною теорією

5.7.1. Роздільна здатність тесту (РЗТ)

Точність остаточних результатів багато в чому залежить від об'єму вибірки, тобто від кількості n учасників тестування і кількості k завдань в тісті. Оскільки k , як правило, значно менше n , то саме параметр k чинить основний

вплив на роздільну здатність тесту при визначенні рівня підготовленості θ . При заданому k первинні бали b_i учасників тестування приймають скінченне число значень $0, 1, 2, \dots, k$ із стандартним кроком $\Delta b = 1$. Назвемо **роздільною здатністю тесту (РСТ)** довжину того проміжку $\Delta\theta$ на латентній шкалі в логітах, який відповідає кроку $\Delta b = 1$.

Оскільки різні значення θ_1 і θ_2 тест не в змозі розрізнити, якщо

$$|\theta_2 - \theta_1| \leq \text{РСТ},$$

тому РСТ визначає собою *поріг чутливості тесту*.

1). Якщо обчислення щодо оцінювання латентних параметрів виконані, то залежність між b і $\Delta\theta$ очевидна, оскільки $\Delta\theta_b$ є скінченними різницями 1-го порядку :

$$\Delta\theta_b = \theta_b - \theta_{b-1}, \quad \text{где } b = 1, 2, \dots, 20 \quad \text{де } b = 1, \dots, k$$

Зазвичай, величина $\Delta\theta$ зменшується в середній частині шкали первинних балів і збільшується на краях цієї шкали.

2). На практиці величину РСТ бажано знати апіорі, тобто до завершальних обчислень, зокрема, в процесі складання тесту. Оскільки математичне сподівання первинного балу

$$M\{b_i\} = \sum_{j=1}^k p_{ij},$$

де
$$p_{ij} = \frac{1}{1 + \exp(\theta_i - \beta_j)},$$

то після диференціювання $M\{b_i\}$ по θ_i та інших перетворень отримуємо, що диференціал $d\theta_i$ та частинна похідна

$$d\theta_i = \left(\sum_{j=1}^k p_{ij} q_{ij} \right)^{-1} db_i, \quad \frac{\partial \theta_i}{\partial b_i} = \left(\sum_{j=1}^k p_{ij} q_{ij} \right)^{-1}$$

співпадають, оскільки $db_i=1$. Отже,

$$\text{РСТ} \approx \frac{\partial \theta_i}{\partial b_i} = \left(\sum_{j=1}^k p_{ij} q_{ij} \right)^{-1},$$

тобто роздільна здатність тесту в околі тим вище, чим більше інформації міститься в i -му рядку матриці відповідей.

Мінімальне значення РСТ рівне $4/k$ при $p_{ij}=q_{ij}=1/2$ для $\forall j = 1, 2, \dots, k$.

Максимальне значення PCT теоретично не існує, але практично обмежуються величиною $11/k$, що відповідає малоймовірному випадку $p_{ij}=0,1$ для $\forall j = 1, 2, \dots, k$. У реальному тісті середнє значення PCT задовольняє

$$\frac{4}{k} < PCT < \frac{11}{k},$$

для наближених розрахунків можна користуватися формулою

$$PCT \approx \frac{7}{k} \text{логит.}$$

5.7.2. Оцінка дискримінаційної здатності тесту

Оцінка дискримінаційної здатності визначається, як зв'язок між результатом виконання даного завдання тесту (1 - виконано вірно, 0 - виконано невірно) учасниками тестування і первинним балом, отриманим цими учасниками відповідно до результатів всього тесту. В якості тісноти такого зв'язку використовується вибірковий коефіцієнт точково-бісеріальної кореляції

$$\rho_{pb} = \frac{\text{cov}\{X, Y\}}{\sigma_Y \cdot \sigma_X},$$

де X – результат виконання одного завдання тесту, Y - первинні бали учасників. Вибірковий коефіцієнт точково-бісеріальної кореляції у вказаних умовах називається **коефіцієнтом дискримінації** даного завдання. Чим ближче цей коефіцієнт до 1, тим ефективніше це завдання дозволяє диференціювати учасників за рівнем їх підготовленості. Якщо $\hat{\rho}_{pb} < 0$, то завдання сформульоване явно невдало.

Методом контрастних груп зручно користуватися при визначенні дискримінаційних можливостей завдання тесту. При цьому в кожену групу відбирається по 27% учасників тестування з найбільш високими і найбільш низькими первинними балами, що і дозволяє сформувати таблицю зв'язаних ознак 2×2 . У простому випадку **коефіцієнт дискримінації** оцінюється

$$d = \frac{n_{11}}{n_1} - \frac{n_{12}}{n_2} = \frac{n_{11} - n_{12}}{0,27n},$$

де n_1 - кількість учасників, що потрапили в групу з високими балами ; з них вірно виконали аналізоване завдання тільки n_{11} чоловік. Аналогічно, n_2 -

кількість учасників, що потрапили в групу з низькими балами; з них вірно виконали аналізоване завдання n_{12} чоловік; n - загальна кількість учасників.

5.8. Визначення надійності тесту сучасною теорією

5.8.1. Загальні відомості про надійність тесту

Одне із фундаментальних понять теорії вимірювань є поняття надійності тесту, яке було достатньо докладно висвітлено у попередньому розділі. Сутність поняття надійності тесту порівняно з класичною теорією не зазнала змін з точки зору сучасної теорії. Коротко наведемо основні відомості про те, як можна оцінити надійність тесту з використанням сучасних моделей тестування.

Нагадаємо, що якщо невелика зміна умов тестування і стану випробовуваних призводить до несуттєвої зміни кінцевих результатів, то відповідна характеристика тесту називається його **надійністю**.

Надійність тесту з точки зору статистики є величиною, що характеризує, *ефективність оцінок тесту*.

Для отримання даних для оцінки надійності існують такі методи:

- 1-й метод – тестування двома паралельними тестами;
- 2-й метод – повторне тестування одним тестом;
- 3-й метод – розщеплювання.

Метод 1 потребує доведення паралельності тестів, методи 1 та 2 складно реалізуються на практиці, метод 3 потребує дослідження на паралельність частин тесту після розщеплення.

Дисперсійна методика полягає у виділенні «факторної дисперсії», що породжується в даному випадку впливом завдань та учасників тестування та «залишкової дисперсії», обумовленої випадковими причинами. При цьому коефіцієнт надійності визначає долю дисперсії «істинного» балу в загальній дисперсії $r = \frac{D\{b^*\}}{D\{b\}} = \frac{D\{b\} - D\{\Delta\}}{D\{b\}}$, тут b^* - «істинний бал»; випадкові величини b - результат тестування, Δ - випадкова помилка, $b^* = b + \Delta$.

Кореляційна методика полягає у вимірюванні коефіцієнту кореляції між результатами двократного тестування (або розщеплення). У цьому випадку коефіцієнтом надійності виступає коефіцієнт кореляції.

5.8.2. Дисперсійний аналіз тесту, що містить декілька розділів

Розглянемо проблему оцінки надійності тесту, який складається з k розділів, кожний з яких оцінюється окремо. Нехай цей тест виконували n осіб, результати виконання зібрані у таблицю

$j \backslash i$	1	...	k	Σ
1	b_{11}	...	b_{1k}	$\sum_{j=1}^k b_{1j}$
2	b_{21}	...	b_{2k}	$\sum_{j=1}^k b_{2j}$
...
N	b_{n1}	...	b_{nk}	$\sum_{j=1}^k b_{nj}$
Σ	$\sum_{i=1}^n b_{i1}$...	$\sum_{i=1}^n b_{ik}$	$\sum_{j=1}^k \sum_{i=1}^n b_{ij}$

Перш за все потрібно обробити кожний розділ тесту, визначити стійкі оцінки рівня підготовки учасників тестування, розмістити всі результати на єдиній шкалі.

$j \backslash i$	1	...	k	Σ
1	θ_{11}	...	θ_{1k}	$\sum_{j=1}^k \theta_{1j}$
2	θ_{21}	...	θ_{2k}	$\sum_{j=1}^k \theta_{2j}$
...
n	θ_{n1}	...	θ_{nk}	$\sum_{j=1}^k \theta_{nj}$
Σ	$\sum_{i=1}^n \theta_{i1}$...	$\sum_{i=1}^n \theta_{ik}$	$\sum_{j=1}^k \sum_{i=1}^n \theta_{ij}$

Обчислимо величини

$$S_{\text{загал}} = \sum_{i=1}^n \sum_{j=1}^k \theta_{ij}^2 - \frac{\theta^2}{nk},$$

$$S_{\text{учасн}} = \frac{1}{k} \sum_{i=1}^n \theta_i^2 - \frac{\theta^2}{nk},$$

$$S_{\text{розд}} = \frac{1}{n} \sum_{j=1}^k \theta_j^2 - \frac{\theta^2}{nk},$$

$$S_{\text{ост}} = S_{\text{загал}} - S_{\text{розд}} - S_{\text{учасн}},$$

де $\theta_i = \sum_{j=1}^k \theta_{ij}$, $\theta_j = \sum_{i=1}^n \theta_{ij}$, $\theta = \sum_{i=1}^n \theta_i = \sum_{j=1}^k \theta_j$.

Оцінимо дисперсію, яка характеризує розсіювання між учасниками тестування $D_{\text{учасн}} = \frac{S_{\text{учасн}}}{n-1}$, дисперсію, яка характеризує розсіювання між

розділами тесту $D_{\text{розд}} = \frac{S_{\text{розд}}}{k-1}$ та дисперсію, яка виникає за рахунок

внутрішнього розсіювання $D_{\text{ост}} = \frac{S_{\text{ост}}}{(n-1)(k-1)}$.

Значення $\sqrt{D_{\text{ост}}}$ дає середню квадратичну помилку вимірів.

- Якщо усі завдання *однакові* для усіх учасників, то дисперсія помилки $D_e\{b^*\} = D_{\text{учасн}} - D_{\text{ост}}$,

$$\text{коефіцієнт надійності } \hat{r} = \frac{D_{\text{учасн}} - D_{\text{ост}}}{D_{\text{учасн}}}.$$

- Якщо учасникам пропонувалися *різні* завдання, то

$$\text{дисперсія помилки } \hat{D}_e = \frac{S_{\text{загал}} - S_{\text{учасн}}}{n(k-1)};$$

$$\text{середня квадратична помилка вимірів } \hat{\sigma}\{\Delta\} = \sqrt{\hat{D}_e},$$

$$\text{коефіцієнт надійності } \hat{r} = \frac{D_{\text{учасн}} - \hat{D}_e}{D_{\text{учасн}}}.$$

5.8.3. Визначення коефіцієнта надійності дисперсійним аналізом

Середнє значення коефіцієнта надійності при всляких розщеплюваннях тесту на дві половини може бути обчислене формулою Кьюдера-Ричардсона. Наведемо її у такій формі

$$r = \frac{k}{k-1} \cdot \frac{D\{b\} - \sum_{j=1}^k \rho_j \gamma_j}{D\{b\}},$$

де ρ_j та γ_j - частки учасників тестування, які можуть дати вірну відповідь та невірну відповідь на j -те завдання тесту. При цьому само розщеплювання практично не потрібно.

За реальними даними тестування після знаходження незміщених ефективних оцінок латентних параметрів θ'_i та β'_j у рамках однопараметричної моделі коефіцієнт надійності можна знайти так

$$r = \frac{k}{k-1} \cdot \frac{\hat{D}\{b\} - \sum_{j=1}^k \hat{\rho}_j \hat{\gamma}_j}{\hat{D}\{b\}},$$

де $\hat{\rho}_j = \sum_{i=1}^n \hat{p}_{ij}$, $\hat{\gamma}_j = \sum_{i=1}^n (1 - \hat{p}_{ij})$, $\hat{p}_{ij} = \frac{1}{1 + \exp[-(\theta'_i + \beta'_j)]}$

Важливо пам'ятати, що формула Кюдера-Ричардсона застосовна, якщо елементи a_{ij} матриці відповідей можуть набувати тільки двох значень 0 та 1.

Існує узагальнення формули Кюдера-Ричардсона у вигляді так званого коефіцієнта альфа Кронбаха, яка наведена у попередньому розділі. Обчислення коефіцієнта альфа у рамках параметричних моделей проводиться аналогічно до формули Кюдера-Ричардсона.

Зауважимо, що методи підвищення надійності, представлені у класичній теорії можуть застосовуватись на основі результатів параметричного моделювання. Можна, зокрема, збільшити кількість завдань, при цьому кратність збільшення обчислюється $K = \frac{r_{tk}(1-r_t)}{r_t(1-r_{tk})}$, де r_t – коефіцієнт надійності до зміни довжини тесту; r_{tk} – коефіцієнт надійності після зміни. Очевидно, що коефіцієнти надійності у цьому випадку обчислюються на основі незміщених ефективних оцінок латентних параметрів θ'_i та β'_j у рамках відповідної параметричної моделі.

5.8.4. Кореляційна методика оцінювання надійності. Ретестова надійність

Для встановлення степеню зв'язку між варіантами тесту або двох частин розщепленого тесту аналізується зв'язок між оцінками латентних параметрів, отриманих у рамках однієї з параметричних моделей, для кожного з варіантів тесту або для двох частин тесту. Позначимо оцінки латентних параметрів першого варіанту $\theta_i^{(1)}$, $\beta_j^{(1)}$ та другого - $\theta_i^{(2)}$, $\beta_j^{(2)}$. Будемо вважати варіанти паралельними (доведення цього факту, взагалі кажучи, потребує окремого дослідження). У такому випадку коефіцієнтом надійності тесту може виступати коефіцієнт кореляції між $\theta_i^{(1)}$ та $\theta_i^{(2)}$

$$r = \frac{n \sum_{i=1}^n \theta_i^{(1)} \theta_i^{(2)} - \sum_{i=1}^n \theta_i^{(1)} \sum_{i=1}^n \theta_i^{(2)}}{\sqrt{\left[n \sum_{i=1}^n (\theta_i^{(1)})^2 - \left(\sum_{i=1}^n \theta_i^{(1)} \right)^2 \right] \left[n \sum_{i=1}^n (\theta_i^{(2)})^2 - \left(\sum_{i=1}^n \theta_i^{(2)} \right)^2 \right]}}$$

Для перевірки гіпотези про відсутність кореляції між даними двох тестів у припущенні нормального розподілу використовують критерій Стюдента.

Коефіцієнт надійності $r = \frac{2r'}{1+r'}$ повного тесту пов'язаний з коефіцієнтом r' надійності половини тесту в результаті його розщеплення. Ця формула є частковим випадком формули Спірмена – Брауна $r = \frac{\lambda r'}{1+(\lambda-1)r'}$, де λ - відношення числа завдань тесту до числа завдань після розщеплення.

Ретестова надійність тесту – це ступень збереження рангових позицій випробовуваних, визначається на порядковій і номінальній шкалах. Початковими даними слугують результати повторного тестування випробовуваних за допомогою одного і того ж тесту. Кількісною мірою є коефіцієнти рангової кореляції Спірмена або Кендалла.

Коефіцієнт рангової кореляції Спірмена $\hat{r} = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^n d_i^2$ є коефіцієнтом лінійної кореляції Пірсона між рангами оцінок рівня підготовленості учасників тестування $r\{\theta_i^{(1)}\}$ та $r\{\theta_i^{(2)}\}$ за результатами паралельних тестів або розщеплення тесту.

Рангова кореляція оцінюється також коефіцієнтом Кендалла τ , для чого треба розмістити учасників тестування у порядку зростання рангів $r\{\theta_i^{(1)}\}$ та підрахувати в відповідному стовпці рангів $r\{\theta_i^{(2)}\}$ кількість R_i рангів, що перевищують $r\{\theta_i^{(1)}\}$ для кожного учасника ($i = 1..n$) та знайти суму $R = \sum_{i=1}^{n-1} R_i$. Вибірковий коефіцієнт рангової кореляції Кендалла і його дисперсія знаходяться по формулах

$$\hat{\tau} = \frac{4R}{n(n-1)} - 1, \quad D\{\hat{\tau}\} = \frac{2(2n+5)}{9n(n-1)}.$$

Обчисливши вибірковий коефіцієнт рангової кореляції Кендалла $\hat{\tau}$, можна виявити степінь зв'язку між результатами, для яких обчислений цей коефіцієнт: між результатами повторного тестування, або між результатами тестування паралельними тестами, або між половинами одного тесту.

Сформулюємо основну гіпотезу $H_0: \tau = 0$ про рівність нулю генерального коефіцієнта рангової кореляції Кендалла τ при фіксованому рівні значущості α та протилежну гіпотезу $H_1: \tau \neq 0$.

Критичне значення $\tau_{\text{крит.}} = U_{\text{крит.}} \sqrt{\frac{2(2n+5)}{9n(n-1)}}$, де $U_{\text{крит.}}$ визначається умовою $\Phi(U_{\text{крит.}}) = \frac{1-\alpha}{2}$, функція Лапласа $\Phi(u)$; n - об'єм вибірки. Якщо $|\hat{\tau}| < \tau_{\text{крит.}}$, то немає підстав для відмови від нульової гіпотези, і рангова

кореляція визнається незначущою, а тест недостатньо надійним. Якщо $|\hat{r}| > r_{\text{крит.}}$, то зв'язок слід вважати значущим, а тест достатньо надійним.

Аналогічно можна перевірити інші гіпотези у сфері вимірювань в освіті: про зв'язок між результатами різних тестів, про зв'язок між рівнем складності тестів, про зв'язок між рівнями підготовки різних груп учасників тестування, тощо.

5.9. Валідність тесту та її визначення сучасною теорією

Властивість валідності тесту не втрачає своєї важливості у сучасній теорії освітніх вимірювань. Нагадаємо основні відомості про цю базову характеристику тесту.

Валідністю називають здатність самій генеральної сукупності завдань оцінити саме те, в чому зацікавлені організатори тестування, наприклад, рівень підготовленості випробовуваних в конкретно певній області знань. Висока валідність тесту означає, що тест вільний від суб'єктивних чинників і неупереджено вимірює саме те, для чого він призначений.

Валідність за змістом означає, що завдання тесту повністю охоплюють - і притому в потрібній пропорції - усі основні аспекти тієї області знання, підготовленість в якій цей тест оцінює.

Тому основну роль в оцінці валідності грають усереднені оцінки незалежних досвідчених експертів. Особливо важливою валідність за змістом є для абсолютного тестування.

Валідність критеріальна є показником того, наскільки за результатами тестування можна судити про рівень підготовленості випробовуваних в сьогоденні і їх можливостях продовжувати навчання в майбутньому.

При цьому судження робиться на основі зіставлення результатів тестування з незалежним і об'єктивним критерієм. Наприклад, результати тестування абітурієнтів та показники успішності цих учасників тестування в ролі студентів.

Кількісною мірою валідності зазвичай служить коефіцієнт кореляції між показниками тесту і критерійною мірою, аналогічно судять про валідності окремих завдань.

Валідність з точки зору статистики слід розглядати як єдине поняття, що характеризує міру незміщеності оцінок того вимірювального інструменту, яким є тест.

Мірою валідності окремого завдання тесту використовують коефіцієнт кореляції. Позначимо Θ_i – значення оцінок рівнів підготовленості учасників тестування за їх критеріальним результатом, які виконали і не виконали досліджуване j -те завдання; \hat{p}_{ij} – оцінки ймовірностей вірної та невірної відповіді i -го учасника на досліджуване j -те завдання тесту

$$R_j = \frac{n \sum_{i=1}^n \Theta_i \hat{p}_{ij} - \sum_{i=1}^n \Theta_i \sum_{i=1}^n \hat{p}_{ij}}{\sqrt{\left[n \sum_{i=1}^n (\Theta_i)^2 - \left(\sum_{i=1}^n \Theta_i \right)^2 \right] \left[n \sum_{i=1}^n (\hat{p}_{ij})^2 - \left(\sum_{i=1}^n \hat{p}_{ij} \right)^2 \right]}}$$

Аналогічно до міри валідності окремого тестового завдання визначається **міра валідності усього тесту**, яку визначають за ступенем зв'язку між виявленими цим тестом оцінками рівнів підготовки учасників тестування Θ_i та відповідними значеннями Θ_i – оцінками рівнів підготовленості учасників тестування за їх критеріальним результатом.

Для перевірки гіпотези про відсутність зв'язку між такими даними у припущенні нормального розподілу використовують критерій Стюдента.

Зауважимо, що у значній кількості сучасних досліджень міра надійності визначається через коефіцієнт бісеріальної кореляції, якщо відповіді на окреме тестове завдання розглядають як реалізацію дихотомічної випадкової змінної. Якщо результати досліджуваного тесту та критеріального тесту є дихотомічними, то мірою надійності може виступати точково-бісеріальний та дихотомний коефіцієнти кореляції. Такий підхід не передбачає застосування апарата параметричного моделювання, тому він віднесений до класичної теорії та описаний у попередньому розділі.

5.10. Порядкові шкали у педагогічному тестуванні

1. **Відсоткові шкали:** виставковий бал B прямо пропорційний первинному балу. Остаточний бал учасника прирівнюється первинному балу b , тобто кількості вірно виконаних завдань. Якщо відношення b/k виразити у відсотках, то утвориться 100- бальна шкала. Аналогічно, складність кожного завдання тесту характеризується відсотковим відношенням $100c/n_1$, де c – первинний бал завдання.

2. Процентильна шкала (шкала первинних процентилей) – це шкала побудована на основі перерахування первинних балів відповідно до розподілу їх частот (часток). Методика перерахування складається з чотирьох етапів:

- первинні бали ранжують від вищого до нижчого;
- кількість учасників тестування, які отримали кожен бал, вносять у графу частот (якщо за кількістю) або часток (якщо за відсотком учасників тестування у загальному підсумку);
- усі частки в напрямку збільшення значень ознаки послідовно підсумовують, тобто визначають кумулятивні частки;
- визначають процентиль, який відповідає значенню кожного первинного бала.

Пентилі можуть бути використані як оцінка остаточного балу B за 5-бальною шкалою, децилі - за 10-бальною шкалою, центилі - за 100-бальною шкалою і так далі. У такому разі виставковий бал B_i певного учасника тестування з номером i виражає процентну частку випробовуваних, первинний бал в яких нижчий за первинний бал b_i даного індивідуума:

$$B_i = 100p_i, p_i = p(b < b_i), \quad B_i = 100p_i, p_i = p(b < b_i)$$

3. Нормалізовані шкали типу $N(\mu; \sigma^2)$, де $\mu \geq 0$ і $\sigma > 0$ - параметри шкали: емпіричний розподіл частот первинних балів нормалізується, тобто перетворюється (лінійно або нелінійно) до стандартного модельного нормального розподілу $N(\mu; \sigma^2)$, і виставковий бал визначається відхиленням від математичного сподівання в одиницях стандарту. Основоположною є стандартна модель з параметрами $\mu = 0, \sigma = 1$.

Потрібно перетворити первинні бали b до стандартної змінної $z \in N(0; 1)$, визначальною S -бальною шкалу B з цілими індексами $1, 2, \dots, S-1$.

Крок сітки визначимо величиною $\Delta_z = \frac{6}{S-2}$,

Нормалізовані бали $B = 1, 2, \dots, S-1$ визначаються так: $z_B = (B - \bar{B})\Delta_z$, де \bar{B} - математичне сподівання нормалізованих балів.

Імовірність попадання випадкової величини z в окіл

$$(z_B - \Delta_z/2, z_B + \Delta_z/2)$$

точки z_B визначається величиною $f(z_B)\Delta_z$, де $f(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ - щільність нормального розподілу.

Знаходимо такі пари значень початкової змінної b і стандартної змінної z (або нормалізованої змінної B , що те ж саме), які розміщені по один бік від свого математичного сподівання та відповідають рівності

$$\frac{n_b}{n} \approx f(z_b) \Delta_z.$$

4. **Шкали нормалізованих процентилів.** Індексами шкал іншого типу виступають проценти нормалізованої стандартної моделі $N(0; 1)$. Відмітимо, що зараз мова йде про проценти (квантили) модельного розподілу $N(0; 1)$ (модельні проценти) на відміну від процентилей (квантилів) початкового емпіричного розподілу (первинних процентилей).

Значення центилей z , відповідні балам (процентилям) 51, 52, ..., 99 додатні і симетричні по модулю від'ємним центилям балів 49, 48... 1, відповідно. Центром симетрії служить нульове значення центили, відповідне п'ятдесятому процентилю і балу $B = 50$.

5. **Нормалізовані шкали з постійним кроком:** індекси шкали виставкового балу відповідають рівновіддаленим значенням стандартної змінної $z \in N(0; 1)$, які конструюються так, щоб індекси шкали відповідали рівновіддаленим значенням стандартної змінної $z \in N(0; 1)$.

Відзначимо, що все вищесказане можна перефразувати і стосовно оцінювання рівня складності кожного завдання тесту на тій або іншій порядковій шкалі.

Можливість ранкувати завдання тесту за рівнем їх складності дає можливість врахувати це при обчисленнях виставкового балу. І навпаки: для визначення складності завдань природно спробувати врахувати рівень підготовленості тих учасників тестування, які виконували ці завдання. Наведемо ітераційну процедуру для обчислення виставкових балів тестових завдань та остаточних балів учасників тестування

$$C_{jh} = D \frac{\sum_{i=0}^n (1 - a_{ij}) B_{i,h-1}}{\sum_{i=0}^n B_{i,h-1}} = D - D \frac{\sum_{i=0}^n a_{ij} B_{i,h-1}}{\sum_{i=0}^n B_{i,h-1}};$$

$$B_{ih} = D \frac{\sum_{j=0}^n a_{ij} C_{jh}}{\sum_{j=0}^n C_{jh}}.$$

Тут n - кількість учасників тестування; $i = 1, 2, \dots, n$ - номер учасника; k - кількість завдань в тесті; $j = 1, 2, \dots, k$ - номер завдання; $h = 1, 2, \dots$ - номер ітерації; B_{ih} - остаточний бал i -го учасника в h -ій ітерації; $B_{i0} = b_i$ - первинний бал i -го учасника; C_{jh} - виставковий бал j -го завдання в h -ій ітерації за тією ж шкалою, по якій фіксується B_{ih} , $C_{j0} = c_j$ - первинний бал j го завдання; a_{ij} - елемент матриці відповідей; D - довжина шкали.

5.11. Побудова єдиної шкали

5.11.1. Побудова єдиної метричної шкали.

Метричні шкали, які визначаються різними варіантами тесту, не зв'язаними між собою, але відрізняються один від одного тільки на константу. Кожен варіант тесту повинен мати не менше трьох *спільних* для всіх тестів завдань (*вузлові, якірні завдання*), причому складність першого $\delta_{jl}^{(1)}$ знаходиться на лівій частині метричної шкали, складність другого $\delta_{jl}^{(2)}$ - приблизно в середині шкали, і складність третього $\delta_{jl}^{(3)}$ - на правій частині.

Для перенесення латентних параметрів на єдину метричну шкалу:

1) За допомогою критеріїв згоди слід перевірити статистичні гіпотези про узгодженість моделі Раша з наявними емпіричними даними і про паралельність варіантів тесту.

2) З оцінок латентних параметрів θ_{il} , β , отриманих з обробки результатів тестування по варіанту з номером l , віднімається значення $\beta_{jl}^{(2)}$, задається умовний початок метричної шкали кожного варіанту $\beta_{jl}^{(2)} = 0$.

3) Складності першого і третього вузлових завдань, отримані по різних варіантах, усереднюються з урахуванням відповідної точності, тобто обчислюються середні вагові значення

$$\bar{\beta}^{(1)} = \frac{\sum_{l=1}^m W_{jl}^{(1)} \delta_{jl}^{(1)}}{\sum_{l=1}^m W_{jl}^{(1)}}, \quad \bar{\beta}^{(3)} = \frac{\sum_{l=1}^m W_{jl}^{(3)} \beta_{jl}^{(3)}}{\sum_{l=1}^m W_{jl}^{(3)}},$$

тут $W_{jl}^{(1)} = \frac{C}{\sigma^2\{\beta_{jl}^{(1)}\}}$ та $W_{jl}^{(3)} = \frac{C}{\sigma^2\{\beta_{jl}^{(3)}\}}$ - ваги оцінок $\beta_{jl}^{(1)}$ і $\beta_{jl}^{(3)}$, відповідно; j - номер вузлових завдань у варіантах тесту (фіксований), l - номер варіанту (мінюється від 1 до m), C - довільна константа.

4) Для кожного варіанту з номером l обчислюються відхилення

$$V_l^{(1)} = \beta_{jl}^{(1)} - \bar{\beta}^{(1)} \quad \text{та} \quad V_l^{(3)} = \beta_{jl}^{(3)} - \bar{\beta}^{(3)}.$$

		Вузол №1		Вузол №3		
№		Вага W	Відхилення V	δ	Вага W	Відхилення V
1	$\beta_{j1}^{(1)}$	$W_{j1}^{(1)}$	$V_1^{(1)} = \beta_{j1}^{(1)} - \bar{\beta}^{(1)}$	$\beta_{j1}^{(3)}$	$W_{j1}^{(3)}$	$V_1^{(3)} = \beta_{j1}^{(3)} - \bar{\beta}^{(3)}$
2	$\beta_{j2}^{(1)}$	$W_{j2}^{(1)}$	$V_2^{(1)} = \beta_{j2}^{(1)} - \bar{\beta}^{(1)}$	$\beta_{j2}^{(3)}$	$W_{j2}^{(3)}$	$V_2^{(3)} = \beta_{j2}^{(3)} - \bar{\beta}^{(3)}$
...
m	$\beta_{jm}^{(1)}$	$W_{jm}^{(1)}$	$V_m^{(1)} = \beta_{jm}^{(1)} - \bar{\beta}^{(1)}$	$\beta_{jm}^{(3)}$	$W_{jm}^{(3)}$	$V_m^{(3)} = \beta_{jm}^{(3)} - \bar{\beta}^{(3)}$
	$\bar{\beta}^{(1)}$	$\sum_{l=1}^m W_{jl}^{(1)}$	$D \{ \beta_j^{(1)} \}$	$\bar{\beta}^{(3)}$	$\sum_{l=1}^m W_{jl}^{(3)}$	$D \{ \beta_j^{(3)} \}$

5) Рекомендуються наступні правила для обчислення виправлених значень латентних параметрів β_{jl}^* і θ_i^* , віднесених до єдиної метричної шкали в логітах з умовним початком відліку:

$$\text{якщо } \beta_{jl} < 0, \text{ то } \beta_{jl}^* = \beta_{jl} + \frac{V_l^{(1)}}{\beta_{jl}^{(1)}} \beta_{jl} = \frac{\bar{\beta}^{(1)}}{\beta_{jl}^{(1)}} \beta_{jl};$$

$$\text{якщо } \beta_{jl} \geq 0, \text{ то } \beta_{jl}^* = \beta_{jl} + \frac{V_l^{(3)}}{\beta_{jl}^{(3)}} \beta_{jl} = \frac{\bar{\beta}^{(3)}}{\beta_{jl}^{(3)}} \beta_{jl};$$

$$\text{якщо } \theta_{il} < 0, \text{ то } \theta_i^* = \theta_{il} + \frac{V_l^{(1)}}{\theta_{il}^{(1)}} \theta_{il} = \frac{\bar{\theta}^{(1)}}{\theta_{il}^{(1)}} \theta_{il};$$

$$\text{якщо } \theta_{il} \geq 0, \text{ то } \theta_i^* = \theta_{il} + \frac{V_l^{(3)}}{\theta_{il}^{(3)}} \theta_{il} = \frac{\bar{\theta}^{(3)}}{\theta_{il}^{(3)}} \theta_{il}.$$

$\bar{\beta}^{(1)}$ та $\bar{\beta}^{(3)}$ - складності вузлових завдань, усереднені по варіантах, відношення $\bar{\beta}^{(1)} / \beta_{jl}^{(1)}$ та $\bar{\beta}^{(3)} / \beta_{jl}^{(3)}$ є константами для кожного варіанту з номером l .

5.11.2. Побудова єдиної нормованої шкали

Обрання нуля на отриманій єдиній метричній шкалі рекомендується здійснювати на основі нормативної вибірки учасників тестування. Всі значення підготовленості θ_i^* випробовуваних, таких, що потрапили в

нормативну вибірку, усереднюються з урахуванням їх середніх квадратичних помилок. Отриманого середньозваженого значення позначимо $\bar{\theta}_{\text{н.в.}}$. Підраховується також відповідна дисперсія підготовленості учасників з нормативної вибірки відносно $\bar{\theta}_{\text{н.в.}}$. Далі оцінки всіх латентних параметрів зміщуються на величину $\bar{\theta}_{\text{н.в.}}$.

Значення $\theta_i^* - \bar{\theta}_{\text{н.в.}}$ і $\beta_i^* - \bar{\theta}_{\text{н.в.}}$ є остаточними оцінками латентних параметрів складності завдань і підготовленості учасників на **єдиній нормованій шкалі**. Початок цієї шкали закріплюється нормативною вибіркою.

Формула для перерахунку на 100 - бальну шкалу

$$B_i = 50 + \frac{50}{t\sigma} (\theta_i^* - \bar{\theta}_{\text{н.в.}}),$$

тут σ - середнє квадратичне відхилення підготовки випробовуваних нормативної вибірки відносно $\bar{\theta}_{\text{н.в.}}$, t - коефіцієнт, значення якого вибирається так, щоб можливості отримати вищі бали були регламентовані.

Якщо нормативна вибірка відсутня, то

$$B_i = 50 + \frac{50}{\theta_{\text{max}}^* - \bar{\theta}_{\text{н.в.}}} (\theta_i^* - \bar{\theta}_{\text{н.в.}}),$$

тут $\bar{\theta}_{\text{н.в.}}$ - середній рівень підготовленості учасників тестування в оброблюваній вибірці, а θ_{max}^* позначає константу в логітах, яка вибирається так, щоб регламентувати можливості отримати вищі результати.

ПРАКТИЧНІ ЗАВДАННЯ

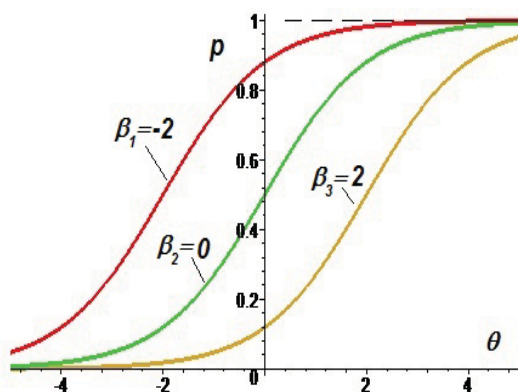
Приклад 1. Побудувати характеристичні криві для складності тестових завдань (ICC) з рівнями складності: (-2) логіта, 0 логітів, +2 логіта. Побудувати інформаційні функції цих завдань та інформаційну функцію системи трьох завдань.

Розв'язання: Формула для побудови графіків характеристичних кривих для складності тестових завдань (ICC) $p = \frac{1}{1 + \exp(\theta - \beta_j)}$.

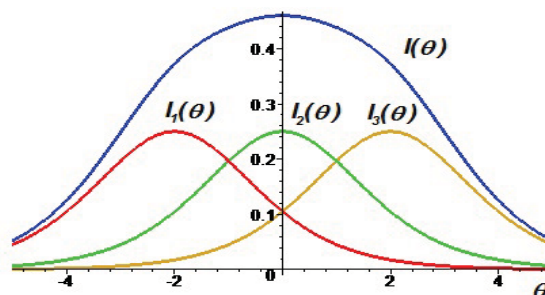
Підставляємо значення $\beta_j = -2, 0, 2$ та отримуємо відповідні ICC

Формула, для побудови графіків інформаційних функцій тестових завдань з рівнем складності β_j за моделлю Раша має вигляд $I_j = \frac{\exp(\theta - \beta_j)}{(1 + \exp 1,7(\theta - \beta_j))^2}$.

Інформаційна функція системи трьох завдань $I(\theta) = I_1 + I_2 + I_3$.



Характеристичні криві тестових завдань



Графіки інформаційних функцій завдань та всього тесту

Приклад 2. Побудувати характеристичні криві для досягнень учасників тестування (PCC) з рівнями підготовки: (-1) логіт, 0 логітів, +1 логіт. Побудувати інформаційні функції для цих учасників та інформаційну функцію групи учасників.

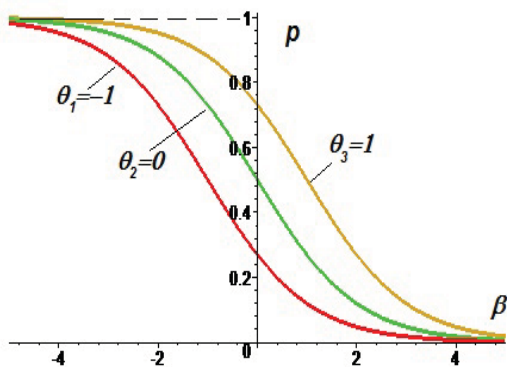
Розв'язання: Формула для побудови графіків характеристичних кривих досягнень учасників тестування (РСС)

$$p = \frac{1}{1 + \exp(\theta_0 - \beta)}$$

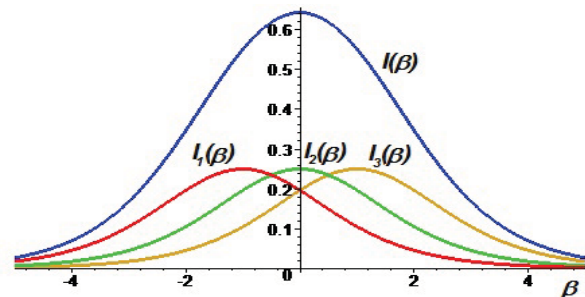
Підставляємо значення $\theta_0 = -1, 0, 1$ та отримуємо відповідні РСС.

Формула, для побудови графіків інформаційних функцій учасників з рівнем підготовки θ_i за моделлю Раша має вигляд $I_i = \frac{\exp(\theta_i - \beta)}{(1 + \exp 1,7(\theta_i - \beta))^2}$.

Інформаційна функція групи учасників $I(\theta) = I_1 + I_2 + I_3$.



Характеристичні криві учасників тестування



Графіки інформаційних функцій учасників тестування та всієї групи

Приклад 3. Учасник тестування підготував відповіді на 100 завдань із 250 завдань бази тестових завдань. Тест містить 25 завдань, ціна одного тестового завдання дорівнює 1 балу. Знайти ймовірність того, що цей учасник буде віднесений до групи задовільно підготовлених учасників, тобто таких, що отримають первинний бал від 10 до 15.

Розв'язання: Імовірність попадання реального первинного балу b_i i -го учасника у заданий проміжок $[m_1; m_2]$, якщо у тесті міститься k тестових завдань, наближено обчислюється за допомогою інтегральної теореми Лапласа

$$P_{ki}(m_1 \leq b_i \leq m_2) \approx \Phi(x_2) - \Phi(x_1), \text{ де}$$

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{u^2}{2}} du - \text{функція Лапласа}; \quad x_1 = \frac{m_1 - kp_i}{\sqrt{kp_i q_i}}, \quad x_2 = \frac{m_2 - kp_i}{\sqrt{kp_i q_i}}.$$

Тут $m_1 = 10$, $m_2 = 15$, $k = 25$.

Оцінка імовірності успіху $\tilde{p}_i = \frac{100}{250} = 0,4$ та невдачі $\tilde{q}_i = 0,6$.

$$\text{Обчислимо } x_1 = \frac{10 - 25 * 0,4}{\sqrt{25 * 0,4 * 0,6}} = 0, \quad x_2 = \frac{15 - 25 * 0,4}{\sqrt{25 * 0,4 * 0,6}} = 2,0412$$

$$P_{25i}(10 \leq b_i \leq 15) \approx \Phi(2,04) - \Phi(0) = 0,4793$$

Відповідь: Імовірність того, що учасник, який може дати правильні відповіді на 100 питань із 250-ти тестових завдань, буде віднесений до групи задовільно підготовлених учасників (отримає первинний бал від 10 до 15) дорівнює 0,4793.

Приклад 4. Правильну відповідь на тестове завдання знають 60% випускників. Тестування проходили 1000 осіб. Знайти ймовірність того, що це тестове завдання не зможуть розв'язати від 370 до 410.

Розв'язання: Імовірність попадання реального первинного балу c_j j -го тестового завдання у заданий проміжок $[n_1; n_2]$, якщо у тестуванні брали участь n учасників, наближено обчислюється за допомогою інтегральної теореми Лапласа

$$P_{kj}(n_1 \leq c_j \leq n_2) \approx \Phi(y_2) - \Phi(y_1), \text{ де}$$

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{u^2}{2}} du - \text{функція Лапласа}; \quad y_1 = \frac{n_1 - n\rho_j}{\sqrt{n\rho_j\gamma_j}}, \quad y_2 = \frac{n_2 - n\rho_j}{\sqrt{n\rho_j\gamma_j}}.$$

Тут $n_1 = 1000 - 200 = 800$, $n_2 = 1000$, $n = 1000$.

Оскільки за умовою правильну відповідь на тестове завдання знають 60% випускників, то оцінка імовірності успіху для тестового завдання це імовірність невдачі учасників тестування, тобто $\tilde{\rho}_j = 0,4$ та невдачі $\tilde{\gamma}_j = 0,6$.

$$\text{Обчислимо } y_1 = \frac{370 - 1000 * 0,4}{\sqrt{1000 * 0,6 * 0,4}} = -1,94, \quad y_2 = \frac{410 - 1000 * 0,4}{\sqrt{1000 * 0,6 * 0,4}} = 0,4738$$

$$P_{1000j}(370 \leq c_j \leq 410) \approx \Phi(0,65) - \Phi(-1,94) \approx 0,2422 + 0,4738 = 0,7260$$

Відповідь: Ймовірність того, що тестове завдання, яке знають 60% випускників, не зможуть розв'язати від 370 до 410 осіб із 1000 учасників тестування, дорівнює 0,7260.

Приклад 5. Учасник тестування набрав $b = 10$ балів при виконанні тесту, що складається з $k = 25$ завдань (ціна одного завдання дорівнює 1 балу). Знайти імовірність того, що реальне значення бала цього учасника не відхиляється від значення, отриманого при тестуванні більше ніж на один бал.

Розв'язання: Побудуємо оцінку імовірності розв'язання одного завдання із 25-ти i -м учасником тестування $\tilde{p}_i = \frac{10}{25} = 0,4$.

Імовірність того, що реальне значення бала цього учасника не відхиляється від значення, отриманого при тестуванні більше ніж на один бал дорівнює

$$\text{імовірності } P_{ki}\{|b_i - M(b_i)| \leq \varepsilon\} = 2\Phi\left(\frac{\varepsilon}{\sqrt{D(b_i)}}\right), \quad \text{де } M(b_i) = 10, \varepsilon = 1,$$

$$D(b_i) = k\tilde{p}_i\tilde{q}_i = 25 * 0,4 * 0,6 = 6, \quad \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{u^2}{2}} du - \text{функція Лапласа.}$$

$$\text{Отже, } P_{25i}\{|b_i - 10| \leq 1\} = 2\Phi\left(\frac{1}{\sqrt{6}}\right), \text{ або}$$

$$P_{25i}\{|b_i - 10| \leq 1\} = 2\Phi(0,41) = 2 * 0,1591 = 0,3182$$

Відповідь: Імовірність того, що реальне значення бала учасника, який набрав 10 балів з 25 можливих балів, не відхиляється від значення 10 більше ніж на один бал, дорівнює 0,3182.

Приклад 6. Вірну відповідь на j -те тестове завдання дали 900 абітурієнтів із 3000 учасників ЗНО. Визначити ймовірність того, що реальна кількість учасників, які не знали відповідь на це завдання, відрізняється від 900 не більше за 50 осіб.

Розв'язання: Побудуємо оцінку ймовірності не розв'язання j -го тестового завдання (успіх) 900 особами із 3000 учасників тестування $\tilde{p}_j = \frac{900}{3000} = 0,3$ та ймовірність його розв'язання (невдача) $\tilde{\gamma}_j = 0,7$.

Ймовірність того, що реальне кількість учасників тестування, які не знали правильну відповідь відрізняється від 900 не більше за 50 осіб дорівнює

ймовірності $P_{nj} \{ |c_j - M(c_j)| \leq \varepsilon \} = 2\Phi\left(\frac{\varepsilon}{\sqrt{D(c_j)}}\right)$, де $M(c_j) = 900$, $\varepsilon = 50$,

$D(c_j) = k\tilde{p}_j\tilde{\gamma}_j = 3000 * 0,3 * 0,7 = 630$, $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{u^2}{2}} du$ - функція Лапласа.

Отже, $P_{3000j} \{ |c_j - 900| \leq 50 \} = 2\Phi\left(\frac{1}{\sqrt{630}}\right)$,

або $P_{3000j} \{ |c_j - 900| \leq 50 \} = 2\Phi(0,0398) = 2 * 0,0160 = 0,0320$

Відповідь: Ймовірність того, що реальна кількість учасників ЗНО із 3000, які невірно відповіли на j -те завдання відрізняється від 900 не більше ніж на 50 осіб, дорівнює 0,0320.

Приклад 7. Учасник тестування набрав $b = 10$ балів при виконанні тесту, що складається з $k = 25$ завдань (ціна одного завдання дорівнює 1 балу). Знайти довірчий інтервал з рівнем значущості 0,8 того, що у цьому випадку реальне значення бала учасника дорівнює отриманому при тестуванні значенню первинного балу.

Розв'язання: Довірчий інтервал $(m_1; m_2)$ з рівнем значущості d для реального значення балу b_i i -го учасника можна отримати, врахувавши формулу $P_{ik} \{ m_1 < M(b_i) < m_2 \} = d$, де

$$m_1 = \frac{k}{k+l^2} \left(\tilde{b}_i + \frac{l^2}{2} - l \sqrt{\frac{\tilde{b}_i(k-\tilde{b}_i)}{k} + \frac{l^2}{2}} \right);$$

$$m_2 = \frac{k}{k+l^2} \left(\tilde{b}_i + \frac{l^2}{2} + l \sqrt{\frac{\tilde{b}_i(k-\tilde{b}_i)}{k} + \frac{l^2}{2}} \right);$$

l - значення аргументу функції Лапласа, при якому $2\Phi(l) = d$; d - рівень значущості інтервалу (m_1, m_2) .

У нашому випадку $2\Phi(l) = 0,8$, тому із таблиці функції Лапласа маємо $l = 1,28$. Таким чином, межі довірчого інтервалу мають вигляд

$$m_1 = \frac{25}{25+1,28^2} \left(10 + \frac{1,28^2}{2} - 1,28 \sqrt{\frac{10 \cdot (25-10)}{25} + \frac{1,28^2}{2}} \right) = 7,02;$$

$$m_2 = \frac{25}{25+1,28^2} \left(10 + \frac{1,28^2}{2} + 1,28 \sqrt{\frac{10 \cdot (25-10)}{25} + \frac{1,28^2}{2}} \right) = 13,29.$$

Округливши m_1 з надлишком і m_2 з недостатчею до цілих значень, одержуємо проміжок (8; 13).

Відповідь: Довірчий інтервал з рівнем значущості 0,8 для реального значення балу учасника тестування, який набрав 10 балів при виконанні тесту, що складається з 25 завдань, є інтервал (8; 13) на шкалі первинних балів учасників тестування.

Приклад 8. Тестове завдання було розв'язане 75 особами із 100 учасників тестування. Знайти довірчий інтервал з рівнем значущості 0,9 того, що у цьому випадку реальне кількість учасників тестування, які вірно виконують це завдання, дорівнює 75.

Розв'язання: Довірчий інтервал $(n_1; n_2)$ з рівнем значущості d для реальної кількості учасників, які знають відповідь на завдання, при загальній кількості учасників тестування можна отримати, врахувавши формулу

$$P_{jn} \{n_1 < M(c_j) < n_2\} = d, \text{ де}$$

$$n_1 = \frac{n}{n+l^2} \left(c_j + \frac{l^2}{2} - l \sqrt{\frac{c_j(n-c_j)}{n} + \frac{l^2}{2}} \right);$$

$$n_2 = \frac{n}{n+l^2} \left(c_j + \frac{l^2}{2} + l \sqrt{\frac{c_j(n-c_j)}{n} + \frac{l^2}{2}} \right);$$

l - значення аргументу функції Лапласа, при якому $2\Phi(l) = d$; d - рівень значущості інтервалу (n_1, n_2) .

За умовою $2\Phi(l) = 0,9$, тому із таблиці функції Лапласа маємо $l = 1,65$. Таким чином, межі довірчого інтервалу мають вигляд

$$n_1 = \frac{100}{100+1,65^2} \left(75 + \frac{1,65^2}{2} - 1,65 \sqrt{\frac{75 \cdot (100-75)}{100} + \frac{1,65^2}{2}} \right) = 67,13;$$

$$n_2 = \frac{100}{100+1,65^2} \left(75 + \frac{1,65^2}{2} + 1,65 \sqrt{\frac{75 \cdot (100-75)}{100} + \frac{1,65^2}{2}} \right) = 81,54.$$

Округливши n_1 з надлишком і n_2 з недостатчею до цілих значень, одержуємо проміжок (68; 81).

Відповідь: Довірчий інтервал з рівнем значущості 0,9 для кількості учасників тестування, які розв'язали j -те тестове завдання, в обсязі 75 осіб зі 100 учасників тестування, є інтервал (68; 81) на шкалі первинних балів тестових завдань.

Приклад 9. Тест з української мови містить 10 завдань по кожній з трьох тем. Результати виконання цього тесту десятьма учасників тестування наведені в таблиці. При цьому номер учасника тестування $i = 1, 2, 3$, номер теми завдань $j = 1, 2, \dots, 10$, величина b_{ij} – кількість завдань j -ої теми, вірно виконаних i -м учасником. Розглядаючи наявні бали b_{ij} як реалізації випадкової величини b , виконати дисперсійний аналіз даних, обчислити коефіцієнти надійності у випадку, якщо завдання тем однакові для кожного учасника тестування та у випадку, якщо завдання тем різні для кожного учасника тестування.

	Тема 1	Тема 2	Тема 3
1	5	6	2
2	4	3	3
3	7	9	7
4	5	6	4
5	6	9	7
6	2	3	4
7	4	3	5
8	3	4	5
9	4	5	6
10	5	9	5

Розв'язання: Для проведення дисперсійного аналізу результатів тестування знайдемо суми балів, отриманих кожним учасником за все три розділи $b_{i*} = \sum_{j=1}^3 b_{ij}$, та загальну кількість розв'язаних завдань у кожній темі всіма десятима учасниками $b_{*j} = \sum_{i=1}^{10} b_{ij}$, а також суму всіх балів $b_{**} = \sum_{i=1}^n \sum_{j=1}^k b_{ij}^2 = 150$. Після обчислимо квадрати названих сум $(b_{i*})^2$ та $(b_{*j})^2$, суми цих квадратів $\sum_{i=1}^{10} b_{i*}^2 = 2462$ та $\sum_{j=1}^3 b_{*j}^2 = 7578$. Дані наведено у таблиці 1.

Обчислимо квадрати даних результатів тестування b_{ij}^2 , суми квадратів $\sum_{j=1}^3 b_{ij}^2$ та $\sum_{i=1}^{10} b_{ij}^2$, а також загальну суму квадратів $\sum_{i=1}^n \sum_{j=1}^k b_{ij}^2 = 858$. Дані наведені у таблиці 2.

Таблиця 1.

b_{ij}	Тема 1	Тема 2	Тема 3	b_{i*}	b_{i*}^2
1	5	6	2	13	169
2	4	3	3	10	100
3	7	9	7	23	529
4	5	6	4	15	225
5	6	9	7	22	484
6	2	3	4	9	81
7	4	3	5	12	144
8	3	4	5	12	144
9	4	5	6	15	225
10	5	9	5	19	361
b_{*j}	45	57	48	$b_{**} = \sum_{i=1}^n \sum_{j=1}^k b_{ij} = 150$	$\sum_{i=1}^{10} b_{i*}^2 = 2462$
b_{*j}^2	2025	3249	2304	$\sum_{j=1}^3 b_{*j}^2 = 7578$	

Таблиця 2.

b_{ij}^2	Тема 1	Тема 2	Тема 3	$\sum_{j=1}^3 b_{ij}^2$
1	25	36	4	65
2	16	9	9	34
3	49	81	49	179
4	25	36	16	77
5	36	81	49	166
6	4	9	16	29
7	16	9	25	50
8	9	16	25	50
9	16	25	36	77
10	25	81	25	131
$\sum_{i=1}^{10} b_{ij}^2$	221	383	254	$\sum_{i=1}^n \sum_{j=1}^k b_{ij}^2 = 858$

Виділимо "факторну дисперсію", що породжується впливом трьох різних тем завдань та різним рівнем десяти учасників тестування, і "залишкової дисперсії", обумовленої випадковими причинами. Для цього обчислимо загальну суму квадратів відхилень спостережуваних значень балів від загального середнього \bar{b} $S_{\text{загал}} = \sum_{i=1}^n \sum_{j=1}^k b_{ij}^2 - \frac{b_{**}^2}{nk}$; факторну суму квадратів відхилень середніх значень по стовпцях від загального середнього $S_{\text{розд}} = \frac{1}{n} \sum_{j=1}^k b_{*j}^2 - \frac{b_{**}^2}{nk}$, яка характеризує розсіяння між розділами тесту; факторну суму квадратів відхилень середніх значень по рядках від загального середнього $S_{\text{учасн}} = \frac{1}{k} \sum_{i=1}^n b_{i*}^2 - \frac{b_{**}^2}{nk}$, яка характеризує розсіяння між учасниками тестування; а також залишкову суму квадратів $S_{\text{залишк}} = S_{\text{загал}} - S_{\text{розд}} - S_{\text{учасн}}$, яка характеризує внутрішнє розсіяння.

Підставимо значення з таблиць у формули при $n = 10$, $k = 3$;

$$S_{\text{загал}} = 858 - \frac{22500}{30} = 108;$$

$$S_{\text{учасн}} = \frac{1}{3} 2462 - \frac{22500}{30} = 70,67;$$

$$S_{\text{розд}} = \frac{1}{10} 7578 - \frac{22500}{30} = 7,8;$$

$$S_{\text{залишк}} = 108 - 70,67 - 7,8 = 29,53.$$

Тепер оцінимо відповідні дисперсії

$$D_{\text{учасн}} = \frac{S_{\text{учасн}}}{n-1} = \frac{70,67}{9} = 7,85,$$

$$D_{\text{розд}} = \frac{S_{\text{розд}}}{k-1} = \frac{7,8}{2} = 3,9,$$

$$D_{\text{залишк}} = \frac{S_{\text{залишк}}}{(n-1)(k-1)} = \frac{29,53}{18} = 1,64$$

Обчислимо коефіцієнт надійності, за умови, що завдання всіх розділів однакові для кожного учасника тестування і тому не вносять додаткових змін у результат

$$r = \frac{D_{\text{учасн}} - D_{\text{залишк}}}{D_{\text{учасн}}} = \frac{7,85 - 1,64}{7,85} = 0,7910$$

Обчислимо коефіцієнт надійності, за умови, що завдання розділів різні для кожного учасника тестування, в результаті чого вносяться додаткові зміни у результат голосування. У такому випадку дисперсію від фактора різних завдань у розділах віднесемо до залишкової, тоді

$$S_{\text{залишк}}^{(1)} = S_{\text{загал}} - S_{\text{учасн}} = 108 - 70,67 = 37,33$$

$$D_{\text{залишк}}^{(1)} = \frac{S_{\text{залишк}}^{(1)}}{n(k-1)} = \frac{37,33}{20} = 1,87.$$

А коефіцієнт надійності $r^{(1)} = \frac{D_{\text{учасн}} - D_{\text{залишк}}^{(1)}}{D_{\text{учасн}}} = \frac{7,85 - 1,64}{7,85} = 0,7622$.

Отже, надійність тестування зменшується, якщо враховувати фактор різних завдань.

Приклад 10. У завданні закритого типу на вибір пропонувались 5 відповідей, одна з яких є вірною. Тестування проходили 3000 випробовуваних, з яких 2400 осіб дали вірні відповіді на це завдання. В таблиці 1 наведені емпіричні частоти вибору кожного дистрактора.

Частоти	Дистракторы			
	1	2	3	4
Емпірична частота вибору дистрактора π	130	175	120	175

Перевірити гіпотезу про рівномірний розподіл частот вибору дистракторів з рівнем значущості $\alpha = 0,05$.

Розв'язання: відповідно до умови невірних відповідей при тестуванні було $3000 - 1400 = 600$. Обчислимо теоретичні частоти, які в умовах рівномірного розподілу однакові і дорівнюють $600 : 4 = 150$. Знайдемо різниці між емпіричними та теоретичними частотами. Результати наведені у таблиці

Частоти	Дистракторы				Σ
Емпірична частота вибору дистрактора n	130	175	120	175	600
Теоретична частота вибору дистрактора n^*	150	150	150	150	600
$n - n^*$	-20	+25	-30	+25	0

Висунемо основну гіпотезу H_0 про рівномірний розподіл частот вибору дистракторів. Протилежна гіпотеза H_1 про те, що розподіл частот дистракторів не є рівномірним.

У якості міри близькості теоретичних та емпіричних частот зручно використовувати статистику χ^2 -квадрат

$$\chi^2_{\text{спостер}} = \sum_{i=1}^4 \frac{(n_i - n_i^*)^2}{n_i}, \text{ або } \chi^2_{\text{спостер}} = \frac{(-20)^2 + 25^2 + (-30)^2 + 25^2}{150} = 17.$$

Критичне значення критерію, що відповідає трьом ступеням свободи і рівню значущості $\alpha = 0,05$, вибираємо з таблиці статистики χ^2 -квадрат. Воно дорівнює $\chi^2_{\text{крит}} = 7,82$. Оскільки $\chi^2_{\text{спостер}} = 17 > 7,8 = \chi^2_{\text{крит}}$, то гіпотезу про рівноімовірний вибір дистракторів слід відкинути.

Приклад 11. Серед учнів 8-х класів було проведено тестування з математики двома тестами різної складності. Використовуючи результати сімох випробовуваних, які взяли участь у тестуванні двох видів, отримати вибіркоче рівняння регресії, перевірити гіпотезу про наявність кореляції. Перерахувати на єдину метричну шкалу результат $\theta_{81} = 0,7$ учасника тестування першим тестом та результат $\theta_{92} = -0,5$ учасника тестування другим тестом.

Номер учасника i	1	2	3	4	5	6	7
θ_{i1} в логітах за наслідками 1-го тесту	0,5	1,4	1,2	1	1,3	0,8	0,8
θ_{i2} в логітах за наслідками 2-го тесту	-1,8	0,2	0,05	0	0,2	-0,2	-0,1

Розв'язання: Нехай рівні підготовки до першого тесту θ_{i1} відповідають горизонтальній осі Ox , а рівні підготовки до другого тесту θ_{i2} відповідають вертикальній осі Ox . Рівняння регресії має вигляд $y = a_0 + a_1x$. Оцінимо коефіцієнти, як корені системи

$$\begin{aligned}\hat{a}_0 + \bar{x}\hat{a}_1 &= \bar{y} \\ \bar{x}\hat{a}_0 + \overline{x^2}\hat{a}_1 &= \overline{xy}\end{aligned}$$

Вибіркові середні обчислюємо формулами

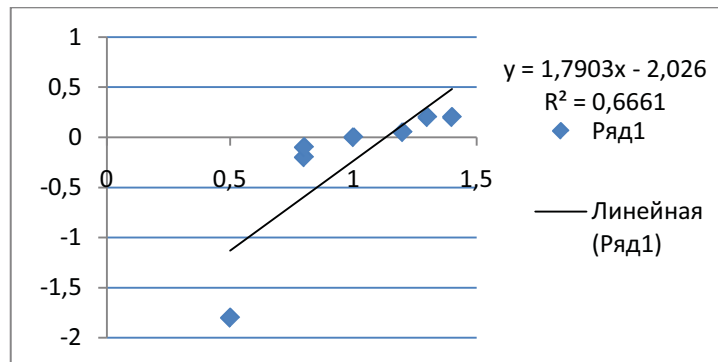
$$\begin{aligned}\bar{x} &= \frac{1}{7} \sum_{i=1}^7 x_i = 1,0, & \bar{y} &= \frac{1}{7} \sum_{i=1}^7 y_i = -0,2357, \\ \overline{xy} &= \frac{1}{7} \sum_{i=1}^7 x_i y_i = -0,0771, & \overline{x^2} &= \frac{1}{7} \sum_{i=1}^7 x_i^2 = 1,0885.\end{aligned}$$

Після їх підстановки у систему та її розв'язання отримаємо оцінки коефіцієнтів рівняння регресії

$$\hat{a} = 1,7903; \hat{a}_0 = -2,026.$$

Вибіркове рівняння лінійної регресії має вигляд

$$y = 1,7903x - 2,026.$$



Вибірковий коефіцієнт кореляції $\hat{\rho} = \frac{\overline{xy} - \bar{x}\bar{y}}{(\overline{x^2} - \bar{x}^2)^{1/2} (\overline{y^2} - \bar{y}^2)^{1/2}} = 0,8161$.

Нехай основна гіпотеза $H_0: \rho = 0$ кореляційна залежність між результатами тестів відсутня. Протилежна гіпотеза $H_1: \rho \neq 0$ кореляційна залежність існує. Рівень значущості 0,05.

Якщо справджується основна гіпотеза, то H_0 , то статистика $t = \hat{\rho} \left(\frac{n-2}{1-\hat{\rho}^2} \right)^{1/2}$ підкоряється розподілу Стюдента з числом степенів волі $n - 2$.

Обчислимо $t_{\text{спостер}} = \hat{\rho} \left(\frac{n-2}{1-\hat{\rho}^2} \right)^{1/2} = 0,8161 \left(\frac{5}{1-0,8161^2} \right)^{1/2} = 3,1577$.

Критичне значення отримуємо із таблиці розподілу Стюдента з рівнем значущості 0,05: $t_{\text{крит}} = 2,26$. Оскільки $|t_{\text{спостер}}| > t_{\text{крит}}$, то приймаємо гіпотезу H_1 про існування кореляційної залежності між результатами двох тестів.

Знайдемо коефіцієнт детермінації

$$R^2 = 1 - \frac{\sum_{i=1}^7 (y_i - f_i)^2}{\sum_{i=1}^7 (y_i - \bar{y})^2} = 0,6661, \text{ де } f_i = 1,7903x_i - 2,026.$$

Видно, що 66,7% розсіювань у пояснюється отриманою лінійною регресією, а 33,4% викликані або випадковими помилками тестування, або неадекватністю лінійної моделі регресії.

Щоб перерахувати на метричну шкалу другого тесту результат $\theta_{81} = 0,7$ учасника тестування першим тестом підставимо його у рівняння регресії

$$y_8 = 1,7903 * 0,7 - 2,026 = -0,7727.$$

Щоб перерахувати на метричну шкалу першого тесту результат $\theta_{92} = -0,5$ учасника тестування другим тестом його також підставимо у рівняння регресії

$$-0,5 = 1,7903 * x_8 - 2,026, \quad \text{звідки } x_8 = 0,8524.$$

Приклад 12. Тест виконували 10 учасників, результати виконання одного завдання та загальний первинний бал подано у таблиці. Проаналізувати дискримінаційні можливості цього тестового завдання

Номер учасника	Відповідь на завдання 1- вірно, 0-невірно	Первинний бал за тест
1	0	2
2	0	12
3	1	17
4	1	15
5	1	6
6	0	7
7	1	13
8	0	8
9	1	19
10	0	6

Розв'язання: Дискримінаційну здатність тестового завдання можна оцінити з допомогою точково-бісеріальної кореляції ρ_{pb} між цим завданням та первинним балом. Якщо дискримінаційна здатність незначна, то точково-бісеріальна кореляція близька до нуля; якщо дискримінаційна здатність велика, то точково-бісеріальна кореляція близька до одиниці. Якщо точково-бісеріальна кореляція від'ємна, то завдання має результат, який у певній мірі суперечить усьому тесту; такі завдання потрібно проаналізувати експертам та вилучити або залишити у залежності від прийнятого ними рішення.

Сформулюємо основну гіпотезу H_0 та протилежну гіпотезу H_1 :

$H_0: \rho_{pb} = 0$ - дискримінаційна здатність тестового завдання відсутня;

$H_1: \rho_{pb} \neq 0$ - має місце дискримінаційна здатність тестового завдання.

Формула для обчислення оцінки точково-бісеріальної кореляції (коефіцієнта дискримінації) має вигляд

$$\hat{\rho}_{pb} = \frac{\bar{x}_1 - \bar{x}_0}{\sigma_x} \left(\frac{n_1 n_0}{n(n-1)} \right)^{1/2}.$$

За умовою можна визначити такі значення

- Загальна кількість учасників $n = 10$.
- Кількість учасників, які вірно відповіли на завдання $n_1 = 5$, а середнє значення їх первинного балу $\bar{x}_1 = 14$.
- Кількість учасників, які невірно відповіли на завдання $n_0 = 5$, а середнє значення їх первинного балу $\bar{x}_0 = 7$.
- Середнє квадратичне відхилення первинного балу $\hat{\sigma}_x = 5,52$.

Обчислимо оцінку точково-бісеріальної кореляції або коефіцієнт дискримінації

$$\hat{\rho}_{pb} = \frac{14 - 7}{5,52} \left(\frac{5 \cdot 5}{10 \cdot 9} \right)^{1/2} = 0,6680.$$

Для перевірки гіпотези про те, що дискримінаційні можливості рівні нулю,

обчислимо статистику $\frac{2}{\sqrt{n}} = \frac{2}{\sqrt{10}} = 0,6324$.

Порівняємо отримані значення

$$\hat{\rho}_{pb} = 0,6680 > 0,6324 = \frac{2}{\sqrt{10}}.$$

Отже, дані цього тестування свідчать, що гіпотезу про те що дискримінаційні можливості даного завдання рівні нулю потрібно відкинути, тобто приймаємо гіпотезу про наявність дискримінаційної здатності завдання.

Приклад 13. У таблиці наведено дані відповідей 10 учасників тестування на два тестові завдання. Встановити степінь відповідності між цими двома завданнями у припущенні нормального розподілу результатів.

Номер Випробовуваного	X:ответ на завдання N 1	Y:ответ на завдання N 2
1	1	0
2	1	1
3	1	1
4	0	0
5	0	0
6	0	1
7	0	1
8	1	1
9	0	0
10	0	0

Розв'язання: Інтерпретуємо елементи кожного стовпчика таблиці як реалізації дихотомної випадкової величини, що набуває значення: 1 - завдання виконане вірно, 0 - невірно. Для перевірки наявності зв'язку між даними відповідей на завдання тестових завдань використаємо коефіцієнт бісеріальної кореляції. Зв'язок існує, якщо значення істотно відрізняється від нуля.

Сформулюємо основну гіпотезу H_0 та протилежну гіпотезу H_1

$H_0: \rho = 0$ - зв'язок між тестовими завданнями відсутня;

$H_1: \rho \neq 0$ - зв'язок між тестовими завданнями існує.

Задамо рівень значущості 0,05.

Формула для обчислення кореляції між дихотомними даними

$$\hat{\rho} = \frac{\hat{p}_{XY} - \hat{p}_X \hat{p}_Y}{(\hat{p}_X \hat{q}_X \hat{p}_Y \hat{q}_Y)^{1/2}}$$

Підрахуємо оцінки імовірностей, користуючись даними таблиці

$$\hat{p}_x = \frac{4}{10} = 0,4, \quad \hat{p}_y = \frac{5}{10} = 0,5, \quad \hat{p}_{xy} = \frac{3}{10} = 0,3.$$

Значення бісеріальної кореляції

$$\hat{\rho} = \frac{0,3 - 0,4 \cdot 0,5}{(0,4 \cdot 0,6 \cdot 0,5 \cdot 0,5)^{1/2}} = 0,4082.$$

Якщо кореляція між даними тестування у припущенні нормального розподілу відсутня, то значення статистики $t_{\text{спостер.}} = \hat{\rho} \left(\frac{n-2}{1-(\hat{\rho})^2} \right)^{1/2}$ буде менше за значення статистики Стьюдента $t_{\text{крит.}} = 2,23$, при $\alpha = 0,05$.

$$t_{\text{спостер.}} = 0,4082 \left(\frac{10-2}{1-(0,4082)^2} \right)^{1/2} = 1,26 < 2,23 = t_{\text{крит.}}$$

і тому гіпотеза про відсутність кореляції не суперечить експериментальним даним. Отже, приймаємо гіпотезу про відсутність зв'язку між результатами виконання двох тестових завдань.

Приклад 14. У таблиці наведено результати тестування жінок та чоловіків на знання правил дорожнього руху для отримання водійських прав, які оцінюються за дихотомним принципом «залік-незалік». Оцінити значущість відмінності жінок та чоловіків у знаннях правил дорожнього руху.

	«Залік»	«Незалік»	Σ
Жінки	975	102	1077
Чоловіки	1261	511	1772
Σ	2236	613	2849

Розв'язання: Як міра зв'язку відмінність в підготовці двох груп учасників тестування між ознаками можна користуватися коефіцієнтом асоціації

$$\varphi = \frac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$$

Коефіцієнт асоціації рівний нулю, якщо зв'язок відсутній; рівний 1, якщо існує повний прямий зв'язок; рівний -1, якщо існує повний обернений зв'язок.

Сформулюємо основну гіпотезу H_0 та протилежну гіпотезу H_1

$H_0: \varphi = 0$ - відмінності жінок та чоловіків у знаннях правил дорожнього руху не існують;

$H_1: \varphi \neq 0$ - відмінності жінок та чоловіків у знаннях правил дорожнього руху існують.

Задамо рівень значущості 0,05.

Якщо основна гіпотеза справджується, то статистика $\varphi\sqrt{n} \sim N(0,1)$.

Виконаємо обчислення

$$\varphi = \frac{975 * 511 - 1261 * 102}{\sqrt{2236 * 613 * 1077 * 1772}} = 0,2282$$

$$\varphi\sqrt{n} = 0,2282 * \sqrt{2849} = 12,1976.$$

При цьому із таблиці нормального розподілу $N(0,1)$ критичне значення $t_{\text{крит.}} = 1,64$. Із нерівності

$$t_{\text{спостер.}} = 12,1976 < 1,64 = t_{\text{крит.}}$$

видно, то наявні дані дають підстави рахувати відмінність в підготовці жінок та чоловіків значущим.

Питання для самоконтролю:

1. Дайте означення функції успіху; назвіть латентні параметри, від яких вона залежить. Чому параметри називають латентними?
2. Яка властивість функції успіху робить безнадійним успіх абсолютно непідготовленого учасника тестування?
3. Яка властивість функції успіху гарантує успіх абсолютно підготовленого учасника тестування?
4. Запишіть формулу функції успіху за основною логістичною однопараметричною моделлю Раша, за умови того, що латентні параметри вимірюються в логітах.
5. Від якого одного параметру (в логітах) залежить ймовірність успіху у моделі Раша?
6. Дайте означення характеристичної функції рівня підготовки учасника тестування та характеристичної функції рівня складності тестового завдання?

7. Який вигляд має логістична модель Раша, яка добре апроксимує функцію розподілу імовірності $N(0,1)$?
8. Дайте означення дискримінаційної здатності тестового завдання?
9. Запишіть функцію успіху за двох-параметричною моделлю Бірнбаума. Від яких параметрів вона залежить?
10. Яку характеристичну криву можна отримати фіксуванням рівня підготовки учасника тестування у двох-параметричній моделі Бірнбаума?
11. При якому значенні дискримінаційної здатності тестового завдання двох-параметрична модель Бірнбаума вироджується у однопараметричну модель Раша?
12. Запишіть функцію успіху за трьох-параметричною моделлю Бірнбаума. Від яких параметрів вона залежить?
13. Яка з трьох моделей сучасної теорії тестування не є логістичною?
14. Чому дорівнює імовірність успіху абсолютно непідготовленого учасника тестування за двох- та трьох-параметричними моделями Бірнбаума?
15. Дайте означення матриці відповідей? Назвіть, із яких елементів вона складається.
16. Які величини називають первинними балами учасників тестування та первинними балами тестових завдань?
17. Який розподіл мають первинні бали за умови різних за складністю тестових завдань?
18. Який розподіл мають первинні бали за умови однакових за складністю тестових завдань?
19. Як обчислюються математичні сподівання та дисперсії первинних балів за умови різних за складністю тестових завдань?
20. Як обчислюються математичні сподівання та дисперсії первинних балів за умови однакових за складністю тестових завдань?
21. Як визначити ймовірність $P_{ik}(b)$ отримання i -м учасником тестування b балів із k завдань?
22. Запишіть формулу для обчислення імовірності попадання первинного бала b_i - в заданий проміжок $[m_1; m_2]$.
23. Запишіть формули для обчислення імовірнісних характеристик первинних балів тестових завдань.
24. Як зміниться розподіл оцінок рівня складності завдань при зменшенні однорідності рівня підготовленості складу учасників тестування.
25. Які достатні статистики можна побудувати для елементів матриці відповідей?
26. Опишіть процес впорядкування матриці відповідей.

27. Як здійснити підрахунок долей правильних і неправильних відповідей учасника тестування на усі завдання тесту та виявити оцінку рівня підготовки окремого учасника тестування?
28. Як здійснити підрахунок долі правильних і неправильних відповідей на одне тестове завдання всіма учасниками тестування та виявити оцінку рівня складності окремого тестового завдання?
29. Як визначити середні значення логітів рівня підготовки учасників та логітів складності завдань тесту?
30. Як перенести оцінки латентних параметрів на єдину шкалу?
31. Запишіть формули для оцінок стандартних похибок оцінок латентних параметрів.
32. Як побудувати ефективні оцінки латентних параметрів, застосувавши метод найбільшої правдоподібності?
33. Які значення беруть в якості початкових значень для підстановки у ітераційний цикл для знаходження максимуму логарифмічної функції правдоподібності?
34. За яких умов переривається ітераційний цикл послідовних підстановок оцінок параметрів θ_i і β_j ?
35. Як побудувати характеристичні криві учасників тестування (PCC) та тестових завдань (ICC) на основі отриманих ефективних оцінок з використанням функції успіху логістичної моделі Раша?
36. Охарактеризуйте два підходи до формування моделей обробки тестування.
37. Опишіть математичний процес побудови однопараметричної моделі Раша для «ідеальних» тестів, які б відповідали теорії.
38. Як перевірити основну гіпотезу про відповідність отриманих емпіричних даних однопараметричній моделі Раша?
39. Що рекомендується робити для зменшення числа степенів волі статистики χ^2 ?
40. За яких умов вважається, що гіпотетична модель Раша не узгоджується з експериментальними даними?
41. За яких умов вважається, що гіпотетична модель Раша узгоджується з експериментальними даними?
42. Дайте означення кількості інформації, яка забезпечується завданням тесту.
43. Якою формулою визначається інформаційна функція тестового завдання?
44. Що таке інформаційна функція сукупності тестових завдань?
45. Що можна сказати про тест, інформаційна функція якого має два екстремуми? Один екстремум?

46. Як побудувати побудувати інформаційні функції тестових завдань для двох-параметричної моделі Бірнбіума?
47. До яких наслідків призводить збільшення дифереціюючої здатності тесту?
48. До чого призводить введення параметру відгадування?
49. Як збільшити інформативність завдання у трьох- параметричній моделі Бірнбаума?
50. Дати означення роздільної здатності тесту?
51. Що таке поріг чутливості тесту?
52. Як визначити роздільну здатність тесту, якщо обчислення щодо оцінювання латентних параметрів виконані?
53. Як визначити роздільну здатність тесту до обчислення латентних параметрів?
54. Яка залежність спостерігається між роздільною здатністю тесту та кількістю інформації?
55. Як визначається оцінка дискримінаційної здатності?
56. Яка величина виступає мірою зв'язку між результатом виконання тестового завдання та первинним балом, отриманим учасниками?
57. Яка величина називається коефіцієнтом дискримінації?
58. Як застосувати «метод контрастних груп» для визначення дискримінаційних можливостей завдання тесту?
59. Що називають дистрактором? Яким принципам повинні відповідати дистрактори?
60. Який знак має точково-бісеріальна кореляція дистрактора тестового завдання?
61. Чи можна назвати тест надійним, якщо невелика зміна умов тестування і стану випробовуваних призводить до несуттєвої зміни кінцевих результатів?
62. Як поняття надійності інтерпретується теорією статистики?
63. Як визначається коефіцієнт надійності за дисперсійною методикою?
64. Як визначити середню квадратичну помилку вимірів? Як зміниться формула для обчислення у випадку однакових завдань?
65. Як обчислити середнє значення коефіцієнта надійності за формулою Кьюдера-Річардсона? Чи потрібне при цьому розщеплення?
66. Як обчислюється кратність збільшення кількості завдань для підвищення надійності тесту?
67. Як визначається коефіцієнт надійності за кореляційною методикою?
68. Як пов'язаний коефіцієнт надійності тесту з коефіцієнтом його половини, отриманої в результаті його розщеплення?

- 69.Що називають ретестовою надійністю тесту? Яка величина є її кількісною мірою?
- 70.Яку величину називають валідністю тесту?
- 71.Що називають валідністю за змістом та критерійною валідністю?
- 72.Як поняття валідності інтерпретується теорією статистики?
- 73.Яка величина є мірою валідності окремого тестового завдання?
- 74.Як перевірити гіпотезу про відсутність зв'язку між даними одного тестового завдання та всім тестом? Що можна сказати про валідність тестового завдання за результатом такого аналізу?
- 75.У якому випадку мірою валідності може виступати коефіцієнт дихотомічної кореляції?
- 76.Як визначити валідність тесту, якщо об'єктивний критерій носить неперервний характер
- 77.Який принцип побудови процентної шкали?
- 78.Назвіть методик перерахування балів відповідно до їх частот на процентильну шкалу.
- 79.Як формується нормалізована шкала?
- 80.Як перетворити первинні бали на нормалізовану шкалу?
- 81.Чим відрізняється шкала нормалізованих процентилів від нормалізованої шкали?
- 82.Опишіть ітераційну процедуру для обчислення ітераційної процедури для обчислення виставкових балів.
- 83.Яке завдання називають якірним? З якою метою вводять його у різні варіанти тесту?
- 84.Скільки якірних завдань рекомендується вводити у варіанти тесту? Яку складність вони повинні мати?
- 85.Опишіть процедуру переведення латентних параметрів на єдину метричну шкалу.
- 86.Назвіть правила для обчислення виправлених значень латентних параметрів, віднесених до єдиної метричної шкали в логітах з умовним початком відліку.
- 87.На основі чого рекомендується здійснювати обрання нуля на єдиній метричній шкалі?
- 88.Які величини називають остаточними оцінками латентних параметрів?
- 89.За якою формулою виконують перерахунок остаточних оцінок латентних параметрів на 100-бальну шкалу? На 200-бальну?

Список джерел, використаних у розділі V

1. Ким В.С. Тестирование учебных достижений. Монография.- Уссурийск: Изд-во УГПИ, 2007. -214 с.
2. Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов.- М.: Логос, 2010.- 668 с.
3. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов.- М.: Прометей, 2000.- 168 с.
4. Чельшкова Н. Б. Теория и практика конструирования педагогических тестов : учебное пособие / Н. Б. Чельшкова. – М. : Логос, 2002. – 432 с.
5. Andrich D. Advanced Social and Educational Measurement. Perth: Murdoch University, 2001.- 128 pp.
6. Baker F.B. The Basics of Item Response Theory.- Portsmouth NH: Heinemann Educational Books, 1985.- 131 pp.
7. Baker F.B., Kim S.-H. Item response theory: parameter estimation techniques (2nd ed.). New York: Marcel Dekker, 2004.
8. Embretson S.E., Reise S.P. Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
9. Fischer G.H., Molenaar I.W. Rasch models: Foundations, recent developments, and applications. New York: Springer-Verlag, 1995.
10. Hambleton R.K., Swaminathan H. Item Response Theory: Principles and Applications. Norwell, MA: Kluwer-Nijhoff Publishing, 1985.
11. Hambleton R.K., Swaminathan H., Rogers H. J. Fundamentals of Item Response Theory. Newbury Park, CA: Sage, 1991.
12. Lord F.M. Application of Item Response Theory to Practical Testing Problems. Hillsdale N-J. Lawrence Erlbaum Ass., 1980. – 266 pp.
13. Smith E.V., Smith M.S. Introduction to Rasch Measurement. Theory, Models and Applications. Maple Grove, Minnesota: JAM Press, 2004. – 689 pp.
14. van der Linden W.J., Hambleton R.K. (Eds.) Handbook of modern Item Response Theory (IRT). New York, New York: Springer-Verlag, 1997.- 510 pp.

VI. КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ У ТЕСТУВАННІ

Проведення масових тестувань з метою моніторингу, вступних випробувань тощо не могли б бути реалізованими без комп'ютерних технологій тестування. Не можливі без комп'ютерних засобів також деякі види критеріально-орієнтованих тестувань. Контроль знань, пробні та навчальні тестування в умовах дистанційної освіти також потребують різноманітних мережевих тестових програмних засобів.

Цей розділ розпочинається з представлення видів, типів, етапів у комп'ютерному тестуванні. Далі подані необхідні відомості з інформаційної безпеки процедур масового комп'ютерного тестування. Особлива увага приділена вільно розповсюджуваному програмному забезпеченню, зокрема, сервісам Web 2.0. Описано систему Hot Potatoes, як приклад програмного засобу для локального тестування. Мережеве програмне забезпечення для проведення тестування представлено системою MyTestX. Описані можливості веб орієнтованої системи тестування TCEexam. Значна увага приділена системі VLE Moodle, зокрема, описане створення курсів, банк тестових завдань, створення тесту, аналізу тестових завдань та засоби оцінювання в цьому середовищі.

Розглянуто такі види комп'ютерних моделей педагогічного тестування: лінійні, ієрархічні та мережні, які пов'язані зі структурною організацією бази завдань.

6.1. Основні відомості про комп'ютерне тестування

Автоматизація контролю тісно пов'язана з автоматизацією всього освітнього процесу і забезпечує зворотний зв'язок в автоматизованих системах управління освітнім процесом.

6.1.1. Види автоматизованого контролю:

1. Вхідний контроль призначений для:

- визначення початкового рівня ЗУН учнів (студентів) з метою відбору кандидатів для подальшого навчання;

- підтримка високого рівня пізнавального інтересу під час навчання на основі визначення стилю навчання, оптимального для того, що навчається;

- складання індивідуального маршруту навчання.

2. Поточний контроль призначений для:

- перевірки ступеня засвоєння студентами змісту навчання.
- перевірки ступеня готовності студентів до вивчення нового матеріалу;
- корекції індивідуальних маршрутів навчання;
- забезпечення регулятивної функції.

3. Періодичний контроль призначений для:

- перевірки ступеня засвоєння студентами змісту навчання. Проводиться рідше, ніж поточний контроль, охоплює цілі розділи учбової програми. Основна форма – залікові тести. На додаток використовуються курсові роботи, контрольні-залікові завдання і так далі. Для періодичного контролю, як правило, використовуються ті ж тести, що і для підсумкового контролю.

4. Підсумковий контроль призначений для:

- перевірки якості виконання студентами учбового плану і проводиться у формі екзамену зі всієї дисципліни;
- вирішення питання про присвоєння випускникам відповідної кваліфікації.

5. Самоконтроль призначений для:

- визначення студентами успішності свого навчання;
- встановлення студентами пропусків у вивченні дисципліни;
- корегування студентами індивідуального стилю навчання.

6. Взаємоконтроль призначений для:

- формування у студентів оцінних алгоритмів і якостей;
- підвищення об'єктивності оцінювання за рахунок участі в цьому процесі самих студентів.

Автоматизація дозволяє значно підвищити роль всіх видів контролю, особливо самоконтролю і взаємоконтролю, відкрити нові аспекти їх застосування. Не всі методи контролю можуть бути в рівній мірі піддані автоматизації, але з розвитком ІТ з'являються нові можливості для цього.

6.1.2. Етапи комп'ютерної атестації

Процедура комп'ютерної атестації – досить тривалий багатоетапний захід, що складається з етапів:

1. Реєстрація студентів в спеціалізованій БД;

2. Тренувальне тестування з елементами навчання, що включає автоматичну генерацію проміжних результатів і рекомендацій студентам;
3. Основне тестування;
4. Статистична обробка результатів;
5. Публікація рейтингів;
6. Підготовка атестаційної документації (відомостей) на затвердження викладачеві;
7. Підготовка і представлення викладачеві інформації про якість засвоєння розділів дисципліни;
8. Поповнення банку результатів новими даними;
9. Статистична обробка результатів, зокрема – порівняння їх з отриманими раніше і паралельно;
10. Представлення результатів у вигляді, що сприяє зростанню рівня якості освіти (у вигляді статей, наказів, звітів і ін.).

6.1.3. Вимоги до комп'ютерної атестації

Для досягнення цілей процедури атестації, необхідне:

1. Забезпечення валідності тесту;
2. Забезпечення варіативності тесту;
3. Забезпечення оновлюваної бази тестових завдань;
4. Забезпечення максимальної роздільної здатності тесту;
5. Забезпечення достовірності результатів (виключення зовнішніх дій);
6. Забезпечення повного і доступного представлення результатів;
7. Включення елементів ділової гри в процедуру атестації і інтерфейс комп'ютерного тестового комплексу;
8. Забезпечення ефективності сеансу тестування (максимально повний і достовірний результат при мінімальних витратах часу і ресурсів);
9. Включення в процедуру атестації елементів змагання (перманентна публікація списків претендентів, що отримали вищі рейтинги у пресі і на Інтернет-сайтах);
10. Забезпечення затребуваності рекомендацій, які отримують студенти та викладачі за наслідками тестувань.

6.1.4. Типи сценаріїв комп'ютерного тестування

Комп'ютерні комплекси для тестування ЗУН можуть бути побудовані із застосуванням різних сценаріїв. Немає загальноприйнятої їх класифікації.

1. *«Стандарт»*. Завдання з фіксованого списку пред'являються на екрані послідовно, всі варіанти виконання завдання видно одночасно, і претенденту в кожному випадку пропонується вибрати один з пред'явлених варіантів виконання. Тут велика вірогідність

- простого списування;
- випадкового вгадування;
- вгадування шляхом порівняння варіантів.

2. *«Стандарт. Випадковий вибір»*. Завдання з фіксованого списку пред'являються претендентам у випадковому порядку, а їх варіанти виконання також перемішуються при кожному пред'явленні. Такий підхід значно знижує вірогідність

- простого «списування».

3. *«Стандарт. Множинний вибір»*. Потрібно вказати в списку не один, а всі вірні варіанти виконання завдання. Такий підхід значно знижує вірогідність

- випадкового вгадування.

4. *«Єдиний вибір»*. Разом з текстом завдання демонструється тільки один з можливих варіантів виконання. Від претендента потрібно оцінити цей єдиний варіант за двозначною шкалою (правильно – не правильно). У цьому варіанті дуже низка вірогідність вгадування шляхом порівняння, проте велика вірогідність простого вгадування. Для виправлення останнього недоліку застосовується система «штрафів». Такий підхід дозволяє

- не тільки констатувати знання, але і контролювати їх глибину; але і значно знижує вірогідність
- випадкового вгадування;
- вгадування шляхом порівняння варіантів;
- простого «списування».

5. *«Пошук однозначної відповідності»*. Потрібно однозначно зіставити терміни (зображення, звукозапис). У такому підході дуже низка вірогідність

- вгадування шляхом порівняння варіантів;
- випадкового вгадування.

6. *«Пошук неоднозначної відповідності»*. Потрібно побудувати таку відповідність, коли одному терміну (зображенню, звукозапису) з першого стовпця можуть відповідати 2-3 терміни з другого, і навпаки. Такий підхід дозволяє

- не тільки констатувати знання, але і контролювати їх глибину; але і робить незначною вірогідність
- вгадування шляхом порівняння варіантів;
- випадкового вгадування.

7. *«Створення повідомлення»*. Потрібно ввести у вказаному місці слово, словосполучку або сукупність цифр і букв, – тобто закодувати повідомлення, що є рішенням поставленої в завданні задачі. У такому підході дуже низка вірогідність

- вгадування шляхом порівняння варіантів;
- випадкового вгадування.

8. *«Пошук і виправлення помилки»*. Місце введення послідовностей знаків не визначене, і учню пропонується його знайти (немає явної вказівки на місце в таблиці або тексті на місце розташування помилки). Такий підхід дозволяє

- не тільки констатувати знання, але і контролювати їх глибину; але і робить незначною вірогідність
- вгадування шляхом порівняння варіантів;
- випадкового вгадування.

9. *«Диктант»*. Мультимедійний сценарій, в якому потрібно написати диктант, слухаючи звук в навушниках, або перекласти на іншу мову. Такий підхід позбавлений більшості недоліків, пов'язаних з випадковістю і дозволяє

- контролювати знання правил написання математичних виразів, словосполук, хімічних формул і ін.

10. *«Подорож»*. Ігровий мультимедійний сценарій, де для виконання завдання потрібно проробляти складні маніпуляції з клавіатурою і мишею. Наприклад, «рухатися» по анімованому приміщенню, вводити з клавіатури

словосполучення в спеціальних «вікнах» і «розшукувати» об'єкти на основі інструкцій. Такий підхід дозволяє

- не тільки констатувати знання, але і контролювати їх глибину; але і робить незначною вірогідність
- вгадування шляхом порівняння варіантів;
- випадкового вгадування.

11. *«Визначення оцінки»*. Сценарій близький сценарію «Єдиний вибір» і ускладнений тим, що претендент повинен оцінити ступінь вірності єдиної пред'явленої відповіді по якій-небудь з багатозначних шкал. Наприклад:

- «так» – не знаю – ні»;
- «вірно – не точно – не вірно»;
- «вірно – не повно – не точно – не вірно».

Такий підхід

- дозволяє не тільки констатувати знання, але і контролювати їх глибину;
- робить незначною вірогідність вгадування шляхом порівняння варіантів;

Сценарій може бути ігровим, імітувати учбові дії вчителя і учнів.

12. *«Адаптивне тестування»*. Тестування ведеться формально по одному з приведених вище сценаріїв, але послідовність пред'явлення завдань залежить від того, як претендент справився з попередніми завданнями. Психологічний стан претендента оцінюється автоматично, залежно від оцінки стану міняється складність і трудомісткість завдань, включаються блоки психологічного розвантаження. Інформація поступає переважно по слуховому каналу. Цей сценарій вважається за найбільш перспективний для застосування в практиці індивідуальних і масових тестувань.

6.2. Інформаційна безпека процедур комп'ютерного тестування

6.2.1. Основні вимоги інформаційної безпеки масових тестувань

Основні вимоги інформаційної безпеки тестувань та забезпечення їх виконання

- Секретність баз тестових завдань;

- Секретність переліків осіб що мають доступ до тестових завдань;
- Недосяжність осіб, що мають доступ до інформації про тести, в період конструкторських робіт;
- Заборона несанкціонованого доступу до баз тестових завдань і тестових програмних оболонок під час їх переміщень, установок, зберігання, використання;
- Заборона несанкціонованого доступу до результатів тестувань під час їх створення, пересилки, обробки і зберігання;
- Недопущення використання несанкціонованих довідкових матеріалів;
- Застосування спеціальних методик розміщення претендентів в аудиторіях, що унеможливають взаємного підказування, підглядання і ін.
- Недопущення несанкціонованих підказок і пояснень з боку персоналу;
- Недоступність сторонніх осіб в аудиторії, коридори, туалети і ін., у тому числі завчасна перевірка приміщень на наявність шпаргалок;
- Запобігання і ліквідація наслідків порушення названих вище пунктів.

Саме ж забезпечення припускає:

- створення документаційного супроводу (інструкції, положення, накази);
- відбір персоналу;
- навчання і інструктаж персоналу;
- підвищення винагорода персоналу за забезпечення особливого режиму;
- проведення інспекцій, у тому числі і шляхом залучення осіб – «інкогніто»;
- оснащення технікою відеоспостереження і сейфами;
- залучення сил міліції або позавідомчої охорони;
- заходи по виявленню порушень і адміністративному покаранню винних осіб.

6.2.2. Шляхи забезпечення інформаційної безпеки

1. Не слід «засекречувати» відповіді на питання комп'ютерного тесту. Велику частину текстів бази тестових завдань слід видавати друкарським способом. Це приведе до зникнення об'єкту несанкціонованої купівлі-продажу.
2. Питань і відповідей в базі тестових завдань повинно бути так багато, щоб неможливо було швидко і повноцінно скористатися паперовою або

електронною шпаргалкою. Для відображення в базі матеріалу об'ємом 36 учбових годин достатні 300-400 завдань (10 завдань на годину).

3. Претенденти, що сидять поряд, повинні бачити на своїх моніторах різні тестові завдання. При повторному тестуванні кожен претендент отримує новий варіант. Виконання цих умов досягається шляхом випадкової генерації варіантів з бази тестових завдань.
4. Варіанти тестів не повинні повторюватися, проте програма – генератор варіантів повинна складати списки завдань для кожного претендента не абсолютно випадковим чином, а виходячи з принципу рівномірності покриття учбового матеріалу з кожної теми курсу.
5. Слід робити заходи захисту не тільки на програмному рівні. Наприклад, поточні результати тестування необхідно контролювати в інтерактивному режимі, а програмні модулі повинні зберігатися на фізично незалежних носіях і оновлюватися з періодом меншим, ніж час сеансу. Цим підвищується рівень захисту від несанкціонованого проникнення.
6. Слід періодично оновлювати бази тестових завдань. Можна не тільки доповнювати базу кількісно, але і, наприклад, просто видозмінювати словарні обороти в завданнях. Цим усувається багато з можливостей застосування шпаргалок.
7. «Змова» претендентів з представниками персоналу, які «закриють очі на шпаргалки», буде виключена, якщо дозволити використання будь-яких паперових навчальних посібників під час тестування. При цьому слід обмежити час сеансу тестування так, щоб було неможливо встигнути скористатися цією інформацією повною мірою. Цим не тільки усувається можливість «злочинної» змови претендентів з персоналом комп'ютерного класу, але і створюється стимул у додатковому вивченню матеріалу саме тієї допомоги, яка рекомендована викладачем курсу.
8. Контрольне тестування повинне відбуватися для всієї учбової групи одночасно. Захист від зовнішніх дій можна забезпечити присутністю зацікавленого викладача, наприклад того, який вестиме подальші учбові курси для цих студентів в майбутньому семестрі.
9. Слід виключати на час сеансу тестування зв'язок класу із зовнішніми мережами.
10. Слід використовувати процедуру випадкового вибору номера комп'ютера і псевдоніма для кожного претендента. Претенденти в класі повинні

сидіти у випадковому порядку і мати достатньо довгі псевдоніми, якими вони не зможуть швидко обмінятися.

11. Тестування повинне оцінюватися за рейтинговою системою. Це впливає як на якість освіти, так і на встановлення внутрішньої системи контролю. Якщо студенти знатимуть, що кількість п'ятірок і четвірок, які вони можуть отримати обмежено, поступово усунуться підказки.
12. Результати рейтингу (кращі) повинні публікуватися в Інтернет і бути доступними для вірогідних працедавців випускників, що стане додатковим стимулом до навчання. При цьому відповідно до законодавства, кожен тестуючий комплекс повинен надати студентові право вибору – дозволити чи ні подібну публікацію.

6.2.3. Кроки для досягнення безпечної технології тестування

1. Програмне забезпечення, що поступило завчасно, встановлюється в комп'ютерному класі. Воно використовується як для тренувальних, так і для атестаційних тестувань. Доступ претендентів на тренування обумовлений інструкцією.
2. Адміністратор мережі регулярно, за встановленим розкладом, отримує ключ, необхідний для формування кодів претендентів.
3. Контроль документів претендентів при реєстрації проводиться незалежним представником адміністрації закладу чи органів правопорядку. Незалежний контроль перекриває один з вірогідних шляхів фальсифікації – підміну претендента.
4. Претенденти, що приступають до тестування, протягом нормативного проміжку часу вводять в комп'ютери свої дані. На основі цієї інформації формується код претендента і відповідний варіант завдань тесту, які можна виконувати сидячи перед екраном комп'ютера.
5. Оголошується початок сеансу. Виконання завдань контролюється черговими. Контроль в даному випадку спрощений, оскільки варіанти завдань у всіх учасників різні; вага і об'єм шпаргалки, наявної в продажі в найближчому магазині, достатня для того, щоб її вчасно помітити.
6. Результати тестування залишаються невідомими протягом обумовленого інструкцією періоду.

- Після закінчення технологічного періоду обробки, публікуються результати. Претенденти з ними знайомляться на стендах.

Десять відомих в практиці тестування способів *фальсифікації*

- Підміна претендента при реєстрації шляхом підробки документів або змови.
- Підміна претендента при тестуванні шляхом зміни місця або часу виконання завдань.
- Підміна бланків з результатами роботи претендента в процесі транспортування.
- Підміна бланків з результатами роботи претендента шляхом змови.
- Підміна файлу результатів шляхом змови.
- Злам і виправлення файлу результатів протягом часу зберігання шляхом змови. Для зламу потрібно досить багато часу.
- Підміна файлу результату на сервері або у момент пересилки між серверами шляхом хакреської атаки.
- Шпаргалка. Використання шпаргалки протягом сеансу тестування. При цьому якщо варіантів тесту достатні багато, використовується змова.
- Списування. Відбувається, якщо претенденти сидять в аудиторії так, що бачать результати дій один одного.
- Обмін файлами. Відбувається, коли один претендент реєструється з даними і паролем іншого претендента.

6.3. Сервіси Web 2.0 в освіті та в освітніх вимірюваннях

6.3.1. Короткі відомості про Web 2.0

Web 2.0 (визначення Тіма О'Рейлі) — методика проектування систем, які шляхом врахування мережевої взаємодії стають тим краще, чим більше людей ними користуються.

Веб 2.0 (англ. Web 2.0) — поняття, яким користуються для позначення ряду технологій та послуг Інтернету, точніше його частини — всесвітньої павутини, відомої також як Веб. Окрім цього, цим поняттям описують зміну сприйняття Інтернету користувачами.

Принциповою відмінністю Веб 2.0 від традиційної мережі є можливість створювати вміст Інтернету будь-якому користувачу.

Web 1.0	Web 2.0	Нові риси
1. розробник і користувач; 2. автор контенту і читач.	1. користувач як спів розробник; 2. читач як співавтор; 3. товариство.	Право на участь; скасування сторонньої регламентуючої сторони (модерції).
1. Данні організуються таксономічно (ієрархія рубрик); 2. Засоби збереження даних — каталог, бібліотека, сховище; 3. Статичний сайт; 4. Джерело — розум автора контенту; 5. Копірайт; 6. Для сприйняття контенту потрібне відвідування сайту, переходячи по посиланню чи закладці.	1. Данні організуються фолксономічно («народна» класифікація на основі тегів); 2. Форма представлення — блоги; 3. Динамічний сайт; 4. Джерело — колективний розум; 5. «Вільна» ліцензія GNU FDL; 6. Для сприйняття контенту не потребується відвідування сайту — можливість читати RSS — стрічки.	Мережа як єдиний колективний розум, атомізація контенту, агрегація, синдикація.

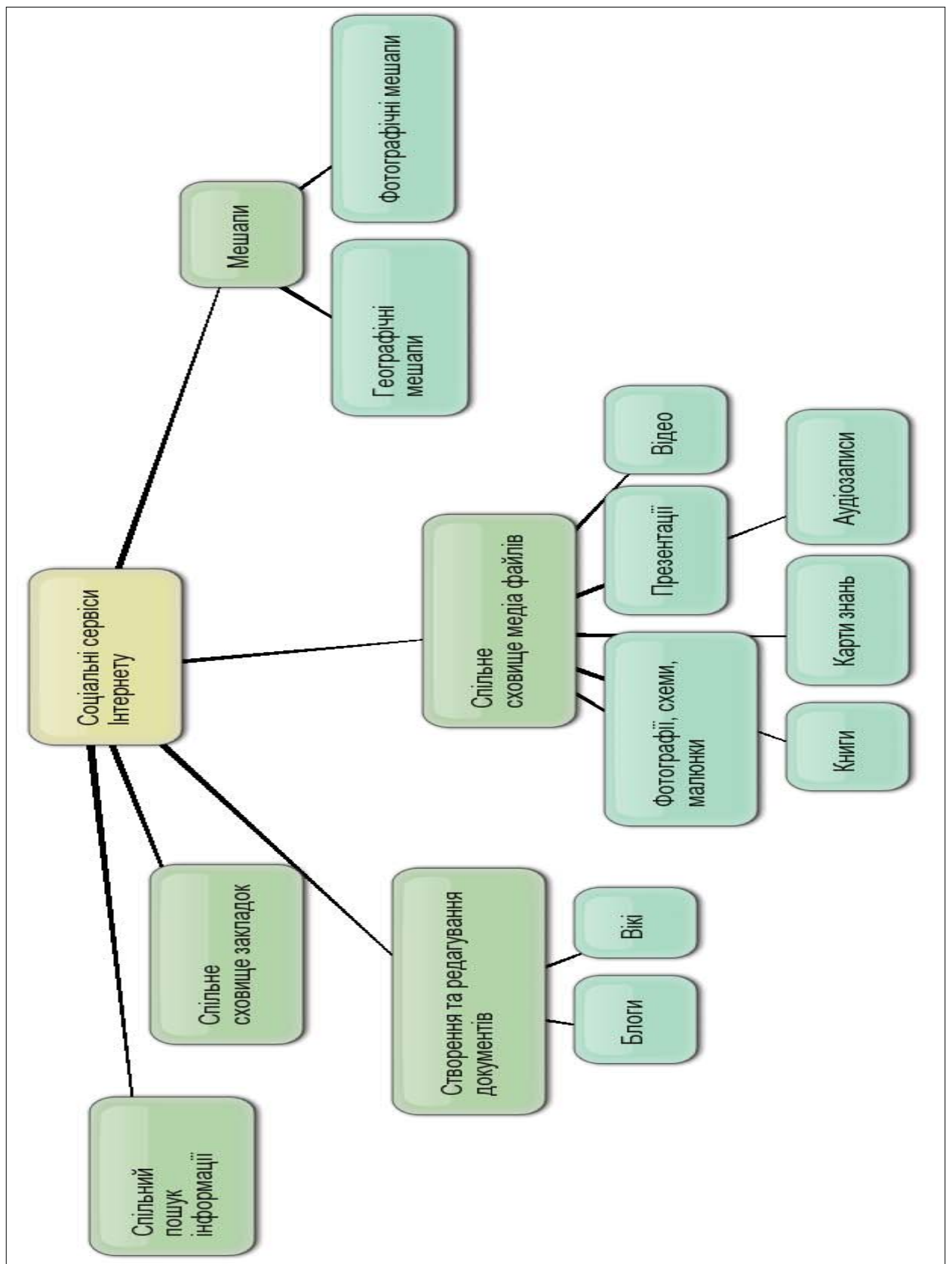
Web 2.0, або соціальні мережеві сервіси – сучасні засоби, мережеве програмне забезпечення, що підтримує групову взаємодію.

Ці групові дії включають:

- персональні дії учасників: запис думок, замітки та анотування текстів інших користувачів, розміщення мультимедійного контенту;
- комунікації учасників між собою.

6.3.2. Можливості використання Web 2.0

- Використання відкритих, безкоштовних ресурсів, що вільно розповсюджуються.
- Самостійне створення мережевого навчального матеріалу.
- Засвоєння інформаційних концепцій, знань, навичок, завдяки участі у нових формах діяльності.
- Спостереження за діяльністю співтовариств.



Соціальне сховище закладнок

<http://www.bobrdobr.ru>

<http://Del.icio.us>

Соціальне сховище мультимедіа даних**Фотосервіси**

<http://flickr.com>

<http://picasaweb.google.com>

<http://www.panoramio.com/>

Відеосервіси

<http://www.teachertube.com>

<http://www.youtube.com/education>

<http://www.rutube.ru>

Аудіосервіси

<http://audacity.sourceforge.net>

<http://www.podomatic.com>

<http://rpod.ru/>

Мережеві щоденники (блоги)

<http://Livejournal.ru>

<http://www.blogger.com>

<http://Liveinternet.ru>

<http://Blogs.mail.ru>

Соціальні геосервіси

<http://maps.google.com/>

<http://maps.yandex.ru>

<http://wikimapia.org>

<http://earth.google.com>

Wiki-сервіси

<http://www.wikipedia.org/>

<http://www.wiktionary.org/>

<http://www.wikibooks.org/>

<http://www.wikiversity.org/>

<http://commons.wikimedia.org>

Соціальні сервіси для спільної роботи з документами

<http://www.slideshare.net>

<https://docs.google.com/>

<https://www.zoho.com/docs/>

<https://www.lotuslive.com>

Карти знань

<http://bubbl.us>

<http://freemind.sourceforge.net/wiki/>

<http://www.graphviz.org/>

6.4. Hot Potatos - програмний засіб для локального тестування

6.4.1. Загальні відомості про систему Hot Potatoes

Hot Potatoes - інструментальна програма-оболонка, що надає викладачам можливість самостійно створювати інтерактивні завдання і тести для контролю і самоконтролю.

За допомогою програми можна створити 10 типів завдань і тестів з використанням текстової, графічної, аудіо- і відеоінформації.

Особливістю цієї програми є те, що створені завдання зберігаються в стандартному форматі веб-сторінки: для їх використання учням необхідний тільки веб-браузер (наприклад, Internet Explorer).

Учням не потрібна програма Hot Potatoes: вона потрібна тільки викладачам для створення і редагування завдань.

Програма широко використовується у всьому світі для створення завдань для вивчення будь-яких дисциплін.

До складу Hot Potatoes входять 5 блоків програм для складання завдань і тестів різних видів. Кожен блок може бути використаний як самостійна програма.

1. JQuiz – Вікторина – питання з множинним вибором відповіді (4 типи завдань).
2. JCloze – Заповнення пропусків.
3. JMatch – Встановлення відповідностей (3 типи завдань).
4. JCross – Кросворд.
5. JMix – Відновлення послідовності.

Завдання або тест		Піктограма	Розширення імені файлу
JQuiz	Вікторина (множинний вибір відповіді)		*.jqz
JCloze	Заповнення пропусків		*.jcl
JCross	Кросворд		*.jcw
JMix	Відновлення послідовності		*.jmx
JMatch	Встановлення відповідностей		*.jmt
The Masher	Інструменти		*.jms

Всі завдання виконуються в режимі самоконтролю (режим тестування передбачений тільки для питань з множинним вибором відповіді). Результат виконання завдань оцінюється у відсотках. Невдалі спроби призводять до зниження оцінки.

Остання версія програми містить також додатковий блок The Masher (Інструменти), що дозволяє об'єднувати створені завдання і інші учбові матеріали в тематичні блоки, уроки і учбові курси.

6.4.2. Блоки системи Hot Potatoes

Вікторина (JQuiz)

За допомогою цієї програми можуть бути створені тести на основі питань з різними варіантами представлення відповідей:

- альтернативний вибір – одна правильна відповідь з декількох варіантів відповідей;
- множинний вибір – декілька правильних відповідей;
- коротка відповідь – коротка відповідь відкритого типу (вводиться за допомогою клавіатури);

- змішаний варіант відповіді – об'єднує питання з відповіддю відкритого типу і множинним вибором відповіді: після введення невірної відповіді (можлива кількість невірних відповідей встановлюється при створенні завдання) завдання з короткою відповіддю трансформується в завдання з множинним вибором.

У один тест або завдання можна включити декілька питань різних типів.

Кількість питань у вправі не обмежена.

Зауваження. У питанні з короткою відкритою відповіддю може бути як один, так і декілька варіантів правильних відповідей.

Заповнення пропусків (JCloze)

Це питання закритого типу, які вимагають вставку слів або букв в залишені місця – пропуски.

Вводиться текст, в якому треба буде зробити пропуски. Текст вводиться у велике текстове поле з клавіатури або копіюється з будь-якого документа.

Потім Ви повинні виділити слово або букву, які мають бути пропущені або приховані і натиснути кнопку «Пропуск».

У діалозі, що відкрився, «Альтернативні слова» Ви повинні вставити всі слова, які можуть вважатися за вірні відповіді. Порядковий номер слова в тексті вказується в полі «Пропуск №», а саме слово – в поле – «Слово».

Слова або букви, відмічені як пропуски, автоматично виділяються в тексті червоним кольором і підкреслюються.

Послідовність вибору слів для пропусків в тексті не має значення, оскільки всі пропущені слова нумеруються автоматично від початку тексту.

Кросворд (JCross)

Програма JCross призначена для швидкого створення кросвордів. Цей вид завдань дуже популярний в середній школі.

Для розміщення слів в кросворді існують два режими:

- ручне розміщення слів;
- автоматичне розміщення слів.

При ручному розміщенні слів положення слів в сітці визначається користувачем: слова буква за буквою вводяться в поле кросворду.

Для автоматичного створення кросворду треба клацнути на кнопці «Автоматичне створення» в панелі інструментів або вибрати пункт меню «Управління сіткою – Автоматичне створення». При автоматичному

розміщенні слів всі слова вводяться списком в полі введення слів в діалозі, що відкрився, при цьому всі слова вводяться в окремі рядки, а потім програма підбирає варіанти їх розміщення. У полі «Максимальний розмір» вказується максимальне число осередків в кросворді. Клацання на кнопці «Створити сітку» приведе до створення кросворду.

Відновлення послідовності (JMix)

Цей вид завдання зручний для викладачів-лінгвістів, хоча можна створити цікаві вправи і для інших предметів.

Матеріалом для вправи можуть бути слова і речення. Для кожного слова або речення створюється окреме завдання. Сенс завдання – розташувати для речення слова в потрібному порядку, а для слова - розташувати букви або склади в потрібному порядку.

Поля «Альтернативні» заповнюються для речень, в яких можливі різні варіанти порядку слів.

Для цих альтернативних пропозицій в робочому вікні встановлюються додаткові налаштування – Налаштування альтернативного речення:

- дозволити використовувати речення, в яких не використовуються всі слова і розділові знаки початкового речення;
- попереджати, якщо у варіанті відповіді не використовуються всі слова і знаки пунктуації початкового речення.

Встановлення відповідностей (JMatch)

Завдання можуть бути представлені в трьох форматах:

- встановлення відповідностей за допомогою переміщення елементів мишею;
- вибір варіанту відповідності із списку, що розкривається;
- картки для запам'ятовування відповідностей.

Тип завдання вибирається на етапі збереження у вигляді сторінок HTML.

У стовпчик зліва вводиться фіксований список елементів, а в стовпчик справа – елементи, які можуть бути переплутані, і з них виконуватиметься підбір відповідностей до елементів фіксованого списку.

Елементи вводяться в правильній послідовності. Вони перемішуватимуться автоматично при завантаженні сторінки.

6.5. Система MyTestX – засіб для проведення мережевого тестування

MyTestX це - система програм для створення і проведення комп'ютерного тестування, збору і аналізу результатів, виставлення оцінки за вказаною в тесті шкалою.

За допомогою програми MyTestX можлива організація і проведення тестування, іспитів в будь-яких освітніх установах (вузи, коледжі, школи) як з метою виявити рівень знань по будь-яких учбових дисциплінах, так і з повчальними цілями. Підприємства і організації можуть здійснювати атестацію і сертифікацію своїх співробітників.

Особливості:

- Окремі функціональні модулі.
- Простота створення тестів.
- Зручний інтерфейс.
- Дев'ять типів завдань.
- Локальне і мережеве тестування.
- Необмежена кількість завдань.
- П'ять формулювань питання.
- Система оцінювання, що настроюється.
- Різні режими тестування.
- Багаті можливості форматування тексту.
- Можливість прикріплювати зображення і звук.
- Зручна функція вибірки завдань.
- Можливість перемішування завдань і варіантів.
- Обмеження часу і кількості запусків.
- Величезна кількість налаштувань процесу тестування і роботи програми.
- Допомога в аналізі результатів.
- Один тест – один файл.
- Файл з тестом зашифрований і стиснений.
- Багатомовний інтерфейс.
- Безкоштовна для некомерційного використання.

Програма складається з трьох модулів:

- Модуль тестування (MyTestStudent).
- Редактор тестів (MyTestEditor).
- Журнал тестування (MyTestServer).

Типи завдань:

- Одиночний вибір.
- Множинний вибір.
- Встановлення порядку.
- Зіставлення варіантів.
- Вибір істинності або помилковості тверджень (MCQ).
- Ручне введення числа.
- Ручне введення тексту.
- Місце на зображенні.
- Перестановка букв.
- + Так само тип «Да/Ні» може бути отриманий з Одиночного вибору, додавши всього два варіанти 1) так і 2) ні.

Режими тестування:

Програма підтримує декілька незалежних один від одного режимів:

- *Навчальний* (тестованому виводяться повідомлення про його помилки, може бути показане пояснення до завдання).
- *Штрафний* (за неправильні відповіді у тестованого віднімаються бали і можна пропустити завдання).
- *Вільний* (тестований може відповідати на питання в будь-якій послідовності, переходити (повертатися) до будь-якого питання самостійно).
- *Монопольний* (вікно програми займає весь екран і його неможливо звенути).

Всі режими незалежні один від одного і їх можна комбінувати довільним чином.

Параметри тесту можна налагоджувати через однойменне меню «*Параметри тесту*».

Можна задати:

- автора тесту;
- заголовок і опис тесту;
- інструкцію тестованому;
- замітки до тесту;
- порядок проходження завдань і варіантів (у завданнях з декількома варіантами);
- використовуване формулювання питання;
- змінити спосіб оцінювання або додати свою систему оцінки;
- створити і накопичувати теми (групи) завдань, задати обмеження на кількість завдань, що задаються (тобто робити вибірку з тих, що є);
- обмеження часу і кількості запусків програми без перезавантаження;
- паролі для відкриття, редагування файлу, початку тестування, відкриття захищених результатів;
- параметри виведення результату тестування, записи результатів у файл і відправки по мережі.

При виборі деяких пунктів відкриватимуться додаткові вікна, де ви можете ввести текст або інші дані.

У рядку стану «*Редактора*» присутні індикатори значень деяких параметрів (порядок завдань і варіантів, включені режими). Наприклад, якщо в рядку стану підсвічуються С і ЗС, то включений навчальний режим і випадковий порядок завдань.

Параметри оцінювання.

Коли ви створюєте новий тест і додаєте декілька завдань, то він вже, практично, готовий до використання. Але у ньому задані параметри по-замовчуванню. У тому числі і параметри оцінювання. Система оцінювання по-замовчуванню - п'ятибальна (на "5" - від 85%, на "4" - від 70%, на "3" - від 50%, така шкала найчастіше рекомендується в методичній літературі по оцінюванню тестів в школі). Але, якщо вас не влаштовує дана шкала оцінювання, або у вашій школі використовується, наприклад, двенадцятибальна система оцінювання, то ви можете її змінити обравши команду «*Параметри тесту - Оцінювання...*»

У вікні, що відкрилося, для зручності ви можете вибрати одну з готових шкал оцінювання або створити свою. Система оцінювання може бути від 2-х бальної до 100 бальної. Для кожного балу шкали вам необхідно вказати мінімальний відсоток отриманих за тест балів для отримання даної оцінки. Якщо потрібно, то кожного балу можна задати назву, яка замінить цифрову оцінку.

6.6. Система TSExam - веб орієнтована система тестування

TSExam - платформи- та мово-незалежне програмне забезпечення призначене для створення, проведення та управління тестами. Система вільнорозповсюджувана (ліцензія GNU - AGPL v.3) з відкритим вихідним кодом написаним мовою PHP та використовує для збереження MySQL бази даних. Таким чином ні сама система, ні необхідне для її функціонування програмне забезпечення не вимагає фінансових вкладень для не комерційного використання.

Системні вимоги.

Для використання системи Ви повинні мати повнофункціональний Веб-сервер. Для ознайомлення із системою, чи для використання в невеликих мережах, можна скористатися так званим віртуальним сервером – Denver, але в такому разі неможливо гарантувати відповідний рівень безпеки.

Система вільнорозповсюджувана з відкритим вихідним кодом.

Відкритість забезпечує надійність програмного забезпечення і якість, підтримуючи незалежний огляд і швидку еволюцію початкового коду. Загальні переваги відкритого вихідного коду:

- **Відкритість:** доступ до коду дозволяє легко вирішувати проблеми, чи створювати модифікації, тому що немає одного розробника в руках якого зосереджено процес розробки.
- **Стабільність:** доступність коду дозволяє проводити незалежний та всебічний моніторинг системи та усунення її недоліків, таким чином з часом досягається значно більша стабільність та надійність програмного забезпечення.

- Простота модифікацій: доступність вихідного коду дозволяє легко модифікувати систему під власні вимоги.
- Якість: широке коло користувачів чи незалежних розробників не тільки забезпечують стабільність, а і якість системи створюючи нові можливості та виправляючи помилки.
- Іновації.
- Безпека.

«Паперове» тестування.

Одна з головних переваг системи підтримка «паперового» тестування. Створений тест може бути роздрукований і розданий для виконання студентам, якщо неможливо організувати одночасне комп'ютерне тестування великої групи студентів. Результати скануються та завантажуються в систему. Оцінювання проводиться автоматично з відсканованих бланків. Таким чином забезпечується об'єктивність та ведення загальної статистики по кожному учню незалежно від способу виконання тестових завдань.

Імпорт та експорт даних.

Інформація про користувачів може бути експортована та імпортована використовуючи XML або CSV формат, що полегшує адміністрування системи. Також можна імпортувати питання в наведених форматах, та експортувати в форматі PDF.

Основні розділи системи.

Користувачі

У даному розділі містяться засоби для створення та корегування користувачів, груп користувачів, призначення користувачів до груп, імпорту користувачів, перегляданню списку он-лайн користувачів та перегляданню статистики роботи користувача.

Створення груп передбачає можливість надання права проходження тесту визначеній групі і у визначений час.

У системі можлива робота тільки зареєстрованих користувачів. Також передбачено розподіл прав користувачів на різні рівні, що дає можливість надавати різним користувачам/групам доступу до різних компонентів системи.

ТСExam використовує потужну систему захисту, що базується як на кукі так і на перевірці IP адреси для запобігання несанкціонованого доступу.

Модулі

У системі ТСExam існує ієрархічна система банку питань. Вона має наступну структуру: модуль - тема - питання - відповідь. Всі структурні елементи створюються та корегуються саме в даному розділі. Також є можливість перегляданню списку питань. Імпорту питань та управління файлами. Що використовуються в тестах.

Ви можете створити необмежену кількість тем, які можуть використовуватися в межах одного чи декількох тестів, та необмежену кількість питань в кожній темі. У тесті можливе використання як всіх питань з теми так і випадкової їх кількості.

Модулі Опис, теми, питання та відповіді підтримують розширене форматування з використанням зображень, форматowanego тексту, формул у форматі TEX.

Тести

Цей розділ включає форму, щоб створювати і змінювати різні тести, форма оцінювання питання з відкритою відповіддю; а також форми для відображення або генерування PDF-версій результатів, деталізоване представлення тестів користувача і відображення статистики питань і відповідей.

ТСExam випадковим чином вибирає різні питання і, пов'язані з ними відповіді так, щоб усі тести для різних користувачів були б різними.

Також можливе генерування довільної кількості тестів в PDF форматі так, щоб їх можна було роздрукувати і використовувати для звичайного іспиту.

Для кожного тесту може бути визначений загальний час на його виконання; час початку та закінчення тестування; група користувачів яким дозволено виконання даного тесту; базова кількість балів за правильну, частково правильну та неправильну відповідь. Тести можуть бути як контролюючі, так і навчаючі, тобто передбачено можливість багаторазового проходження тесту та система коментарів до тестових завдань та відповідей.

6.7. Освітнє середовище Moodle: тестування та оцінювання

6.7.1 Загальна характеристика VLE Moodle, створення курсів

MOODLE (www.moodle.org) є платформою, дистрибутив якої розповсюджується безкоштовно за принципами ліцензії Open Source.

Кожний користувач має доступ до вихідного коду платформи і може його змінювати залежно від своїх цілей та бажань.

Незважаючи на те, що MOODLE, як правило, асоціюється з дистанційним навчанням, найбільш точно призначення системи відображає клас систем до яких вона відноситься – VLE, що означає віртуальне освітнє середовище.

Автор концепції платформи MOODLE австралієць - Martin Dougiamas. Головною його метою було створення платформи, відмінної від доступних на ринку, а саме такої, яка враховувала б педагогічні аспекти, що базуються на основах пізнавальної психології, а особливо одній з її течій, яка іменується конструктивізмом.

Конструктивізм припускає, що учень – активний суб'єкт, який самостійно створює свою власну систему знань, користуючись при цьому доступними йому джерелами знань. Роль вчителя (тут тьютора), в мотивуванні і підтримці своїх підопічних, полягає головним чином в постановці завдань і формулюванні запитань, що становлять для учнів дилеми. Результатом роботи є – рішення цих дилем, що сприяє виникненню в розумовому потенціалі учнів нових знань. Відповідно до основ суспільного конструктивізму, конструйоване знання найбільш ефективно, коли ми навчаємося в співпраці. Це можливе тоді, коли учень працює в групі, ділячись своїм власним досвідом і думками, і будучи відкритим для досвіду і думки інших.

Достоїнством платформи e-learning MOODLE є той факт, що почавши від її появи, тобто 1999 року, вона була модифікована і доповнена новими засобами і інструментами. Програмне забезпечення платформи написано на мові PHP і робить можливим використання безкоштовних, загальнодоступних баз даних (MySQL, PostgreSQL, і т.п.). Платформу MOODLE можна інсталювати в довільному операційному середовищі (MS Windows, Unix, Linux).

Відповідно до цих основ платформа MOODLE була оснащена рядом інструментів, які дають можливість співпрацювати на рівнях учень - учень і учень - вчитель. До цих інструментів належать:

- Голосування (опитування).
- Анкети.
- Чати.
- Опитування.
- Форуми.
- Уроки.
- Журнали.
- Тести.
- Тести HOT POTATOES QUIZ.
- Пакети SCORM.
- Словники.
- Семінари.
- Вікі.
- Завдання і т.д. (всього 35 модулів).

Основні налаштування курсу

Усі індивідуальні налаштування мають поряд “?” для пояснення їх призначення. Деякі з цих полів передбачають спеціальні коментарі:

- *Короткий опис* – це опис курсу, що буде відображено в списку доступних курсів. Якщо Вам доступні HTML редактори, Ви можете використати повне форматування, включаючи надрядкові індекси, підрядкові індекси, смайлики та ін.

- *Формат*. Курс може мати декілька різних форматів:

- формат-структура,
 - формат-календар,
 - формат-форум,
 - SCORM format,
 - LAWS course format,
 - Weekly format – CSS/No tables.
- *Формат-календар* розподіляє заняття по тижнях, із завданнями, дошками дискусій (форумами, тестами і т. д.), усі вони належать до

потижневого блоку. Організується розклад (тиждень за тижнем) з точним терміном початку та закінчення курсу.

- *Формат-структура* усе розподіляє за темами (чи уроками), незалежно від їхньої тривалості. Зовні нагадує Формат-календар за винятком того, що не має тимчасових обмежень і не прив'язаний до календаря.
- *Формат-форум* (Соціальний формат) побудований навколо форумів (дошки оголошень), що сприяє оголошенням та дискусіям. Курс організується на основі одного великого форуму. Може використовуватися не тільки як курс, але і як одна велика дошка повідомлень.

Потижневий та тематичний формати є більш корисними, але певна користь може бути й від творчого соціального формату.

Зауважимо, що потижневий та тематичний формати мають схожий вигляд, але в їх організації є велика різниця. *Тематичний* формат пристосований до занять, які структуровані у звичайному форматі, а *Потижневий* – до занять, у яких є уроки, що мають хронологічну динаміку.

- *Дата початку курсу*

- *Course enrollable (Hi, Tak, Date range)* – Доступність курсу (Hi, Tak), по замовчуванню Так. Якщо вибираємо опцію *Date range* – треба визначити дату початку доступу та закінчення.
- *Enrolment duration*: - Тривалість доступу (від 1 дня до 365, чи опція – необмежений). По замовчуванню доступ є необмежений (*unlimited*)
- *Enrolment expiry notification (Tak, Hi)*: Сповіщення закінчення Enrolment.
- *Threshold (1-30 днів)*: Поріг (Межа) сповіщення закінчення Enrolment.

- *Кількість тижнів/тем (від 1 до 52)* – відображує кількість тижнів чи тем, що є на сторінці Вашого курсу (за замовчуванням встановлено 10 тижнів чи 10 тем).

- *Груповий метод* – Це налаштування за замовчуванням для груп курсу. У Вас є три налаштування для груп:

- *Жодної групи* – заняття проводиться в одній великій групі. Усі один одного бачать.

- *Окремі групи* – кожна група відокремлена – групи не бачать одна одну.
- *Видимі групи* – студенти належать до груп, але групу можуть бачити одна одну.
- *Примусово* (налаштування, пов'язане з режимом групи) – якщо налаштовано “Ні,” тоді група може бути розподілена за кожним доданим модулем (окремий розподіл). У цьому випадку, групи занять налаштовуються за замовчуванням, але можуть бути змінені. Якщо налаштовано “Так,” тоді налаштування групи не можуть бути змінені на рівні розподілу – налаштування рівня занять завжди налаштоване.
- *Кодове слово* – пароль входу до курсу. Якщо Ви заповните це поле, студентам потрібно буде вводити пароль, коли вони уперше входять до курсу. Це для того, щоб до занять не приєднувалися незареєстровані особи. Ключем до реєстрації може бути що-завгодно – слово, число чи комбінація.
- *Доступ для гостя* – контролює чи можуть люди без створених профілів заходити до Вашого курсу. Тут налаштовано “Не допускати гостя” за замовчуванням, але це може бути змінено, щоб дозволити гостям входити, якщо вони мають ключ до реєстрації на занятті (пароль), чи усім гостям, навіть без ключа до реєстрації. Зауважте, що гості нічого не можуть змінювати у курсі – вони можуть лише читати чи переглядати зроблене.
- *Показати оцінки* – цей значок визначає, чи можуть студенти бачити оцінки чи завдання для оцінювання (такими є більшість із них). За замовчуванням це налаштовано на “Так”, отже, студент може бачити оцінку. Якщо налаштовано “Ні”, тоді студенти не бачать своїх оцінок.
- *Звіт про діяльність*. Звіт про діяльність доступний для кожного учасника і відображає його роботу в даному курсі. Він може містити докладний список дій користувача. Викладач завжди має доступ до звітів кожного студента із сторінки персональної інформації студента. Доступом студентів до звітів їх діяльності управляє викладач за допомогою налаштувань курсу.
- *Максимальний розмір завантаженого файлу* (до 2 Мб). Ця установка визначає максимальний розмір файлу який може завантажити студент в межах курсу. Вона не може бути більше значення, встановленого

адміністратором для всього сайту. Ви можете надалі зменшити це значення для окремих елементів вашого курсу.

- Курс *Meta (Так, Ні)* – можна вибрати один з варіантів Так чи Ні. Метакурс є оболонкою, об'єднує в собі декілька інших курсів.

6.7.2. Середовище Moodle, банк тестових завдань, створення тесту

Модуль тестів

Цей модуль дає можливість викладачу створювати завдання у тестовій формі, що містять питання з декількома варіантами відповіді, питань типу Так/Ні, питань, що вимагають коротких відповідей та інше. Тести можуть допускати декілька спроб знаходження відповіді. Кожна спроба автоматично перевіряється і викладач може урахувати відповіді, представити свій коментар або правильний варіант відповіді.

Цей модуль включає і пропонує інструменти, що створюють можливість оцінювання.

Час виконання тесту також може бути обмежений чи ні.

- Викладачі можуть визначати базу даних питань для використання у різних тестах.

- Тести можуть зберігатися у категоріях для полегшення доступу, і ці категорії можуть бути доступними для будь-якого курсу на сайті.

- Тести автоматично оцінюються і переоцінюються, якщо питання змінюються.

- Тести можуть бути обмежені часом, поза яким доступу до них немає.

- На вимогу викладачів тести можуть проводитися багаторазово, у них можуть показуватися правильні відповіді.

- Тестові питання і відповіді можуть бути рандомізовані (генеруватися випадковим чином) для захисту від списування. Тести підтримують HTML, малюнки, звукові та відео-файли, анімації та інші мультимедійні об'єкти.

- Тестові питання можуть бути імпортовані із зовнішнього текстового файлу. Кількість спроб проходження тестів може бути обмеженою.

Існують

- Питання в закритій формі (з одною чи багатьма правильними відповідями).
- Питання відкриті, що потребують короткої відповіді (слово чи фразу).

- Питання-альтернатива (Так/Ні).
- Випадкові питання.
- Числові питання.
- Питання вкладені відповіді.
- Питання на відповідність.
- Випадкове питання на відповідність.
- Ессе.

Питання може містити *описовий текст і графіку чи посилання до мультимедійних об'єктів*.

Викладач, при відповідному налагодженні, може зробити можливим студенту перегляд своїх результатів негайно ж після виконання тесту з переліком правильних і помилкових відповідей.

Важливі можливість тестування через Інтернет та виконання тесту кілька разів, наприклад перед, в час і після курсу, завдяки чому як студент так і викладач можуть спостерігати прогрес в навчанні.

Тести (Тест) також можуть мати коментар, де пояснено студентам, чому відповідь саме така.

Щоб додати *Тест (Контрольну роботу)*, оберіть «Тести» із меню «Додати елемент курсу». Ви опинитесь на сторінці «Редагування Тесту»:

До Ваших послуг є кнопки допомоги (“?”).

- Тест(и) має *дати (день, місяць, рік, година, хвилина) відкриття (початку тестування) та закриття (закінчення тестування)*. Студент не може виконати контрольну роботу перед часом її відкриття чи після часу закриття. Викладач у будь-який час може змінити ці налаштування.
- *Обмеження часу*. За замовчуванням тести не обмежені в часі і дозволяють студенту використовувати стільки часу, скільки необхідно для завершення тесту. Якщо Ви все ж обмежуєте час, то деякі додаткові можливості дозволять вам бути упевненими, що тест завершений протягом цього часу:
 - Підтримка Javascript в браузері стає обов'язковою - це дозволяє таймеру працювати правильно.
 - В спливаючому вікні відображається зворотний відлік часу.
 - Коли час закінчується, результати тесту передаються на сервер автоматично незалежно від того на скільки питань одержані відповіді.

- Якщо студенту вдалося «обдурити» систему і витратити на 60 секунд більше часу чим відведено, то спроба тестування автоматично оцінюється в нуль.
- *Питання на одній сторінці.* В більш довгих тестах (що мають 5 та більше питань), можна обмежити число питань на одній сторінці. При додаванні питань до модулю тестів межа кількості тестів на сторінці буде автоматично вставлена згідно з вашим налагодженням. Проте ви будете також завжди здатні пізніше змінити значення цього параметру на сторінці параметрів.
- *“Випадковий порядок питань”* (опції Так/Ні) змінює порядок питань тесту під час кожного його виконання студентом (для кожного окремого студента). Це запобігає списуванню студентами один в одного.
- *“Випадковий порядок відповідей”* - дуже схоже налаштування, але воно змінює порядок відповідей для питань множинного вибору чи питань на відповідність.
- *«Дозволено» («Кількість спроб»).* Налаштовує кількість разів виконання тесту одним студентом. Це може бути дуже корисно, якщо тест є вправою на закріплення, і студент може виконувати його стільки разів, скільки потрібно викладачеві (та кожна оцінка не повідомляється викладачеві).
- *“Залежність спроб”* - налаштовує, чи тест базується на попередніх. Якщо спроби у виконанні тесту дозволені, і налаштовано “Так”, тоді результати минулого тесту будуть уключені у цю спробу (включаючи коментар, якщо його увімкнено). Якщо ж налаштовано “Ні”, тоді тест буде “свіжим”(поновленим) кожного разу, коли студент буде його виконувати.
- *“Оцінювання”* дозволяє Вам налаштувати оцінювання тестів, які студент може виконувати багато разів. Ви можете обрати оцінку, дотримуючись
 - *найкращої оцінки,*
 - *середньої з оцінок,*
 - *першої оцінки (оцінки, отриманої при виконанні першої спроби),*
 - *останньої оцінки (оцінки, отриманої при виконанні останньої спроби).*
- *“Adaptive mode”* – Адаптивний метод (Навчальний режим). Якщо ви вибираєте для цього налаштування «Так» - студент має можливість відповідати багато разів. Проте набраний бал буде зменшено на величину визначену параметром штрафу.

Нарахування штрафів. («Apply penalties» - Застосуйте покарання). Ця установка не має ніякого результату, якщо прапорець в адаптивному методі не є встановлений. Вартість штрафу (Кількість покарання) вказується індивідуально для кожного питання під час його створення, налаштування параметрів або редагування. Це ніяк не впливає на тести обмежені єдиною спробою.

- *“Дозволити перегляд” (Студенти можуть переглянути)* дозволить студентам переглянути всю виконану контрольну роботу (тести). Якщо налаштовано “Так”, тоді студент може подивитися (переглянути) тест та всі колишні спроби після його закриття.
- *“Після відповідей показати коментар”* дозволяє відображати студенту коментар до питання. Ви набираєте коментар разом із питанням.
- *«У коментарі показати правильні відповіді»* може показати студентові правильні відповіді за допомогою коментаря. Щоб активувати цю опцію, налаштуйте її на “Так”.
- *Показ в "безпечному" вікні.* "Безпечне" вікно дозволяє забезпечити небагато більшої безпеки для тестів (запобігає копіюванню з інших вікон, чи перегляд інших документів).

Затримка часу між першою і другою спробою (від 30 хвилин до 7 діб, чи Пусто). Якщо ви встановлюєте затримку часу, то студентам доведеться чекати протягом того часу перед тим, як вони зможуть спробувати протестувати себе після першої спроби.

Вимагати пароль. Це поле необов'язкове. Якщо ви введете пароль, потім студенти також повинні ввести той самий пароль перед проходженням тесту.

Метод групи (Режим групової роботи)

Ви можете вибрати один з трьох групових режимів:

- *Немає груп* - немає ніяких груп, кожний студент - це учасник одної групи (студенти не діляться на групи, кожний є частиною одного великого співтовариства).
- *Окремі групи великого співтовариства* - студенти кожної групи можуть тільки бачити їх власну групу, інші є невидимими (студентам кожної групи здається що їх група - єдина, робота студентів інших груп для них не видима).

- *Видимі групами (доступні групи)* - кожен студент групи навчається в своїй власній групі, але має можливість також бачити інші групи (студенти кожної групи працюють тільки в межах своєї групи, але можуть бачити що відбувається в інших групах).

6.7.3. Аналіз тестових завдань та засоби оцінювання в середовищі Moodle

При використанні автоматизованих тестових систем виникає можливість реального застосування для оцінки якості використовуваних контрольних вимірювальних матеріалів формалізованих підходів, заснованих на методології теорії педагогічних вимірів. Стають можливими накопичення і статистична обробка результатів випробувань і отримання кількісних характеристик, що дозволяють оцінювати якість тестових контрольних вимірювальних матеріалів.

Один з відомих підходів для такого аналізу базується на відомій в теорії педагогічних вимірів однопараметричній моделі Раша. Практичне використання запропонованих цієї теорією підходів полегшується тим, що в системі управління навчанням Moodle є розвинені вбудовані засоби для автоматизації обчислення за результатами виконання тестових завдань статистичних показників, що дозволяють здійснювати об'єктивну оцінку якості тестових завдань з точки зору їх здатності служити засобом виміру рівня підготовки випробовуваних.

Для здійснення Moodle такої обробки результатів тестування необхідно на вкладці перегляду результатів тестування перейти до пункту «*Аналіз питань*».

Обробка результатів випробувань, необхідна для отримання характеристик якості тестових завдань, полягає в наступному.

Відповідно до педагогічної теорії вимірів отримані випробовуваними результати випробувань представляються у вигляді матриці, зображеної нижче на рисунку.

		Испытуемые ($i = 1, 2, \dots, N$)					$p_j = \sum x_{ij}$
Задания ($j = 1, 2, \dots, K$)		x_{11}	x_{12}	...	x_{1N-1}	x_{1N}	p_1
		x_{21}	x_{22}	...	x_{2N-1}	x_{2N}	p_2
	
		$x_{K-1 1}$	$x_{K-1 2}$...	$x_{K-1 N-1}$	$x_{K-1 N}$	p_{K-1}
		x_{K1}	$x_{K 2}$...	x_{KN-1}	x_{KN}	p_K
	$s_i = \sum x_{ij}$	s_1	s_2	...	s_{N-1}	s_N	

Матриця результатів тестових завдань

У стовпцях цієї матриці знаходяться результати, отримані конкретними випробовуваними за кожним тестовим завданням, а по рядках розміщуються результати кожного випробовуваного за відповідним тестовим завданням. У нижньому рядку представлені сумарні значення балів, набраних кожним випробовуваним за тест в цілому ($S_i = \sum x_{ij}$). Ці значення використовуються для формування оцінки рівня підготовки випробовуваних з певного розділу навчального матеріалу. У правому стовпці приведені суми балів, отриманих усіма випробовуваними для кожного тестового завдання (для j -го завдання - $P_j = \sum x_{ij}$). За величиною значень P_j вже можна судити про міру реальної легкості (труднощі) конкретних тестових завдань для випробовуваних в цій групі - чим більше випробовуваних правильно відповіло на конкретне завдання, тим це завдання в середньому легше для виконання. Для практичного використання, проте, як міра легкості (труднощі) завдань придатніші відносні значення такого роду показників. Це індекс легкості завдання (ІЛ) і індекс труднощі завдання (ІТ) :

$$ІЛ_j = \frac{x_{срj}}{x_{максj}} = \frac{\sum_i x_{ij}}{N \cdot x_{максj}} = \frac{p_j}{N \cdot x_{максj}},$$

$$ІТ_j = 1 - ІЛ_j$$

де $x_{срj}$ - середнє значення балів, набраних усіма випробовуваними за виконання j -го завдання; $x_{максj}$ - максимально можлива кількість балів за виконання j -го завдання.

Ця характеристики (ІЛ і ІТ) тестових завдань є мірою легкості (трудності) завдань для конкретної групи випробовуваних. Значення цих параметрів залежать від рівня підготовки випробовуваних і не залежать від їх числа.

Іншою характеристикою, яку дозволяють обчислювати вбудовані в систему Moodle засоби аналізу результатів тестових завдань, є дисперсія цих результатів (чи їх середньоквадратичне відхилення). Дисперсія результатів виконання тестових завдань обчислюється за формулою:

$$\sigma_j^2 = \frac{1}{N} \sum_i (x_{ij} - x_{cpj})^2.$$

Дисперсія результатів (чи середньоквадратичне відхилення) характеризує розкид балів, отриманих усіма N випробовуваними при відповіді на конкретне (j - е) завдання тесту. Якщо усі випробовувані відповідають на завдання однаково, то що характеризується цим параметром розкид отриманих випробовуваними балів дорівнюватиме нулю. Нульове або низьке значення розкиду отримуваних при виконанні конкретних тестових завдань результатів в групі випробовуваних свідчить про низьку диференціюючу здатність цих завдань, тобто їх слабку здатність розділяти випробовуваних в групі по мірі їх підготовки. Такі завдання підлягають виключенню з тесту. Чим вище диференціююча здатність тестових завдань (більше величина розкиду результатів), що входять в тест, тим вище якість тесту.

Ще однією важливою статистичною характеристикою якості тестових завдань, яку дозволяють обчислювати засоби Moodle, являється коефіцієнт кореляції значень відповідей, отриманих випробовуваними при виконанні конкретного завдання, з результатами виконання цими ж випробовуваними тіста в цілому, який розраховується по наступній формулі:

$$КД_j = \frac{1}{N \cdot \sigma_j \cdot \sigma_s} \sum (x_{ij} - x_{cpj})(s_i - s_{cp}),$$

де s_{cp} - середнє значення балів, отриманих усіма N випробовуваними за тест в цілому; s_i - сума балів і -го випробовуваного за виконання усіх завдань тесту.

Цей показник може набувати значень - 1 і +1 і в системі Moodle називається Коефіцієнтом Диференціації (КД). Його використання як характеристика здатності конкретного тестового завдання правильно диференціювати випробовуваних по мірі їх підготовки, засновано на тому,

що в якісному тісті в середньому вищі бали при відповіді на кожне конкретне завдання тесту повинні отримувати випробовувані, набираючи вищу суму балів за виконання тесту в цілому. Тобто між цими значеннями має бути позитивна кореляція. Негативні ж значення цього коефіцієнта свідчать про те, що при відповіді на це завдання "слабкіші" випробовувані в середньому отримують вищі бали, ніж "сильніші" випробовувані. Очевидно, що такі завдання, можливо, через допущені помилки в їх формулюванні або які-небудь інші причини не є справжніми тестовими завданнями, і їх також слід видаляти з тесту.

6.8. Комп'ютерні моделі педагогічного тестування

6.8.1. Основні вимоги та класифікація

Сучасний стан розвитку інформаційно-комп'ютерних технологій сприяє впровадженню в навчальних закладах систем автоматизованої діагностики, зокрема комп'ютерного педагогічного тестування. Тому доцільно охарактеризувати моделі комп'ютерного тестування.

Комп'ютерна система тестування повинна забезпечувати такі основні можливості: підтримувати створення та редагування бази тестових завдань різних форм; здійснювати вибірку завдань з бази за деяким алгоритмом/алгоритмами; забезпечувати оцінювання відповідей на завдання; інформування про результати тестування. Форма представлення бази тестів у комп'ютері є різною. Це може бути або окремий файл певного виду, або база даних. Окрім того, тестові завдання можуть групуватися за певними властивостями або залежати одне від одного. У цьому випадку між завданнями встановлюються зв'язки. Форма представлення бази та зв'язки між завданнями визначають організацію бази завдань, яку ми назвемо *структурою бази* тестових завдань.

Генерування завдань здійснюється за певним алгоритмом. Вибір алгоритму залежить від структури бази. Структурна організація бази визначає можливі варіанти алгоритмів. Одночасно, структура бази й алгоритм впливають на методи оцінювання відповідей на завдання.

Модель педагогічного тестування визначається алгоритмом генерування та методом оцінювання. Можливість реалізації певної моделі

на комп'ютері залежить від способу організації бази завдань. Тому, при визначенні комп'ютерної моделі слід враховувати структуру бази. Отже, *комп'ютерну модель педагогічного тестування* визначаємо як схему структури бази завдань, вибірки завдань із бази та оцінювання результатів тестування.

Виокремимо такі **види комп'ютерних моделей** педагогічного тестування: лінійні, ієрархічні та мережні. Ці назви моделей пов'язані зі структурною організацією бази завдань, адже саме структура бази є визначальною у виборі конкретної моделі комп'ютерного тестування. У праці О. Колгатіна пропонується така класифікація методів діагностики за структурою бази завдань: фіксована послідовність завдань; множина рівноправних завдань; кожне завдання з визначеним рівнем складності; структура за рівнями складності; структура за рівнями розумової діяльності; структура за темами (елементами навчального матеріалу) і рівномірним вибором завдань з кожної теми; послідовність завдань за зростанням складності з автоматичним поділом послідовності завдань на блоки й випадковим вибором завдань з блоку; структурування бази завдань за елементами навчального матеріалу, рівнями розумової діяльності, мінімальна одиниця структури – блок паралельних завдань, тест будується як послідовність блоків, випадковий вибір здійснюється тільки у межах блоку; семантична мережа (дерево за структурою навчальної дисципліни) [1]. Визначені нами назви видів комп'ютерних моделей охоплюють усі методи діагностики, запропоновані О. Колгатіним, а їх назва пов'язана з формою організації структури бази завдань.

6.8.2. Лінійні моделі педагогічного тестування

Лінійні моделі є найпростішими, що найчастіше реалізуються в комп'ютерних системах тестування. Будь-яка лінійна модель створюється таким чином:

1. База складається зі скінченної кількості завдань або груп завдань. Між завданнями (групами завдань) не встановлено ніяких зв'язків, тобто формально завдання (групи завдань) вважаються рівноправними.
2. Кожне завдання має певну вагу.
3. Алгоритми вибірки завдань із бази: строга послідовність, випадкова вибірка, комбінування випадкової вибірки із строгою послідовністю.

4. Завдання, що вибрані для тестування, оцінюються залежно від їх ваги та виставляється підсумкова оцінка за тест.
5. Додатково можуть накладатися обмеження на час виконання завдань або всього тесту.

Розглянемо детальніше вимоги, що пред'являються до лінійної моделі. Завдання у базі можуть бути різними за формою. Важливо, що формально вони вважаються незалежними. Кожному завданню обов'язково приписується вага – це деяке додатне число, найчастіше – натуральне. Найпростіший варіант – вважати вагу кожного завдання рівною 1. Але вага може визначати рівень складності, рівень засвоєння, оцінку умінь або навичок, тощо. Загалом, вага завдання може визначати його оцінку, якщо воно виконане правильно. Алгоритми вибірки завдань із бази прості. Можна пропонувати всім учасникам тестування завдання, вибираючи їх із бази в деякій фіксованій послідовності. Краще використовувати випадкову вибірку. Тоді всі особи, що тестуються, отримають однаковий набір завдань, але в різних послідовностях.

Наведемо приклад лінійної моделі, яку доцільно використовувати при тестуванні. Нехай база завдань складається із n груп, а в кожній групі по k завдань. Розглядається випадок, коли в кожній групі однакова кількість завдань, хоч це не обов'язково. Доцільно, щоб в межах групи завдання були рівноцінними й охоплювали якусь одну навчальну тему або перевіряли засвоєння певного поняття, уміння, тощо. Під час тестування випадковим чином вибирається номер групи та з цієї групи теж випадковим чином пропонується s завдань ($1 \leq s \leq k$), причому слід випадково перебрати всі номери груп. Таким чином з кожної групи буде вибрано s завдань, а загальна кількість вибраних завдань дорівнює $n \cdot s$.

Кожне вибране завдання має вагу V_{ij} , $i=1,2,\dots,n$, $j=m_{i1},m_{i2},\dots,m_{is}$, де $m_{i1},m_{i2},\dots,m_{is}$ – номери завдань, вибрані з i -ої групи. Завдання можна оцінювати числом R_{ij} так: $R_{ij}=V_{ij}$, якщо відповідь правильна; $R_{ij}=k \cdot V_{ij}$, якщо відповідь частково правильна (k – деякий коефіцієнт корегування); $R_{ij}=0$, якщо відповідь повністю неправильна. Загальний результат тестування обчислюється за формулою

$$R = \frac{\sum R_{ij}}{\sum V_{ij}}, \quad R \in [0;1].$$

Залежно від величини R виставляється відповідна оцінка. У запропонованому прикладі можливо додатково встановити обмеження на час виконання завдання. Тоді оцінка за виконане завдання залежатиме від його ваги й використаного часу.

Розглянутий приклад лінійної моделі дозволяє забезпечити паралельність вибору завдань. Усі учасники тестування виконують однаковий за складністю набір завдань. Це дозволяє забезпечити досить високу надійність результатів тестування.

На основі лінійної моделі можлива реалізація всіх відомих класичних моделей педагогічного тестування з урахуванням ваги завдання та часу відповіді на нього.

6.8.3. Ієрархічні моделі педагогічного тестування

Ієрархічна модель будується таким чином:

1. Створюється база зі скінченної кількості завдань або груп завдань.
2. Кожне завдання має певну вагу.
3. База тестів упорядковується за зростанням ваг завдань або ваг груп завдань.
4. Вибірка завдань із бази повинна підпорядковуватися встановленому порядку. Це означає, що за потреби можна рухатися по базі в двох напрямках – за зростанням або спаданням ваг.
5. Алгоритми вибірки завдань із бази: повна або часткова вибірка за зростанням ваг, вибірка завдання залежно від відповіді на попереднє.
6. Завдання, що вибрані для тестування, оцінюються залежно від їх ваг та виставляється підсумкова оцінка за тест.
7. Додатково можуть накладатися обмеження на час виконання завдань або всього тесту.

В ієрархічних моделях за вагу завдання доцільно брати рівень складності. Однак, це може бути інший показник, зокрема, рівень засвоєння навчального матеріалу. Якщо база складається з окремих завдань, то в принципі необов'язково, щоб усі завдання мали різну вагу. Але немає сенсу створювати ієрархічну модель для завдань з однаковою вагою. Доцільно створювати базу з груп завдань і кожній групі призначити вагу (усі завдання групи мають однакову вагу, і вона вважається вагою групи). Ваги груп повинні бути різними, і тоді база упорядковується за зростанням ваг груп.

Наведемо приклад ієрархічної моделі. Нехай вагою кожного завдання є рівень складності, а база складається з n груп завдань. Кожна група містить k завдань, усі завдання в межах групи мають однакову вагу V_i , $i=1,2,\dots,n$. Число V_i буде також вагою i -ої групи. Базу створюємо так, щоб усі числа V_i були різними. Упорядковуємо цю базу за зростанням ваг груп. Для таким чином побудованої бази розглянемо деякі алгоритми вибірки завдань.

1. Обходимо всі групи за зростанням їхніх ваг та з кожної групи випадковим чином вибираємо s завдань ($1 \leq s \leq k$). Такий алгоритм дозволяє реалізувати модель тестування за зростанням складності. Кожному студенту буде запропоновано $n \cdot s$ завдань, а різні студенти отримають різні набори завдань. Оцінювання відповідей може бути таким, як у наведеному вище прикладі лінійної моделі.

2. Випадковим чином вибираємо s завдань із групи з найменшою вагою. Якщо студент правильно відповів на всі завдання, пропускаємо декілька груп, наприклад r , і знову випадковим чином вибираємо s завдань із $(r+1)$ -ої групи. Якщо студент завжди дає правильні відповіді на вибрані завдання, продовжуємо обхід з кроком r усієї бази, поки не досягнемо її кінця. У цьому випадку вважаємо, що студент успішно пройшов тестування й зараховуємо йому правильні відповіді на $n \cdot s$ тестів, тобто ніби він проходив тестування за першим алгоритмом.

Припустимо, що у процесі тестування з кроком r ми в h -ій ($1 \leq h < n$) групі випадково вибрали s завдань, і вперше студент не відповів правильно на всі завдання. Попередні s завдань, на які студент відповів правильно, були вибрані з $(h-r)$ -ої групи. Тоді повертаємося до $(h-r+1)$ -ої групи завдань і, починаючи з неї, працюємо за алгоритмом 1. Зараховуємо студенту правильні відповіді на $s \cdot (h-r)$ завдань із груп $1, \dots, h-r$, а відповіді на решту завдань отримуємо за алгоритмом 1.

Цей алгоритм теж реалізує модель тестування за зростанням складності, але він дозволяє швидше отримати результати тестування при правильних відповідях студента. Можна також раніше припинити тестування, якщо студент не дав правильної відповіді на жодне з завдань із декількох сусідніх груп. Тоді існує висока ймовірність, що ще складніші завдання він теж не зможе виконати.

3. *Визначення рівня складності.* У базі n рівнів (груп завдань), вага кожного дорівнює V_i , $i=1,2,\dots, n$. Позначимо через ρ номер шуканого рівня

складності, ($1 \leq \text{nom} \leq n$). Під шуканим рівнем розуміємо таку групу завдань, на які студент з високою ймовірністю відповідає вірно і на складніші завдання в основному відповідає невірно. В алгоритмі на кожному кроці шукається номер nom потрібної групи в інтервалі $[l, r]$, який є частиною загального інтервалу $1, \dots, n$. Якщо відсоток правильних відповідей $g \geq 0.75$, то наближено шуканий рівень знайдений. Далі здійснюється уточнення рівня. Для відсотка $g < 0.75$ пошук здійснюється серед рівнів меншої складності. Якщо кожного разу відсоток правильних відповідей буде менший 0.75 , то шуканого рівня не існує.

Ієрархічну модель можна створювати за рівнями засвоєння навчального матеріалу. Для кожного рівня створюється блок, який складається з груп завдань. Кожна група складається з тестів однакової ваги. У межах блоку групи упорядковуються за зростанням ваги, а в цілому вся база упорядкована за зростанням рівнів блоків. До такої бази можна застосовувати вищенаведені алгоритми 1 і 2, внівши в них необхідні зміни. Також можна визначати рівень засвоєння, який відповідає рівню підготовленості студента.

На основі ієрархічної моделі можна також створити базу, яка буде складатися із блоків, що упорядковані за послідовністю вивчення навчального матеріалу.

Загалом, за допомогою ієрархічних моделей можна реалізувати модель тестування за зростання складності та частково-адаптивні моделі:

- кожному студенту генерувати набір завдань, який відповідає його рівню підготовленості;
- вибірка завдань, яка відповідає послідовності вивчення навчального матеріалу;
- вибірка завдань на основі відповіді студента.

6.8.4. Мережні моделі педагогічного тестування

Мережна модель будується таким чином:

1. Створюється база зі скінченної кількості завдань або груп завдань.
2. Кожне завдання має певну вагу.
3. Встановлюються зв'язки між парами тих завдань, для яких вони існують.
4. Алгоритми вибірки завдань з бази залежать від встановлених зв'язків.

5. Завдання, що вибрані для тестування, оцінюються залежно від їх ваг i , можливо, від існуючих зв'язків між завданнями.

6. Додатково можуть накладатися обмеження на час виконання завдань.

Мережні моделі створювати складно. Але, напевне, лише на їх основі можна реалізувати адаптивне тестування. Мережні моделі доцільно створювати для перевірки знань з усієї навчальної дисципліни або її розділу або з певної галузі знань.

Розглянемо приклад нескладної мережної моделі. Нехай створений набір із n тестів для перевірки якогось розділу навчальної дисципліни. Кожний тест має певну складність. Структуру бази представимо орієнтованим графом. Вершинами графа вважатимемо створені завдання, тобто кожне завдання визначає деяку вершину. Припустимо, що деякі завдання залежні між собою. Залежність двох завдань будемо тлумачити таким чином: для правильної відповіді на друге завдання необхідні знання чи вміння, що перевіряються першим завданням. Тоді вважаємо, що друге завдання залежить від першого. Проведемо дугу (орієнтовану стрілку) від вершини першого завдання до вершини другого. Якщо провести дуги між усіма парами залежних завдань бази, то структура бази буде орієнтованим графом.

Занумеруємо вершини графа числами від 1 до n . Кожне завдання бази матиме свій номер. Тоді інформацію про структуру бази, а саме про орієнтований граф, можна зберігати у вигляді матриці суміжностей графа. Це квадратна матриця порядку n , елементами якої є 0 або 1.

У графа бази будуть завдання, у вершини яких не входить жодна дуга. Такі завдання назвемо початковими. Також є завдання, з вершин яких не виходять дуги. Їх назвемо кінцевими. Можуть зустрічатися вершини, в які не входять та з яких не виходять дуги. Відповідні завдання віднесемо до початкових.

Оцінка відповідей на завдання визначатиметься їхньою вагою, якщо не враховувати залежностей між ними. Нехай завдання з номерами r і f залежні, тобто з r проведена дуга до f . Припустимо, що під час тестування студент відповів невірно на завдання r і вірно на f . Оскільки f залежить від r , то оцінка завдання f повинна залежати не тільки від його ваги, а також від відповіді на r . У цьому випадку можна оцінити f як добуток деякого множника корегування на вагу f .

Наведемо можливий алгоритм вибірки завдань зі створеної бази, вважаючи, що початковим завданням відповідають вершини з номерами 1,2,..., k. Спочатку пропонуємо для тестування всі завдання, які можна обійти по дугах від вершини 1 аж до кінцевих. Паралельно заповнюється таблиця оцінювання відповідей лише на основі ваг завдань та створюється список номерів вибраних тестів. Далі, аналогічно, вибираємо всі завдання, які можна обійти по дугах від вершини 2 аж до кінцевих, і т. д. При цьому, ті завдання, які вже були вибрані раніше, ми пропускаємо. Алгоритм завершується, коли вибрані всі завдання в напрямку дуг, починаючи з k-го. Студенту будуть запропоновані для тестування всі завдання бази та буде заповнена таблиця оцінок відповідей. Далі потрібно перерахувати таблицю оцінок, врахувавши залежності між завданнями.

Розглянутий приклад мережної моделі організації бази тестів дозволяє досягти досить високої надійності результатів тестування.

Загалом, мережні моделі бази завдань будуються по-різному, і зв'язки між тестами встановлюються на основі різних параметрів. Методично доцільно, щоб вершинами мережі були не окремі завдання, а групи завдань об'єднані за певною властивістю. Можливо розробляти різні алгоритми генерування завдань для руху по мережі та на основі мережних моделей реалізовувати елементи адаптивного контролю.

Питання для самоконтролю:

1. Які основні вимоги інформаційної безпеки тестування?
2. Як забезпечується виконання вимог інформаційної безпеки тестування?
3. Чи слід засекречувати відповіді на питання тесту і чому?
4. Основні способи фальсифікації при комп'ютерному тестування.
5. Як можна запобігати списуванню при комп'ютерному тестуванні?
6. Чи можна дозволити використовувати паперові носії інформації при комп'ютерному тестування? Обґрунтуйте свою думку.
7. Чому для тестування «вигідне» рейтингове оцінювання?
8. Чи потрібен контроль документів при комп'ютерному тестуванні? Чому?
9. Яких способів фальсифікації більше, технічних чи обумовлених людським фактором?
10. Хто ввів поняття Web 2.0? І яке початкове його визначення?
11. Яка основна відмінність Web 2.0 порівняно з Web 1.0?
12. Що таке фолксономічна класифікація і в чому її переваги над таксономією?
13. Структура сервісів Web 2.0
14. Основні соціальні сховища закладенок та їх призначення?
15. Найпоширеніші відео сервіси та їх використання
16. Блоги чи форуми, що краще в освітніх проектах?
17. Особистий блог в Вашій професійній діяльності як може бути використаний?
18. Переваги он-лайн офісних пакетів, наведіть приклади використання
19. Які тип завдань можна створити за допомогою JQuery?
20. Які тип завдань можна створити за допомогою JCloze?
21. Які тип завдань можна створити за допомогою JCross?
22. Які тип завдань можна створити за допомогою JMix?
23. Які тип завдань можна створити за допомогою JMatch?
24. В чому особливість змішаного варіанту тестового завдання в JQuery?
25. Чи можна створювати альтернативні правильні відповіді в Jcloze?
26. Як автоматично можна згенерувати кросворд у Hot Potatoes?
27. В яких форматах можна створювати питання на встановлення відповідності у Hot Potatoes?
28. Як використовується поле альтернативи в питаннях на встановлення послідовності у Hot Potatoes?
29. Яке призначення вхідного контролю?
30. Який етап комп'ютерної атестації мусить бути першим?
31. Який етап слідує відразу за основним тестуванням і чому?

32. Які вимоги до комп'ютерної атестації можуть покращувати мотивацію до навчання?
33. Яким чином можна забезпечити у TCExam варіативність комп'ютерного тесту?
34. Які відмінності у TCExam між наступними типами сценаріїв комп'ютерного тестування:
 - «Стандарт. Множинний вибір» та «Стандарт. Випадковий вибір»;
 - «Пошук однозначної відповідності» та «Пошук неоднозначної відповідності»;
35. Опишіть та наведіть приклад застосування типу сценарію тестування «Визначення оцінки» у TCExam
36. З яких програм складається система MyTest X?
37. Які типи тестових завдань можна створювати в системі MyTest X і?
38. Які режими тестування передбачені в системі MyTest X?
39. Яка складова MyTest X відповідає за ведення статистики?
40. Чи можна створювати власні шкали оцінювання у MyTest X?
41. Чи можна використовувати систему MyTest X для локального тестування?
42. Як можна обмежувати доступ учнів до проходження тесту у MyTest X?
43. Як створити інструкцію до тесту у MyTest X?
44. Чи можна систему використовувати без повноцінного веб-серверу?
45. Які типи тестових завдань можна створювати в даній системі?
46. Які переваги відкритого програмного коду?
47. Чи є, і які, засоби підтримки «паперового» тестування?
48. В які формати можна експортувати тестові завдання?
49. Чи можна і в якому вигляді імпортувати списки студентів?
50. Яка структура банку тестових завдань?
51. Які обмеження можна встановити для тесту?
52. Чи ведеться статистика проходження тесту і де її можна переглянути?
53. Чи можна додавати в тест випадкові питання?
54. Що означає абревіатура VLE?
55. В чому суть конструктивізму в навчанні?
56. За якими правилами розповсюджується система Moodle?
57. Які формати курсів підтримуються системою Moodle?
58. В чому відмінність між форматом-структурою і форматом-календарем у Moodle?
59. Як визначається кількість блоків (складових частин) курсу у Moodle?
60. Як можна обмежити доступ до курсу у Moodle?
61. Що таке мета курсу у Moodle?
62. Навіщо створюються категорії питань у Moodle?

63. Чи можна переоцінити спробу учня у Moodle?
64. Які типи питань підтримуються системою Moodle?
65. Як обмежити час на проходження тесту у Moodle?
66. Чи можна у Moodle визначати строки проходження тесту і як?
67. Які у Moodle з параметрів тесту запобігають списуванню?
68. Які методи оцінювання можливі у Moodle, якщо дозволено декілька разів виконувати тест?
69. Що означає адаптивний режим в Moodle?
70. Чи можна у Moodle обмежувати час між спробами проходження тесту?
71. Що передбачає параметр «Режим групової роботи» у Moodle?
72. Які статистичні характеристики питань автоматично обчислюються в Moodle?
73. Що таке рівень легкості питання і як вона обчислюється у Moodle?
74. Як співвідносяться індекси легкості і трудності і як вони обчислюються у Moodle?
75. Що показує середньоквадратичне відхилення і як воно обчислюється у Moodle?
76. Що таке коефіцієнт дискримінації і як він обчислюється у Moodle?
77. Що таке диференціююча здатність питання і як вона визначається у Moodle?
78. Що розуміють під комп'ютерною моделлю педагогічного тестування?
79. Як будується лінійна модель?
80. Наведіть приклади лінійних моделей.
81. Як будується ієрархічна модель?
82. Наведіть приклади ієрархічних моделей.
83. Як будується мережна модель?
84. У чому переваги і недоліки мережних моделей?

Список джерел, використаних у розділі VI

1. А. М. Анисимов. Работа в системе дистанционного обучения Moodle. Учебное пособие. – Харьков: ХНАГХ, 2009. – 292 с.
2. Бесплатные тесты онлайн. [Електронний ресурс].- Режим доступу: <http://tests.specialist.ru/> .- Назва з екрану
3. Булах І.Є. Комп'ютерна діагностика навчальної успішності.- К.: ЦМК МОЗ України, 1995.- 221с.
4. Булах І.Є. Теорія і методика комп'ютерного тестування успішності навчання: Автореф. дис. ... доктора пед. наук.- Київ, 1995.
5. Булах І.Є. Теорія комп'ютерного тестування.- К.: ЦМК МОЗ України, 1994.- 63с.
6. Колгатін О.Г. Автоматизована педагогічна діагностика у сучасному університеті [Електронний ресурс]. – <http://www.ime.edu-ua.net/em8/content/08kogumi.htm>.
7. Колгатін О.Г. Педагогічна діагностика та інформаційно-комунікаційні технології:

- монографія / О.Г.Колгатін.- Х.: ХНПУ, 2009.- 324 с.
8. Майстер-Тест [Електронний ресурс].- Режим доступу: <http://master-test.net/> .- Назва з екрану
 9. Морев И. А. Образовательные информационные технологии. Учебное пособие для учащихся педагогических специальностей вузов и слушателей курсов повышения квалификации педагогических и управленческих кадров. – Владивосток: Издательство Дальневосточного университета, 2004
 10. Наследие АБВУ FormReader [Електронний ресурс].- Режим доступу: <http://www.abbyu.ru/formreader/> .- Назва з екрану
 11. О программе для тестирования знаний и конструкторе тестов x-TLS [Електронний ресурс].- Режим доступу: <http://xtls.org.ua/about.php> .- Назва з екрану
 12. Построение тестовых заданий в системе компьютерного тестирования знаний OpenTEST2 0 [Електронний ресурс].- Режим доступу: <http://opentest.com.ua/postroenie-testovyx-zadaniy-v-sisteme-kompyuternogo-testirovaniya-znaniy-opentest2/> .- Назва з екрану
 13. Разработка компьютерных заданий. Инструментальные программы. [Електронний ресурс].- Режим доступу: <http://www.itlt.edu.nstu.ru/hotpothelp.php> .- Назва з екрану
 14. УСАТИК 2 [Електронний ресурс].- Режим доступу: <http://www.usatic.narod.ru/files-desc.html> .- Назва з екрану
 15. Чельшкова М.Б. Адаптивное тестирование в образовании (теория, методология, технология)". - М: Исследовательский центр проблем качества подготовки специалистов, 2001. - 165 с.
 16. АБВУ TestReader [Електронний ресурс].- Режим доступу: <http://www.abbyu.ru/testreader/> .- Назва з екрану
 17. Computer-Adaptive Testing: A Methodology Whose Time Has Come. By John Michael Linacre, Ph.D. MESA Psychometric Laboratory University of Chicago MESA Memorandum No. 69. Published in Sunhee Chae, Unson Kang, Eunhwa Jeon, and J. M. Linacre, 2000. – 58 p.
 18. easyQuizzy [Електронний ресурс].- Режим доступу: <http://easyquizzy.ru/> .- Назва з екрану
 19. Hot Potatoes Home Page [Електронний ресурс].- Режим доступу: <http://hotpot.uvic.ca/index.php> .- Назва з екрану
 20. Quizful – тесты онлайн [Електронний ресурс].- Режим доступу: <http://www.quizful.net/test> .- Назва з екрану
 21. SunRav TestOfficePro [Електронний ресурс].- Режим доступу: <http://www.sunrav.ru/testofficepro.html> .- Назва з екрану
 22. TestBOX [Електронний ресурс].- Режим доступу: <http://www.testbox.softvea.ru/> .- Назва з екрану

VII. МОНІТОРИНГ ЯКОСТІ ОСВІТИ

У роділі розкрито поняття моніторингу та його окремих типів - моніторинг якості освіти, як кваліметричний інструмент. Представлені етапи проведення моніторингового дослідження, розкрито сутність моніторингу якості освіти, його мета та завдання, а також принципи моніторингових досліджень. Окремим пунктом коротко висвітлена історія розвитку моніторингу якості освіти, описані рівні функціонування системи моніторингу, види моніторингових досліджень та інформаційне забезпечення моніторингу якості освіти. Моніторинг якості освіти розглянутий, як системне дослідження освітнього процесу. Наведено інформацію про міжнародні моніторингові порівняльні дослідження якості освіти та про організації, що їх проводять (TIMSS, PISA, PIRLS, SITES).

7.1. Показники моніторингу та характеристики якості освіти

7.1.1. Моніторинг та моніторингові показники

Дефініція «моніторинг» уперше стала використовуватися в екологічних науках для постійного спостереження з метою контролю, прогнозування розвитку та охорони довкілля. Зазвичай розрізняють глобальний, регіональний, локальний екологічний моніторинг, провідною функцією якого є вивчення стану та прогнозування розвитку екологічного об'єкта спостереження.

Найзагальнішим чином **моніторинг** можна визначити як постійне спостереження за деяким процесом з метою виявлення його відповідності бажаному результату або вихідному стану.

Об'єкт моніторингу – система, на яку спрямовані конкретні моніторингові процедури. Основна особливість об'єктів моніторингу – динамічність. Усі об'єкти, вивчення яких здійснюється із застосуванням інформаційного моніторингу, постійно змінюються, розвиваються, зазнають зовнішніх впливів, що можуть викликати небажані зміни у функціонуванні об'єкта.

Предметом інформаційного моніторингу є стан системи, яка досліджується, у певні періоди часу і конкретні зміни в рамках цієї системи.

Суб'єктами моніторингу є носії моніторингових функцій, що здійснюють ці функції. Вони умовно поділяються на дві великі групи: суб'єкти, що надають інформацію, та суб'єкти, що збирають і обробляють інформацію. Тобто, суб'єктами в системі моніторингу виступають як установи, структури, так і окремі люди.

Комплекс моніторингових показників (індикаторів) – сукупність первинних і вторинних показників, здатних забезпечити цілісне уявлення про стан системи, якісні та кількісні зміни в ній.

Інструментарієм моніторингової діяльності є сукупність форм статистичної звітності, інформаційних стандартів, анкет, опитувальних аркушів та ін.

Інструментами моніторингової діяльності є різні технічні засоби, які використовуються суб'єктами моніторингу у своїй діяльності.

Моніторинг розглядають у зв'язку з якістю освіти. У педагогіці поняття "якість" – це системна методологічна категорія. Вона відображає ступінь відповідності результату поставленій меті.

7.1.2. Якість освіти та її основні характеристики

Якість освіти можна тлумачити як ступінь задоволення учасників освітнього процесу наданими навчальним закладом освітніми послугами або ступінь досягнення поставлених освітніх цілей і завдань. Також якість освіти можна розглядати як відповідність певній нормі, стандарту. Тобто якість визначає корисність, цінність об'єктів, їх здатність задовольнити певні потреби або реалізувати певні цілі, норми, тобто виражає адекватність вимогам, потребам, нормам. Відомо, що поняття норми не є абсолютним, незмінним, фіксованим. Отже, категорія «якість освіти» переглядається, змінюється залежно від умов, вимог суспільства та часу.

Структура якості освіти:

- якість навчального процесу як результату педагогічної діяльності;
- якість освітніх програм, навчальної літератури, підручників і посібників;
- якість професійної підготовки і кваліфікації педагогічних і науково-педагогічних кадрів;

- якість ресурсного забезпечення та навчального середовища, у якому відбувається освітній процес (правового, фінансового, кадрового, науково-методичного, матеріально-технічного);
- якість особистісних рис і здібностей учнів;
- якість державно-громадського управління системою освіти;
- результативністю та розгалуженістю інституту зовнішнього оцінювання якості освіти (національної системи моніторингу якості освіти);
- якість проведення, інтерпретації результатів моніторингових досліджень у системі освіти;
- якість та ефективність державного управління освіти.

Цілі оцінювання якості освіти:

- удосконалення педагогічних засобів;
- визначення результативності навчання та виховання учнів;
- порівняння навчальних закладів, встановлення рейтингу;
- визначення ефективності використання коштів і ресурсів;
- планування та прогнозування розвитку освітньої галузі певного регіону;
- формування освітньої політики держави, регіону;
- визначення престижності та конкурентоспроможності національної системи освіти.

Основні групи показників якості освіти:

- соціально-економічний потенціал розвитку освіти;
- демографічні умови розвитку освітита фізична і фінансова доступність освіти;
- характеристика матеріально-технічного потенціалу освіти;
- характеристики управлінського і педагогічного кадрового потенціалу (вікові, кваліфікаційні, наявність вакансій за регіонами, навчальними предметами тощо) системи освіти;
- нормативно-правове забезпечення системи управління якістю освіти;
- джерела фінансування системи освіти, обсяги та розподіл коштів;
- соціальна захищеність учасників освітнього процесу;
- соціальна адаптація випускників навчальних закладів;
- особистісний потенціал учнівської молоді (учнівські досягнення в інтелектуальній сфері, рівень фізичного та морального здоров'я);
- узагальнені показники ефективності функціонування освіти.

7.2. Моніторинг якості освіти як кваліметричний інструмент

У педагогіці моніторинг використовується у двох напрямках: управлінський – для прийняття відповідних рішень (як системи збирання, накопичення інформації про стан освіти, обробки показників динаміки її розвитку) та наукового дослідження – для визначення і створення оптимальних умов здійснення освітнього процесу з метою досягнення максимально позитивного результату.

Під **моніторингом в освіті** розуміється спеціально організована і постійно діюча система збирання, зберігання та аналізу інформації, проведення додаткових інформаційно-аналітичних обстежень і оцінки стану, тенденцій, ризиків і прогнозу розвитку освіти для прийняття управлінських рішень у сфері освіти. Сутнісними характеристиками моніторингу в освіті є інформаційна система обстеження стану освіти, прийняття на основі отриманої інформації управлінських рішень. Тобто моніторинг базується на чіткій основі, якою виступають відповідність будь-якому стандарту, еталону, закону, або порівняльні дані для двох чи кількох подібних за характеристиками систем, або тимчасові динамічні ряди, що характеризують розвиток об'єкта моніторингу.

Водночас моніторинг розглядають у зв'язку з якістю освіти. **Моніторинг якості освіти** визначають як неперервну, науково обґрунтовану, діагностико-прогностичну комплексну систему збирання, зберігання, аналізу і розповсюдження інформації про якість освітніх послуг на основі їх відповідності визначеним цілям та отриманим результатам (комплексне відстеження якості за всіма показниками освітньої діяльності). Разом із тим моніторинг можна розглядати і як метод дослідження в межах контрольної-діагностичної діяльності, для визначення відповідності стану об'єкта його попередньо заданим показникам.

Завдання моніторингу в освіті:

- визначати якість навчальних досягнень учнів, рівень їх соціалізації;
- вивчати зв'язок між успішністю учнів і соціальними умовами їх життя, результатами роботи педагогів, рівнем їх соціального захисту, моральними установками, запитами, цінностями тощо;

- оцінювати якість кадрового, навчально-методичного, матеріально-технічного, лабораторного забезпечення й оснащення навчальних закладів;
- оцінювати величину впливу на навчальний процес державних освітніх стандартів, навчальних програм, організації шкіл і класів, методичного та технічного обладнання та інших факторів;
- досліджувати педагогічну практику та успіхи учнів в залежності від соціального статусу та аналізувати політику держав у галузі забезпечення гарантій доступності освіти та поліпшення її якості;
- виявляти фактори, які чинять вплив на хід і результати освітніх реформ з метою зменшення негативного їх впливу (або, навіть, і нейтралізації);
- порівнювати результати функціонування закладів освіти, систем освіти з метою визначення найбільш оптимальних шляхів їх розвитку.

Моніторинг як кваліметричний інструмент передбачає:

- визначення системи критеріїв, показників та індикаторів якості освіти;
- обрання методів оцінювання, які адекватні інструментарію та процедурам вимірювання;
- проведення моніторингових досліджень з метою управління якістю освіти.

Етапи проведення моніторингового дослідження:

I етап: Цілепокладання та планування дослідження

1. Визначення мети та завдань дослідження.
2. Визначення об'єкта дослідження.
3. Розрахунок та формування вибірки.
4. Побудова графіку дослідження:
 - визначення термінів і процедур дослідження;
 - добір та підготовка (навчання) координаторів дослідження.
5. Визначення критеріїв та показників оцінювання.
6. Вибір методів дослідження.

II етап: Розробка інструментарію

1. Розробка тестів та їх апробація.
2. Розробка анкет та їх апробація.
3. Підготовка інструктивно-методичних матеріалів для координаторів дослідження всіх рівнів, учасників дослідження.

4. Вибір статистичних і математичних методів обробки та обрахунку одержаних результатів дослідження.

III етап: Проведення дослідження

1. Пілотне дослідження (підготовка учасників, проведення інструктажу).
2. Основне дослідження.

IV етап: Збирання та обробка результатів.

V етап: Аналіз та інтерпретація результатів дослідження.

1. Узагальнення статистичної інформації.
2. Виявлення факторів впливу.
3. Підготовка рекомендації щодо корекційної роботи, усунення негативних факторів, формування освітньої політики тощо.

7.3. Моніторинг якості освіти та моніторингові дослідження

7.3.1. Об'єкти і суб'єкти моніторингу якості освіти

Якість освіти є тим показником, за яким визначається ефективність функціонування системи освіти будь-якої держави на будь-якому етапі її розвитку. Якість освіти – це сукупність властивостей системи освітньої галузі, що відповідає сучасним вимогам педагогічної теорії та практики й спроможна задовольнити освітні потреби особистості, суспільства, держави. Підвищення якості освіти потребує її постійного дослідження, налагодження системи моніторингу освіти, головною метою якого є збирання, обробка, зберігання та поширення інформації про освітню систему або її окремі елементи, оцінювання й аналіз її показників на всіх рівнях функціонування, поширення та доступ до цієї інформації громадськості, різних користувачів освітніх послуг, посилення управлінських дій щодо якісних показників в освіті, прогнозування її розвитку.

Поняття «моніторинг» у педагогічній літературі й освітянській практиці тлумачиться неоднозначно. Нині широко і рядоположно вживаються такі терміни і поняття, як «моніторинг» – аналіз (самоаналіз), «моніторинг» – педагогічна діагностика, «моніторинг» – анкетування, опитування, «моніторинг» – перевірка, контроль, оцінювання тощо. У широкому значенні,

моніторинг якості освіти – спеціальна система збору, обробки, зберігання та розповсюдження інформації про стан освіти, інтерпретація зібраних фактів та прогнозування на їх основі динаміки і основних тенденцій її розвитку та розробка науково обґрунтованих рекомендацій для прийняття управлінських рішень стосовно підвищення ефективності функціонування освітньої галузі.

Об'єктами моніторингу якості освіти можуть бути: а) система освіти (загальна середня, професійна, вища, національна або загальнодержавна, регіональна система, муніципальна); б) процеси (управлінський на різних рівнях державного управління освітою; навально-виховний процес, педагогічний процес у цілому та окремі його складові: процес морального та трудового виховання, процес підготовки дітей до школи; в) зміст освіти; г) рівень і якість сформованості навчальних досягнень, процес і результати сформованості життєвих принципів та здатності до самостійного подальшого життя після закінчення школи тощо; д) діяльність (трудова, навчальна, управлінська); е) явища (вихованість, освіченість, майстерність педагога та інші). Таким чином, можна говорити, про безліч видів моніторингу в системі освіти.

Серед найпоширеніших підходів, які застосовують країни в побудові об'єктивної моделі моніторингу, тобто визначення категорій об'єктів, відкритих для моніторингових процедур, передусім можна назвати такі:

- *Результативний підхід*, що передбачає моніторинг лише категорій, яка об'єднує об'єкти, які належать до результатів освітнього процесу;
- *Системний підхід*, який розширює об'єкти, що підпадають під моніторингові процедури в освіті, об'єднуючи їх три категорії: категорія, що об'єднує ресурсні аспекти, тобто ресурси (фінансові, людські тощо), що їх вкладають в освіту; категорія, яка об'єднує аспекти, що стосуються освітнього процесу; категорія, до якої входять об'єкти, що належать до результатів освітнього процесу.

Суб'єкти моніторингу: а) органи державного контролю за діяльністю навчальних закладів; б) моніторингова служба відповідного рівня; в) адміністрація навчального закладу, рада закладу, піклувальна рада; г) органи громадського та місцевого самоврядування.

7.3.2. Принципи моніторингових досліджень.

Основними принципами, на яких ґрунтується моніторинг в освіті є:

- ✓ **узгодженість** нормативно-правового, організаційного та методичного забезпечення складових моніторингу;
- ✓ **об'єктивність** одержання та обробки інформації, що передбачає максимальне виключення суб'єктивних оцінок, врахування всіх результатів, створення для всіх учасників дослідження рівних умов у процесі перевірки якості підготовки;
- ✓ **комплексність** дослідження різноманітних аспектів навчально-виховного процесу, обробки та аналізу одержаних результатів ;
- ✓ **безперервність** і тривалість спостережень за станом освіти;
- ✓ **своєчасність** отримання, обробки та використання об'єктивної інформації про якість освіти ;
- ✓ **перспективність** запланованих моніторингових досліджень, спрямованість їх на розв'язання актуальних завдань розвитку освіти ;
- ✓ **гуманістична спрямованість** моніторингу – створення обстановки доброзичливості, довіри, поваги до особистості, максимально сприятливих умов, позитивного мікроклімату, неможливості використання результатів досліджень для застосування будь-яких репресивних дій ;
- ✓ **відкритість і оперативність** доведення результатів досліджень до відповідних органів управління, громадськості, зацікавлених міжнародних установ.

7.3.3. Види та напрямки моніторингу

На сучасному етапі дослідники пропонують різноманітні критерії класифікації моніторингових досліджень. Наведемо деякі з них, наприклад, О.М.Майоров висуває такі критерії: область використання; засоби проведення; засоби збору інформації; час проведення; масштабність; орієнтація. С.Є Шишов пропонує класифікацію видів моніторингових досліджень за: масштабністю цілей освіти; етапами освіти; часом; циклічністю; масштабністю об'єкта. Український науковець Г.В.Єльнікова виділяє такі види моніторингових досліджень: державний; регіональний;

локальний, який може бути керівний, адміністративний, педагогічний. Канадський професор Д.Уїлмс пропонує – моніторинг узгодження управління; діагностичний моніторинг; моніторинг діяльності. Також, вважаємо за доцільне, подати види моніторингу на рівні навчального закладу.

- **За масштабами цілей освіти:** оперативний, тактичний, стратегічний.
- **За етапами навчання:** вхідний або вибіркового, проміжний або навчальний, вихідний або підсумковий.
- **За охопленням об'єкта спостереження:** локальний, вибіркового, суцільний.
- **За організаційними формами:** індивідуальний, груповий, фронтальний.
- **За рівнями управління навчально-виховним процесом:** керівний, адміністративний, педагогічний, учнівський.
- **За формами суб'єкт-об'єктних відносин:** соціальний (зовнішній), взаємоконтроль, самоаналіз.
- **За часовою залежністю:** ретроспективний, попереджувальний, поточний.
- **За частотою процедур:** разовий, періодичний, систематичний.
- **За інструментарієм, що використовується:** стандартизований, не стандартизований, матричний.

Залежно від мети дослідження науковці виділяють різноманітні напрямки моніторингу. Визначимо сутність деяких з них.

1. Моніторинг узгодження управління (автор Д.Уїлмс) передбачає, що якщо школа відповідає певним стандартам в освіті, автоматично забезпечується адекватний рівень її діяльності.
2. Діагностичний моніторинг (автор Д.Уїлмс), основною метою якого є визначення рівня академічних навичок учнів незалежно від їх особистості.
3. Статичний моніторинг (автор В.М.Лізінський), що надає можливість одночасно зняти показники за одним або кількома напрямками діяльності освітнього закладу, порівняти отриманий результат з нормативом і визначити відхилення від стандарту, здійснити аналіз і прийняти управлінське рішення.

4. Моніторинг освітніх систем (автор С.І.Подмазін) – дозволяє вироблення актуальних знань про стан системи, в якій відбуваються зміни, з подальшим прийняттям управлінського рішення.
5. Освітній моніторинг (автор Г.В.Єльнікова) передбачає супровідне оцінювання і поточну регуляцію будь-якого процесу в освіті.
6. Освітній моніторинг (автор Н.Ладнушкіна) досліджує форми організації, збору, збереження, обробки і розповсюдження інформації про діяльність педагогічної системи, що забезпечує безперервне спостереження за її станом і прогнозування її розвитку.
7. Моніторинг результативності навчального процесу (автори Н.Вербицька, В.Бодряков) показує загальну картину дії усіх факторів, що впливають на навчання, і визначає напрямки, які потребують більш детального дослідження.

Таким чином, моніторингові дослідження є надійним інструментом аналізу різноманітних аспектів освітнього процесу. Вони дозволяють отримати реальну картину результативності нововведень, напрямків плану розвитку навчального закладу, якості знань учнів тощо.

7.4. Історія розвитку моніторингу якості освіти

Аналіз психолого-педагогічної літератури свідчить, що поняття «моніторинг» введено в науковий обіг у 60-х рр. ХХ ст. і одне з перших його нормативних визначень подано в 1977 р. у Міжнародній енциклопедії освіти, як технологія постійного спостереження конкретного явища, його оцінки і прогнозування розвитку. Проте, моніторингові дослідження проблем освіти як феномен управління освітою, як новий напрям прикладних наук зародився понад сто років тому в США. У листопаді 1900 р. президенти дванадцятиох університетів північного сходу США створили організацію, яку вони назвали Радою коледжів із вступних іспитів (СЕЕВ). Завданням цієї організації було адміністрування спільного для її засновників вступного іспиту.

Історію становлення моніторингових досліджень в освіті можна умовно поділити на декілька етапів.

Перший етап охоплює 30–50-ті рр. ХХ ст.. Він ознаменувався тим, що американська Асоціація прогресивної освіти (Progressive Education Association in the United States) за допомогою моніторингу дослідила рівень підготовки та ефективні риси (суспільну позицію та критичне мислення) випускників 30 шкіл. Це дослідження мало важливе значення, оскільки започаткувало проведення систематичного вивчення досягнутого освітньою системою стану, порівняння рівня навчальних досягнень учнів різних шкіл залежно від умов навчання. Дослідження розробив американський учений Р.Тайлер. 1952–1959 рр. Шведською національною радою освіти було здійснено зіставлення ефективності роботи нових об'єднань шкіл у порівнянні з традиційними. 1952 рік ознаменувався заснуванням Інституту освіти ЮНЕСКО в Гамбурзі, який започаткував регулярне проведення конференцій з питань діагностики якості освіти.

Другий етап в історії розвитку моніторингу охоплює 60-70-ті рр. ХХ ст. і характеризується виходом цих досліджень на міжнародний рівень. Основними подіями цього періоду були:

- 60 – 70-ті рр. Міжнародною асоціацією з шкільної успішності (International Education Association) були започатковані міжнародні порівняльні моніторингові дослідження у сфері освіти;
- Розробка індикаторів оцінювання освітньої сфери ЮНЕСКО та ОЕСР;
- 1959 – 1961 рр. – в дванадцяти країнах світу проведено дослідження якості освіти з математики, природознавства та читання (МАОНД);
- 1959 – 1967 рр. проводилося перше моніторингове порівняльне міжнародне дослідження – FIMS у 13 країнах світу, яке було присвячено оцінюванню успішності з математики у початковій і середній школі.
- 1973 р. – документ ОЕСР “Схема освітніх індикаторів для керівництва в прийнятті урядових рішень” (46 індикаторів);
- 1876 – 1989 рр. – друге міжнародне дослідження – SIMS, що передбачало вивчення успіхів школярів у засвоєнні математики, природничих наук, англійської та французької мов як іноземних. А також дослідження навчального середовища в межах одного класу. У дослідженні брало участь 24 країни.

Третій етап моніторингових досліджень (1980 – 1990 рр.) характеризується підвищенням інтересу до проблем визначення

рентабельності, ефективності освітніх систем, оцінювання їхньої продуктивності. Якість освіти стає економічною і політичною категорією оскільки неякісний її рівень загрожує національному добробуту. Новими характеристиками цього етапу моніторингу стає збір даних не тільки про витрати на освіту та навчальні досягнення учнів, але й такі результати, як кількість випускників, безробітних тощо. В цей період розроблялися і використовувалися єдина методологія та стандартизований інструментарій, визначалася система показників і критеріїв якості й результативності, що давало змогу порівнювати й узагальнювати одержані результати як на національному, так і на міжнародному рівнях. 1983 р. – друге дослідження якості освіти з математики, природознавства та читання (МАОНД). Міжнародні моніторингові дослідження активізували національні дослідження у багатьох країнах і допомогли визначити власні орієнтири розвитку системи освіти.

Четвертий етап в історії моніторингу розпочався наприкінці 90-х рр. і триває донині. Закладаються основи нового етапу порівняльної педагогіки, а саме спрямованість на політичну підтримку та наукову обґрунтованість досліджень. Освітня політика й освітні вимірювання вийшли на міжнародний рівень і це потребує створення спеціального міжнародного банку порівняльної інформації. Створено методологічну основу для проведення різноманітних досліджень, розроблено нові статистичні методи вимірювань, відпрацьовано ефективні технології тощо. У 1998 році завершилося міжнародне дослідження MONEE в галузі освіти, що було проведено міжнародним дитячим фондом ЮНІСЕФ. Це дослідження під назвою «Освіта для всіх?» охопило 27 країн Європи. Аналіз результатів проекту дозволив виявити чинники економічного характеру, що впливають на стан системи середньої освіти, рівень освіченості учнівської і студентської молоді, а також визначають доступність освіти для всіх членів суспільства. Механізми контролю та оцінювання ефективності діяльності систем середньої освіти європейських країн вивчалися в рамках міжнародного проекту «Загальна середня освіта» (1994 – 1998 рр). Додаткову інформацію про моніторингові дослідження, що проводилися протягом четвертого етапу, можна знайти у відповіді на питання «Міжнародні порівняльні дослідження моніторингу якості освіти».

7.5. Міжнародні моніторингові дослідження якості освіти

7.5.1. Організація перших міжнародних моніторингових досліджень

У 1959 році Міжнародною асоціацією з оцінювання шкільної успішності (IEA) уперше проведено міжнародне порівняльне моніторингове дослідження успішності з математики учнів початкової і середньої школи (FIMS). Надалі міжнародні дослідження рівня навчальних досягнень з окремих предметів, інформаційно-комунікативної, функціональної грамотності учнів здійснюють Міжнародна асоціація з оцінювання якості освіти (IAEA), Міжнародна асоціація з оцінювання шкільної успішності (IEA), Міжнародний Інститут планування освіти (MIPO), Інститут освіти ЮНЕСКО, Організація економічної співдружності та розвитку (OECD), Міжнародний дитячий фонд ЮНІСЕФ, Інститут економічного розвитку при Всесвітньому банку та інші організації. Вищеназвані організації діють на правах науково-дослідних незалежних установ, які за допомогою офіційних державних органів управління освітою проводять різноманітні дослідження. Вони не є контролюючими органами, їхня діяльність ґрунтується на принципах гуманістичної спрямованості. Вони не мають права втручатися у навчально-виховний процес і змінювати його, розголошувати інформацію стосовно компетентності викладачів, рівня підготовки окремих учнів. За результатами міжнародних порівняльних досліджень створюються банки даних, визначаються загальні тенденції розвитку освітніх систем і якості загальноосвітньої підготовки учнів.

Охарактеризуємо деякі моніторингові порівняльні міжнародні дослідження. Як вже було зазначено, перше моніторингове порівняльне дослідження **FIMS** проводилося IEA протягом 1959 – 1967 рр. у 13 країнах світу і було присвячено оцінюванню успішності з математики учнів початкової та середньої школи. У дослідженні інформація збиралася і систематизувалася за такими блоками: 1) організація навчання та диференціація учнів; 2) зміст і методи навчання; в) соціальні, економічні та матеріально-технічні характеристики суспільства, школи, родини. Друге міжнародне дослідження – **SIMS** – проводилося у декілька етапів протягом 1976 – 1989 рр. у 24 країнах. Воно передбачало вивчення досягнень школярів з математики, природничих наук, англійської та французької мов як іноземних, а також дослідження навчального середовища в межах одного класу.

7.5.2. Міжнародне моніторингове дослідження TIMSS та інші моніторингові дослідження кінця ХХ ст.

Велике значення мали результати третього міжнародного моніторингового дослідження стану вивчення математики й природничих наук **TIMSS** (Trends in International Mathematics and Science Studies), яке розпочалося у 1991 році (з 1995 року регулярно проводиться через кожні 4 роки). У ньому взяли участь понад 50 країн світу і воно дало можливість визначити своєрідний рейтинг держави у галузі освіти за цими напрямками. TIMSS – найбільш масштабний проект у галузі освіти, спрямований на вивчення досвіду країн світу з метою вдосконалення національних систем навчання математики та природничих дисциплін. Україна взяла участь у даному міжнародному вимірюванні рівня якості освіти школярів у 2007 і 2011 рр. Дослідження має на меті:

- оцінити якість математичної та природничої освіти учнів 4-х та 8-х класів;
- виявити динаміку результатів (1995, 1999, 2003, 2007, 2011 рр.);
- виявити фактори, що дозволяють пояснити відмінність у результатах.

Завдання дослідження реалізуються поетапно:

- 1) збір, аналіз та узагальнення інформації стосовно систем освіти в країнах;
- 2) аналіз та порівняння систем природничо-математичної освіти на рівні визначення цілей, планування змісту освіти та вимог щодо підготовки учнів (аналіз навчальних планів, програм та підручників природничо-математичних предметів);
- 3) аналіз навчального процесу з математики та природничонаукових предметів (організація навчального процесу; методи викладання; можливості, які надаються учням під час навчання забезпечення навчального процесу літературою та іншими засобами навчання тощо);
- 4) оцінка результатів навчання, яка включає оцінку навчальних досягнень учнів та виявлення її залежності від попередніх факторів;
- 5) аналіз взаємозв'язку між запланованими та реалізованими рівнями освіти і результатами навчання.

Під час дослідження аналізується зміст шкільної математичної та природничонаукової освіти, оцінюється математична та природничонаукова підготовка учнів початкової школи (3–4 класи), основної школи (7–8 класи), а

також випускників середньої школи. Участь у тестуванні обов'язкова для всіх країн, що беруть участь в проекті. Водночас оцінюються навчальні досягнення під час виконання учнями практичних робіт. Результати досліджень TIMSS для початкової, основної та старшої школи з аналізом результатів практичного тесту відображаються у міжнародних звітах.

Також, наприкінці ХХ ст. були здійснені наступні моніторингові порівняльні дослідження:

- ✓ **LES** – проводилося протягом 1995-1997 рр. і було присвячено вивченню стану викладання іноземних мов.
- ✓ **CIVICS** – розпочалося у 1996 р. і проводилося у декілька етапів з фіксацією результатів у 1999, 2000 роках. Досліджувалася громадянська суспільствознавча освіта. Взяло участь 28 країн.
- ✓ **MONEE «Освіта для всіх?»** (1998 р.) – проводилося Міжнародним дитячим фондом ЮНІСЕФ; охопило 27 країн Східної, Центральної Європи та Балтії. Основна мета проекту полягала в аналізі соціальних умов життя дітей, їх родин та державної політики постсоціалістичних країн.
- ✓ **IEAP** – дослідження порівняльної оцінки математичної підготовки учнів.
- ✓ **SITES** – порівняльне дослідження інформаційних та комунікаційних технологій в освіті.

7.5.3. Міжнародна програма оцінки знань та умінь учнів PISA

Організація з економічного співробітництва та розвитку (OECD – Organisation for Economic Cooperation and Development) впроваджує Міжнародну програму оцінки знань та умінь учнів **PISA** (Programme for International Student Assessment), основною метою якої є отримання надійних відомостей про результати навчання у різних країнах світу, які можна порівняти на міжнародному рівні. Передбачається, що отримана інформація надасть можливість країнам-учасникам приймати обґрунтовані рішення для визначення перспектив розвитку галузі освіти.

Вимірювання знань проводиться серед учнів 15-річного віку. Такий вибір визначено тим, що в багатьох країнах до цього віку закінчується обов'язкове навчання в школах і навчальні програми мають багато спільного. Особливим інтересом дослідження є визначення стану знань та вмінь, які можуть бути корисними для учнів у майбутньому, а також уміння самостійно

здобувати знання, потрібні для успішної адаптації в сучасному світі. Оцінка підготовки 15-річних підлітків стандартизована. Матеріали розробляються країнами–учасницями міжнародної програми спільно.

Щодо змісту дослідження, то воно проводиться за трьома напрямками: *“грамотність читання”*, *“математична грамотність”* та *“природничонаукова грамотність”*. Розглянемо визначення та стислі характеристики кожного з трьох напрямків даного дослідження. *“Грамотність читання”* – здатність людини розуміти письмові тексти та розмірковувати над ними, використовувати їх зміст для досягнення власних цілей, розвитку знань і можливостей, для активної участі в житті суспільства. Таким чином, термін *“грамотність читання”* має широкий зміст. Він не передбачає простої перевірки техніки читання. Мета дослідження висвітлює сучасне уявлення стосовно вміння *“грамотно читати”*. Згідно з таким підходом, випускник основної школи повинен розуміти тексти, розмірковувати над змістом, оцінювати їх смисл та значення, викладати свої думки стосовно того, що він прочитав. Основна увага приділяється перевірці вміння *“грамотно читати”* в різних ситуаціях. Учням пропонуються тексти різних жанрів: уривки з художніх творів, біографії, розважальні тексти, особисті листи, документи, статті з газет і журналів, інструкції, рекламні оголошення, географічні карти тощо.

“Математична грамотність” – здатність людини визначати й розуміти роль математики у світі, висловлювати добре обґрунтовані математичні судження та використовувати математику таким чином, щоб задовольняти сьогоденні й майбутні потреби, властиві творчому, зацікавленому та мислячому громадянину. Термін *“грамотність”* використовується для того, щоб показати, що вивчення стану математичних знань та вмінь, які зазвичай визначені в шкільній програмі, не є першочерговим завданням цього дослідження. Основна ситуація через використання різних підходів, які потребують міркування та інтуїції. Очевидно, що для цього треба мати значний обсяг математичних знань та вмінь, які вивчаються в школі. Учням в основному пропонуються не навчальні, а практичні ситуації із повсякденного життя (медицина, житло, спорт тощо).

“Природничонаукова грамотність” – здатність використовувати природничонаукові знання для визначення у реальних ситуація проблем, які

досліджуються та вирішуються за допомогою наукових методів, а також для отримання висновків, заснованих на спостереженнях та експериментах. Ці висновки потрібні для розуміння навколишнього світу, змін, які відбуваються завдяки діям людей, та прийняття відповідних рішень. При цьому остаточне рішення в багатьох випадках приймається з урахуванням суспільно-політичних або економічних умов. Ці знання та вміння формуються у школах під час вивчення предметів природничонаукового циклу: фізики, біології, хімії та географії. Метою дослідження є комплексна перевірка вмінь та понять. Головна увага приділяється перевірці вмінь визначати серед запропонованих питань ті, на які природничі науки можуть дати відповідь, робити науково обґрунтовані висновки на основі запропонованої інформації тощо. Реальні ситуації, запропоновані учням, пов'язані з актуальними проблемами, які виникають в особистому житті кожної людини (наприклад, використання продуктів харчування під час дотримання дієти), у житті людини як члена якогось колективу або суспільства (наприклад, визначення місця розташування електростанції відносно міста), як «громадянина світу» в кращому розумінні цього слова (наприклад, усвідомлення наслідків глобального потепління).

Особлива увага приділяється виявленню рівня вищезазначених напрямків та вміння використовувати свої знання в різних ситуаціях. Тобто вимірюється не рівень опанування конкретним змістом навчальних дисциплін, а рівень більш широких знань та умінь, які накопичені під час вивчення шкільних предметів і потрібні для дорослого життя. Велика увага також приділяється оцінці *міжпредметних* компетентностей учнів. Дослідження проводять трирічними циклами, останній був у 2004–2006 роках. У кожному циклі головна увага приділяється одному з напрямків дослідження, за двома іншими отримують інформацію лише стосовно окремих умінь. Так, у 2000 році основним напрямком була “грамотність читання”, у 2003-му – “математична грамотність”, у 2006-му – “природничонаукова грамотність”. На 2012 рік заплановано участь України у дослідженні оцінки навчальних досягнень учнів 15-річного віку за напрямками “природничо-математична грамотність” і “грамотність читання”.

Програма здійснюється консорціумом, який складається з провідних міжнародних дослідницьких організацій за участю національних центрів та організації ОЕСР. Керує роботою консорціуму Австралійська рада

педагогічних досліджень (The Australian Council for Educational Research – ACER). До складу консорціуму входять Нідерландський національний інститут педагогічних вимірювань (Netherlands National Institute for Educational Measurement – Cito), Служба педагогічного тестування (Educational Testing Service, ETS, США), Національний інститут досліджень у галузі освіти (NIER, Японія), Вестат США (Westat, USA).

У світлі широкомасштабного реформування освіти розвинуті суспільства сьогодні змінюють свій погляд на те, якою повинна бути підготовка випускника основної школи. Поруч із формуванням предметних знань та вмінь школа повинна розвивати в учнів уміння використовувати свої знання в різноманітних ситуаціях, наближених до реальності, сприяти набуттю необхідних життєвих компетентностей. У подальшому житті вони будуть стимулювати активну участь випускника школи в житті суспільства, допоможуть йому навчатись протягом усього життя.

7.5.4. Міжнародний проект PIRLS

Міжнародний проект **PIRLS** (Progress in International Reading Literacy Study) – це моніторингове дослідження, організоване IEA, що передбачає вивчення якості читання та розуміння тексту учнів 4-х класів. За результатами дослідження можна отримати відповіді на запитання:

- наскільки добре читають випускники початкової школи порівняно зі своїми однолітками з інших країн;
- якими рівнями читацької грамотності володіють учні й чи люблять четвертокласники читати;
- як сім'я сприяє розвитку грамотності дітей;
- як на сьогодні організовано навчання читанню в школах;
- чи має навчання читанню національні особливості порівняно з іншими країнами і якщо так, то в чому вони полягають;
- чи відрізняються методи навчання, які використовують учителі в початковій школі у різних країнах.

Разом із дослідженням TIMSS це оцінювання охоплює три основні предмети у початковій школі – читання, математику та природничі науки. В Україні дослідження проводилося в лютому-березні 2011 року за підтримки проекту “Рівний доступ до якісної освіти в Україні”, а пілотне тестування (у лютому–березні 2010 року) і було спрямоване на вміння дітей точно

відтворювати прочитану інформацію, робити висновки, оцінювати зміст й елементи тексту, пояснювати інформацію. Дані з оцінювання досягнень дітей у читанні, математиці та природничих науках разом з інформацією, отриманою від учителів, директорів шкіл та батьків, допоможуть країнам-учасницям визначити прогрес у даних предметах, важливих для подальшого навчання, та вплив на цей прогрес оточення у школі, класі та вдома.

Узагальнюючи інформацію, яка характеризує основні напрямки розвитку міжнародної системи моніторингу якості освіти, зазначимо, що участь країн у міжнародних дослідженнях дозволяє: не тільки одержати об'єктивні та порівнювані дані про рівень підготовки учнів за визначеним напрямком, а й отримати актуальну інформацію про сучасний стан розвитку систем освіти інших країн; порівняти зміст освіти та освітніх стандартів різних країн світу, світові тенденції розвитку освіти взагалі. Результати будь-яких моніторингових досліджень дають змогу урядам країн реально оцінити свої досягнення та недоліки, визначити стратегії змін у державній політиці, забезпечити конкурентноздатність національних освітніх систем та підвищити їх міжнародний авторитет. Участь України в таких дослідженнях дозволить отримати об'єктивну інформацію про систему освіти в країні та визначити ефективні способи вирішення проблем.

7.6. Моніторинг якості освіти як системне дослідження освітнього процесу

Моніторинг якості освіти – це науково обґрунтована, неперервна, діагностико-прогностична комплексна система збору, зберігання, аналізу і розповсюдження інформації про якість освітніх послуг на основі їх відповідності визначеним цілям та отриманим результатам (комплексне відстеження якості за всіма показниками освітньої діяльності). Сутнісними характеристиками моніторингу в освіті є інформаційна система обстеження стану освіти, прийняття управлінських рішень на основі отриманої інформації. Тобто моніторинг базується на чіткій основі, якою виступають відповідність будь-якому стандарту, еталону, закону, або порівняльні дані для двох чи кількох подібних за характеристиками систем, або тимчасові динамічні ряди, що характеризують розвиток об'єкта моніторингу.

Моніторинг якості освіти використовується в двох напрямках: управлінський – для прийняття відповідних управлінських рішень (як система збору, накопичення інформації про стан освіти, обробки показників динаміки її розвитку) та наукового дослідження – для визначення і створення оптимальних умов здійснення освітнього процесу з метою досягнення максимально позитивного результату.

Освітній моніторинг має свою специфіку – він не лише надає інформацію про стан освітньої системи, але запускає механізми поточного регулювання (саморегулювання). Моніторинг має комплексний, системний характер і створює умови для планування – річного, перспективного, стратегічного, дозволяє оцінити ефективність функціонування та розвитку навчального закладу не формально, а за досягнутими показниками.

Система моніторингу якості освіти, на думку науковців, має включати в себе наступні компоненти: мету і зміст освіти; вимоги освітніх стандартів; контроль та прогнозування; освітній моніторинг; модель учня (студента), вчителя (викладача), закладу освіти та органів управління освітою.

Системне моніторингове дослідження передбачає оцінювання таких складових:

- ресурси, що вкладаються в освіту;
- освітній процес;
- результати, що продукує освітня система і зворотній зв'язок.

Система моніторингу якості освіти підпорядкована ієрархічним зв'язкам освітньої системи. **Система моніторингу** виявляється на різних рівнях функціонування освіти:

- **індивідуальному рівні** — самооцінки учнями й студентами якості своєї загальноосвітньої та професійної підготовки, суспільної, професійної і життєвої компетентності, досвіду оволодіння алгоритмічними й евристичними способами діяльності, навичками критичного мислення тощо;

- **локальному рівні** — оцінювання навчальним закладом учасників навчально-виховного процесу щодо якості освіти, досягнення ними поставленої мети в опануванні вимог державного стандарту відповідного рівня освіти, коригування стратегії розвитку за соціальними, педагогічними, економічними та іншими показниками;

- **муніципальному рівні** — оцінювання місцевими органами управління суб'єктів освітньої діяльності (рівня навчальних досягнень учнів і студентів,

кваліфікації та педагогічної майстерності вчителів і викладачів, компетентності керівників навчальних закладів тощо) та порівняння результатів діяльності власної мережі навчальних закладів;

- **регіональному рівні** — оцінювання ефективності функціонування місцевої системи освіти, зокрема підпорядкованих органів управління, забезпечення ними державної освітньої політики в регіоні, вибіркоче вивчення ефективності роботи окремих навчальних закладів і органів управління, вдосконалення мережі навчальних закладів на підставі аналізу одержаних даних;

- **державному рівні** (всеукраїнському, національному) — акцентованого й узагальненого оцінювання якості функціонування національної системи освіти та порівняння її показників з міжнародними індикаторами й системами, аналіз і порівняння стану реалізації державної освітньої політики в різних регіонах України;

- **міжнародному рівні** — оцінювання якості функціонування систем освіти різних країн світу; порівняння показників якості освіти в різних країнах; узагальнення інформації про якість освіти у світі в цілому; визначення тенденцій розвитку освіти у світі в цілому та в окремих країнах.

Наведемо приклади моделі системного моніторингового дослідження на рівні регіональної (міської) системи управління загальною середньою освітою (автор О.О.Байназарова) та приклад побудови системи освітнього моніторингу на рівні вищого навчального закладу (автор С.С.Кретович).

Моніторингова модель регіональної (міської) системи управління загальною середньою освітою складається з чотирьох підсистем, які відображають різні види та сторони діяльності закладів освіти. У кожній з них передбачається комплекс моніторингових досліджень.

У першій педагогічній підсистемі проводиться моніторинг різних видів діяльності учнів: аналіз відвідування навчальних занять; рівень навчальних досягнень; тематична атестація; навчальні практики; участь у Всеукраїнських учнівських олімпіадах з базових дисциплін, конкурсах, турнірах; працевлаштування випускників.

Друга підсистема — науково-методична — передбачає моніторинг: методичної та експериментальної роботи; кадрового складу вчителів; науково-методичного забезпечення навчального процесу; навчально-методичне забезпечення викладання інваріативних та варіативних курсів;

рівня викладання базових дисциплін; виконання навчального плану та навчальних програм.

Моніторинг **третьої** – соціально підсистеми – включає оцінку діяльності закладів освіти з соціальних питань: охорона здоров'я учнів та вчителів; дотримання в навчальних кабінетах правил техніки безпеки; соціальний та професійний захист вчителів.

У **четвертій**, економічній підсистемі, оцінюється матеріально-технічне забезпечення навчальних закладів.

Зазначена система моніторингу якості освіти на регіональному рівні може бути використана і при проектуванні внутрішкільного моніторингу. Подальша реалізація представленої моделі потребує розробки технології моніторингових досліджень, загальний алгоритм яких повинен включати в себе: аналіз діяльності підсистем моделі, обґрунтування критеріїв для оцінки діяльності закладу освіти, діагностику їх роботи.

Розглянемо приклад побудови системи освітнього моніторингу вищого навчального закладу. Системний підхід при проведенні моніторингового дослідження дозволяє розглядати діяльність вищого навчального закладу як складну систему взаємодії елементів, що до неї входять. Такий підхід дозволяє представити роботу ВНЗ в повному обсязі, виявити слабкі місця, їх причини та визначити комплекс заходів щодо їх усунення. Проведення освітнього моніторингу передбачає на початку чітке визначення мети проведення моніторингу, змісту, структури, критеріїв якості освіти, а саме – що очікується в результаті дослідження і як це вплине на діяльність ВНЗ. На цьому етапі створюється нормативний еталон (стандарт) діяльності ВНЗ; визначаються критерії, показники, індикатори якості для вимірювання стану. Наступним кроком є розробка технології здійснення моніторингу та його інструментарію, обираються методики вимірювання. Цей етап залежить від завдань моніторингу та обсягу необхідної інформації. Після добору інструментарію визначаються етапи дослідження: складається план проведення моніторингу, визначаються терміни проведення дослідження та відповідальні особи. За планом проводиться збір даних (результати тестів, спостережень за обраними методиками тощо). Потім дані обробляються та аналізуються, на основі інтерпретації отриманих даних здійснюється рефлексивний аналіз та виробляються рекомендації для подальшого розвитку об'єкта дослідження (програма корекційних заходів). Наступний

етап – процес коригування (самокоригування) та контролю (самоконтролю). Коригування передбачає унесення певних змін в процес функціонування ВНЗ на основі розроблених рекомендацій, а контроль – констатування правильного виконання розроблених рекомендацій. Далі відбувається оприлюднення результатів.

Таким чином, моніторинг якості освіти є системним процесом і з позицій дослідження складної системи, якою являється система освіти і якість освіти, і з позицій системності самої процедури моніторингового дослідження

Питання для самоконтролю:

1. Що називають моніторингом?
2. У якій сфері людської діяльності вперше був застосований моніторинг?
3. Назвіть складові системи моніторингу.
4. Як можна тлумачити поняття «якість освіти»?
5. Якою є структура якості освіти?
6. Як Ви розумієте якість державно-громадського управління системою освіти?
7. Назвіть цілі оцінювання якості освіти.
8. Назвіть основні групи показників якості освіти.
9. Конкретизуйте таку групу показників якості освіти як соціально-економічний потенціал розвитку освіти.
10. Що називають моніторингом якості освіти?
11. У яких напрямках використовується моніторинг у педагогіці?
12. Назвіть завдання моніторингу в освіті.
13. Який комплекс моніторингових показників може бути визначений для з'ясування якості навчальних досягнень учнів?
14. Який комплекс моніторингових показників може бути визначений для вивчення зв'язку між успішністю учнів і соціальними умовами їхнього життя?
15. Охарактеризуйте моніторинг як кваліметричний інструмент.
16. Назвіть етапи проведення моніторингового дослідження.
17. Розкрийте сутність поняття «моніторинг якості освіти».
18. Визначте суб'єктів та об'єкти моніторингу якості освіти.
19. Назвіть основні види моніторингових досліджень якості освіти. Чому складною є проблема їх класифікації?
20. Сформулюйте основні принципи, на яких ґрунтується моніторинг в освіті. Обґрунтуйте принципи об'єктивності і комплексності.

21. Поясніть сутність основних напрямів моніторингових досліджень
22. Що характерно для першого етапу становлення моніторингу якості освіти?
23. Визначте основні події в історії розвитку моніторингу, що сталися в 60-70-ті рр. ХХ ст. Які події ознаменували вихід моніторингу якості освіти на міжнародний рівень?
24. Які проблеми в освіті актуалізувалися наприкінці у 80-90 рр. ХХ ст.? Чому якість освіти стає економічною і політичною категорією?
25. Визначте характерні риси кожного етапу моніторингу якості освіти.
26. Коли і з якої проблеми було проведено перше порівняльне моніторингове дослідження якості освіти?
27. Охарактеризуйте мету і завдання моніторингового дослідження стану вивчення математики й природничих наук TIMSS.
28. Визначте напрями дослідження за Міжнародною програмою оцінки знань та умінь учнів PISA.
29. Розкрийте сутність понять “грамотність читання”, “математична грамотність” та “природничонаукова грамотність”.
30. Що досліджується у рамках Міжнародного проекту PIRLS?
31. У яких міжнародних проектах брала участь Україна?
32. Покажіть взаємозв'язок між управлінським і науковим напрямом моніторингу якості освіти.
33. Охарактеризуйте моніторингову модель регіональної (міської) системи управління загальною середньою освітою. Запропонуйте варіанти як вона може бути використана і при проектуванні внутрішкільного моніторингу.
34. Проаналізуйте основні етапи системи освітнього моніторингу вищого навчального закладу.

Список джерел, використаних у розділі VII

1. Байназарова О.О., Ракчеєва В.В. Моніторинг та оцінювання якості освіти: Методичний посібник. – Харків: ХНУ імені В.Н. Каразіна, 2009. – 58 с.
2. Касянова О.М., Волобуєва Т.Б. Моніторинг в управлінні навчальним закладом. Управлінський супровід моніторингу якості освіти.– Х.: Вид. гр. «Основа», 2004.– 96с
3. Приходько В.М. Моніторинг якості освіти і виховної діяльності навчального закладу. Навчально-методичний посібник для вчителя. – Х.: Вид. група «Основа»: «Тріада+», 2007. – 144 с.
4. Моніторинг у ЗНЗ / Упорядник М.Голубенко. – К.: Шк. світ, 2007. – 128 с.
5. Моніторинг якості освіти: становлення та розвиток в Україні: Рекомендації з освітньої політики / Під заг. ред. О.І.Локшиної – К.: “К.І.С.”, 2004. – 160 с.
6. Кретович С.С. Поняття, сутність та характеристика моніторингу розвитку вищого навчального закладу // Вісник ТІМО. – № 5-6. – 2011. – С.45 – 49.
7. <http://testportal.gov.ua/>
8. www.test-center.od.ua

VIII. ОСВІТНІ ВИМІРЮВАННЯ: СТАНДАРТИ, ФОРМИ ТА ЗАСОБИ НАВЧАННЯ

Освітні вимірювання як окрема галузь знаходиться на етапі формування. Тим не менше потреба суспільства у фахівцях з освітніх вимірювань уже сьогодні досить гостра. Тому існує нагальна необхідність у розробці як програм, так і методик підготовки відповідних фахівців.

У даному розділі розглянуто питання розвитку освітніх вимірювань як окремої галузі, зв'язку її з іншими галузями науки і освіти, методів досліджень у галузі освітніх вимірювань. Приділяється увага проблемам освітніх вимірювань.

Значне місце у розділі присвячене власне питанням методики навчання дисциплін, що становлять зміст освітніх вимірювань: методиці навчання математико-статистичних основ освітніх вимірювань, методиці навчання розробки та використання тестів, методиці навчання розробки та використання тестів. Детально розглядаються компоненти методичної системи навчання освітніх вимірювань, програма підготовки та освітньо-кваліфікаційна характеристика фахівця з освітніх вимірювань, вимоги до професійної відповідальності при оцінюванні та тестуванні.

8.1. Освітні вимірювання як галузь

8.1.1. Розвиток освітніх вимірювань як окремої галузі

Навчання в інформаційному суспільстві стає провідною діяльністю та способом існування для більшості громадян: потреба у підвищенні кваліфікації або отриманні ліцензії на виконання певної діяльності, потреба у перекваліфікації в зв'язку зі скороченням або зникненням попиту на фахівців певного фаху, розвиток нових технологій та інструментів у традиційних широко застосовних галузях, – усе це спонукає людей до «навчання протягом життя», а отже значно розширює аудиторію тих, хто навчається і потребує оцінювання знань, кваліфікації та професійної придатності. Разом з розвитком сучасних засобів навчання, орієнтованих на використання інформаційно-комунікаційних технологій, це потребує нових, більш технологічних підходів до організації системи навчання в цілому та системи оцінювання навчальних та кваліфікаційних досягнень зокрема. У цьому

контексті формування та розвиток такої галузі як освітні вимірювання є надзвичайно актуальним для вирішення протиріччя між тенденцією переформування освітнього процесу (технологізацією) та суб'єктивними, заснованими переважно на оцінках експертів-вчителів засобами оцінювання.

У розвитку освітніх вимірювань як самостійної галузі можна виділити декілька етапів. Ще у стародавньому Китаї застосовували систему конкурсного відбору претендентів на певні посади. На жаль на довгі роки цей досвід було втрачено. У середні віки у різних країнах виникли бальна та рейтингова системи оцінювання навчальних досягнень, засновані на експертному оцінюванні викладачем. Наступним етапом можна визнати кінець 19 ст., коли у США для перевірки грамотності учнів започаткували використання тестових завдань з арифметики та письма, які стали поштовхом для розвитку тестів на здібності. На початку 20 ст. розвиток психометрії спонукав до розвитку математичних моделей та методів аналізу результатів тестування, було започатковано розробку та теоретичне обґрунтування більшості із сучасних методів статистичного аналізу даних. А з розвитком комп'ютерних та комунікаційних технологій, технологій комп'ютерного та дистанційного навчання стали розвиватися сучасні засоби оцінювання, засновані на застосуванні кваліметричного підходу та розробці інструментів, власне, вимірювання рівня знань, здібностей, кваліфікації тощо, таких як тести, шкали.

На сьогоднішньому етапі – освітні вимірювання це потужна галузь, яка формувалася в рамках науки психометрії і нині існує на стику різних дисциплін, передусім психології, педагогіки, математики та інформатики.

Саме тому програма підготовки фахівця з освітніх вимірювань охоплює такі дисципліни циклу професійної та практичної підготовки: вища математика; теорія ймовірностей та математична статистика; вимірювання в освіті; математично-статистичні методи в освітніх вимірюваннях; класичні тестові моделі та технології; моделі і методи IRT; основи педагогічного оцінювання; інформаційно-телекомунікаційні технології в освіті та науці; моніторинг якості освіти; конструювання тестів; прикладна статистика; когнітивна психологія та психометрія; психологія та педагогіка вищої школи.

Об'єктом освітніх вимірювань є знання, уміння, навички, компетенції, інші кваліфікаційні характеристики, які підлягають оцінюванню в освітній діяльності та в галузі сертифікації та ліцензування професійної діяльності.

Предметом освітніх вимірювань є теоретичні засади, методи і засоби оцінювання та вимірювання, доцільні для прийняття рішень в галузі освіти, ліцензування та сертифікації професійної діяльності.

8.1.2. Проблеми освітніх вимірювань

Слід розрізняти використовувані в галузі освітніх вимірювань терміни «вимірювання» та «оцінювання».

Під вимірюванням (англ. *measurement*) розуміють стандартний інструмент для визначення характеристик досліджуваного об'єкта та саму процедуру, в результаті якої значення вимірюваної характеристики об'єкта 1) порівнюється з еталоном (одиницею вимірювання) та 2) перетворюється у зручний для використання вигляд (найчастіше фіксується у вигляді числового значення). Вимірювання широко застосовують до фізичних характеристик об'єктів, таких як довжина, температура, швидкість і тому подібне. Розроблено спеціальні інструменти для вимірювання (лінійка, термометр, спідометр), визначено одиниці та шкали вимірювання таких характеристик.

Ключовими проблемами освітніх вимірювань є, по-перше, визначення об'єкта вимірювань (визначення моделі компетенції в будь-якій галузі діяльності є нетривіальною задачею) та, по-друге, розробка та стандартизація шкал і визначення одиниць вимірювання: що? чим? як?

Оцінювання (*assessment*) в англomовній літературі – це процес збору даних для опису або кращого розуміння питання. Термін найбільше відповідає прийнятому у нас терміну «педагогічне оцінювання», тобто оцінювання знань та навичок, набутих в процесі навчання, оцінювання прогресу у навчанні. Найчастіше оцінювання здійснюється за допомогою бальної шкали або рейтингу вчителем, який виступає у ролі експерта. Одним з видів *assessment* є тести.

Проблемами оцінювання є 1) низька об'єктивність – експерт може бути упередженим по відношенню до учнів і оцінювати не стільки їхні досягнення, як своє ставлення до них; 2) кваліфікація експерта – кожен вчитель має власну «модель» компетентності учня, згідно якої здійснює оцінювання; 3) далеко не всі види діяльності учня можна діагностувати – якщо володіння навичками легко продемонструвати, то розуміння продемонструвати і перевірити набагато складніше; 4) оцінювання потребує часу та достатньої кількості експертів, що в умовах технології навчання також становить

проблему. Перелічені проблеми в основному вирішуються застосуванням стандартизованих тестів, які по-суті є інструментом об'єктивного, неупередженого вимірювання. Однак розробка таких тестів також становить проблему.

Каменем спотикання при розробці стандартизованих тестів, придатних для оцінювання та вимірювання, є їхня валідність та надійність. Надійний тест дозволяє отримати точну оцінку вимірюваної характеристики, відтворювану при повторних вимірюваннях. Під валідністю розуміють адекватність теста меті вимірювання.

Оцінювання в іншому розумінні (англ. *evaluation*) – це результат судження на основі визначених заздалегідь критеріїв. Наприклад, результатом оцінювання може бути висновок про те, що працівник заслуговує або ні присвоєння більш високого рівня кваліфікації; про те, що учень може бути переведений до наступного класу або залишений на повторний курс, зарахований на навчання до ВНЗ і т.п.

Проблемами оцінювання у такому сенсі є розробка критеріїв для прийняття зважених, обґрунтованих рішень.

Самі ж освітні вимірювання повинні розглядатися в широкому розумінні: не лише як тестові технології, адже вони глибоко укорінені в контекст педагогічного (і не тільки) оцінювання, загальних методів психолого-педагогічних досліджень, моніторингу якості освітніх програм та послуг, критеріального професійного відбору.

8.2. Методи досліджень у галузі освітніх вимірювань

8.2.1. Методи емпіричного дослідження

Освітні вимірювання як галузь науки має потужний інструментарій для дослідження явищ та процесів, як на емпіричному, так і на теоретичному рівні. На емпіричному рівні встановлюють нові наукові факти і на основі їх узагальнення формують закономірності. На теоретичному рівні – ви сувають і формулюють основні загальні закономірності, що дають змогу пояснити раніше відкриті факти, а також передбачити майбутні факти і явища. В галузі освітніх вимірювань емпіричне дослідження спрямоване безпосередньо на конкретний об'єкт і спирається на дані спостережень та експериментів.

Основними методами емпіричного дослідження є спостереження, анкетування, інтерв'ю, тестування, аналіз продуктів діяльності, шкалування, експеримент.

З перелічених методів найбільш широко в галузі освітніх вимірювань застосовується саме тестування. Однак інші вище перелічені методи також застосовуються на різних етапах дослідження. Так спостереження за ходом проведення процедури пробного тестування та детальний аналіз робіт, виконаних учнями, може допомогти виявити помилки та неточності у формулюваннях, незручності, що виникають у людей з особливими потребами, і таке інше. Анкетування дозволяє з'ясувати ставлення фахівців та широкої громадськості до процедури освітніх вимірювань, правових, технічних та інших аспектів їх проведення. Інтерв'ю дозволяє ознайомитися з думками фахівців стосовно моделі компетенцій у певній предметній галузі, специфічних засобів оцінювання тощо.

У психологічних дослідженнях *тести* – це об'єктивне і стандартизоване визначення вибору поведінки. Метою тестів є визначення різниці між індивідами або між реакціями одного індивіда в різних умовах. Призначення тесту як методу дослідження полягає у виявленні відносного рівня розвитку особи, аналізу його змін під впливом певних дій; визначення рівня розумового розвитку, ступеня обдарованості, психологічного стану людини; структурних якостей особистості, прогнозованої межі можливостей подальшого розвитку людини і т.п.

Шкалування означає визначення масштабу, одиниці вимірювання. Це діагностичний метод, за допомогою якого реальні якісні явища отримують своє числове відображення у формі кількісних оцінок. Він може бути використаний для оцінювання якостей особистості, виявленні ставлення учнів до різних видів діяльності (навчальної, трудової, ігрової) тощо.

Різновидом шкалування є побудова *рейтингу* – діагностичний метод визначення досягнень учнів шляхом збирання суджень і різнобічних оцінок, даних компетентними експертами, або шляхом ранжування (впорядкування за зростанням або спаданням) деяких комплексних показників, наприклад, рейтинг успішності може враховувати результати навчання з усіх дисциплін. Застосування певної шкали потребує теоретичного обґрунтування.

При розробці нових шкал, виявленні рівнів прояву психолого-педагогічних характеристик перспективним є застосування методів

експертного оцінювання. Завданням методу є визначення об'єктивних рівнів прояву досліджуваної характеристики у об'єктів дослідження шляхом опитування думки висококваліфікованих фахівців-експертів у питаннях, з яких поки не існує формалізованих процедур оцінювання; виявлення розбіжностей у думках експертів та з'ясування причин неузгодженостей. У подальшому результати експертного оцінювання використовують для розробки критеріїв та шкал оцінювання. Прикладом експертного оцінювання може бути традиційне оцінювання досягнень учня вчителем, оцінювання конкурсних робіт та виступів на спортивних змаганнях групою експертів (журі, суддями) і т.п.

8.2.2. Педагогічний експеримент

Теоретичне дослідження пов'язане з розвитком і вдосконаленням поняттєвого апарату відповідної науки і спрямоване на різнобічне пізнання об'єктивної дійсності разом з її суттєвими зв'язками і закономірностями. До теоретичних методів дослідження відносять: аналіз і синтез, індукцію та дедукцію, логічні методи (порівняння, розрізнення, подібність), а також моделювання й частково- експеримент. Теоретичні методи дослідження передбачають глибокий аналіз фактичного матеріалу на рівні сутності, абстрагування від усього другорядного, виявлення процесу в чистому вигляді, розкриття його істотних закономірностей, пояснення зовнішнього внутрішнім тощо.

Основним методом наукового педагогічного дослідження є педагогічний експеримент– комплекс методів, що забезпечує науково - об'єктивну і доказову перевірку правильності обґрунтованої на початку дослідження гіпотези. Це цілеспрямований, заздалегідь організований дослід із розробленою системою впливів, умов і дій, які дозволяють планово втручатися в процес, змінювати і кількаразово повторювати особливості його протікання, стежити за особливостями прояву досліджуваних якостей особистості, виявляти та пояснювати зміни прояву досліджуваних якостей особистості під впливом умов проведення експерименту.

Проведення експерименту відбувається у декілька етапів:

1. На підготовчому етапі вивчають стан розвитку відповідної галузі освіти, виявляють актуальні проблеми, що потребують вирішення за допомогою експерименту, визначають проблеми, предмет та об'єкт, цілі і

завдання дослідження, вивчають науково-методичну літературу, будують гіпотези та конструюють план-програму експерименту, добирають методи та розробляють методики дослідження відповідно до обраних методологічних принципів.

2. Дослідницький етап розподіляється на три стадії. Основним завданням констатувального експерименту є встановлення фактичного стану та рівня досліджуваних характеристик учасників експерименту на момент проведення дослідження, наприклад, здійснення діагностики наявного рівня знань.

Формувальний експеримент спрямований на вивчення досліджуваного явища безпосередньо в процесі спеціально організованого експериментального навчання та активного формування досліджуваних особливостей.

Контрольний експеримент проводиться з метою порівнянні отриманих результатів експериментальної групи, яка брала участь у формувальному експерименті, з результатами контрольної групи (зазвичай з тою, яка брала участь у констатувальному експерименті).

3. На етапі обробки даних дослідження передбачене застосування математико-статистичних методів узагальнення та зведення в систему отриманих даних, виявлення прихованих закономірностей серед кількісних показників, вимірюваних в ході експерименту. Методи статистичної обробки дають можливість підтвердити чи спростувати висунуті на початку експерименту гіпотези та зробити висновки про достовірність здобутих результатів.

4. На етапі інтерпретації даних дослідження та формулювання висновків здійснюється тлумачення та наукове обґрунтування досліджуваного явища.

5. На етап впровадження основні результати експерименту висвітлюються у фахових виданнях, на конференціях та школах; з'ясовуються практичне та теоретичне значення отриманих висновків, шляхи для їхнього впровадження у практику.

В галузі освітніх вимірювань особливе місце посідають моніторингові дослідження та притаманні прикладним математичним дисциплінам методи розробки та дослідження математичних моделей.

8.3. Методична система навчання освітніх вимірювань

8.3.1. Складові методичної системи

Системний підхід є одним з провідних методологічних принципів дослідження в кожній галузі. Під системою розуміємо єдність елементів, які функціонують за внутрішньо властивими їй законами.

Педагогічна система являє собою множину взаємопов'язаних структурних та функціональних компонентів, підпорядкованих цілям навчання, освіти та виховання. Структурними компонентами педагогічної системи є навчальна інформація; засоби педагогічної комунікації (форми, методи, прийоми, засоби); учні; педагоги; мета. Зв'язки між структурними компонентами забезпечують функціональні компоненти: конструктивні, комунікаційні, організаційні, прогностичні, гностичні. Таким чином, під педагогічною системою необхідно розуміти динамічно функціонуючий комплекс діалектично пов'язаних між собою компонентів і чинників, які створюють оптимальні умови для розв'язання завдань навчання і виховання людей залежно від соціальних завдань.

Підсистемою педагогічної системи є методична система навчання певної дисципліни або галузі знань. Кожна методична система являє собою сукупність п'яти компонентів: цілі, зміст, методи, засоби та організаційні форми, – притаманних їй та таких, що не суперечать загальній педагогічній системі.

Методична система навчання освітніх вимірювань формується разом з формуванням самої галузі.

8.3.2. Цілі навчання освітніх вимірювань

Метою навчання освітніх вимірювань є підготовка фахівця, здатного на професійному рівні

- розробляти і вибирати методи оцінювання та вимірювання, доцільні для прийняття рішень в галузі освіти, ліцензування та сертифікації професійної діяльності;
- використовувати результати оцінювання та вимірювання для прийняття рішень щодо окремих осіб та груп осіб, планувати навчальний процес, розробляти нові та поліпшувати існуючі навчальні та освітні програми;

- доводити результати освітніх вимірювань, їхню інтерпретацію та прийняті на її основі рішення до зацікавлених осіб та широких соціальних верств населення;
- знати існуючі теорії та методи оцінювання та вимірювання, уміти застосовувати їх та інтерпретувати отримані результати;
- знати технічні, юридичні та етичні вимоги до засобів оцінювання та вимірювання, уміти визначати відповідність цим вимогам як самих засобів, так і процедури проведення оцінювання та вимірювання, оцінювати інформацію щодо результатів їх застосування;
- розробляти і проводити тестування та експертизу рівнів підготовленості, знань та навичок учасників тестування;
- виконувати статистично-математичне та комп'ютерне моделювання процесу тестування;
- виконувати моніторинг якості освітніх програм та оцінювання якості освітніх послуг, що надаються навчальними закладами України;
- розробляти та застосовувати програми професійного відбору для реального та фінансового секторів економіки;
- знати методи та засоби комп'ютеризації оцінювання та вимірювання і аналізу їх результатів, використовувати інформаційно-комунікаційні технології у фаховій практичній діяльності.

8.3.3. Типові задачі діяльності фахівця з освітніх вимірювань

Типові задачі діяльності фахівця зі спеціальності «Освітні вимірювання» можна умовно поділити на дослідницькі, проектувальні, організаційні, управлінські та технологічні.

1. *Дослідницькі задачі*: аналіз та опис культурного середовища, інтересів, навичок та здібностей учнів¹ при плануванні навчального процесу; аналіз інформації, зібраної до та під час навчального процесу, для розуміння реального прогресу учнів та для прийняття рішень щодо подальшого змісту навчання; статистичний аналіз результатів підсумкового оцінювання навчальних досягнень осіб та груп осіб; дослідження якісних характеристик навчального процесу на рівні окремих осіб та їх груп, навчальних закладів, груп навчальних закладів за їх типом, регіонів;

¹ Далі в тексті учнями будемо називати усіх осіб, які проходять навчання, незалежно від віку та категорії навчального закладу, тобто як учнів шкіл, курсів підвищення кваліфікації, так і студентів університетів, коледжів тощо.

аналіз валідності та надійності результатів освітніх вимірювань; дослідження якості освітніх програм.

2. *Проектувальні задачі*: планування навчального процесу для окремих осіб та їх груп; розробка та відбір інструментів і методів освітнього вимірювання та оцінювання, професійного відбору, ліцензування та сертифікації професійної діяльності; розробка та відбір методів дослідження якості освітніх програм; консультування з питань професійного відбору, ліцензування та сертифікації професійної діяльності.
3. *Організаційні задачі*: розвиток особистості фахівця та організація його професійної діяльності; планування діяльності, формування організаційної структури колективу з оцінювання та моніторингу якості освіти; планування та організація заходів стандартизованого оцінювання.
4. *Управлінські задачі*: прийняття рішень на основі даних вимірювання навчальних досягнень або здібностей особи чи групи осіб; керування колективом з оцінювання та моніторингу якості освіти.
5. *Технологічні задачі*: здійснення формівного оцінювання досягнень учнів; здійснення підсумкового оцінювання досягнень учнів; адміністрування заходів стандартизованого тестування; документування результатів оцінювання та ознайомлення з ними зацікавлених сторін; здійснення збору даних для дослідження якості освітніх програм.

8.3.4. Зміст підготовки фахівця з освітніх вимірювань

Цілі навчання визначають його зміст. Для успішної діяльності фахівцеві у галузі освітніх вимірювань необхідно знати основні поняття, категорії, методи психології, зокрема когнітивної психології та психометрії; знати та вміти застосовувати основні методи психолого-педагогічного дослідження; знати понятійний апарат і основні постулати теорії педагогічних вимірювань (рівні педагогічних вимірювань, види шкал, види розподілів даних педагогічних вимірювань), методи оцінювання якості та рівня навчальних досягнень учнів, дидактичні принципи тестового контролю навчальних досягнень, форми і методи його проведення і організації; знати основні напрями досліджень в галузі педагогічних вимірювань, сучасні аспекти проблеми тестування, методологічні засади конструювання і застосування тестових вимірників; знати класичні та сучасні тестові моделі; знати математико-статистичні

методи, які застосовуються для обробки результатів педагогічних, психологічних та соціологічних досліджень, мати навички організації та проведення наукових досліджень; знати основні положення цивільного захисту, охорони праці в галузі, інтелектуальної власності, володіти діловою іноземною мовою; знати прикладне програмне забезпечення, що використовується в освітніх вимірюваннях, та вміти його застосовувати.

Все це знайшло відображення у освітньо-кваліфікаційній характеристиці, освітньо-професійній програмі та навчальному плані підготовки фахівців спеціальності «Освітні вимірювання», у яких приділено належну увагу теоретичним засадам педагогічних вимірювань, основам педагогічного оцінювання, спеціальній математико-статистичній підготовці та опануванню сучасних комп'ютерних технологій. Згідно навчального плану на опанування математичної складової відведено 35% навчального часу, психологічної – 17%, на опанування організаційних форм та технологій освітньої діяльності – 36% та на інші, включаючи інформаційно-комунікаційні, – 12%.

Основним питанням, які становлять зміст освітніх вимірювань, присвячено попередні розділи.

8.3.5. Принципи, методи та організаційні форми навчання освітніх вимірювань

При навчанні дисциплін підготовки фахівця з освітніх вимірювань застосовуються загальні принципи та методи навчання. Однак варто враховувати специфіку та особливості методик викладання дисциплін, включених до навчального плану. Так у викладанні математичних дисциплін особливу увагу слід приділяти доказовості та наочності викладення матеріалу; при викладанні гуманітарних дисциплін, зокрема дисциплін психолого-педагогічного блоку, варто застосовувати проблемні та дискусійні методи; при викладанні комп'ютерно-орієнтованих дисциплін – слід більше уваги приділяти навичкам роботи з відповідним програмним забезпеченням та формувати цілісне уявлення учнів про сучасні інформаційно-комунікаційні технології та засоби їх реалізації.

Разом з тим основними принципами, яких слід дотримуватися при викладанні дисциплін, що становлять ядро галузі освітніх вимірювань, є принципи науковості, системності та систематичності, зв'язку теорії з практикою, міжпредметних зв'язків.

Основними формами навчальної діяльності у вищій школі є лекція, семінар, практичне та лабораторне заняття, самостійна робота та навчальна практика. Усі ці форми використовуються також і при підготовці фахівця з освітніх вимірювань. Крім того, обов'язковим є написання курсової та дипломної роботи.

Методи навчання (гр. *methodos* – шлях пізнання, спосіб знаходження істини) – це впорядковані способи взаємопов'язаної, цілеспрямованої діяльності педагога й студентів, спрямовані на ефективне розв'язання навчально-виховних завдань. Вони реалізуються через систему прийомів і засобів навчальної діяльності. У практиці організації навчального процесу у вищій школі методи навчання тісно пов'язані з організаційними формами навчального процесу. Основними серед них залишаються лекція, лабораторне, практичне та семінарське заняття, самостійна робота студентів та виробнича практика.

Лекція – це метод, за допомогою якого педагог у словесній формі розкриває сутність наукових понять, явищ, процесів, логічно пов'язаних, об'єднаних загальною темою. Вона передбачає систематичний, послідовний виклад навчального матеріалу, будь-якого питання, теми, розділу, предмета, методів науки. Сучасна лекція при використанні мультимедійних засобів дозволяє у повній мірі проілюструвати матеріал, що розглядається під час лекції, продемонструвати у динаміці різноманітні його аспекти. Однак прийоми та методи, використовувані на лекціях, диктуються специфікою дисципліни.

Для дисциплін психолого-педагогічного циклу основними методами є вербальні проблемні методи, які реалізуються на семінарських заняттях. *Семінар* – вид навчальних занять практичного характеру, спрямованих на поглиблене опрацювання теоретичного матеріалу. Семінарські заняття покликані сприяти активізації пізнавальної діяльності студентів, формуванню самостійності суджень, умінню обстоювати власні думки, аргументувати їх на основі наукових фактів. Формами проведення семінарів є теоретична конференція, дискусія, обговорення рефератів, «мозковий штурм» тощо.

Практичні роботи є однією з провідних форм організації навчання математичних дисциплін. Практична робота спрямована на використання набутих під час лекції знань у розв'язуванні практичних вправ та завдань. Вправа – це метод навчання, сутність якого полягає у цілеспрямованому,

багаторазовому повторенні студентами окремих дій чи операцій з метою формування умінь та навичок. Завданням практичного заняття математичного циклу є також опанування основних алгоритмів розв'язування задач та методів прийняття рішень, дослідження математичних моделей, перевірка та доведення їх ефективності.

Лабораторний метод є провідною формою опанування прикладного програмного забезпечення: його інструментарію та технології застосування. Традиційно лабораторна робота передбачає використання спеціального обладнання та певної технології для набуття нових знань або перевірки наукових гіпотез на рівні досліджень. Комп'ютерні лабораторні роботи передбачають використання спеціального програмного забезпечення та проведення дослідження за його допомогою.

Значне місце у підготовці фахівця посідає виробнича практика, на якій відпрацьовуються та перевіряються рівні опанування відповідними професійними компетенціями в умовах, наближених до умов майбутньої діяльності.

Запорукою якісної підготовки фахівця при опануванні усіх дисциплін навчального плану є самостійна робота. Найбільша її частка при написанні курсової та дипломної роботи. В організації самостійної роботи студентів найбільше прояву мають проблемний, частково-пошуковий та дослідницький методи навчання, які передбачають створення проблемних ситуацій, дослідницьких завдань та спонукання студентів до їх розв'язання.

8.4. Галузевий стандарт та програми підготовки фахівців з освітніх вимірювань

8.4.1. Галузевий стандарт підготовки магістра

Спеціальність 8.18010022 Освітні вимірювання галузі знань 1801 Специфічні категорії забезпечена стандартом вищої освіти освітньо-кваліфікаційної характеристики та освітньо-професійної програми освітньо-кваліфікаційного рівня «магістр». На даний час кожен університет розробляє внутрішній варіант нормативної компоненти, а також його варіативну компоненту. До складу стандарту спеціальності 8.18010022 Освітні

вимірювання входять освітньо-кваліфікаційна характеристика (ОКХ) та освітньо-професійна програма (ОПП), які тимчасово діють до введення галузевого стандарту вищої освіти України, мають бути погодженими з Науково-методичною комісією Міністерства освіти і науки, молоді та спорту України, Департаментом вищої освіти Міністерства освіти і науки, молоді та спорту України та Інститутом інноваційних технологій і змісту освіти Міністерства освіти і науки, молоді та спорту України.

На основі ОКХ та ОПП в університеті розробляється навчальний план підготовки магістра за типовою формою, затвердженою Міністерством освіти і науки, молоді та спорту України. Навчальний план затверджується на засіданні вченої ради університету. Навчальним планом передбачений перелік нормативних, вибіркових дисциплін, види аудиторних занять та форми підсумкового контролю.

Освітньо-кваліфікаційна характеристика (ОКХ) випускників вищого навчального закладу є галузевим нормативним документом, в якому узагальнюється зміст вищої освіти, тобто відображаються цілі вищої освіти та професійної підготовки, визначається місце фахівця в структурі галузей економіки держави і вимоги до його компетентності, інших соціально важливих властивостей та якостей.

ОКХ є складовою галузевих стандартів вищої освіти, в якій узагальнюються вимоги з боку держави, світового співтовариства та споживачів випускників до змісту вищої освіти. ОКХ відображає соціальне замовлення на підготовку фахівця з урахуванням аналізу професійної діяльності та вимог до змісту вищої освіти з боку держави та окремих замовників фахівців.

ОКХ установлює галузеві кваліфікаційні вимоги до соціально-виробничої діяльності випускників вищого навчального закладу з певних спеціальностей та освітньо-кваліфікаційного рівня і державні вимоги до властивостей та якостей особи, яка здобула певної освітній рівень відповідного фахового спрямування.

ОКХ використовується для:

- визначення цілей освіти та професійної підготовки;
- розроблення та корегування складових державного стандарту вищої освіти (перелік кваліфікацій за відповідними освітньо-кваліфікаційними

рівнями та перелік напрямів та спеціальностей, за якими здійснюється підготовка фахівців за відповідними освітньо-кваліфікаційними рівнями) та галузевих стандартів вищої освіти (освітньо-професійна програма підготовки фахівців, засоби діагностики якості вищої освіти);

- визначення первинних посад випускників вищих навчальних закладів та умов їх використання;
- розроблення та корегування складових стандартів вищої освіти вищих навчальних закладів (варіативні частини освітньо-кваліфікаційної характеристики, освітньо-професійної програми підготовки фахівців та засобів діагностики якості вищої освіти, навчальний план, програми навчальних дисциплін);
- визначення змісту навчання в системі перепідготовки та підвищення кваліфікації;
- професійної орієнтації здобувачів фаху та визначення критеріїв професійного відбору;
- прогнозування потреби у фахівцях відповідної спеціальності та освітньо-кваліфікаційного рівня, плануванні їх підготовки і під час укладання договорів або контрактів щодо підготовки фахівців;
- розподілу та аналізу використання випускників вищих навчальних закладів.

Загальні вимоги до властивостей і якостей випускників вищого навчального закладу як соціальних особистостей подаються у вигляді переліків компетенцій щодо вирішення певних проблем і задач соціальної діяльності, інструментальних, загально-наукових і професійних компетенцій та системи умінь, що забезпечують наявність цих компетенцій.

Компетентність – інтегрована характеристика якостей особистості, результат підготовки випускника вузу для виконання діяльності в певних професійних та соціально-особистістних предметних областях (компетенціях), який визначається необхідним обсягом і рівнем знань та досвіду у певному виді діяльності.

Компетенція включає знання й розуміння (теоретичне знання академічної області, здатність знати й розуміти), знання як діяти (практичне й оперативне застосування знань до конкретних ситуацій), знання як бути (цінності як невід'ємна частина способу сприйняття й життя з іншими в соціальному контексті).

Вищі навчальні заклади готують випускників як соціальних особистостей, здатних вирішувати певні проблеми і задачі діяльності за умови оволодіння системою умінь та компетенцій, що визначені в ОКХ.

Відповідно до посад, які можуть обіймати випускники вищого навчального закладу, вони повинні бути здатними до виконання виробничих функцій та типових для даної функції задач професійної діяльності. Кожній типовій задачі відповідає компетенція, яка формується системою умінь щодо вирішення цієї задачі діяльності.

Вищі навчальні заклади забезпечують опанування випускниками системою умінь та набуття відповідних компетенцій, які дозволять вирішувати типові задачі діяльності під час здійснення певних виробничих функцій.

Освітньо-професійна програма (ОПП) є галузевим нормативним документом, у якому визначається нормативний термін та зміст навчання, нормативні форми державної атестації, встановлюються вимоги до змісту, обсягу й рівня освіти та професійної підготовки фахівця відповідного освітньо-кваліфікаційного рівня певного напрямку.

Цей стандарт є складовою галузевого стандарту вищої освіти і використовується під час:

- розроблення складової галузевого стандарту вищої освіти (засоби діагностики якості вищої освіти);
- розроблення складових стандартів вищої освіти вищих навчальних закладів (варіативні частини освітньо-професійної програми підготовки фахівців та засобів діагностики якості вищої освіти);
- розроблення навчального плану, програм навчальних дисциплін і практик.

Варіативна частина освітньо-професійної програми (ОПП) є документом, в якому визначається зміст навчання за вибірковими навчальними дисциплінами, встановлюються вимоги до якості, обсягу та рівня освітньої і професійної підготовки фахівця освітньо-кваліфікаційного рівня магістр спеціальності «Освітні вимірювання».

Варіативна частина (ОПП) є доповнюючою складовою галузевої компоненти державних стандартів вищої освіти і використовується під час:

- розроблення відповідних навчальних планів і програм навчальних дисциплін;

- розроблення засобів діагностики рівня освітньо-професійної підготовки фахівця;
- визначення змісту навчання студентів спеціальності «Освітні вимірювання» як бази для опановування новими спеціальностями, кваліфікаціями;
- визначенні змісту навчання в системі перепідготовки та підвищення кваліфікації.

8.4.2. Ключові компетенції фахівця з освітніх вимірювань

Основними компетенціями фахівця з освітніх вимірювань визначені компетенції соціально-особистісні, загальнонаукові, інструментальні, загально-професійні та спеціалізовано-професійні.

Соціально-особистісними компетенціями визначено 1) здатність застосовувати свої знання й розуміння, здатність до вирішення проблем в нових або незнайомих ситуаціях у більш широкому (або міждисциплінарному) контексті, ніж освітні вимірювання; 2) здатність інтегрувати знання й долати труднощі, а також формулювати судження на основі неповної або обмеженої інформації, з урахуванням етичної та соціальної відповідальності за їх застосування; 3) здатність брати на себе відповідальність за управління і перетворення в умовах обраної роботи та навчання, які характеризуються як складні, непередбачувані і вимагають нових стратегічних підходів.

Загальнонауковими компетенціями визначено 1) знання основ філософії, психології, педагогіки, що сприяють розвитку загальної культури та соціалізації особистості, спрямовують її до етичних цінностей; 2) знання фундаментальних розділів математики, в обсязі, необхідному для володіння математичним апаратом психології, педагогіки, здатність використовувати математичні методи в освітніх вимірюваннях; 3) знання фундаментальних розділів статистики, в обсязі, необхідному для володіння статистичним апаратом психології, педагогіки, здатність використовувати статистичні методи в освітніх вимірюваннях; 4) базові знання з інформатики, необхідні для засвоєння загально-професійних дисциплін освітніх вимірювань.

Інструментальними компетенціями визначено 1) здатність до ділової усної та письмової комунікації іноземною мовою; 2) здатність до написання іноземною мовою наукових доповідей та статей; 3) здатність до проведення

наукових та прикладних досліджень за допомогою інформаційних технологій.

Загально-професійними компетенціями бакалавра визначено:

- 1) знання методології досліджень, методів дослідження та аналізу освітніх систем та процесів, розуміння складності освітніх систем та процесів, їх різноманіття, багатофункціональності, взаємодії та умов існування для розв'язання прикладних і наукових завдань в галузі освіти;
- 2) знання математичних методів побудови та аналізу моделей освітніх систем та процесів, розробки нових математично обґрунтованих інструментів та методів досліджень в галузі освіти;
- 3) знання вимог чинних державних та міжнародних стандартів в галузі освіти, методів і засобів їх розробки та контролю за їх дотриманням;
- 4) знання основних методів та підходів щодо організації, планування, керування та контролю робіт з проектування, розроблення та впровадження освітніх проектів;
- 5) знання вимог чинних державних та міжнародних стандартів в галузі ліцензування та сертифікації професійної діяльності, методів і засобів контролю за їх дотриманням, методами та засобами їх розробки;
- 6) знання основних теоретичних підходів, теорій, фундаментальних і прикладних галузей педагогіки для розв'язання прикладних і наукових завдань в галузі освіти;
- 7) знання основних теоретичних підходів, теорій, фундаментальних і прикладних галузей психології для розв'язання прикладних і наукових завдань в галузі освіти.

Спеціалізовано-професійними компетенціями бакалавра визначено:

- 1) уміння вибирати та застосовувати методи оцінювання та вимірювання, доцільні для прийняття рішень в галузі освіти та професійного відбору;
- 2) уміння інтерпретувати результати процедур оцінювання та вимірювання;
- 3) уміння на основі інтерпретації результатів процедур оцінювання та вимірювання приймати рішення щодо окремих осіб та груп осіб, планування навчального процесу, розробки нових та поліпшення існуючих навчальних та освітніх програм;

- 4) уміння шкалювати та доводити результати освітніх вимірювань, їх інтерпретацію та прийняті на їх основі рішення для зацікавлених осіб та широких соціальних верств;
- 5) знання технічних, юридичних та етичних вимог до засобів оцінювання та вимірювання, уміння визначати відповідність цим вимогам як самих засобів оцінювання та вимірювання, так і процедури їх проведення та інформації щодо результатів їх застосування;
- 6) знання методів та засобів комп'ютеризації оцінювання та вимірювання і аналізу їх результатів, уміння їх використовувати;
- 7) уміння вибирати та розробляти процедури моніторингу освітніх програм і процесів.

8.4.3. Програма підготовки бакалавра

Нагальною проблемою сьогодення є недостатня підготовка педагогічних працівників, психологів, науковців, які залучаються до створення тестових завдань, для їх перевірки та інтерпретації отриманих результатів, оскільки процес тестування вимагає роботи фахівців у галузі освітніх вимірювань, озброєних комплексом знань, умінь та досвідом в оцінюванні досягнень учнів. Для підготовки таких фахівців в Україні групою супроводження Проекту TEMPUS «Освітні вимірювання, адаптовані до стандартів ЄС» було запропоновано поєднувати підготовку майбутніх учителів різного профілю з додатковою спеціалізацією «Освітні вимірювання» на рівнях бакалавра та спеціаліста.

Програма підготовки бакалавра покликана забезпечувати підготовку фахівців за додатковою спеціалізацією «Освітні вимірювання» і може бути включена до варіативного блоку навчальних планів підготовки бакалаврів за напрямками 6.010103 Технологічна освіта, 6.020302 Історія*, Українська мова та література*, 6.040101 Хімія*, 6.040102 Біологія*, 6.040104 Географія*, 6.040201 Математика*, 6.040203 Фізика*.

Для реалізації програми підготовки фахівців розроблено навчальний план для спеціалізації «Освітні вимірювання», у якому виділено час на теоретичну частину педагогічний вимірювань, на вивчення основ педагогічного оцінювання, на спеціальну математично-статистичну підготовку, а також на опанування сучасних комп'ютерних технологій.

По закінченню навчання фахівець у галузі освітніх вимірювань повинен володіти інформацією про різні аспекти проблеми тестування: сучасні напрями досліджень в області педагогічних вимірювань; методологічні основи процесів конструювання і застосування тестових вимірників; методи оцінки якості навчальних досягнень тих, що навчаються. Крім того, для подальшої професійної діяльності фахівець у галузі педагогічних вимірювань повинен знати основні поняття, категорії, методи психології і педагогіки; дидактичні принципи тестового контролю навчальних досягнень, форми і методи його проведення і організації; понятійний апарат і основні постулати теорії педагогічних вимірювань; рівні педагогічних вимірювань, види шкал; види розподілів даних педагогічних вимірювань.

*Навчальний план спеціалізації «Освітні вимірювання»
на освітньому рівні бакалавр*

№	Навчальні дисципліни	Кредити нац.	Кредити ECTS	Кількість годин						Розподіл Годин			
				Всього	Самост. роб.	Ауд.- робота	Лекції	Лабораторні	Практичні	III к.		IVк.	
										5	6	7	8
										Кільк. тижнів			
17	18	17	10										
	Цикл дисциплін самостійного вибору ВНЗ	13,5	20,25	729	483	246	120	34	92	2	5	4	5
1	Вступ до педагогічних вимірювань та моніторингу якості освіти	2,0	3,00	108	72	36	20		16	2			
2	Основи педагогічного оцінювання	3,0	4,50	162	108	54	20	16	18		3		
3	Математично-статистичні методи в педагогічних вимірюваннях	5,0	7,50	270	180	90	42		48		2	3	
4	Педагогічні вимірювання та моніторинг якості освіти	2,0	3,00	108	70	38	28		10			1	2
5	Комп'ютерні технології у тестуванні	1,5	2,25	81	53	28	10	18					3
	Цикл дисциплін самостійного вибору студента	3,0	4,50	162	104	58	10	32	16	0	0	3	0
6	Комп'ютерні статистичні пакети	2,0	3,00	108	70	38	6	32				2	
7	Вибіркові обстеження у психології, соціології та педагогіці	1,0	1,50	54	34	20	4		16			1	
	Кредитів та годин (за планом теоретичних занять)	16,5	24,75	891	587	304	130	66	108	2	5	7	5
8	Практика з освітніх вимірювань (без відриву від навчання)	1,5	2,25	81	81								X
	Загальний обсяг кредитів та годин	18,0	27,00	972	668	304	130	66	108	2	5	7	5

Сферами професійної діяльності фахівців у галузі педагогічних вимірювань є освітні установи, а також інші установи і організації, пов'язані з функціонуванням системи освіти.

На жаль випускник освітньо-кваліфікаційного рівня бакалавр, який навчався за додатковим планом бакалавра на спеціалізації «Освітні вимірювання», разом з кваліфікацією вчителя відповідно до напрямку підготовки не може отримати кваліфікацію «2433.2 Професіонал в галузі інформації та інформаційні аналітики», хоча навчальний план складено так, що він міг би на професійному рівні розробляти та проводити тестування та експертизу рівня підготовки учнів основної школи. Таку кваліфікацію випускник отримує тільки після отримання освітньо-кваліфікаційного рівня «спеціаліст», після опанування особливостей педагогічних вимірювань у старшій школі.

8.4.4. Програма підготовки магістра

Однією із задач Проекту TEMPUS «Освітні вимірювання, адаптовані до стандартів ЄС» є розробка навчально-методичної документації, підготовка професорсько-викладацького складу для здійснення підготовки магістрів за спеціальністю «Освітні вимірювання».

Підготовка магістрів за спеціальністю 8.18010022 Освітні вимірювання має бути спрямована на формування готовності майбутніх випускників до таких видів діяльності: підготовка та проведення тестування (розробка педагогічних тестів різних видів, застосування педагогічних тестів у навчальному процесі); експертна діяльність (експертиза якості тестових вимірників, дослідження у сфері педагогічних вимірювань); просвітницька діяльність (ознайомлення педагогічної громадськості з можливостями педагогічних тестів, проведення занять з проблем педагогічних вимірювань).

Навчальний план названої спеціальності розроблений відповідно до стандарту вищої освіти галузі знань 1801 Специфічні категорії спеціальності 8.18010022 Освітні вимірювання. Він розрахований на 1 рік (2 семестри) підготовки, передбачає вивчення 2-х циклів – цикл професійно-орієнтованої гуманітарної та соціально-економічної підготовки та цикл професійно-орієнтованої математичної та природничо-наукової підготовки, містить нормативну (47 кредитів) та варіативну (13 кредитів) частини. Завершальна (узагальнююча та систематизуюча) стадія підготовки фахівців-магістрів з

освітніх вимірювань згідно затвердженого навчального плану передбачає виконання курсової роботи (1-й семестр), організацію асистентської практики, оформлення магістерської роботи та проведення державної атестації (2-й семестр). Всього узагальнюючий та систематизуючий етап підготовки фахівців має обсяг 19 кредитів ECTS, а теоретичний блок навчального плану має обсяг 41 кредит ECTS.

Структурно-логічна схема програми підготовки магістрів.

Спеціальність «Освітні вимірювання» відноситься до галузі знань «Специфічні категорії», це означає, що випускник кожної спеціальності може вступити до цієї магістратури. При цьому фахівець освітніх вимірювань за короткий час навчання – один рік - має опанувати великий обсяг різнопланового матеріалу, основою якого становить ґрунтовна підготовка в галузі теорії ймовірностей, статистики, психології, педагогічних та комп'ютерних технологій. Стандартом підготовки магістра передбачені дисципліни як для поглибленої підготовки у кожному із названих напрямів, так і суто фахові дисципліни, пов'язані з теорією та практикою вимірювань в освіті.

Перший досвід навчання майбутніх магістрів освітніх вимірювань виявив необхідність удосконалення графіку навчання шляхом викладання у першій половині I семестру фундаментальних математичних дисциплін «Математико-статистичні методи в освітніх вимірюваннях», «Прикладна статистика», паралельно з якими викладається дисципліна «Основи педагогічного оцінювання», яка є вступом до спеціальності. У середині I семестру необхідно провести контроль знань з названих дисциплін, щоб можна було розпочинати навчання базових дисциплін: «Класичні тестові моделі», «Вимірювання в освіті» та «Конструювання тестів». І тільки наприкінці першого семестру з'являється можливість розпочати ефективне викладання дисципліни «Комп'ютерні технології в тестуванні».

Другий семестр розпочинається з асистентської практики, до якої магістранти підійдуть у достатній мірі підготовленими, отримають завдання щодо проведення тестування, пов'язаного з тематикою магістерської роботи. Після практики магістранти продовжують теоретичне навчання ще шість тижнів, продовж яких викладаються курси поглибленого рівня «Моделі і методи IRT», «Когнітивна психологія та психометрія», продовжується вивчення процесу комп'ютерної обробки результатів тестування. Магістранти повинні відвідувати «Науковий семінар», кожний з них має підготувати

матеріал за результатами самостійного дослідження. Також до навчального плану внесено дисципліну «Методика навчання освітніх вимірювань», що дозволить магістрантам отримати кваліфікацію «Викладач університету». Графік викладання таких дисциплін, які читаються в університеті на всіх спеціальностях магістратури у першому семестрі «Ділове спілкування іноземною мовою», «Інтелектуальна власність», «Психологія вищої школи» та «Педагогіка вищої школи», а також у другому семестрі «Цивільний захист» та «Охорона праці в галузі». Послідовність викладання змістових модулів наведено нижче у вигляді структурно-логічної схеми підготовки фахівців.

У підготовці за напрямом «Організаційні форми та технології освітньої діяльності» (13,975 кредитів, 34,09 % загального обсягу підготовки) основний внесок забезпечують три дисципліни: «Комп'ютерні технології в тестуванні», «Конструювання тестів», «Вимірювання в освіті», також «Основи педагогічного оцінювання», «Ділове спілкування іноземною мовою», «Моніторинг якості освіти», «Педагогіка вищої школи» та «Класичні тестові моделі».

Основним напрямком у підготовці магістра є напрямок математичний (16,925 кредитів або 41,28 % загального обсягу підготовки). Вагомий внесок до математичної складової вносять як суто математичні дисципліни: «Математико-статистичні методи в освітніх вимірюваннях» та «Прикладна статистика», – так і професійно-орієнтовані: «Моделі і методи IRT», «Вимірювання в освіті», «Класичні тестові моделі», «Комп'ютерні технології в тестуванні», «Конструювання тестів», «Комп'ютерні статистичні пакети».

Психологічний напрям (6,41 кредитів, 15,63 % загального обсягу підготовки) в основному зосереджений у чотирьох дисциплінах «Когнітивна психологія та психометрія», «Конструювання тестів», «Вимірювання в освіті», «Психологія вищої школи».

Інші напрями підготовки (3,69 кредитів, 9 % загального обсягу підготовки) представлені переважно в дисциплінах гуманітарного напрямку, обов'язкових до вивчення магістрами усіх спеціальностей. Це «Ділове спілкування іноземною мовою», «Інтелектуальна власність», «Цивільний захист», «Охорона праці в галузі». Вивчення іноземної мови необхідне магістрам для опрацювання наукових та практичних світових досягнень у галузі освітніх вимірювань.

Структурно-логічна схема підготовки фахівців з вищою освітою за спеціальністю
8.18010022 «Освітні вимірювання»

Державна атестація										
Виконання магістерської роботи										
3. Цивільний захист	↑	4. Охорона праці в галузі	18.3. Діагностика інтелекту	12.2. Основні математичні моделі та методи сучасної теорії тестування в освіті	10.3. Апробація тестових завдань та тесту	11.2. Обробка результатів тестування	13.1. Науково-теоретичні основи моніторингових досліджень якості освіти	17. Методика навчання освітніх вимірювань	19.2. Моделювання при дослідженнях у сфері освітніх вимірювань	22. Тестування у галузі природничих наук / Спецкурс з методів IRT
			18.2. Психологія інтелектуальної діяльності						19.1. Освітні вимірювання як сучасний напрям науки	
			18.1. Когнітивна психологія пізнавальних процесів	12.1. Вступ до Item Response Theory					19.1. Освітні вимірювання як сучасний напрям науки	
Асистентська практика										
2-й семестр										
1. Ділове спілкування іноземною мовою	↑	2. Інтелектуальна власність	5. Психологія вищої школи	11.1. Комп'ютерні технології у тестуванні			14.2. Організація наукових досліджень	15. Педагогіка вищої школи	20. Комп'ютерні статистичні пакети / Аналіз даних в освіті / Методи досліджень в освіті	21. Організація дистанційної освіти у навчальному закладі / Вибіркові дослідження у психології, соціології та педагогіці / Інформаційно-телекомунікаційні технології в освіті та науці
				6.4. Основи моніторингу якості освіти	10.2. Конструювання педагогічних тестів	8.4. Валідність тесту				
				6.3. Аналіз тестових завдань, розрахунок та інтерпретація балів		8.3. Надійність тесту				
				6.2. Надійність та валідність тестів	10.1. Основні поняття теорії тестів	8.2. Статистична обробка емпіричних даних в рамках класичної теорії тестів				
				6.1. Основи педагогічних вимірювань		8.1. Класична модель істинної оцінки				
				9.2. Математико-статистичні методи аналізу закономірностей та тенденцій розвитку соціально-педагогічних явищ і процесів	7.2. Методика проведення педагогічного оцінювання	16.4. Перевірка непараметричних статистичних гіпотез				
				9.1. Повторення і узагальнення відомостей з теорії ймовірностей		7.1. Науково-дидактичні основи педагогічних оцінювань				
						16.2. Основні закони розподілу статистичних даних				
						16.1. Повторення і узагальнення відомостей з теорії статистики				
						14.1. Методологія наукових досліджень				
1-й семестр										

Підготовку магістрантів до науково-дослідницької діяльності в галузі, а також до обґрунтування освітніх вимірювань, виховання магістрантів самостійними науковими дослідниками та ініціативними компетентними практиками, підготовленими до роботи у різних сферах соціальної діяльності забезпечують дисципліни «Науковий семінар» та «Методологія та організація наукових досліджень».

Термін навчання в магістратурі – один рік. Тому існує необхідність подальшого вдосконалення системи підготовки за рахунок, наприклад, об'єднання близьких за змістом курсів або перехід від семестрового графіку навчального процесу до четвертного, що відображено на структурно-логічній схемі.

8.5. Вимоги до професійної відповідальності при оцінюванні

8.5.1. Стандарти освітнього та психологічного тестування

Становлення нової системи оцінювання, якою, наприклад, для нашої країни є система зовнішнього незалежного тестування, завжди супроводжується певними проблемами, як то надмірне «захоплення» окремими методами оцінювання, некоректне їх використання, недопустимі способи інтерпретації результатів та використання їх не за призначенням, необґрунтоване піднесення або занепад у використанні певних методів та підходів до оцінювання, негативний вплив, який починає створювати на систему навчання нова система оцінювання тощо. У практиці зарубіжних країн для запобігання вище переліченим проблемам та забезпечення якості педагогічного оцінювання і вимірювання розроблено низку документів для регулювання в першу чергу етичних, а також технічних і юридичних питань. Такими є «Технічні рекомендації для психологічних тестів та діагностичних методик», розроблені у 1954 р. Американською психологічною асоціацією, «Технічні рекомендації для тестів досягнень», розроблені Американською асоціацією педагогічних досліджень спільно з Національною радою вимірювань в освіті. На їх основі створено найдавніші з існуючих донині «Стандарти педагогічного та психологічного тестування».

У широкому розумінні стандарти – це документ, який містить вимоги, рекомендації, кодекси та інші засоби регулювання та сприяння належній практиці в оцінюванні. Стандарти є основою політики оцінювання і

тестування, яка охоплює питання належного використання тесту; охорону та безпеку тестових матеріалів і результатів; вимоги до персоналу, який може та має право адмініструвати тести, розраховувати тестові бали та інтерпретувати результати тестування; конфіденційність доступу до результатів тестування та розрахунку тестових балів; кваліфікаційні вимоги для користувачів тестів та їх тренінг; підготовку экзаменованих до тестування та зворотній зв'язок щодо повідомлення результатів; відповідальність перед экзаменованими перед, підчас та після тестування; відповідальність та підзвітність кожного окремого користувача. Згідно рекомендаціям Міжнародної тестової комісії (International Test Committee) політика тестування розробляється для того, щоб гарантувати використання тестів згідно мети, уникати їх застосування не за призначенням, гарантувати використання лише кваліфікованим персоналом на основі повної та доцільної інформації, не використовувати для несправедливої дискримінації.

Політика оцінювання та тестування регулює діяльність учасників процесу. Основними категоріями учасників є розробники тестів (розробляють, видають, адмініструють та обробляють тестові матеріали), користувачі тестів (зацікавлені у результатах тестування для прийняття рішень) та особи, яких оцінюють. У зарубіжній практиці також виділяють замовників (спонсорів), експертів-рецензентів, адміністраторів та інш.

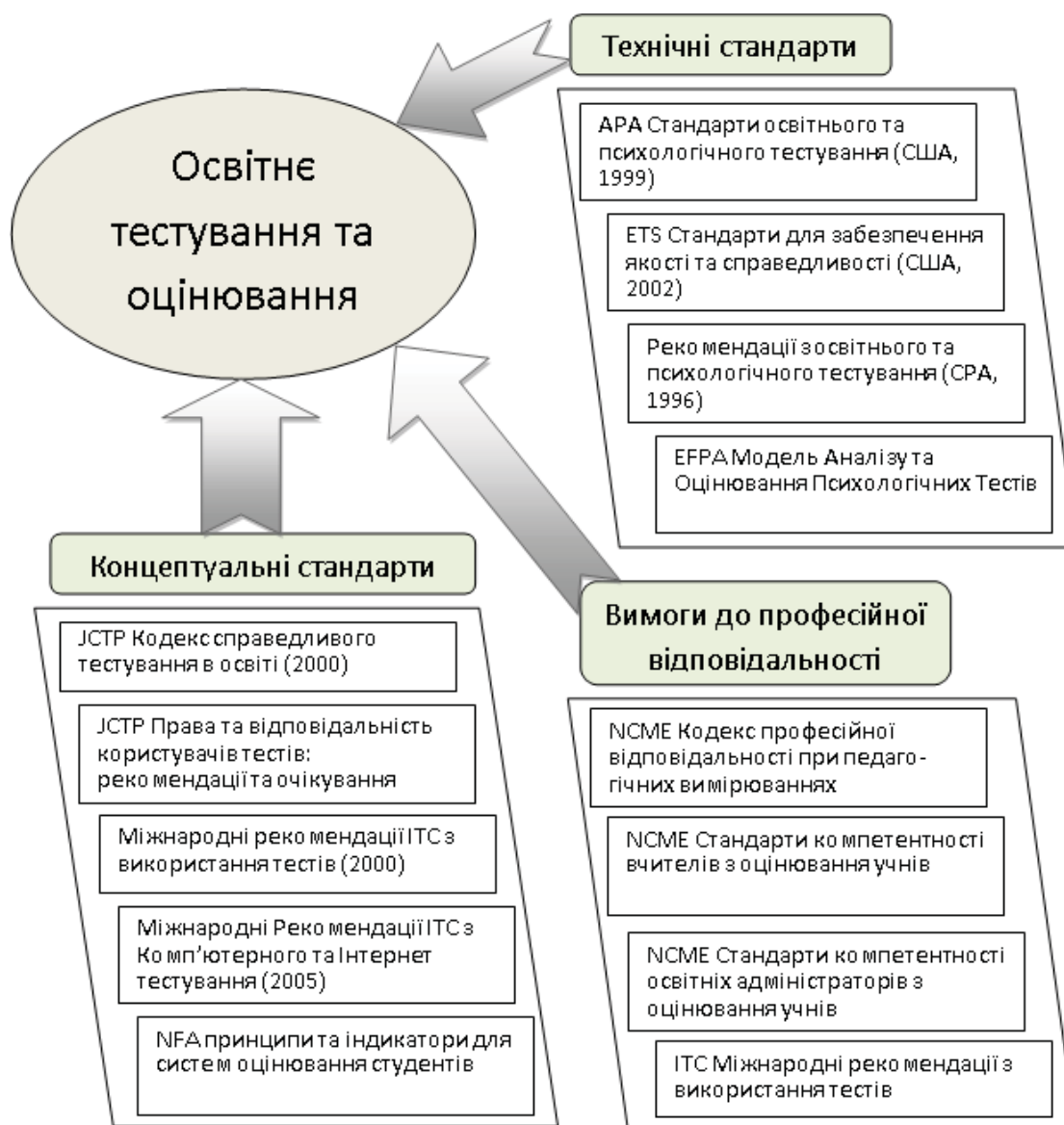
8.5.2. Зарубіжні стандарти

Стандарти якості тестування та оцінювання, які нині функціонують у зарубіжній практиці, умовно можна поділити на три групи: технічні, концептуальні стандарти та вимоги до професійної відповідальності.

Технічні стандарти – спрямовані на визначення характеристик якості продукту (тесту) з технічними деталями та інструкціями; використовуються для регулювання використання тестів залежно від контексту, для формулювання технічних вимог до високоякісної професійної практики, визначення валідності, надійності та інших властивостей тестового інструменту та результатів. Найбільш впливовим та відомим документом цієї групи є APA Стандарти освітнього та психологічного тестування (APA Standards for Educational and Psychological Testing, 1999) розроблені Американською асоціацією освітніх досліджень (AERA) спільно з

Американською психологічною асоціацією (АРА) та Національною радою вимірювання в освіті (NCME).

Зарубіжні стандарти якості в галузі педагогічного оцінювання



Концептуальні стандарти стосуються принципів, на яких базується система оцінювання та діяльність різних учасників процесу; більше уваги в них приділяють етичному використанню оцінювання та взаємовідносинам між різними учасниками в системі оцінювання. Основна мета концептуальних стандартів – коригувати акценти у сфері оцінювання та тестування: поставити навчання у центр процесу оцінювання. Прикладом концептуальних стандартів є Міжнародні рекомендації ІТС з використання

тестів (2000) та з Комп'ютерного та Інтернет тестування (2005), які на відміну від попередніх документів, розроблених у США, є універсальними шаблонами без національної специфіки. Також рекомендації ІТС містять опис не лише того, що мають робити особи, задіяні у оцінювання, але й те, які знання, вміння та навички повинні мати ці особи.

Вимоги до професійної відповідальності різних категорій учасників процесу оцінювання ставлять чіткі вимоги до осіб, задіяних в оцінюванні. Такими є, наприклад, NCME Кодекс професійної відповідальності при педагогічних вимірюваннях, NCME Стандарти компетентності вчителів з оцінювання учнів.

8.5.3. Кодекс справедливого тестування в освіті

У «Кодексі професійної відповідальності при педагогічних вимірюваннях» (далі «кодексі»), визначено типи загальної відповідальності причетних до освітніх оцінювань та специфічні види відповідальності для тих, хто розробляє оцінювання, для тих, хто займається продажем, маркетингом та відбором інструментів та послуг з оцінювання, для тих, хто проводить оцінювання, розраховує тестові бали та інтерпретує результати оцінювання, для тих, хто навчає та проводить дослідження у сфері освітніх оцінювань.

Кодекс виконує освітню функцію: інформування усіх причетних про їхні обов'язки сумлінно ставитися до розробки, використання, визначення якості та маркетингу інструментів оцінювання та функцію стимулювання ґрунтовного обговорення складових професійно відповідальної практики оцінювання на усіх освітніх рівнях.

До загальних видів відповідальності віднесено, наприклад, такі:

1. Захищати здоров'я та безпеку усіх екзаменованих.
2. Діяти згідно з місцевими та федеральними законами щодо здійснення професійної діяльності.
3. Підтримувати та підвищувати рівень професійної компетентності в галузі освітнього оцінювання.
4. Надавати послуги з оцінювання лише у сфері своєї компетентності та дослвіду.
5. Пропагувати розуміння надійних практик оцінювання в освіті.

6. Дотримуватися найвищих стандартів діяльності та пропагувати професійно відповідальність в освітніх закладах та установах, що надають освітні послуги.
7. Виконувати професійні обов'язки чесно, надійно та сумлінно.

До розробників інструментів оцінювання висуваються крім того такі вимоги:

1. Забезпечувати розробку інструментів та послуг з оцінювання за відповідними професійними, технічними та правовими стандартами.
2. Розробляти інструменти та послуги з оцінювання без упередженості щодо характеристик, які не мають відношення до вимірюваного конструкту, таких як стать, вік, соціоекономічний статус, віросповідання, обмежені можливості і таке інше.
3. При розробці враховувати можливості модифікації для груп екзаменованих з обмеженими можливостями чи іншими особливими потребами.
4. Повідомляти зацікавленим сторонам про можливі конфлікти інтересів.
5. Використовувати в інструментах та послугах з оцінювання матеріали із застереженим авторським правом лише згідно із діючим законодавством.
6. Надавати відповідальним особам інформацію про докази надійності, валідності та інші характеристики оцінювання.
7. Забезпечувати захист права екзаменованих на конфіденційність інформації, визнаючи це як частину процесу розробки інструментів оцінювання.
8. Чітко та однозначно попереджати користувачів інструментів оцінювання про найбільш імовірні помилки в інтерпретації та використанні даних, що впливають з процесу розробки інструментів оцінювання.
9. Уникати неправильних чи непідтверджених висловлювань щодо оцінювання, його використання та інтерпретацію результатів у матеріалах для підготовки до тестування, супровідних матеріалах до програми та послуг.
10. Виправляти будь-які суттєві неточності в інструментах оцінювання чи супровідних матеріалах до них в найкоротші можливі терміни.
11. Розробляти форми звітності та супровідні матеріали в доступній для розуміння результатів оцінювання формі.

8.5.4. Стандарти компетентності вчителів з оцінювання учнів

Оцінювання учнів є важливою складовою навчального процесу. Тому вчителів слід готувати до застосування різноманітних методів оцінювання. Розроблені у США в 1990 році «Стандарти компетентності вчителів з оцінювання учнів» покликані бути орієнтиром при підготовці та сертифікації вчителів. Згідно до стандартів до початку навчального процесу діяльність учителя полягає у розумінні культурного середовища, інтересів, навичок та здібностей учнів; розумінні мотивації учнів; відборі та формулюванні очікуваних від учнів результатів. Під час навчального процесу вчитель має здійснювати моніторинг прогресу учнів у досягненні навчальних цілей, визначати досягнення учнів та аналізувати труднощі, які виникають в процесі навчання; коригувати навчальний процес; мотивувати учнів до навчання. Після завершення відповідного сегмента навчального процесу (уроку, семестру і т.п.) – описувати ступінь досягнення учнем короткострокових та довгострокових навчальних цілей; повідомляти учням та батькам результати оцінювання; оцінювати ефективність навчального процесу, навчального плану та використаних матеріалів; документувати та аналізувати результати оцінювання.

Основними компетентностями у стандартах названі такі:

учителі повинні вміти

- 1) вибирати та розробляти методи оцінювання, доцільні для прийняття навчальних рішень;
- 2) застосовувати як зовнішні готові, так і власно розроблені методи оцінювання, розраховувати за ними тестовий бал та інтерпретувати результати;
- 3) використовувати результати оцінювання для прийняття рішень стосовно окремих учнів, для планування навчального процесу, розробки навчального плану та поліпшення функціонування школи;
- 4) розробляти валідні процедури визначення рівня учнів за допомогою оцінювання;
- 5) повідомляти результати оцінювання учням, батькам та іншим непрофесійним аудиторіям, а також іншим працівникам освіти;
- 6) визначати неетичні, незаконні та інші невідповідні методи оцінювання та способи використання інформації щодо оцінювання.

В основі всіх видів діяльності, пов'язаних з оцінюванням учнів, повинні лежати справедливість, етичне поведіння та дотримання прав усіх зацікавлених сторін. Учителі повинні добре усвідомлювати свою етичну та правову відповідальність в оцінюванні та не допускати використання неналежних методів оцінювання іншими.

8.5.5. Відповідальність при тестуванні та оцінюванні в Україні

В Україні питання якості методів, що застосовуються при оцінюванні та вимірюваннях в освітній галузі, наприклад, у зовнішньому незалежному оцінюванні, обговорюються менше ніж адміністративні аспекти тестування. Зауважимо, що на даний час ситуація поступово виправляється.

Результатами розробки, розповсюдження та впровадження відповідних стандартів якості в Україні можуть бути такі наслідки:

- наявність узгоджених і прийнятних критеріїв дозволить об'єктивно визначити якість послуг та продуктів, пов'язаних з оцінюванням;
- за умови більш наполегливого запровадження державою такі стандарти можуть слугувати певним стимулом для професійного розвитку фахівців, що займаються оцінюванням та використовують його результати;
- за умови впровадження механізмів державного контролю щодо дотримання стандартів якості принаймні для найкритичніших програм оцінювання, де ризик несправедливих рішень є найбільш критичним (наприклад, національні програми оцінювання, вступні університетські програми), стандарти можуть слугувати реальною рушійною силою у напрямку забезпечення якості та підвищення прозорості систем оцінювання.

Однак стандарт якості в Україні, як і в західних країнах, має бути рекомендаційним стандартом, щоб не спричинити формалізований вплив на розвиток системи оцінювання. Стандарти якості педагогічного оцінювання в Україні є потрібними як орієнтир для встановлення якості систем оцінювання, але наявність таких стандартів не є достатньою умовою для забезпечення високої якості оцінювання.

Необхідно створити попит на оцінювання високої якості та забезпечити доступ до вітчизняних експертів, які можуть встановити якість існуючих систем оцінювання. Освітнє оцінювання має підтримуватись широкими науковими дослідженнями, розвиваючись у окрему наукову галузь.

Дослідження мають донести до широкого загалу освітян, що застосовувані методи, зокрема, тестування, є лише одним з інструментів оцінювання, що мають свої переваги та обмеження, що мають відповідати меті, і що в певних випадках відсутність оцінювання є кращою ніж неякісне оцінювання.

Для сприяння підвищенню якості оцінювання визначено та реалізуються такі заходи: 1) проведення широких дискусій серед громадськості та педагогів щодо зв'язку між якістю оцінювання та його результатами в засобах масової інформації, у пресі, на зібраннях, формуючи таким чином громадську думку щодо потреби у якісному оцінюванні; 2) забезпечення українського перекладу міжнародних стандартів із забезпечення якості та розповсюдження цих документів серед освітян; 3) видання фахових видань з питань педагогічного оцінювання; 4) розробка державної програми поступового запровадження механізмів державного контролю за процедурами формативного оцінювання в освіті (спочатку для національних іспитів, а пізніше для іспитів на рівні навчальних закладів); 5) сприяння утворенню недержавних організацій, що діють у сфері педагогічного оцінювання і тестування; 6) забезпечення навчальних матеріалів з педагогічного оцінювання; 7) забезпечення проведення систематичних тренінгів та курсів підвищення кваліфікації вчителів; 8) посилення наукових досліджень в сфері педагогічного оцінювання.

8.6. Дистанційне навчання

Дистанційне навчання стало однією з рушійних сил сучасного розвитку освітніх вимірювань. З одного боку саме у дистанційному навчанні застосування об'єктивних автоматизованих засобів контролю якості навчання є абсолютно дещим. Тому розробники систем дистанційного навчання зацікавлені у розробці таких методів, форм і засобів освітніх вимірювань, які б забезпечували якісний, швидкий, по-можливості точний, надійний і валідний контроль результатів навчання. З іншого боку, дистанційне навчання надає розробникам засобів освітніх вимірювань аудиторію та матеріал для статистичних досліджень і моделювання та технічні засоби, включаючи мережеві та телекомунікаційні, а також замовлення на удосконалення існуючих та розробку нових моделей та методів освітніх вимірювань.

Дистанційне навчання (ДН) є формою освіти, поряд з очною та заочною, за якою в освітньому процесі використовуються кращі традиційні та інноваційні засоби і форми навчання, що ґрунтуються на комп'ютерних і телекомунікаційних технологіях. Дистанційне навчання є цілеспрямованим інтерактивним асинхронним процесом взаємодії суб'єктів і об'єктів навчання між собою та із засобами навчання, причому процес навчання індиферентний до їхнього просторового розташування. Освітній процес проходить у специфічній педагогічній системі. Елементами цієї системи є підсистеми: мета навчання; зміст навчання; засоби навчання; організаційні форми навчання ідентифікаційно-контрольна, навчально-матеріальна, фінансово-економічна, нормативно-правова, маркетингова.

8.6.1. Ознаки дистанційного навчання

Основою освітнього процесу в ДН є цілеспрямована й контрольована інтенсивна самостійна робота студента, котрий може навчатися в зручному для себе місці, за індивідуальним розкладом, маючи комплект спеціальних засобів навчання й погоджену можливість контакту з викладачем по телефону, електронною та звичайною поштою, а також очно.

Дистанційне навчання покликане вирішувати специфічні завдання стосовно розвитку творчої складової освіти й проблеми, складні для розв'язання у звичайному навчанні, зокрема:

а) підсилення активної ролі студента у власній освіті: у постановці освітньої мети, виборі домінуючих напрямків, форм і темпів навчання в різноманітних освітніх царинах;

б) різке збільшення обсягу доступних освітніх масивів, культурно-історичних досягнень людства, доступ до світових культурних та наукових скарбів для дітей з будь-якого населеного пункту, де є телезв'язок;

в) отримання можливості спілкування учня з педагогами-професіоналами, з ровесниками-одномудцями, консультування у фахівців високого рівня незалежно від їхньої територіальної розташованості;

г) збільшення евристичної складової навчального процесу за рахунок застосування інтерактивних форм занять, мультимедійних навчальних програм;

г) більш комфортні, порівняно з традиційними, умови для творчого самовираження студента, можливість демонстрації продуктів своєї творчої

діяльності для всіх, широкі експертні можливості оцінювання творчих досягнень;

д) можливість змагання з великою кількістю ровесників, що знаходяться у різних містах та країнах, за допомогою участі у дистанційних проектах, конкурсах, олімпіадах.

Отже, дистанційне навчання – це технологія, що базується на принципах відкритого навчання, широко використовує комп'ютерні навчальні програми різного призначення та сучасні телекомунікації для доставки навчального матеріалу та спілкування. Для цієї технології характерна сильна пізнавальна мотивація, що відрізняє дистанційне навчання від заочного.

8.6.2. Характерні риси дистанційного навчання

Характерними рисами, які відрізняють дистанційне навчання від традиційного, є:

Гнучкість. Студенти, що навчаються в системі дистанційної освіти (СДО), в основному не відвідують регулярних занять, а працюють у вигідний для себе час, у зручному місці й темпі, що дає перевагу тим, хто не може або не хоче змінювати свій звичайний спосіб життя. Студентові формально не треба дотримуватися якихось освітніх цenzів. Кожний може навчатися стільки, скільки йому індивідуально потрібно для засвоєння навчального предмета й отримання необхідних заліків за обраними курсами.

Модульність. В основу програм дистанційного навчання покладено модульний принцип. Кожний окремий курс створює цілісне уявлення про певну предметну сферу. Це дає змогу із сукупності незалежних курсів-модулів формувати навчальну програму, що відповідає індивідуальним або груповим (приміром, персоналу окремої фірми) потребам.

Економічна ефективність. Досвід показує, що витрати на підготовку фахівця становлять приблизно 60 % від витрат на його підготовку за денною формою. Відносно низька собівартість навчання забезпечується за рахунок концентрації та уніфікації змісту, орієнтації технологій ДН на велику кількість студентів, а також більш ефективного використання наявних навчальних площ і технічних засобів, наприклад, у вихідні дні.

Нова роль викладача. На нього покладаються такі функції, як координація пізнавального процесу, коригування курсу, що викладається, консультування при укладанні індивідуального навчального плану,

керівництво навчальними проектами та інше. Він керує навчальними групами взаємопідтримки, допомагає студентам у їхньому професійному самовизначенні. Асинхронна, як правило, взаємодія студентів та викладача припускає обмін повідомленнями шляхом їхнього взаємного посилання на адреси кореспондентів. Це дає змогу аналізувати інформацію, що надходить, і відповідати на неї у будь-який час. Засобами асинхронної взаємодії є електронна голосова пошта або ж електронні комп'ютерні мережі.

Нова роль учня. У системі дистанційного навчання учень є «центром» навчальної парадигми: саме він обирає методи, технології та засоби навчання, він є активним учасником процесу навчання (ставить проблеми, визначає час навчання тощо).

Спеціалізований контроль якості освіти. Як форми контролю в дистанційному навчанні використовуються дистанційні іспити, співбесіди, практичні, курсові і проектні роботи, екстернат, комп'ютерні інтелектуальні тестувальні системи. Варто підкреслити, що розв'язання проблеми контролю якості дистанційного навчання, його відповідності освітнім стандартам має принципове значення для успіху всієї системи. Від успішності її вирішення залежить академічне визнання дистанційних курсів, можливість їх прийняття у традиційних навчальних закладах. Ось чому для здійснення контролю в системі дистанційної освіти має бути створена єдина система державного тестування.

Використання спеціалізованих технологій і засобів навчання. Як уже зазначалося, технологія дистанційного навчання це сукупність методів, форм і засобів взаємодії з людиною у процесі самостійного, але обов'язково контрольованого засвоєння певного масиву знань. Навчальна технологія будується на фундаменті відповідного змісту й має задовольняти вимоги його подання. Зміст, що пропонується для засвоєння, акумулюється в спеціальних курсах і модулях, призначених для дистанційного навчання з урахуванням наявних у країні освітніх стандартів, а також у банках даних і знань, бібліотеках відеосюжетів і т. д.

8.6.3. Принципи дистанційного навчання

Дистанційне навчання ґрунтується на таких принципах:

Принцип гуманістичності навчання. Його суть полягає в спрямованості навчання та освітнього процесу загалом на людину; у створенні максимально

сприятливих умов для оволодіння студентами соціально накопиченим досвідом, структурованим у зміст навчання; в опануванні обраною професією для розвитку та вияву творчої індивідуальності, високих громадянських, моральних, інтелектуальних якостей, які б забезпечували особистості соціальну захищеність, безпечно та комфортне існування.

Принцип пріоритетності педагогічного підходу при проектуванні освітнього процесу в СДО. Сутність цього принципу полягає в тому, що проектування СДО необхідно починати з розробки теоретичних концепцій, створення дидактичних моделей явищ, запланованих для реалізації. Досвід дистанційного навчання дає змогу стверджувати, що коли пріоритет на боці педагогічної сторони, система є більш ефективною.

Принцип педагогічної доцільності застосування нових інформаційних технологій. Він потребує педагогічної оцінки ефективності кожного кроку проектування та створення СДО. Тому на перший план треба ставити не впровадження техніки, а відповідне змістове наповнення навчальних курсів і освітніх послуг.

Принцип вибору змісту освіти. Зміст дистанційної освіти має відповідати нормативним вимогам державного освітнього стандарту й вимогам ринку.

Принцип забезпечення безпеки інформації, що циркулює в СДО. Згідно з цим принципом слід передбачати організаційні й технічні засоби безпечного та конфіденційного зберігання, передачі й використання потрібних відомостей.

Принцип стартового рівня освіти. Ефективне навчання в СДО потребує відповідного набору знань, умінь, навичок. Наприклад, для продуктивної навчальної діяльності кандидат на навчання повинен бути обізнаним з науковими основами самостійної роботи, володіти певними навичками роботи з комп'ютером та ін.

Принцип відповідності технологій навчання. Технології навчання мають бути адекватними моделям дистанційного навчання. Так, у традиційних дисциплінарних моделях як організаційні форми навчання використовуються лекції, семінарські й практичні заняття, імітаційні або ділові ігри, лабораторні заняття, самостійна робота, виробнича практика, курсові й дипломні роботи, контроль засвоєння знань і умінь. У процесі становлення СДО можуть з'явитися нові моделі, що у разі необхідності будуть включені в неї.

Прикладом таких моделей можуть слугувати об'єктивно-інформаційні або проектно-інформаційні моделі. Як організаційні форми навчання в цих моделях використовуватимуться комп'ютерні конференції, телеконференції, телеконсультації, проектні роботи і т. ін.

Принцип мобільності навчання. Полягає у створенні інформаційних мереж, баз і банків знань та даних для дистанційного навчання, що дають змогу студенту коригувати або доповнювати свою освітню програму в потрібному напрямку за відсутності відповідних послуг у вищому навчальному закладі, де він навчається. При цьому вимагається збереження інформаційної інваріантності освіти, що забезпечує можливість переходу з одного ВНЗ до іншого за спорідненими або іншими напрямками.

Принцип неантагоністичності дистанційного навчання наявним формам освіти. Проектована СДО зможе дати необхідний соціальний та економічний ефект за умови, якщо створювані та впроваджувані інформаційні технології стануть не чужорідним елементом у традиційній системі вищої освіти, а природно інтегруватимуться в нього.

8.6.4. Педагогічні вимоги до дистанційного навчання

Якість дистанційного навчання залежить від дотримання певних педагогічних вимог, а саме:

1. *Заохочення контактів між студентами й викладачами.* Ці контакти – найважливіший чинник пізнавальної мотивації. Спілкування з різноманітними викладачами збільшує інтелектуальні здібності студентів і змушує їх замислюватися над своїми цінностями й планами.

Сучасні комунікації полегшують спілкування професорсько-викладацького складу, дають змогу об'єднувати зусилля при розв'язанні різноманітних проблем. Особливо це влаштовує сором'язливих студентів – їм легше ставити запитання в письмовій формі й конфіденційно. Велику роль при цьому відіграє асинхронний зв'язок, що дає змогу організувати ефективну бесіду навіть при невисокому рівні володіння мовою, якою здійснюється навчання.

2. *Розвиток співробітництва студентів.* Навчання в колективі значно ефективніше, ніж самотійне. Не секрет, що робота в колективі збільшує захопленість навчанням, а розподіл ідей поліпшує процес мислення. Спільне навчання, розв'язання проблем, їх обговорення посилюється існуючими

системами комунікацій. Електронна пошта уможлиблює спілкування студентів у будь-який час, навіть якщо вони розділені простором.

3. *Використання активних засобів навчання.* Студенти повинні обговорювати матеріал, що вивчається, обмірковувати, пов'язувати його зі своїм життєвим досвідом.

Діапазон технологій, що використовуються, дуже великий. Усі вони включають три компоненти: програмні засоби, навчальні матеріали й бесіди в реальному масштабі часу. Нині сучасні засоби комунікації можуть їх підтримувати.

4. *Швидкий зворотний зв'язок.* При вивченні матеріалу студенти потребують допомоги. Це оцінка їхніх знань, консультації під час виконання самостійних завдань. Вони повинні знати, про що їм треба ще дізнатися і як оцінити свої знання. Нині таких можливостей багато. Це електронна пошта, відеоконференцзв'язок. Комп'ютери накопичують інформацію про роботу студента, а це дає можливість аналізувати його досягнення.

5. *Ефективне використання часу.* Час, помножений на енергію, є основою навчання. Ефективне використання часу важливе і для студентів, і для викладачів. Успішний розподіл часу забезпечує ефективне учіння та ефективне викладання.

6. *Висока мотивація.* Мотивація потрібна кожному: і тому, хто недостатньо підготовлений або не хоче виявляти себе, і тому, хто здібний та активний. Нові технології істотно підвищують мотивацію навчання. При використанні Інтернету, не лише опановується інформація, а й активізуються пізнавальні вміння студента аналізувати, узагальнювати й оцінювати. Потужним мотивом є подання закінченої студентської роботи в Інтернет.

7. *Урахування здібностей студентів і шляхів навчання.* До вивчення веде безліч шляхів. Урахування здібностей студентів збагачує, урізноманітнює й індивідуалізує їхню навчальну роботу. Студенти мають показати свої здібності в оволодінні методами навчання, а технологічні ресурси якраз і дають змогу забезпечити різноманітні засоби навчання, використовуючи допомогу викладачів, лабораторні роботи, задачі тощо.

8.6.5. Рівні дистанційного навчання

Система дистанційного навчання включає програми й курси різноманітних рівнів:

1. *Початкова освіта.* Програми й курси, розраховані на учнів 5-11 років: уроки читання, письма, арифметики, музики.

2. *Середня освіта.* Курси відповідають типу й рівню середньої освіти. На цьому рівні активно застосовуються освітні ігри, широко представлені в комп'ютерних мережах.

3. *Середня професійна освіта.* Курси й програми для тих, хто закінчив загальноосвітню школу. Вони дають змогу отримати професію, але нижчого рівня кваліфікації, ніж курси для отримання ступеня.

4. *Підготовчі курси.* Курси, що готують до навчання для підвищення кваліфікації. Тут використовуються також різноманітні олімпіади, що вносять змагальний елемент у навчання.

5. *Вища освіта.* Програми для отримання ступенів бакалавра, спеціаліста й магістра. Велике значення має те, що особи, які мешкають в інших країнах, можуть одержувати освіту мовою, не державною в країні їхнього проживання.

6. *Післядипломна освіта.* Науково-освітні програми для отримання ступенів кандидата й доктора наук.

7. *Додаткова освіта.* Скорочені програми для осіб, які вже мають вищу освіту, але в іншій сфері знань. Курси для отримання додаткових знань у галузі, пов'язаній з професійною діяльністю тих, хто навчається. Наприклад, курси підвищення кваліфікації, курси іноземної мови. Важливо, що фахівці, які працюють на міжнародному рівні, можуть отримувати освіту в тій країні, де вони працюють. Особливу вагу це має для фахівців з міжнародного бізнесу.

8. *Професійні курси.* Спрямовані на отримання практичних навиків (курси машинопису, бухгалтерські курси). Ці курси є складовою навчання, при якому проводяться й практичні заняття, під час яких удосконалюються отримані навички. Сюди ж слід віднести й дистанційне тестування. Так, тест на знання російської мови як іноземної може складатися за допомогою комп'ютерних телекомунікацій з будь-якої країни світу.

9. *Курси соціального спрямування.* Це, наприклад, курси навчання населення правил поведінки у випадку стихійних лих, правил вуличного руху тощо. До цієї групи входять і курси ліквідації неграмотності, якщо вони спрямовані на певні групи населення, наприклад, етнічні меншості.

8.6.6. Забезпечення дистанційного навчання

Система дистанційного навчання конкретного ВНЗ включає такі види забезпечення:

1. *Методичне*: методики, рекомендації з технології дистанційного навчання з урахуванням дидактичних і психологічних аспектів.
2. *Програмне*: мережеві системні програми, комп'ютерні навчальні програми, інструментальні середовища для створення навчальних програм. Для створення комп'ютерних навчальних програм можна використати програмні середовища, розміщені на різних серверах (WebCT).
3. *Технічне*: персональні ЕОМ для організації сервера, персональні й мережеві комп'ютери.
4. *Інформаційне*: конспекти лекцій, підручники та інші методичні матеріали на паперових і магнітних носіях, довідники, різноманітні бази даних методичного характеру та ін.
5. *Організаційне*: нормативні документи держави і організацій, визначальні структури організацій, що здійснюють дистанційне навчання.

Засобами дистанційного навчання є друковані видання, електронні видання на гнучких магнітних, лазерних або жорстких дисках; комп'ютерні навчальні системи у звичайному та мультимедійному варіантах; аудіальні навчально-інформаційні матеріали; відеонавчальні матеріали; лабораторні дистанційні практикуми; тренажери; бази даних і знань з дистанційним доступом; електронні бібліотеки з дистанційним доступом; дидактичні матеріали на основі експертних навчальних систем; дидактичні матеріали на основі геоінформаційних систем; комп'ютерні мережі; телевізійні передачі тощо.

8.6.7. Вимоги до викладачів у дистанційному навчанні

Ефективне використання засобів дистанційного навчання потребує від його організаторів – вчителів, або як їх називають у системі ДН, тьюторів, відповідних методичних та технологічних компетенцій. Наведемо деякі з них, важливі також для фахівця з освітніх вимірювань:

Знати:

- основні принципи функціонування телекомунікаційних систем;

- особливості проведення теле- та відео конференцій, форумів;
- основи телекомунікаційного етикету;
- особливості самостійної діяльності учнів у мережі в процесі дистанційного навчання;
- активні методи навчання.

Уміти:

- працювати з різноманітними інформаційними ресурсами;
- користуватися комплексом послуг, які надаються інформаційно-освітнім середовищем;
- інтегрувати очні та дистанційні форми навчання;
- організовувати та проводити телекомунікаційні проекти.

8.7. Засоби навчання освітніх вимірювань

Засоби навчання – це різноманітне навчальне обладнання, що використовується у системі пізнавальної діяльності (книги, письмове приладдя, лабораторне обладнання, технічні засоби та ін.).

Засобами навчання освітніх вимірювань є підручники та навчальні посібники з педагогіки, психології, математичної статистики, періодика та електронні видання українською та іноземними мовами з проблем освітніх вимірювань, мережеві та стаціонарні засоби тестування, комп'ютерні статистичні пакети, віртуальні навчальні середовища тощо. Комп'ютерним засобам тестування та навчання присвячено Розділ VI даного посібника.

8.7.1. Інформаційні джерела з освітніх вимірювань

Наведемо короткий огляд україномовних та російськомовних підручників, посібників та монографій, які присвячені теорії та практиці освітніх вимірювань.

В Україні широко відомі наукові та науково-методичні праці І.Є.Булах зі співавторами, деякі з яких наведені нижче. Принципи комп'ютерної діагностики навчальної успішності, а також теорія та методика комп'ютерного тестування наведені у посібниках [5] та [6]. Основні відомості про систему моніторингу та оцінювання якості освіти представлені у науково-методичному електронному виданні під редакцією Булах І.Є. [23]. Проблеми

організації педагогічного оцінювання, навчально-методичні та інформаційно довідкові матеріали для педагогічних працівників зібрані у виданні [21].

Наукове видання [4] присвячене педагогічному оцінюванню і тестуванню, зокрема, наведені правила, стандарти, відповідальність тощо.

У посібнику І.Є.Булах та М.Р.Мруги [7] обговорюється вітчизняний і закордонний досвід впровадження у практику педагогічної діяльності сучасних методів та технологій педагогічного оцінювання. У першій частині розглянуто основні принципи та правила побудови тесту і тестових завдань. У другій — етапи розробки валідного тесту, а також приклади стандартизованих тестів як закордонних, так і вітчизняних. У посібнику наведено також приклади предметних тестів, розроблених групами викладачів під час проведення тренінгів у міжнародній літній школі у 2005 р. Навчання проводилося у межах міжнародного проекту «Справедливе оцінювання — мережа дистанційного навчання для вчителів — TRAST». Цей проект передбачає поширення серед освітян (вчителів, викладачів та керівників навчальних закладів) нових знань та підходів у педагогічному оцінюванні.

Назвемо також серію україномовних навчальних посібників Вілмута Дж. за редакцією Булах І.Є. та Мруги М.Р. [8 - 10] та методичний посібник Байназарової О.О. та Ракчєєвої В.В. з моніторингу та оцінювання якості освіти [3].

Проблемам педагогічної діагностики та інформаційно-комунікаційним технологіям присвячена монографія О.Г.Колгатіна [16], де обґрунтовано теоретико-методичні засади проектування комп'ютерно-орієнтованої системи педагогічної діагностики майбутніх вчителів природничо-математичних спеціальностей.

У фундаментальному підручнику Л.Крокер та Дж.Алгини [17] розкрито сутність теорії тестів, показані наступність і відмінності її класичної та сучасної версій. Викладено введення в теорію вимірювань. Представлені статистичні концепції в теорії тестування. Дано введення в шкалювання, показаний процес конструювання тестів. Розкрито поняття надійності в контексті класичної концепції істинної оцінки та розробки процедур оцінювання справедливості тестів. Висвітлено проблема валідності тестів і викладені шляхи її вирішення. Охарактеризовано тестові оцінки та методи їх інтерпретації. Підручник складається з семи розділів. У першому розділі

представлено вступ у теорію вимірювань, наведено статистичні концепції в теорії тестів, основи шкалування, описано процес конструювання тестів та тестові оцінки. Другий розділ присвячено одному з наріжних понять вимірювань в освіті – надійності тесту. Наведено процедури для оцінювання надійності, основи теорії генералізації, співвідношення для коефіцієнту надійності для критеріально-орієнтованих тестів. У третьому розділі висвітлено інше важливе поняття – валідність. Зокрема представлено статистичні процедури для прогнозування і класифікації, причини та оцінки системної помилки відбору, застосування методів факторного аналізу до визначення рівня валідності. Аналізу завдань у розробці тестів присвячено четвертий розділ підручника. Аналіз проводиться з використанням теоретичних відомостей, наведено приклади. В останньому розділі наведено тестові оцінки та їх інтерпретація, зокрема, описано поправку на вгадування і інші методи оцінювання, процес встановлення стандартів, норми і стандартні оцінки, вирівнювання оцінок з різних тестів. Підручник містить завдання для самостійного розв'язання теоретичного та прикладного характеру.

Навчальний посібник І.А.Морева [19] містить навчальні і методичні матеріали необхідні для тих, хто вивчає і застосовує інформаційні технології в системі освіти. Посібник без технічних подробиць дає відповіді на питання, які стосуються освітніх інформаційних технологій. Матеріал посібника і додатків містить багато практичних порад для педагогів та керівників освітніх закладів. Посібник містить велику кількість інтернет посилань на російські та зарубіжні освітні сайти. Розділи посібника містять обговорення дійсності тих чи інших методів освіти і контролю ЗУН, застосовності їх в умовах наявності комп'ютерної підтримки. В додатках приведений матеріал для практичного застосування викладених теоретичних розробках, положень, свідчень.

У книзі В.І.Звонникова [12] представлені основні положення теорії педагогічних вимірювань. Викладена історія розвитку вимірювань в науках про людину. Розглянутий понятійний апарат, проблеми розмірності простору вимірювань і компоненти процесу вимірювань. Проаналізовані різноманітні підходи до трактування об'єктивності педагогічних вимірювань які пов'язані з педагогічними аспектами тестування, оцінкою надійності і валідності педагогічних тестів.

В.І.Звонниковим розглянуті становлення, теоретико-методологічні основи та методичні аспекти педагогічних вимірювань в [11]. Проаналізовані

процес управління якістю освіти, його компоненти та наукове забезпечення. Показана роль в ньому педагогічних вимірювань. Охарактеризовані сучасні вимірювачі якості освіти, розкриті вимоги до них в контексті проблем управління якістю освіти.

В Інтернет-виданні В.С.Аванесова [1] подано початки теорії педагогічних вимірювань, наведено означення педагогічного тесту, класифіковано тести на традиційні та нетрадиційні. Автором висвітлені питання про зміст тесту, принципи розробки, степінь складності, вимоги до тестів, докладно описані завдання різних типів.

Підручник Ю.М.Неймана та В.А.Хлебникова [20] є вступом в ТППТ (IRT) і передбачає, що читач знайомий з основними поняттями теорії ймовірностей. Розрахована на широкий круг викладачів та інших працівників освіти, зацікавлених у використанні об'єктивних засобів оцінювання результатів цієї діяльності. В даній роботі викладені тільки найпростіші ідеї і методи ТППТ, для розуміння яких не потрібно глибоких знань математики. Книга складається з 6 розділів (а саме: 1) основні поняття і термінології; 2) оцінювання латентних параметрів; 3) оцінювання точності параметризації; 4) статистична перевірка гіпотез тестування; 5) елементи аналізу регресії та кореляції; 6) шкалювання результатів тестування) основного тесту і додатку. Додаток містить 5 статистичних таблиць, необхідних для найбільш важливих розрахунків.

Монографія В.С.Кима [15] присвячена теоретичним і практичним проблемам тестування навчальних досягнень. Наведено короткий огляд розвитку тестування в Росії і за кордоном. Розглянуто поняття надійності та валідності тесту, структури та форми тестового завдання. Обговорюються питання обробки результатів тестування, як на основі класичної теорії тестів, так і з застосуванням IRT до аналізу якості тестових завдань. Монографія складається з п'яти глав. В першій главі наведено основні відомості про педагогічне тестування, у другій главі представлені форми, структура та принципи формування тестових завдань. Основним методам статистичної обробки результатів тестування у рамках класичної теорії тестів присвячена третя глава монографії. У четвертій главі висвітлені завдання та основні функції тестування навчальних досягнень. Методи сучасної теорії тестування IRT у тестуванні навчальних досягнень та їх практичне застосування

представлено у п'ятій главі, наведено також рекомендації щодо аналізу результатів тестування у середовищі RUMM.

Монографія А.А.Маслака [18] присвячена актуальному науковому напрямку – вимірювання латентних параметрів. Книга складається з чотирьох частин. В першій частині розглядаються теоретичні аспекти вимірювання латентних параметрів. Розглянуті класична теорія тестування і теорія вимірювання латентних параметрів, наведено їх порівняння. В другій частині монографії досліджується точність вимірювання латентних параметрів в залежності від числа індикаторних змінних, числа їх градацій та степені кореляції. В третій частині наведені приклади вимірювання латентних змінних в галузі навчання: рівня знань з педагогіки, якості випускної кваліфікаційної роботи, толерантності. На базі методології вимірювання латентних параметрів аналізується якість текстових завдань з вибором правильної відповіді. Четверта частина присвячена застосуванню системи вимірювання латентних параметрів в соціальних системах. Тут розглядаються і вимірюються латентні змінні в різних соціально-економічних системах: рівень розвитку сфери навчання, якість вищої освіти в країнах світу, рівень соціального положення населення, рівень розвитку сфери культури та інші.

Монографія С.А.Позднякова [22] присвячені дослідженню точності вимірювання латентних параметрів в освіті. Проаналізовані існуючі методи оцінювання інтегральних показників (латентних параметрів) в навчальних системах, такі як – класична теорія тестування, метод зважування, вимірювання на основі моделей Бірнбаума. Представлено метод генерування наборів даних для дослідження точності вимірювання латентних змінних на основі моделі Раша для випадку дихотомічних індикаторних змінних. Досліджено точність вимірювання латентних змінних в залежності від числа дихотомічних індикаторних змінних і розроблені відповідні алгоритми. Проведено розробку структурно-функціональної системи управління якістю тестових завдань у відповідності з моделлю Раша і розроблені практичні рекомендації з управління якістю тестових завдань.

Розв'язанню проблеми вимірювання латентних параметрів з метою збільшення якості освіти в Росії присвячено працю Т.С.Анісімової [2].

Процес розробки педагогічних тестів на основі сучасних математичних моделей викладено в посібнику М.Б.Чельшкової [25].

У підручнику М.Б.Чельшкової [26] висвітлюються теоретичні основи, методи конструювання і практичні питання створення і використання гомогенних педагогічних тестів. Розкриваються роль і функції педагогічного контролю в навчальному процесі. Даються основи теорії педагогічних вимірювань. Аналізуються зміст тестів і форм їх подання. Чільне місце відводиться статистичному обґрунтуванню якості тесту. Наводяться інструктивно-методичні матеріали і математичні формули, які використовуються в ході створення тестів і шкалування результатів.

Підручник М.Б.Чельшкової [24] присвячений теорії, методології та технології адаптивного тестування в освіті.

У навчальному посібнику В.І.Звонникова та М.Б.Чельшкової [13] розглядаються загальні положення теорії педагогічних вимірювань, даються основні визначення і поняття, викладаються підходи до відбору змісту тесту. Окремий розділ присвячений формам тестових завдань, які супроводжуються прикладами і рекомендаціями з розробки завдань. Значна увага приділяється математико-статистичним методам обґрунтування якості тестів. В посібнику міститься коротка інформація про деякі інноваційні засоби аутентичного оцінювання, до яких, в першу чергу, відносять портфоліо. Наводяться різні методи шкалювання даних тестування. Окремий розділ присвячено проблема використання тестових технологій в освіті. Сфери їх застосування пов'язуються з експериментом по введенню ЄДЕ і з моніторингом якості шкільної освіти. Всього в посібнику 15 глав, кожна з яких завершується і вправою для самостійного виконання. В кінці книги наведений список літератури для більш детального вивчення порушених питань. В посібнику робиться акцент на сучасний стан практики використання тестів в Росії.

У посібнику В.І.Звонникова та М.Б.Чельшкової [14] розглянуті проблеми оцінки якості результатів навчання при атестації студентів в контексті компетентнісного підходу. Викладені ключові положення теорії педагогічних вимірювань. Розкриті основні підходи до розробки вимірювачів. Висвітлені класичний та сучасний підходи до аналізу якості тестових завдань і тестів.

Різноманітна цінна інформація про моделювання тестів, обробку їх результатів, практику проведення моніторингових досліджень, поради з підготовки до ЗНО та статистичні дані за результатами ЗНО тощо наводиться у *Віснику «Тестування і моніторинг в освіті»* (головний редактор професор

Раков С.А.). Вісник «ТІМО» - єдине в Україні спеціалізоване видання з актуальних проблем зовнішнього незалежного оцінювання і моніторингу якості освіти.

Наведемо також деякі інші періодичні видання, напрям яких пов'язаний з освітніми вимірюваннями, тестуванням та інформаційними технологіями, - «Інформаційні технології в освіті», «Інформаційні технології і засоби навчання», «Педагогічні вимірювання» (Російська Федерація).

Література, рекомендована до впровадження у навчальний процес

1. Аванесов В.С. Теория и методика педагогических измерений [Материалы публикации в открытых источниках и Интернет]. – ЦТ и МКО УГТУ-УПИ, 2005. – 98 с.
2. Анисимова Т.С. Измерение латентных переменных в образовании.- М.: Исследовательский центр проблем качества подготовки специалистов, 2004.- 148 с.
3. Байназарова О.О., Ракчеева В.В. Моніторинг та оцінювання якості освіти: Методичний посібник. – Харків: ХНУ імені В.Н. Каразіна, 2009. – 58 с.
4. Болюбаш Я.Я., Булах І.Є., Мруга М.Р., Філончук І.В. Педагогічне оцінювання і тестування: правила, стандарти, відповідальність, наук.вид. – К.: Майстер-клас, 2007. – 272 с.
5. Булах І.Є. Комп'ютерна діагностика навчальної успішності.- К.: ЦМК МОЗ України, 1995.- 221с.
6. Булах І.Є. Теорія комп'ютерного тестування.- К.: ЦМК МОЗ України, 1994.- 63с.
7. Булах І.Є., Мруга М.Р. Створюємо якісний тест, навчально-методичний посібник, 2-е видання. – К.: Майстер-клас, 2009. – 176 с.
8. Вілмут Дж. Використання портфоліо для навчання та оцінювання: Навч.посіб. / За ред. І.Є.Булах, М.Р.Мруга. – К.: Майстер-клас, 2007. – 48 с.
9. Вілмут Дж. Оцінювання для навчання: Навч.посіб. / За ред. І.Є.Булах, М.Р.Мруга. – К.: Майстер-клас, 2007. – 170 с.
10. Вілмут Дж. Оцінювання письмових робіт учнів: Навч.посіб. / За ред. І.Є.Булах, М.Р.Мруга. – К.: Майстер-клас, 2007. – 56 с.
11. Звонников В.И. Измерения и качество образования.- М.: Логос, 2006. – 312 с.
12. Звонников В.И. Измерения и шкалирование в образовании.- М.: Логос, 2006.- 134 с.
13. Звонников В.И. Современные средства оценивания результатов обучения: учеб. Пособие для студ. высш. учеб. заведений / В.И. Звонников, М.Б. Чельшкова.- М.: Издательский центр «Академия», 2007.- 224 с.
14. Звонников В.И., Чельшкова М.Б. Контроль качества обучения при аттестации. – М.: Логос, 2011.– 272с.
15. Ким В.С. Тестирование учебных достижений. Монография.- Уссурийск: Изд-во УГПИ, 2007. -214 с.
16. Колгатін О.Г. Педагогічна діагностика та інформаційно-комунікаційні технології: монографія / О.Г.Колгатін.- Х.: ХНПУ, 2009.- 324 с.
17. Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов.- М.: Логос, 2010.- 668 с.
18. Маслак А.А. Измерение латентных переменных в образовании и других социально-экономических системах: теория и практика.- Славянск-на-Кубани: Изд. центр СГПИ, 2007.- 424 с.

19. Морев И. А. *Образовательные информационные технологии. Часть 2. Педагогические измерения: Учебное пособие.*- Владивосток: Изд-во Дальневост. ун-та, 2004.- 174 с.
20. Нейман Ю.М., Хлебников В.А. *Введение в теорию моделирования и параметризации педагогических тестов.*- М.: Прометей, 2000.- 168 с.
21. *Основи педагогічного оцінювання. Частина II. Практика. навчально-методичні та інформаційно довідкові матеріали для педагогічних працівників / Упорядники: Артемчук Л.М., Булах І.Є., Мруга М.Р. – К.: Майстер-клас, 2005. – 54 с.*
22. Поздняков С.А. *Исследование точности измерения латентных переменных в образовании.*- Славянск-на-Кубани: Издательский центр СГПИ, 2007.- 118 с.
23. *Система моніторингу та оцінювання якості освіти, наукове-методичне видання. Під ред. Булах І.Є. / ПРООН-МФВ, 2002. – 142 с. (електронне видання).*
24. Челышкова М.Б. *Адаптивное тестирование в образовании (теория, методология, технология).*– М.:Исследовательский центр проблем качества подготовки специалистов, 2001.– 165 с.
25. Челышкова М.Б. *Разработка педагогических тестов на основе современных математических моделей: Учебное пособие.*– М.: Исследовательский центр проблем качества подготовки специалистов, 1995.– 32 с.
26. Челышкова М.Б. *Теория и практика конструирования педагогических тестов.*- М: Логос, 2002.- 432 с.

8.7.2. Комп'ютерні статистичні пакети

Конструювання якісних педагогічних тестів, а тим більше ґрунтовний статистичний аналіз результатів тестування вимагає від фахівців в галузі освітніх вимірювань крім фундаментальної математико-статистичної підготовки також кваліфікованого володіння інструментарієм сучасних комп'ютерних статистичних пакетів.

У сучасній практиці використовують як програмні продукти, розроблені власноруч, або вузькофункціональні інструменти, корисні лише в певних галузях для вирішення конкретних фахових задач, так і засоби з широким спектром можливостей. Найбільш популярними серед статистичних пакетів впродовж останніх років залишаються такі комерційні продукти:

- 1) SPSS – найвідоміший статистичний пакет, розроблений у Стенфордському університеті з метою обробки соціологічних даних, тому часто використовується саме дослідниками гуманітарних галузей (психології, соціології тощо).
- 2) Statistica – багатоплановий інтегрований пакет, розроблений фірмою StatSoft Inc. з потужними засобами програмування.
- 3) SAS – пакет, що використовується для статистичного бізнес-аналізу.
- 4) MATLAB – відомий універсальний математичний пакет та мова програмування для числового аналізу даних.

Разом з тим збагачується функціональність, а разом з нею і популярність таких вільнорозповсюджуваних статистичних пакетів, як, наприклад, Weka, StarProbe, Microsirir, YALE, RapidMiner, R, PSPP. Вважаючи на високу вартість комерційних пакетів, опанування засобів з вільною ліцензією є цілком актуальним.

Крім того для попереднього аналізу результатів тестування засобів прикладних програм загального призначення, якими є, наприклад, табличний процесор MS Excel з пакету MS Office або табличний процесор Calc з пакета OpenOffice.org, може цілком вистачити.

Важливо, щоб фахівець з освітніх вимірювань міг виконати необхідний статистичний аналіз отриманих результатів та моделей тими засобами, які є у його розпорядженні, або опанувати їх при необхідності.

Питання для самоконтролю

1. Що розуміють під терміном «вимірювання», зокрема, стосовно вимірювань в освіті?
2. Що розуміють під терміном «оцінювання»?
3. Назвіть основні етапи та передумови розвитку освітніх вимірювань як самостійної галузі знань.
4. Які проблеми освітніх вимірювань?
5. Як вирішують пролеми оцінювання?
6. Як можна визначити об'єкт освітніх вимірювань?
7. За якими критеріями методи дослідження поділяють на емпіричні та теоретичні?
8. Які методи можна віднести до емпіричних? Наведіть приклади.
9. Яке місце методів експертного оцінювання в освітніх дослідженнях?
10. Наведіть приклади математичних методів, які застосовують для аналізу результатів експерименту.
11. Назвіть основні риси педагогічного експерименту.
12. Назвіть основні етапи педагогічного експерименту.
13. Що таке моніторинг?

14. Наведіть приклади математичних моделей, що використовуються в освітніх вимірюваннях.
15. Що таке система? Які структурні компоненти методичної системи навчання освітніх вимірювань?
16. В яких документах визначено основні цілі навчання фахівців у галузі освітніх вимірювань?
17. Яке призначення освітньо-кваліфікаційної характеристики фахівця?
18. Які основні компетенції фахівця з освітніх вимірювань?
19. Які типові задачі повинен вміти розв'язувати фахівець в галузі освітніх вимірювань?
20. Які основні напрямки підготовки фахівця з освітніх вимірювань? Яка питома вага кожного з напрямків?
21. В яких дисциплінах навчального плану реалізовано математико-статистичний напрямок підготовки?
22. Які основні форми та методи навчання дисциплін математико-статистичного напрямку?
23. Які основні форми та методи навчання комп'ютерних технологій навчання і тестування?
24. Які основні форми та методи навчання дисциплін психолого-педагогічного та організаційно-педагогічного циклу?
25. Що визначає стандарт якості тестування?
26. Які групи стандартів функціонують у зарубіжній практиці освітніх вимірювань?
27. Які групи учасників визначено у стандартах тестування та вимірювання?
28. Яке призначення технічних стандартів тестування та вимірювання?
29. Які вимоги до розробників тестів висуваються у кодексі професійної відповідальності при педагогічних вимірюваннях.
30. Яких заходів вживають в Україні для підвищення якості оцінювання?
31. Що таке дистанційне навчання?
32. Які основні принципи дистанційного навчання?
33. Якими рисами дистанційне навчання відрізняється від традиційних форм?
34. Які засоби використовують при підготовці фахівців освітніх вимірювань?

Список джерел, використаних у розділі VIII

1. Ковальчук Ю.О. Стандарт підготовки магістрів за спеціальністю «Освітні вимірювання»: компетенції, виробничі функції та типові задачі діяльності // Матеріали III Міжнародної науково-методичної конференції «Освітні вимірювання-2011», 14–16 квітня 2011 р. – [Електронний ресурс] – <http://ea2011.com.ua/materialy-konferentsiji/sektsiya-3-osvitni-vymiryuvannya-i-monitorynh-yakosti-osvity-u-zahalnoosvitnij-ta-vyschij-shkoli/standart-pidhotovky-mahistriv-za-spetsialnistyu-osvitni-vymiryuvannya-kompetentsiji-vyrobnychi-funksiji-ta-typovi-zadachi-diyalnosti.html>
2. Булах І. Є., Мруга М. Р. Застосування міжнародних документів для забезпечення якості педагогічного оцінювання в Україні. // Інформаційні технології і засоби навчання: електронне наукове фахове видання [Електронний ресурс] / Гол. ред.: В.Ю. Биков; Ін-т інформ. технологій і засобів навчання АПН України, Ун-т менеджменту освіти АПН України. – 2008. – № 4(8). – Режим доступу <http://www.nbu.gov.ua/e-journals/ITZN/em8/emg.html>. – Заголовок з екрана.
3. Педагогічне оцінювання і тестування. Правила, стандарти, відповідальність. Наукове видання/ Я.Я.Болюбаш, І.Є.Булах, М.Р.Мруга, І.В.Філончук. – К.: Майстер-клас, 2007. – 272 с.
4. Code of Professional Responsibilities in Educational Measurement – [Електронний ресурс] – <http://www.unl.edu/buros/bimm/html/article2.html>
5. Standards for Teacher Competence in Educational Assessment of Students/ Buros Institute of Mental Measurement, 15.09.2011. – <http://www.unl.edu/buros/bimm/html/article3.html>
6. Кузьмінський А.І. Педагогіка вищої школи: Навчальний посібник / К.: Знання, 2005. 486 с. <http://www.info-library.com.ua/books-book-105.html>
7. Лаврентьева Г.П., Шишкіна М.П. Методичні рекомендації з організації та проведення науково-педагогічного експерименту. - Київ: ІІТЗН, 2007. – 72 с: [Електронний ресурс] – <http://www.experiment.edu-ua.net/HTMLText.htm>
8. Стандарт вищої освіти / Нормативна та варіативна компоненти освітньо-професійної програми магістра галузі знань 1801 Специфічні категорії спеціальності 8.18010022 Освітні вимірювання / Розробники О.В.Авраменко, Ю.О.Ковальчук, В.П.Сергієнко. – Кіровоград. – 2011 р.
9. Стандарт вищої освіти / Освітньо-кваліфікаційна характеристика (нормативна компонента) магістра галузі знань 1801 Специфічні категорії спеціальності 8.18010022 Освітні вимірювання / Розробники О.В.Авраменко, Ю.О.Ковальчук, В.П.Сергієнко. – Кіровоград. – 2011 р.

ПІДРУЧНИК

ВИМІРЮВАННЯ В ОСВІТІ

Редактор Авраменко О.В.

Технічний редактор Лисенко В.Ф.

Комп'ютерний набір та верстка Авраменко О.В.

Підписано до друку 27.12.2011.

Формат 60x84/16. Умовн. друк. арк. 20,9. Обл. вид. арк 15,4. Авт. Лист 15,1. Наклад 300. Зам. 121.

Видавець і виготовлювач Лисенко В. Ф.

25028, м. Кіровоград, вул. Пацаєва, 14, к. 1, кв. 101. Тел. (0522) 322-326.

Свідоцтво суб'єкта видавничої справи: серія ДК № 3904 від 22.10.2010.