

Міністерство освіти і науки, молоді та спорту України
Ніжинський державний університет імені Миколи Гоголя
Національний педагогічний університет імені М.П. Драгоманова
Кіровоградський педагогічний університет імені
Володимира Винниченка

Підготовка фахівців з освітніх вимірювань в Україні

Навчально-методичний комплекс

Частина 2



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.

УДК 371

ББК 74.04(4Укр)я73

Роботу виконано в рамках міжнародного проекту «**Освітні вимірювання, адаптовані до стандартів ЄС**» за програмою Європейського Союзу Темпус

Автори: *О.В. Авраменко, Ю.О. Ковальчук, В.П. Сергієнко, Л.О. Кухар, Р.Я. Ріжняк, С.Д. Паращук, В.В. Котяк, Т.В. Лісова, Л.І. Лутченко, І.В. Лупан, М.П.Малежик, Т.В. Сіткар, І.С. Войтович, Н.В. Янчукова, І. А. Сліпухіна, В.С. Фетісов.*

Рецензент: доктор педагогічних наук, професор *С.А. Раков*

Підготовка фахівців з освітніх вимірювань в Україні: [навчально-методичний комплекс] / О.В. Авраменко, Ю.О. Ковальчук, В.П. Сергієнко та ін.; за заг.ред. О.В. Авраменко. – Ніжин: Видавець ПП Лисенко М.М., 2012. — Частина 2. — 398 с.

До другої частини книги увійшли статті, у яких висвітлено важливі аспекти змісту та методики викладання дисциплін з освітніх вимірювань. Представлено програми вступних фахових випробувань, виробничої практики та державної атестації фахівців освітньо-кваліфікаційного рівня магістр спеціальності *Освітні вимірювання*.

ISBN

ББК 74.04(4Укр)я73

© Ніжинський державний університет імені Миколи Гоголя, 2012
© Національний педагогічний університет імені М.П. Драгоманова, 2012
© Кіровоградський педагогічний університет імені Володимира Винниченка, 2012
© О.В. Авраменко, Ю.О. Ковальчук, В.П. Сергієнко та інші, 2012
© Видавець ПП Лисенко М.М., 2012

Зміст

Розділ 1. Математико-статистичні методи освітніх вимірювань ... 5

Моделі та методи статистичної обробки результатів тестування: огляд монографій та підручників (<i>Авраменко О.В.</i>).....	5
Математична складова системи підготовки магістрів спеціальності 8.18010022 Освітні вимірювання (за планом 2011-2012 н.р.) (<i>Авраменко О.В., Лутченко Л.І., Паращук С.Д.</i>)	25
Курс «Математико-статистичні методи в педагогічних вимірюваннях» для майбутніх вчителів математики (<i>Лісова Т.В.</i>)	45
Можливості комп'ютерної обробки результатів тестування (<i>Лісова Т.В.</i>)	53
Аналіз моделей та методів діагностики знань (<i>Малежик М.П., Сіткар Т.В.</i>)	76

Розділ 2. Організація навчального процесу та методика викладання 92

Удосконалення підготовки фахівців з моніторингу якості освіти (<i>Сергієнко В.П., Войтович І.С.</i>).....	92
Викладання теми «Валідність» в загальному курсі освітніх вимірювань: комплексний підхід (<i>Ковальчук Ю.О.</i>)	96
Ключові підходи до викладання теми «Надійність» у загальному курсі освітніх вимірювань (<i>Ковальчук Ю.О.</i>)	115
Викладання поняття надійності тесту, як основної його статистичної характеристики (<i>Паращук С.Д.</i>)	134
Методичне забезпечення практичних занять за змістовим модулем «Надійність тесту» (<i>Янчукова Н.В., Авраменко О.В.</i>)	149
Конструювання тестів: нова навчальна дисципліна для студентів вищих педагогічних навчальних закладів (<i>Кухар Л.О., Сергієнко В.П.</i>)	162
Теорія і практика конструювання завдань в тестовій формі (<i>Кухар Л.О.</i>)	177
Становлення і розвиток систем оцінювання і моніторингу якості освіти в Україні (<i>Сліпухіна І. А.</i>)	182
Логіка побудови програми вступного фахового випробування на спеціальності 8.18010022 Освітні вимірювання у КДПУ імені В. Винниченка (<i>Різняк Р.Я.</i>)	196
Асистентська практика магістратури 8.18010022 Освітні вимірювання (<i>Лутченко Л.І.</i>)	202

Державна атестація в магістратурі 8.18010022 Освітні вимірювання (Авраменко О.В.).....	207
Методика навчання освітніх вимірювань як дисципліна програми підготовки магістрів (Лупан І.В.)	219
Розділ 3. Комп'ютерні технології в тестуванні	225
Про деякі аспекти викладання дисципліни «Комп'ютерні технології тестування»: web-орієнтовані системи тестування (Котяк В.В.)	225
Лабораторні роботи з курсу «Вибіркові обстеження у психології, соціології та педагогіці» (Лупан І.В.)	259
Особливості викладання дисципліни «Комп'ютерні статистичні пакети» (Лупан І.В.)	267
Розробка комп'ютерних моделей педагогічного тестування – складова підготовки магістрів освітніх вимірювань (Паращук С.Д.)	305
Як вибрати комп'ютерну систему тестування знань (Фетісов В.С.)	316
Додатки	325
<i>Додаток I.</i> Програма вступного фахового випробування для вступників за освітньо-кваліфікаційним рівнем «Магістр» спеціальності 8.18010022 Освітні вимірювання у НПУ імені М.П. Драгоманова	325
<i>Додаток II.</i> Програма виробничої практики магістрів спеціальності 8.18010022 Освітні вимірювання у НПУ імені М.П. Драгоманова	332
<i>Додаток III.</i> Програма комплексного державного екзамену для магістрів спеціальності 8.18010022 Освітні вимірювання у НПУ імені М.П. Драгоманова	346
<i>Додаток IV.</i> Програма вступного фахового випробування для вступників за освітньо-кваліфікаційним рівнем «Магістр» спеціальності 8.18010022 Освітні вимірювання у КДПУ імені В. Винниченка	364
<i>Додаток V.</i> Програма асистентської практики магістрів спеціальності 8.18010022 Освітні вимірювання у КДПУ імені В. Винниченка	371
<i>Додаток VI.</i> Питання до державного екзамену з освітніх вимірювань для магістрантів спеціальності 8.18010022 Освітні вимірювання у КДПУ імені В. Винниченка	392

Розділ 1. Математико-статистичні методи освітніх вимірювань

МОДЕЛІ ТА МЕТОДИ СТАТИСТИЧНОЇ ОБРОБКИ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ: ОГЛЯД МОНОГРАФІЙ ТА ПІДРУЧНИКІВ

Авраменко О.В.

Вступ. Стрімкий розвиток протягом останніх років в Україні практики застосування різноманітних технологій тестування у освіті, психології, медицині та інших галузях діяльності людини викликає необхідність ознайомлення галузевих фахівців з основами математичного моделювання та параметризації тестів та сучасними методами визначення латентних параметрів.

Наведемо деякі фундаментальні праці, в яких наведені загальні поняття теорії освітніх вимірювань, оскільки вони ознайомлюють дослідників з проблематикою освітніх вимірювань, термінологією, сучасними методами, практичними рекомендаціями, тощо.

Видання R.Brennan «*Educational Measurement*» [27], яке називають Біблією у галузі освітніх вимірювань, було опубліковано Американською радою з освіти (ACE) в 1951 році. Метою четвертого видання є оновлення новими поколіннями дослідників та практиків та розширення ними тематики, представленої у трьох попередніх виданнях. Редаговане Робертом Бреннаном під спільною егідою Американської ради з освіти (ACE) та Національна рада з вимірювання в галузі освіти четверте видання надає увагу поглибленому аналізу освітніх вимірювань. Тематику можна розділити на три предметні області: теорія і загальні принципи; конструювання, адміністрування та

шкалювання; застосування вимірювань на практиці. Перша частина книги охоплює теми валідності, надійності, IRT, масштабування і нормування, зв'язків та відповідності, справедливості тесту, а також когнітивної психології. Друга частина включає розділи, присвячені розробці тестів, їх адмініструванню, оцінки діяльності, встановлення стандартів продуктивності і технології в тестуванні. Останній розділ включає в себе розділи, присвячені повторному мовному тестуванню, тестування на підзвітності в школах типу К-12, стандартизовані оцінки індивідуальних досягнень у школах типу К-12, адмініструванню тестів у вищій освіті, моніторингу навчальних досягнень, ліцензування та сертифікації тестування, а також правовим та етичним питанням.

Кожні п'ять років оновлюється та виходить у друці фундаментальна праця *Measurement and assessment in Teaching* (*Measurement and evaluation in teaching*), зараз існує десять її видань. Автором перших п'яти видань був N.E.Gronlund [41], останні п'ять видань були написані спільно з співавторами ([42], [56], [57], [58], [60]).

Широкий спектр наукових досліджень у галузі освітніх вимірювань представлено у збірнику R.L.Linn [55], який всебічно висвітлює проблематику та може використовуватись як посібник для вивчення теоретичних основ всіх етапів тестування та нюансів практичного застосування тестів.

У фундаментальному підручнику Л.Крокер та Дж.Алгини [8] 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 велику кількість завдань для самостійного розв'язання теоретичного та прикладного характеру.

Збірник статей з освітніх вимірювань (W.J. van der Linden, R.K.Hambleton, 1997) [75] надає широкі можливості щодо вивчення сучасної теорії тестування в освіті та психології. Ця книга є одночасно вступом до теорії та керівництвом до практичних досліджень. Підручник містить шість великих розділів, кожна з яких присвячена окремій моделі тестування.

Навчальний посібник І.А.Морева [10] 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 [4] 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 [3]. 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 якості освіти.

Підручник «Вимірювання в освіті» за редакцією О.В. Авраменко [2] підготовлений за Проектом 145029-TEMPUS-2008-SE-JPCR «Освітні вимірювання, адаптовані до стандартів ЄС». 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.

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.

Наведемо також деякі інші видання, які широко представляють проблематику освітніх вимірювань [33], [63], [43].

Наведені вище джерела роблять значний внесок у поглиблення теоретичних знань та поширення практичного застосування прийомів та методів освітніх вимірювань, але при цьому вони не є вузько спеціальними математичними або статистичними виданнями, у яких представлені суто математично-

статистичні методи, на яких базується сучасна теорія освітніх вимірювань.

Нижче наведено огляд монографій, підручників та навчально-методичних посібників, у яких висвітлюються різні аспекти сучасної математичної теорії обробки результатів тестування, зокрема, Item Response Theory (IRT), та деяких її узагальнень.

1. Модель Раша та IRT.

Монографія G.Rasch [65] стала першим важливим внеском у теорію і практику вимірювання в освіті, яка породила величезну кількість наукових досліджень в галузі статистики та педагогічних вимірювань.

Деякі приклади освітніх вимірювань та їх аналіз у рамках моделі Раша наведено O.D.Duncan в [32]. Основи теорії Раша, сучасні розробки у рамках цієї теорії та деякі її застосування представлені у праці [36], яка видана під редакцією G.H.Fischer, I.W.Molenaar у 1995 р. Загальні підходи щодо вимірювання з використанням лог-лінійної моделі Раші подані у посібнику H.Kelderman [49]. Прийоми аналізу складності тестових завдань та рівня підготовки учасників тестування моделлю Раша демонструються у праці R.M.Smith [70]. Окремі застосування вимірювань за Рашем наведені у посібнику R.M.Smith [71].

У підручнику J.M.Linacre [54] Вивчається застосування вимірювання рейтингів в результаті експертного аналізу тестування окремих характеристик та деяких інших ситуацій. Вимірювання виконується у рамках моделі Раша. Виконано узагальнення моделі вимірюванням з застосуванням багаторангового впорядкування. Аналізуються приклади рейтингів та оцінок.

Основам IRT та її застосуванням присвячено книгу R.K.Hambleton & H.Swaminathan [45].

R.K.Hambleton, H.Swaminathan & H.J.Rogers [46] видали цікавий підручник з основ сучасної теорії моделювання та параметризації тестів. Книга також містить обговорення альтернативних методів оцінювання параметрів IRT, таких як методи максимальної правдоподібності, граничну оцінку

максимальної правдоподібності, Байесовські оцінки, тощо. Наведено покрокове розв'язання прикладів.

У книзі A.Boomsma, M.A.J. van Duijn, & T.A.B.Snijders [26] розглянуті деякі аспекти параметричної та непараметричної IRT, в рамках кожної з них представлено такі питання: історичні аспекти та огляди, нові моделі, нові методи, а також застосування. Оскільки у книзі представлено уявлення про сучасний стан моделей і методів IRT, то цю роботу можна розглядати як важливий додаток до фундаментального збірника *Handbook of modern Item Response Theory (IRT)* (1997) [75], а також як роботу, що встановлює напрямок майбутнього розвитку IRT

Праця Ch. de Mars [29] представляє собою баланс концептуального пояснення та математичних деталей щодо різних моделей IRT.

У книзі N.J.Dorans, M.Pommerich & P.W.Holland [30] проводиться порівняння результатів вимірювань, які виконані в різних обставинах різними методами з різними учасниками тестування, що є фундаментальною передумовою для вимірювань в освіті.

Підручник Ю.М.Неймана та В.А.Хлебникова [11] 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 термінологія; 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 [7] 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 [9] 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 [12] та його дисертаційне дослідження [13] 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 [1].

Монографія F.V.Baker & S.-H.Kim [23] докладно описує більшість нещодавно розроблених моделей IRT, дає докладні пояснення алгоритмів, які можуть бути використані для оцінки параметрів при різних моделях IRT. Широко переглянуте і розширене, це видання пропонує три нові глави, присвячені обговоренню оцінки параметрів з множинними групами, оцінками параметрів для тестів зі змішаними типами, а також застосування ланцюгів Маркова та методу Монте-Карло. Вона включає в себе обговорення питань, пов'язаних зі статистичною теорією, чисельними методами і алгоритмами комп'ютерних програм для оцінки параметрів, які сприяють створенню чіткого розуміння вимог процедур IRT оцінки.

Теорія освітніх вимірювань моделлю Раша та основи застосування цієї теорії викладені у фундаментальному підручнику E.V.Smith & M.S.Smith [69]. Зокрема, модель Раша та нові моделі імовірнісної теорії освітніх вимірювань представлено у праці G.Karabatsos [48], яка є його складовою.

Процес розробки педагогічних тестів на основі сучасних математичних моделей викладено в посібнику М.Б.Чельшкової [15].

У підручнику М.Б.Чельшкової [16] 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. В додатках наводяться інструктивно-методичні матеріали і математичні формули, які використовуються в ході створення тестів і шкалювання результатів тих, що тестуються.

Підручник М.Б.Чельшкової [14] присвячений теорії, методології та технології адаптивного тестування в освіті.

У навчальному посібнику В.І.Звонникова та М.Б.Чельшкової [5] 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 цілому в посібнику робиться акцент на сучасний стан практики використання тестів в Росії.

У посібнику В.І.Звонникова та М.Б.Чельшкової [6] 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. Даний словник термінів.

Основні відомості та припущення сучасної теорії тестування IRT наведено в підручнику M.J.Allen & W.M.Yen [17].

Навчальний посібник F.V.Baker [22] побудований за принципом від простого до складного. Кожна тема будується на основі попередньої теми. В цій книзі підкреслюються основи, мінімізується кількість математики та не переслідуються технічні деталі, які становлять інтерес тільки для фахівців. В якомусь сенсі показано тільки «те, що потрібно знати», а не всі деталі теорії. Книга дає змогу інтерпретувати результати випробувань, які були проаналізовані в рамках IRT. Проте для того, щоб застосовувати теорію у практичних наборах потрібно вивчати більш просунуті книги по застосуванні теорії.

У книзі W.J. van der Linden [74] відображено історію розвитку теорії конструювання та моделювання тестів в освіті та психології. Представлено стандартні алгоритми для моделювання тестів на основі багатоцільової оптимізації з обмеженнями. Обговорюються лінійне моделювання тестів для великої кількості різноманітних практичних задач, кожна проблема ілюструється характерними прикладами.

У підручнику D.Thissen & H.Wainer [73] викладено класичну теорію тестування, сучасну теорію тестування та указано шляхи інтеграції цих двох підходів та перспективи розвитку освітніх вимірювань альтернативними методами, зокрема, засобами психометрії.

Підручники M.Wilson [77] та [78] призначені для вивчення сучасної теорії тестування та її практичного застосування, матеріал розрахований на студентів, які цікавляться процесом тестування та іншими методами вимірюваннями в освіті.

Основні правила роботи у середовищі RUMM, призначеного для обробки результатів тестування у рамках моделі Раша, наведені у посібнику Getting Started RUMM 2010 [40]. Комп'ютерне середовище WINSTEPS також обробляє результати тестування за моделлю Раша, принципи роботи у WINSTEPS наведено J.M.Linacre [53].

2. Деякі узагальнення IRT.

Книга M.D.Reckase [66] складається з трьох основних частин. Перші три розділи – інформація, яка корисна для розуміння багато розмірної IRT (MIRT). Наведено загальний концептуальний огляд та резюме однопараметричної IRT. Коротко викладено історичні основи MIRT, описано характеристики MIRT моделей. Описано математичні форми моделей, наведено статистичні дані, які використовуються для опису функцій тестових завдань в MIRT-контексті, описано процедури оцінки параметрів моделей. Наведено також інформацію, необхідну для застосування моделей і наведені приклади застосувань. Визначено число вимірювань, необхідних для опису взаємодії між людьми і тестовими завданнями. Продемонстровано, як визначити систему координат. Наведено методи перетворення ING оцінки параметрів з різних калібрувань MIRT до координатної системи. Показано, як всі процедури можуть бути застосовані в контексті комп'ютеризованого адаптивного тестування.

Підручник K.Sijtsma & I.W.Molenaar [68] є вступом до непараметричної IRT. Праця R.Ostini & M.L.Nering [64] присвячена декільком моделям IRT (упорядкованим та невпорядкованим), багатопараметричним даним, включаючи номінальну модель реагування, багатопараметричну модель Раша, такі як моделі за частково кредитною та рейтинговою шкалою, моделі Самеїма. У вступній частині йде мова про вимірювальні теорії, застосування моделі IRT, причини використання багатопараметричних моделей IRT, типи вірогідностей та типи багатопараметричних моделей. Далі висвітлюється поняття математичної моделі та її зв'язок з іншими моделями IRT. Описана багатопараметричні моделі Раша та моделі Самеїма. Визначено, як здійснити вибір моделі (за якими критеріями та відмінності в моделювальних результатах).

В невеликій за обсягом роботі M.H.Moulton [61] представлено теоретичні основи не-одновимірної моделі шкалування, за результатами аналізу практичних прикладів доведено переваги цієї моделі, висвітлено перспективи подальших досліджень.

Теорії генералізації у освітніх оцінюваннях присвячена робота R.J.Shavelson & N.M.Webb [67].

Інформація про ідентифікацію неоднорідних Differential Item Functions (DIF) у політоміальному шкалуванні тестових завдань наведена у праці J.Spray & T.Miller [72].

Наближені методи обчислення та кластерний аналіз в багатовимірній IRT представлено у докторській дисертації J.-P.Kim [50].

Елементи теорії латентного структурного аналізу подані у праці P.F.Lazarsfeld & N.W.Henry [52].

У книзі H.Wainer, E.T.Bradlow & Wang X. [76] розглядаються не окремі тестові завдання, а тестлети - невеликі групи елементів як взаємозамінні одиниці, з яких побудовані тести. Простежено історію еволюції сучасної теорії тест летів (TRT). Перша частина книги пропонує вступ до TRT та її застосування. Далі здійснюється обговорення моделі сформульованій в рамках байєсівського підходу.

У книзі J.-P.Fox [37] йде мова про байєсовські моделі IRT та їх застосування. Математичні деталі процедури оцінки якості тесту обговорюються в книзі, програми, що пов'язані з книгою, надаються через веб-сайт. Зміст, а також алгоритм з його реалізації роблять цю книгу самодостатньою. У вступі до байєсівського моделювання, розглядаються традиційні моделі. Представлено ієрархічний підхід до моделювання. Наведено загальний вступ до байєсівського моделювання (елементи байєсовської статистики, байєсовські перевірки гіпотез). Описано моделі подання відповідей, а розглянуто методи перевірки гіпотез і вибору моделі. Описано багатопараметричні моделі IRT та випадкові ефекти моделей. Представлено багатопараметричний аналіз відповідей та кілька емпіричних прикладів для ілюстрації методів і корисності байєсовського підходу.

З основами застосування теорії ланцюгів Маркова та методу Монте Карло до стохастичного симулювання байєсовських висновків можна ознайомитись у підручнику D.Gamerman & H.F.Lopes [38]. Елементи байєсовського аналізу даних

представлені у праці A.Gelman, J.B.Carlin, H.S.Stern & D.B.Rubin [39].

Процес комп'ютерного моделювання на основі байесовських підходів описано у посібнику I.Ntzoufras до програмного продукту WinBUGS [62].

У посібнику L.Yao [81] описано користувацькі можливості програмного засобу VMIRT, який використовує теорію Марківських ланцюгів та метод Монте Карло для оцінки латентних параметрів тестів, які містять як дихотомічні так і політоміальні завдання.

3. Застосування моделей та методів теорії тестування.

Для багатьох дослідників модель Раша забезпечує практичний аналіз даних в соціальних науках.

У книзі M.J.Kolen & R.L.Brennan [51] наведено методи випробування для стандартизованих тестів в сфері освіти та психології з метою гарантування їх взаємозамінності при практичному використанні. Ця праця може вважатися вступом до теорії порівняння тестів за допомогою найбільш часто використовуваних методологій.

Елементи математичного моделювання тестів у соціальній психології, зокрема спеціальна логістична модель Раша, представлені у роботі E.V.Andersen [18]. Застосування моделі Раша, а також узагальненні принципи та підходи в соціальних та освітніх вимірювань представлені в працях D.Andrich, [21] та [22]. Прикладні аспекти сучасної теорії тестування у психологічних дослідженнях наведені у роботі S.E.Embretson & S.P.Reise [34]. Можливості застосування IRT щодо вимірювань показників учасників тестування у різних галузях, таких як освіта, психологія та ін., наводяться у виданні R.K.Hambelton [44]. Застосування IRT до розв'язання проблем тестування, які виникають у практичній діяльності описані також у праці F.M.Lord [59]. Можливість деяких застосувань політоміальної IRT у психологічних вимірюваннях висвітлено у спеціальному випуску *Applied Psychological Measurement* за редакцією F.Drasgow [31]. Праця R.J. de Ayala, [28] присвячена застосуванню IRT у соціальних дослідженнях.

У двох посібниках В.Д.Врайт & Г.Н.Мастерс [79] та В.Д.Врайт & М.Н.Стейн [80] представлені поширення моделі Раша освітніх вимірювань для оцінки службової діяльності та деяких інших категоріальних даних.

Праця Г.Дж.Енгельхард & М.Вілсон [35] висвітлює теоретичні та практичні аспекти вимірювань, представлено цікаві приклади використання на практиці основних теоретичних закономірностей.

У підручнику Г.С.Інгебо [47] показано переваги вимірювання за моделлю Раша для регіональних програм шкільного тестування. Протиставлені результати, отримані за методом Раша, зі звичайними статистичними даними для оцінки відповідей учнів при визначенні їх навичок. Результати представлено у доступній формі. Підручник розрахований на учнів, батьків, вчителів, адміністраторів і наукових працівників, тому інформація, надана у популярній формі

У монографії Т.Г.Бонд & С.М.Фокс [25] стверджується, що ця модель знаходиться найближче до реалізації об'єктивних фундаментальних вимірювань, ця книга призначена для дослідників науки про людину, зацікавлених у реалізації моделі Раша з їх власними даними. Застосування моделі Раша у вимірюванні показників здоров'я висвітлено у фундаментальній праці Н.Безруцько [24].

Висновки та перспективи дослідження. У статті зроблено спробу стислого огляду основних наукових та навчально-методичних видань з математичного моделювання в галузі освітніх вимірювань. Відмітимо, що основні сучасні дослідження зосереджено у США; окремі цікаві результати отримано в Великій Британії, Данії, Голландії та інших країнах; в тому числі існує невелика кількість російськомовних монографій та підручників.

Наведений у роботі огляд сучасних досягнень у галузі побудови математично-статистичних моделей для визначення рівня досягнень учасників тестування обмежений тільки основоположними виданнями, такими як монографії, підручники та навчально-методичні посібники. Разом з тим, існує значна кількість статей у наукових журналах, доповідей на конференціях, звітів за результатами моніторингових досліджень та ін., тому цей

короткий огляд, очевидно, не може претендувати на вичерпність та потребує розширення та доповнення.

Література

1. Анисимова Т.С. Измерение латентных переменных в образовании.- М.: Исследовательский центр проблем качества подготовки специалистов, 2004. - 148 с.

2. Вимірювання в освіті: Підручник / За редакцією О.В. Авраменко.- Кіровоград: Лисенко В.Ф., 2011. – 360 с.

3. Звонников В.И. Измерения и качество образования.- М.: Логос, 2006. – 312 с

4. Звонников В.И. Измерения и шкалирование в образовании.- М.: Логос, 2006.- 134 с.

5. 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.

6. Звонников В.И., Чельшкова 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. 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.

10. Морев И. А. Образовательные информационные технологии. Часть 2. Педагогические измерения: Учебное пособие.- Владивосток: Изд-во Дальневост. ун-та, 2004.– 174 с.

11. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов.- М.: Прометей, 2000.- 168 с.

12. Поздняков С.А. Исследование точности измерения латентных переменных в образовании.- Славянск-на-Кубани: Издательский центр СГПИ, 2007.- 118 с.

13. Поздняков С.А. Метод и алгоритмы измерения латентных переменных при управлении в образовательных системах (автореферат диссертации на соискание ученой степени кандидата технических наук). – Курск, 2009.

14. Чельшкова М.Б. Адаптивное тестирование в образовании (теория, методология, технология). – М.: Исследовательский центр проблем качества подготовки специалистов, 2001. – 165 с.

15. Чельшкова М.Б. Разработка педагогических тестов на основе современных математических моделей: Учебное пособие. – М.: Исследовательский центр проблем качества подготовки специалистов, 1995. – 32 с.

16. Чельшкова М.Б. Теория и практика конструирования педагогических тестов. – М: Логос, 2002. – 432 с.

17. Allen M.J., Yen W.M. Introduction to Measurement Theory. Belmont, CA: Wadsworth, 1979.

18. Andersen E.B. The logistic model for m answer categories. In W. E. Kempf & B. H. Repp (Eds.), Mathematical models for social psychology. Vienna, Austria: Hans Huber, 1997.

19. Andersen E.B. Discrete statistical models with social science applications. Amsterdam: North-Holland, 1980.

20. Andrich D. Advanced Social and Educational Measurement. Perth: Murdoch University, 2001. – 128 pp.

21. Andrich D. Rasch Models for Development. London: Sage Publications, Inc., 1988. – 94 pp.

22. Baker F.B. The Basics of Item Response Theory. – Portsmouth NH: Heinemann Educational Books, 1985. – 131 pp.

23. Baker F.B., Kim S.-H. Item response theory: parameter estimation techniques (2nd ed.). New York: Marcel Dekker, 2004.

24. Bezruczko N. Rasch Measurement in Health Sciences. Maple Grove, Minnesota: JAM Press, 2005. – 483 pp.

25. Bond T.G., Fox C.M. Applying the Rasch model. Fundamental Measurement in the Human Sciences. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers, 2001. – 255 pp.

26. Boomsma, A., van Duijn, M. A. J., & Snijders, T. A. B. Essays on item response theory. New York, New York: Springer Verlag, 2001.

27. Brennan R. Educational Measurement.- Westport, CT: Praeger, 2006.- 796 pp.
28. de Ayala R.J. The Theory and Practice of Item Response Theory (Methodology In The Social Sciences).- The Guilford Press, 2008.
29. de Mars Ch. Item response theory.- Oxford University Press, 2010.- 131 pp.
30. Dorans N.J., Pommerich M., Holland P.W. Linking and aligning scores and scales. Springer, New York, 2007.
31. Drasgow F. Polytomous IRT [Special issue]. Applied Psychological Measurement, 19(1), 1995.
32. Duncan O.D. Rasch measurement: Further examples and discussion. In C. F. Turner & E. Martin (Eds.), Surveying subjective phenomena, Volume 2. New York, New York: Russell Sage Foundation, 1984.
33. Ebel R.L., Frisbie D.A. Essentials of educational measurement.- Prentice-Hall, 1986.- 360 pp.
34. Embretson S.E., Reise S.P. Item response theory for psychologists. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
35. Engelhard G.Jr., Wilson M. Objective measurement: Theory into practice, Vol. 3. Norwood, New Jersey: Ablex, 1996.
36. Fischer G.H., Molenaar I.W. Rasch models: Foundations, recent developments, and applications. New York: Springer-Verlag, 1995.
37. Fox J.-P. Bayesian Item Response Modeling: Theory and Applications.- New York: Springer, 2010.- 313 pp.
38. Gamerman D., Lopes H.F. Markov chain Monte Carlo: Stochastic simulation for Bayesian inference (2nd edition). Chapman & Hall, Boca Raton, FL, 2006.
39. Gelman A., Carlin J.B., Stern H.S., Rubin, D.B. Bayesian data analysis (2nd ed.). New York: Chapman & Hall, 2003.
40. Getting Started RUMM 2010. Rasch Unidimensional Measurement Models. Perth: RUMM Laboratory Ltd, 2001.- 87 pp.
41. Gronlund N.E. Measurement and evaluation in teaching.- Macmillan Pub. Co., 1965 - 420 pp.

- 42.Gronlund N.E., Linn R.L. Measurement and evaluation in teaching, Sixth edition. New York: Macmillan, 1990.
- 43.Haladyna T.M. Developing and Validating Multiple-choice Test Items. 3 edition – Routledge, 2004.- 320 pp.
- 44.Hambelton R.K. Application of Item Response Theory. Vancouver. BC: Educ Res Inst B.C., 1983.
- 45.Hambleton R.K., Swaminathan H. Item Response Theory: Principles and Applications. Norwell, MA: Kluwer-Nijhoff Publishing, 1985.
- 46.Hambleton R.K., Swaminathan H., Rogers H. J. Fundamentals of Item Response Theory. Newbury Park, CA: Sage, 1991.
- 47.Ingebo G.S. Probability in the Measure of Achievement. Chicago: MESA Press, 1997.- 148 pp.
- 48.Karabatsos G. The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. In E. Smith & R. M. Smith (Eds.), Introduction to Rasch measurement (Ch. 16). Gainesville, FL: JAM Press, 2004.
- 49.Kelderman H. Common item equating using the log-linear Rasch model. Twente, Netherlands: Department of Education, University of Twente, 1986.
- 50.Kim J.-P. Proximity measures and cluster analyses in multidimensional item response theory. Unpublished doctoral dissertation, Michigan State University, East Lansing, MI, 2001.
- 51.Kolen M.J., Brennan R.L. Test equating, scaling, and linking: Methods and practices (2nd edition) Springer, New York, 2004.- 574 pp.
- 52.Lazarsfeld P.F., Henry N.W. Latent Structure Analyses. Houghton Mifflin Co. Boston. N-Y, 1969.- 292 pp.
- 53.Linacre J.M. Auser's guide and manual to WINSTEPS: Rasch-model computer programs. Chicago: Linacre, 2004.
- 54.Linacre J.M. Many-Facet Rasch Measurement. — Chicago: MESA Press, 1994.- 149 pp.
- 55.Linn R.L. (Ed.) Educational Measurement (Third Edition). New York: Macmilan, 1988.
- 56.Linn R.L., Gronlund N.E. Measurement and assessment in teaching, Seventh edition. New York: Macmillan, 1995.

57.Linn R.L., Gronlund, N.E. Measurement and assessment in teaching, 8th Edition. Upper Saddle River, NJ: Prentice Hall, 2000.

58.Linn R.L., Miller M.D. Measurement and assessment in teaching, 9th Edition. Upper Saddle River, NJ: Prentice Hall, 2005.

59.Lord F.M. Application of Item Response Theory to Practical Testing Problems. Hillsdale N-J. Lawrence Erlbaum Ass., 1980. – 266 pp.

60.Miller M.D., Linn R.L., Gronlund N.E. Measurement and assessment in Teaching, 10th Edition. Upper Saddle River, NJ: Pearson Education, Inc., 2009.

61.Moulton M.H. One use of a non-unidimensional scaling (NOUS) model: Transferring information across dimensions and subscales. Educational Data Systems, Morgan Hill, CA, 2004.- 22 pp.

62.Ntzoufras I. Bayesian Modeling Using WinBUGS. Hoboken, NJ:Wiley, 2009.

63.Osterlind S.J. Constructing Test Items: Multiple-Choice, Constructed-Response, Performance and Other Formats (Evaluation in Education and Human Services).- Springer; 2nd edition, 1997.- 352 pp.

64.Ostini R., Nering M.L. Polytomous item response theory models.- Australia: Measured Progress, 2006.-120 pp.

65.Rasch G. Probabilistic models for some intelligence and attainment tests (Expanded edition, with foreword and afterword by Benjamin D. Wright). -Chicago: University of Chicago Press, 1980.- 199 pp.

66.Reckase M.D. Multidimensional Item Response Theory.- New York: Springer, 2009.- 354 pp.

67.Shavelson R.J., Webb N.M. Generalizability Theory: A Primer. Newbury Park, CA: Sage, 1991.

68.Sijtsma K., Molenaar I.W. Introduction to Nonparametric Item Response Theory. London: SAGE Publications, 2002.- 168 pp.

69.Smith E.V., Smith M.S. Introduction to Rasch Measurement. Theory, Models and Applications. Maple Grove, Minnesota: JAM Press, 2004. – 689 pp.

70.Smith R.M. IPARM: Item and person analysis with the Rasch model. Chicago, Illinois: MESA Press, 1991.

71. Smith R.M. Applications of Rasch measurement. Chicago, Illinois: MESA Press, 1992.

72. Spray J., Miller T. Identifying nonuniform DIF in polytomously scored test items. (RR 941). Iowa City, IA: American College Testing Program, 1994.

73. Thissen D., Wainer H. Test scoring. Lawrence Erlbaum Associates, Mahwah NJ, 2001.

74. van der Linden W.J. Linear models of optimal test design. Springer, New York, 2005.

75. 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.

76. Wainer H., Bradlow E.T., Wang X. Testlet response theory and its applications.- New York: Cambridge University Press, 2007.- 267 pp.

77. Wilson M. Measurement: A constructive approach. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2003.

78. Wilson M. Constructing Measures: An Item Response Modeling Approach. Mahwah, New Jersey: Lawrence Erlbaum associates, 2005.- 228 pp.

79. Wright B.D., Masters G.N. Rating Scale Analysis. Chicago: MESA PRESS, 1982.- 206 pp.

80. Wright B.D., Stone M.H. Best Test Design. Chicago, MESA PRESS, 1979.- 222 pp.

81. Yao L. BMIRT: Bayesian multivariate item response theory. CTB/McGraw-Hill, Monterey, CA, 2003

**МАТЕМАТИЧНА СКЛАДОВА СИСТЕМИ ПІДГОТОВКИ
МАГІСТРІВ СПЕЦІАЛЬНОСТІ 8.18010022 ОСВІТНІ
ВИМІРЮВАННЯ**

(за планом 2011-2012 н.р.)

Авраменко О.В., Лутченко Л.І., Паращук С.Д.

Вступ. Підготовка магістрів спеціальності «Освітні вимірювання» вперше в Україні розпочата в 2011 році в Кіровоградському державному педагогічному університеті імені Володимира Винниченка та Національному педагогічному університеті імені М.П.Драгоманова.

Галузевий стандарт спеціальності 8.18010022 «Освітні вимірювання» вимагає, щоб випускник володів «компетентностями, які необхідні для розробки, впровадження і використання класичних та сучасних тестових моделей», «поняттями, концепціями і фактами теорії та практики освітніх вимірювань, що необхідні у професійній діяльності». Його підготовка повинна забезпечувати «вміння організовувати тестові перевірки знань, умінь та навичок учасників тестування, здійснювати математико-статистичне та комп'ютерне моделювання процесу тестування, тощо». Таким чином, магістр з освітніх вимірювань повинен володіти методикою та інструментарієм вимірювань в освіті.

Аналіз сучасних досліджень з теорії тестів свідчить, що одним із важливих напрямків є математичне обґрунтування якості тесту, а відповідно розвиток і побудова моделей тестування. Без математичного апарату неможливо провести аналіз результатів тестування і провести дослідження надійності та валідності тесту – центральні поняття теорії тестів. Це означає, що сучасний рівень підготовки магістрів з освітніх вимірювань передбачає якісне вивчення математичного апарату теорії тестів.

У навчальних програмах у 2011-2012 н.р. дисциплін «Вимірювання в освіті», «Класичні тестові моделі», «Моделі і методи ІРТ», «Математико-статистичні методи в освітніх вимірюваннях» та «Прикладна статистика» міститься інформація

про основні математичні моделі тестування та математично-статистичні методи в освітніх вимірюваннях, тому на аналіз програм саме цих п'яти дисциплін спрямована дана робота. У списку літератури наведені основні підручники та монографії [1]-[10], [16]-[21], які використовуються при вивченні названих курсів.

Математична складова системи підготовки магістра освітніх вимірювань має забезпечувати вирішення задач: знання основ математико-статистичного апарату, який використовується в теорії тестів; знання математичних методів опрацювання емпіричних результатів тестування та їх візуальне представлення; шкалювання; аналіз тестових завдань; вибір та використання моделі для аналізу надійності та валідності тесту.

Метою даної статті є визначення змісту математичної складової системи підготовки магістрів спеціальності 8.18010022 Освітні вимірювання, проведення порівняльного аналізу навчальних програм п'яти основних дисциплін, які викладались у 2011-2012 р., та визначення взаємозв'язків між ними.

1. Навчальна дисципліна «Вимірювання в освіті», як базова в системі підготовки магістра (у 2011-2012 н.р.) Метою дисципліни «Вимірювання в освіті» [11] є формування компетентностей майбутніх фахівців з питань педагогічного оцінювання, ознайомлення з методиками створення та використання тестового інструментарію для оцінки якості освіти.

Після вивчення дисципліни студенти повинні знати: науково-понятійний апарат педагогічного оцінювання, тестування і моніторингу якості освіти; концептуальні розв'язання проблеми вимірювання в освіті; методи і засоби педагогічного контролю; основи теоретичних і методичних підходів до конструювання педагогічних тестів; поняття надійності та валідності тестів; форми тестових завдань, особливості їх конструювання та оцінювання, переваги і недоліки різних форм тестових завдань; основні етапи розробки педагогічних тестів; особливості організації процесу тестування; особливості організації зовнішнього незалежного оцінювання в Україні, особливості організації моніторингових досліджень в освіті.

Вимірювання в освіті

ЗМІСТОВИЙ МОДУЛЬ I. Основи педагогічних вимірювань

- Поняття педагогічного вимірювання
- Типи вимірювання
- Класифікація навчальних цілей, таксономія Блума

ЗМІСТОВИЙ МОДУЛЬ II. Надійність та валідність тестів

- Надійність тесту
- Валідність тесту

ЗМІСТОВИЙ МОДУЛЬ III. Аналіз тестових завдань, розрахунок та інтерпретація балів

- Форми тестових завдань
- Аналіз тестових завдань
- Шкалування результатів

ЗМІСТОВИЙ МОДУЛЬ IV. Основи моніторингу якості освіти

- Короткий огляд систем вступу до ВНЗ

Рис.1. Структура дисципліни «Вимірювання в освіті»

При цьому студенти також повинні вміти: характеризувати і використовувати на практиці тестові завдання різних форм; виділяти із предметної діяльності ключові компоненти, від яких залежить успішність цієї діяльності; розробляти і аналізувати тестові завдання зі свого фаху та пропонувати рекомендації для підготовки учнів до виконання цих завдань; добирати комплекс моніторингових показників (індикаторів). Схема навчальної програми наведена на рис.1.

Вивчення дисципліни «Вимірювання в освіті» забезпечує знайомство майбутніх фахівців з основами методології освітніх вимірювань. Базові поняття про вимірювання вивчаються у першому змістовому модулі – суть педагогічних вимірювань, їх типи та мета вимірювання. Основним інструментальним засобом вимірювань в сучасних умовах є тест. Тому логічним продовженням змісту навчального курсу є вивчення основних показників тесту, як інструмента для проведення вимірювань. Основними показниками якості тесту виступають його надійність та валідність. Їх вивчення передбачається в другому змістовому модулі, завдання якого є знайомство з основними методами визначення надійності та валідності, а також їх інтерпретація. У третьому модулі вивчаються основні форми тестових завдань, стандартні математичні методи їх аналізу на відповідність меті тестування та шкалювання його результатів. Останній модуль передбачає знайомство з початковими поняттями моніторингу якості освіти.

Таким чином, вивчення дисципліни «Вимірювання в освіті» формує у магістрантів фахові компетентності у сфері освітніх вимірювань.

2. Навчальна дисципліна «Класичні тестові моделі» - фундаментальний курс з математичного моделювання в освітніх вимірюваннях. Метою дисципліни «Класичні тестові моделі» [12] є сформування у магістрів компетентностей, необхідних для опрацювання результатів тестування, зокрема, для визначення надійності та валідності тесту класичними методами.

Класичні тестові моделі

ЗМІСТОВИЙ МОДУЛЬ I. Класична модель істинної оцінки

- Коротка історія тестових досягнень
- Основні положення класичної теорії тестування

ЗМІСТОВИЙ МОДУЛЬ II. Статистична обробка емпіричних даних класичними методами

- Первинна обробка результатів тестування
- Розрахунок основних математико-статистичних характеристик тестових завдань та тесту

ЗМІСТОВИЙ МОДУЛЬ III. Надійність тесту

- Поняття надійності тесту в класичній теорії
- Методи оцінювання надійності

ЗМІСТОВИЙ МОДУЛЬ IV. Валідність тесту

- Основні типи валідності
- Статистичні процедури для прогнозування валідності
- Проблематика систематичної помилки відбору учасників тестування з генеральної сукупності
- Застосування факторного аналізу до проблеми дослідження валідності

Рис.2. Структура дисципліни «Класичні тестові моделі»

Після вивчення курсу магістри повинні знати: основні положення класичної теорії тестів; характеристики тестових завдань та статистичні методи їх оцінки класичними методами; процедури оцінювання надійності тесту. Магістри також повинні вміти: обчислювати основні статистичні характеристики завдань тесту класичними методами; давати оцінку надійності та валідності тесту у рамках класичної теорії тестів. Структура навчальної дисципліни «Класичні тестові моделі» наведена на рис.2.

Навчальна дисципліна «Класичні тестові моделі» складається із чотирьох змістових модулів і є логічним продовженням курсу «Вимірювання в освіті». У ній детальніше вивчаються ключові положення теорії тестів.

Перший змістовий модуль є коротким і в ньому розкривається класична модель істинної оцінки. Вводиться основне співвідношення між істинною оцінкою та оцінкою, що спостерігається.

Загалом обґрунтовуються п'ять основних положень класичної теорії тестів, аналізуються поняття паралельних та еквівалентних форм тестів.

У другому змістовому модулі вивчаються основні етапи обробки емпіричних даних класичними методами. Вони включають методи визначення частоти тестового балу, моди, середньо вибіркового, дисперсії, середньо квадратичного відхилення на основі матриці результатів тестування. Обв'язковою є перевірка гіпотези про нормальний закон розподілу. Завершується етап обробки даних відшукуванням кореляційної матриці та обґрунтуванням різних формул для обчислення коефіцієнтів кореляції. У цьому ж модулі вивчаються також методи розрахунку основних характеристик тестових завдань: складність, дискримінаційна здатність завдань, правдоподібність дистракторів, гомогенність, ваговий коефіцієнт та коефіцієнт контингенції.

Третій змістовий модуль присв'ячений вивченню надійності тесту в класичній теорії тестів. Виводяться формули показника та коефіцієнта надійності, розглядається їхня інтерпретація. Важливим є обґрунтування коефіцієнтів надійності Спірмена –

Брауна та α -Кронбаха для паралельних форм тестів. Отримані формули коефіцієнтів використовуються при вивченні загальноприйнятих методів наближеного обчислення коефіцієнта надійності. Також аналізуються фактори, що впливають на його значущість. Окремо виділена тема обчислення коефіцієнтів надійності для критеріально-орієнтованих тестів.

У четвертому модулі вивчаються методи визначення валідності тесту: змістовна валідність та її оцінювання, валідність для критеріально-орієнтованих тестів та інтерпретація результатів валідації, конструктивна валідність, а також коефіцієнти валідності для істинних оцінок.

Магістранти повинні вивчити статистичні процедури для прогнозування валідності: частинна кореляція, множинна регресія, методи оцінювання точності прогнозування, а також дискримінантний аналіз як метод дослідження валідності щодо категоризації учасників тестування. Важливо ознайомити студентів з проблематикою систематичної помилки відбору учасників тестування з генеральної сукупності та методами такого відбору. Останній розділ пов'язаний із застосуванням факторного аналізу до проблеми дослідження валідності тесту: факторні навантаження, обертання, корельовані фактори.

3. Сучасні підходи у математичному моделюванні освітніх вимірювань у навчальній дисципліні «Моделі і методи IRT». Метою дисципліни «Моделі і методи IRT» [14] є забезпечення поглибленої фахової підготовки магістрів для системи вищої та середньої освіти, а також установ, пов'язаних із зайнятістю населення, в галузі сучасної теорії та практики освітніх вимірювань з використанням тестів.

Дисципліна «Моделі і методи IRT» ознайомлює магістрантів з сучасною теорією моделювання та параметризації тестування в освіті. Представлено основні ідеї і методи цієї теорії, для розуміння яких вимагається достатньо високий рівень знань вищої математики, зокрема теорії ймовірностей та математичної статистики. Передбачається, що попередньо студенти отримали ґрунтовну математичну та психолого-педагогічну підготовку, а

також вивчили дисципліни нормативної частини циклу професійно-орієнтованої математичної та природничо-наукової підготовки «Вимірювання в освіті», «Класичні тестові моделі», «Математично-статистичні методи в освітніх вимірюваннях».

Опанувавши курс, магістри зможуть розробляти контрольні-вимірювальні матеріали, а також інтерпретувати результати вимірювань, спираючись на статистично-математичну базу, що дасть змогу здійснювати керівництво роботою комісій з підготовки та оцінювання тестових завдань.

Після вивчення дисципліни студенти повинні знати: основні математичні моделі сучасної теорії тестування; латентні параметри складності завдання і рівня підготовки випробовуваного; логістичні моделі Раша і Бірнбаума; поняття матриці відповідей, первинних балів; процедури та способи оцінювання надійності тесту; алгоритм усереднювання значень функції вимірювання; точність початкових вимірювань; розподільна здатність тесту, надійність та валідність тесту; алгоритми і методи оцінки рівня підготовленості тестованих та характеристик тестових завдань у рамках сучасної теорії тестування; тестування відносно і абсолютне; регресію і кореляцію на метричній шкалі; кореляцію на порядковій і номінальній шкалах; кореляцію бісеріальну та кореляцію дихотомну; основні методи дисперсійного, кореляційного, дискримінантного та факторного аналізів у педагогічних вимірюваннях; таблицю зв'язаних ознак; розподіл дистракторів; види, критерії та джерела підвищення валідності тесту; порядкову та метричну шкали; поняття остаточного балу учасника тестування; методи диференціації учасників тестування; основні методи шкалювання результатів освітнього вимірювання; інформаційну функцію тесту; характеристичну функцію тесту.

Студенти також повинні вміти: будувати оцінки латентних параметрів; виконувати оцінювання латентних параметрів методом моментів та методом найбільшої правдоподібності; будувати інформаційну функцію окремого завдання тесту; будувати інформаційну функцію тесту для одно- та багатопараметричних моделей; проводити статистичну перевірку гіпотез тестування; виконувати перевірку адекватності моделі

Моделі і методи IRT

ЗМІСТОВИЙ МОДУЛЬ I. Вступ до Item Response Theory

- Основні поняття і термінологія
- Оцінювання латентних параметрів та точності параметризації

ЗМІСТОВИЙ МОДУЛЬ II. Основні математичні моделі та методи сучасної теорії тестування в освіті

- Статистична перевірка гіпотез тестування
- Теоретичні основи шкалювання результатів тестування

Рис.3. Структура дисципліни «Моделі і методи IRT»

Раша за допомогою критерію згоди "Хі-квадрат" Пірсона; проводити порівняння емпіричної і теоретичної імовірності успіху; перевіряти рівномірність розподілу дистракторів; перевіряти значущість розбіжності різних результатів тестування на метричній та на порядковій шкалі; здійснювати перевірку

паралельності варіантів тесту; аналізувати значущість лінійної кореляції; використовувати методи дисперсійного, кореляційного та факторного аналізу при конструюванні тесту; здійснювати шкалювання результатів тестування; переносити латентні параметри, одержані у паралельних варіантах тесту, на єдину метричну шкалу; перетворювати єдину метричну шкалу у нормовану; виконувати порівняння тестів при конструюванні; вирівнювати різні варіанти тесту на єдиній метричній шкалі у процесі математичної обробки результатів.

Структура навчальної дисципліни «Моделі і методи IRT» розміщена на рис. 3. До основних понять IRT відносяться латентні параметри складності завдання і рівня підготовки випробовуваного. У курсі описуються логістичні моделі Раша і Бірнбаума, матриця відповідей, властивості первинних балів. Необхідно показати, що первинні бали є достатніми статистиками, провести редукування матриці відповідей.

Одним із важливих етапів у викладанні курсу є оцінювання латентних параметрів та точності параметризації. Слід дати магістрантам найпростіший алгоритм – усереднювання значень функції вимірювання, засоби оцінювання латентних параметрів методом моментів та методом найбільшої правдоподібності. Базовими поняттями є також інформаційна функція окремого завдання тесту та інформаційна функція тесту для одно- та багатопараметричних моделей. Важливо математично обґрунтувати точність початкових вимірювань, подати матеріал про роздільну здатність тесту, а також про надійність та валідність тесту за сучасною теорією тестування.

У другому змістовому модулі вводяться основні математичні моделі та методи сучасної теорії тестування в освіті. Так, викладається перевірка адекватності моделі Раша за допомогою критерію згоди "Хі-квадрат" Пірсона: аналіз основної моделі, аналіз матриці відповідей. Поглиблюється теоретичний матеріал з таких питань: кореляція бісеріальна, кореляція дихотомна, таблиця зв'язаних ознак. Проводиться порівняння емпіричної і теоретичної імовірності успіху, перевірка рівномірності розподілу дистракторів, перевірка значущості розбіжності різних результатів

тестування на метричній шкалі, перевірка значущості розбіжності різних результатів тестування на порядковій шкалі, перевірка паралельності варіантів тесту. Необхідно дати також огляд інших непараметричних критеріїв аналізу однорідності вибірок, пов'язаних з порядковими шкалами.

З основами теорії шкалювання результатів тестування магістранти ознайомились у попередніх курсах. Тут же узагальнюється класифікація шкал. Вивчаються такі алгоритми: перенесення латентних параметрів, одержаних у паралельних варіантах тесту, на єдину метричну шкалу; перетворення єдиної метричної шкали в нормовану. Виводиться алгоритм для отримання остаточного балу учасників тестування. Розглядаються методи підвищення диференціації учасників тестування. Вивчаються можливості порівняння тестів при конструюванні, а також вирівнювання різних варіантів тесту на єдиній метричній шкалі у процесі математичної обробки результатів.

Таким чином, дисципліна «Моделі і методи IRT» присвячена математико-статистичному моделюванню тестів та обробці їх результатів, при цьому цей курс є обов'язковим і входить до нормативної частини навчального плану. Він забезпечує необхідні знання та навички для продовження поглибленого вивчення різних сучасних узагальнень IRT, а також інших методів вимірювань в освіті, що базуються на байєсівських підходах з використанням методу Монте Карло, тощо. Названі розділи можна включити до тематики спецкурсів та спецсемініарів, а також можна внести до тематики магістерських робіт.

4. Математичний апарат теорії освітніх вимірювань у дисциплінах «Математико-статистичні методи в педагогічних вимірюваннях» та «Прикладна статистика». Метою викладання дисциплін «Математико-статистичні методи в освітніх вимірюваннях» [13] та «Прикладна статистика» [15] є забезпечення поглибленої математико-статистичної підготовки

фахівців системи вищої та середньої освіти в галузі сучасної теорії та практики освітніх вимірювань.

Вивчення дисциплін «Математико-статистичні методи в освітніх вимірюваннях» та «Прикладна статистика» передбачає узагальнення та систематизацію теоретичних відомостей з теорії ймовірностей та теорії статистики, знайомить з основними математико-статистичними методами обробки результатів тестування й інших видів контролю засвоєння навчального матеріалу, визначення законів розподілів емпіричних даних результатів тестування та ін., побудови статистичних гіпотез та їх перевірки. Зокрема, розкривається суть вибіркового методу й статистичного оцінювання, методів вивчення взаємозв'язків явищ і процесів, методів перевірки параметричних і непараметричних гіпотез. Передбачається, що студенти мають достатню математичну підготовку для вивчення дисциплін нормативної частини циклу професійно-орієнтованої математичної та природничо-наукової підготовки.

Після вивчення вказаних дисциплін студенти повинні знати: основні поняття й теореми теорії ймовірностей; елементарні ймовірнісні моделі в дискретних просторах елементарних подій; основні розподіли випадкових величин: гіпергеометричний, геометричний, біноміальний, Пуассона, рівномірний, показниковий, нормальний, логнормальний, Стюдента, χ^2 , Фішера-Снедекора та їх числові характеристики; закон великих чисел (ЗВЧ), центральну граничну теорему (ЦГТ) та їх застосування у практиці вимірювань; основні поняття і задачі математичної статистики; види статистичних рядів, їх числові характеристики, графічне зображення; основні поняття і визначення вибіркового методу; основні способи формування вибіркової сукупності; помилки вибірки та методи їх обчислень; види статистичних оцінок числових характеристик генеральної сукупності, їх властивості, методи визначення й побудови; функцію впливу та її властивості; нерівність Рао-Фреше-Крамера; типи зв'язків між випадковими величинами; кореляційні параметричні методи вивчення зв'язку; вибіркового коефіцієнта кореляції та його властивості; рівняння лінійної та нелінійної

регресії; рангові коефіцієнти зв'язку; коефіцієнт конкордації та його використання при розв'язуванні основних задач статистичного аналізу рангового зв'язку; індекс кореляції; індекс Фехнера і кореляційне співвідношення; часові ряди; тренди часових рядів; методи аналізу основної тенденції (тренду) часових рядів; статистичні гіпотези, їх види; статистичний критерій, потужність критерію; критичні області, їх види, принципи побудови; загальну схему перевірки статистичної гіпотези; параметричні методи перевірки статистичних гіпотез для однієї вибірки, двох та декількох вибірок; непараметричні методи перевірки статистичних гіпотез; критерії узгодження Пірсона, Колмогорова, Романовського та ін.; комплексне застосування математико-статистичних методів аналізу статистичних даних.

Студенти також повинні вміти: знаходити ймовірності випадкових подій; обчислювати числові характеристики випадкових величин, визначати їх розподіл; оцінювати надійність і точність вимірювань, користуючись нерівністю Чебишова, ЗВЧ й ЦГТ, визначати необхідну кількість вимірювань; будувати статистичні ряди з емпіричних даних; знаходити числові характеристики статистичних рядів та функції розподілу статистичних даних, будувати їх графіки; знаходити точкові та інтервальні оцінки параметрів генеральної сукупності; встановлювати причинно-наслідкові, кореляційні зв'язки між ознаками, визначати їх вид, форму та щільність, оцінювати істотність зв'язку; аналізувати лінійну кореляцію за даними випадкової вибірки; оцінювати значущість коефіцієнта лінійної кореляції; визначати рангові коефіцієнти зв'язку; обчислювати показники зміни рівнів часових рядів; аналізувати основну тенденцію (тренд) часових рядів; виконувати регресійний аналіз зв'язних часових рядів, визначати їх кореляцію; здійснювати прогнозування й інтерполяцію; перевіряти параметричні й непараметричні гіпотези; аналізувати помилки, які можуть виникнути при перевірці статистичних гіпотез; оцінювати розподіли з використанням критеріїв узгодженості та ін.; застосовувати математико-статистичні методи для обробки результатів і аналізу даних.

Схеми навчальних дисциплін «Математико-статистичні методи в освітніх вимірюваннях» та «Прикладна статистика» наведені на рис.4.

Опанувавши названі курси, магістранти зможуть систематизувати, обробляти та аналізувати статистичні дані, правильно інтерпретувати їх, виявляти причинно-наслідкові зв'язки та використовувати результати аналізу для наукових і практичних висновків.

Міжпредметні зв'язки між навчальними дисциплінами «Вимірювання в освіті», «Класичні тестові моделі», «Моделі і методи IRT», «Математико-статистичні методи в освітніх вимірюваннях» та «Прикладна статистика» у 2011-2012 н.р. можна представити у вигляді схеми (рис.5).

Дисципліна «Вимірювання в освіті» є центральною дисципліною системи підготовки, вона у співрозмірних частинах наповнена математичним, організаційно-технологічним та психологічним змістом. У програмі цієї дисципліни вводиться широке коло основних понять теорії освітніх вимірювань без глибокого математичного обґрунтування.

Дисципліни «Класична теорія тестів» та «Моделі і методи IRT» містять інформацію про класичний та сучасний підходи у математичному моделюванні тестів та обробки їх результатів. Для якісного вивчення матеріалу цих професійно-орієнтованих дисциплін магістрантам необхідно вивчити дві суто математичних дисципліни «Математико-статистичні методи в освітніх вимірюваннях» та «Прикладна статистика», кожна з яких узагальнює раніше отримані раніше знання та надає нові відомості з теорії ймовірностей та теорії статистики і їх застосувань. при викладанні зазначених дисциплін наводиться значна кількість характерних прикладів із практики вимірювань в освіті, що сприяє глибшому розумінню теорії освітніх вимірювань.

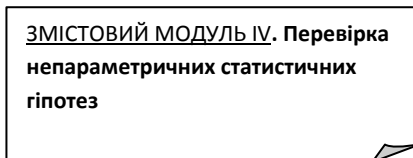
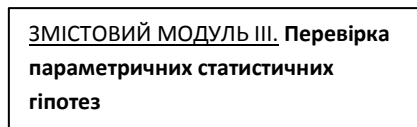
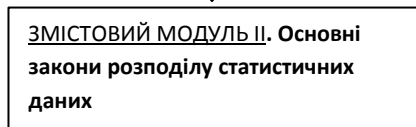
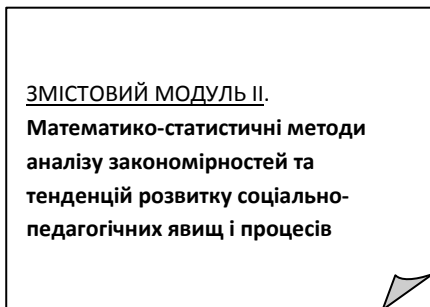
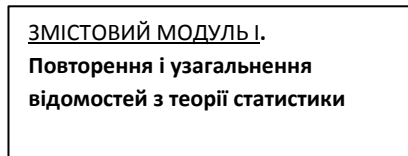
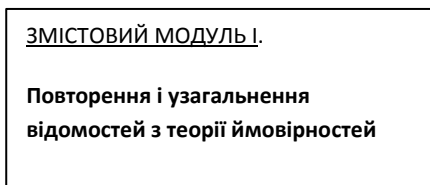
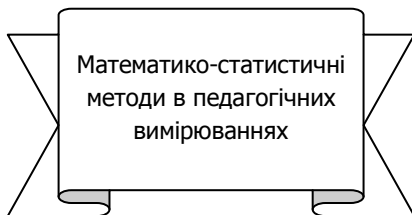


Рис.4. Структура дисциплін «Математико-статистичні методи в освітніх вимірюваннях» та «Прикладна статистика»



Рис. 5. Міжпредметні зв'язки основних дисциплін математичної компоненти за планом 2011-2012 н.р.

5. Зміни у програмі підготовки, затверджені у 2012 році.

Після детального аналізу результатів впровадження у 2011-2012 н.р. тимчасового стандарту підготовки магістра спеціальності «Освітні вимірювання» у Кіровоградському державному педагогічному університеті імені Володимира Винниченка було прийняте рішення про укрупнення навчальних дисциплін при збереженні попередньої кількості та якості змістових модулів.

Замість п'яти дисциплін «Вимірювання в освіті», «Класичні тестові моделі», «Моделі і методи IRT», «Математико-статистичні методи в освітніх вимірюваннях» та «Прикладна статистика» (2011-2012 н.р) пропонується основні змістові модулі, що містять відомості про основні математико-статистичні методи обробки результатів тестування та про класичні та сучасні підходи до моделювання та параметризацію тестів, зібрати у двох дисциплінах «Моделювання та параметризація тестів» та «Статистичні методи в освітніх вимірюваннях».

У оновленому стандарті підготовки магістрів, затвердженому у 2012 році запропоновано новий перерозподіл змістових модулів та нову комплектацію нових блоків змістових модулів, наведені нижче:

Дисципліна	Блок змістових модулів	Змістовий модуль
Моделювання та параметризація тестів	Вимірювання в освіті	Вступ до теорії педагогічних вимірювань
		Аналіз тестових завдань
		Розрахунок та інтерпретація тестових балів
	Класичні тестові моделі та технології	Основні постулати класичної теорії тестування
		Розрахунок основних математико-статистичних характеристик тестових завдань та тесту
		Валідність тесту
		Надійність тесту

	Моделі і методи IRT	Модель Раша
		Оцінювання латентних параметрів та точності параметризації
		Шкалювання результатів тестування
		Сучасна теорія тестування (Item Response Theory)
Статистичні методи в освітніх вимірюваннях	Математичні основи теорії тестування	Імовірісно-статистичні основи класичної теорії тестування
		Елементи математичного моделювання
		Класична теорія тестування
	Елементи прикладної статистики	Багатовимірні випадкові величини
		Варіансний аналіз
		Кореляційний і регресійний аналізи
		Методи багатомірного статистичного аналізу
		Статистична перевірка гіпотез тестування
		Елементи аналізу регресії і кореляції

Спеціальність «Освітні вимірювання» є новою у світі, а в Україні впроваджується вперше, тому процес становлення дієвої та надійної схеми взаємозв'язків між дисциплінами викликає та буде викликати зміни у програмі підготовки магістрів ще не один рік. Сподіваємося, що незабаром буде розроблений та затверджений Державний галузевий стандарт спеціальності «Освітні вимірювання», який у повній мірі врахує досвід підготовки перших магістрів, зокрема, буде надано чільне місце математичній складовій підготовки.

Література

1. Айвазян С.А. и др. Прикладная статистика: Исследование зависимостей. – Москва: Финансы и статистика, 1985. – 487 с.
2. Вимірювання в освіті: Підручник / За редакцією О.В. Авраменко.– Кіровоград: Лисенко В.Ф., 2011. – 360 с.
3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: Учеб. пособие для студентов вузов. – Москва: Высш. шк., 2003. – 405 с.
4. Гмурман В.Е. Теория вероятностей и математическая статистика: Учеб. пособие для вузов. – Москва: Высш. шк., 2003. – 479 с.
5. Жалдак М.І., Кузьміна Н.М., Михалін Г.О. Теорія ймовірностей і математична статистика: Підручник. – Вид.2-е, перероб. і доп. – Полтава: Довкілля-К, 2009. – 500 с.
6. Жлуктенко В.І., Наконечний С.І., Савіна С.С. Теорія ймовірностей і математична статистика: Навч.-метод. посібник: У 2-х ч. – Ч.1. Теорія ймовірностей. – К.: КНЕУ, 2000. – 304 с.
7. Жлуктенко В.І., Наконечний С.І., Савіна С.С. Теорія ймовірностей і математична статистика: Навч.-метод. посібник: У 2-х ч.– Ч.2. Математична статистика. – К.: КНЕУ, 2001. – 336 с.
8. Звонников В.И., Челышкова М.Б. Современные средства оценивания результатов обучения. – М.: Издательский центр «Академия», 2007. – 224 с.
9. Ким В.С. Тестирование учебных достижений. Монография. – Уссурийск: Издательство УГПИ, 2007. – 214 с.
10. Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов / Пер. с англ. под общ. ред. В.И. Звонникова, М.Б. Челышковой. – М. : Логос, 2010. – 668 с.
11. Навчальна програма дисципліни «Вимірювання в освіті» // Укладач Резіна О.В.- Кіровоград: КДПУ ім.Винниченка, 2011.
- 12.Навчальна програма дисципліни «Класичні тестові моделі» // Укладач Парашук С.Д.- Кіровоград: КДПУ ім.Винниченка, 2011.

13. Навчальна програма дисципліни «Математико-статистичні методи в освітніх вимірюваннях» // Укладач Лутченко Л.І.- Кіровоград: КДПУ ім.Винниченка, 2011.

14. Навчальна програма дисципліни «Моделі і методи IRT» // Укладач Авраменко О.В.- Кіровоград: КДПУ ім.Винниченка, 2011.

15. Навчальна програма дисципліни «Прикладна статистика» // Укладач Лутченко Л.І.- КДПУ ім.Винниченка, 2011.

16. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. – М.: Прометей, 2000. – 168 с.

17. Самылкина Н. Н. Современные средства оценивания результатов обучения. – М.: БИНОМ. Лаборатория знаний, 2007. – 172 с.

18. Чельшкова М.Б. Теория и практика конструирования педагогических тестов. – М.: Логос, 2002. – 432 с.

19. Ярошук Л.Г. Основи педагогічних вимірювань та моніторингу якості освіти : навч. посіб. – К. : Видавничий Дім «Слово», 2010. – 304 с.

20. Baker F.B. The Basics of Item Response Theory.- Portsmouth NH: Heinemann Educational Books, 1985.- 131 pp.

21. Hambleton R.K., Swaminathan H., Rogers H. J. Fundamentals of Item Response Theory. Newbury Park, CA: Sage, 1991.

КУРС «МАТЕМАТИКО-СТАТИСТИЧНІ МЕТОДИ В ПЕДАГОГІЧНИХ ВИМІРЮВАННЯХ» ДЛЯ МАЙБУТНІХ ВЧИТЕЛІВ МАТЕМАТИКИ

Лісова Т.В.

Важливим аспектом будь-якої педагогічної діяльності є оцінки, які виставляють викладачі своїм учням, студентам. Наслідки цих оцінок можуть бути різними – від морального ефекту до впливу на долю людини. При цьому всі розуміють, що вони є суб'єктивними і часто дуже наближеними. Принципово змінити ситуацію може лише підхід до оцінювання знань як до процесу об'єктивного вимірювання, результати якого можна обробляти стандартними математичними методами. Але виникають питання: що, чим і як вимірювати, якими шкалами користуватись, як оцінювати точність вимірювань тощо. Відповіді на ці питання може дати наука тестологія, яка вивчає різні тести, в тому числі і педагогічні, які є основним інструментом об'єктивного вимірювання рівня підготовленості учнів.

Тестологія з'явилася на стику психології, соціології, педагогіки та інших наук. Вперше тести виникли наприкінці XIX – початку XX ст., коли для виявлення фізичних, фізіологічних та психічних особливостей людини психологи намагалися застосувати різні способи вимірювання. Одночасно розвивались і методики обробки результатів тестування, у яких використовувались сучасні досягнення статистики, особливо кореляційного та факторного аналізів. З початку XX ст. визначився педагогічний напрям у тестології. Розробка першого педагогічного тесту належить американському психологу Е. Торндайку, який вважається основоположником педагогічних вимірювань. Він говорив, що «якщо щось існує, то воно може бути виміряне», однак висловлював жаль з приводу того, що психолог може пропрацювати все життя, так ніколи і не зіткнувшись з математикою. Але швидкий розвиток природознавства за останні сто років сприяв проникненню точних вимірвальних методів і у такі «неточні» сфери, як психологія та педагогіка. Сьогодні

отримати та обробити необхідну інформацію, яка використовується в теорії та практиці соціальних наук, можна лише на основі грамотного застосування статистичних підходів та методів. Саме тому теорія ймовірностей та статистика стали обов'язковими дисциплінами не лише для математичних, а й для психологічних, соціологічних, економічних спеціальностей у вузах. Для педагогічних спеціальностей вона є нормативною лише для фізико-математичного та природничого напрямів. Саме студенти цих спеціальностей більш підготовлені до вивчення теорії педагогічних вимірювань на рівні, що передбачає присвоєння кваліфікації «фахівець у галузі освітніх вимірювань».

Вивчення дисципліни «Математико-статистичні методи в педагогічних вимірюваннях» при підготовці вчителів за напрямом 6.040201 Математика* зі спеціалізацією «Освітні вимірювання» у Ніжинському державному університеті передбачається обсягом 144 години (4 кредити ECTS).

При розробці програми даного курсу враховується, що студенти прослухають курс теорії ймовірностей та математичної статистики обсягом 216 годин (6 кредитів ECTS). Причому, у зв'язку із введенням нової спеціалізації, програму курсу теорії ймовірностей та математичної статистики планується оновити. Деякі розділи теорії ймовірностей та, особливо, математичної статистики доречно поповнити розглядом задач, що можуть виникати у практиці педагогічних вимірювань. Наприклад, вивчення формул Байеса доцільно супроводжувати застосуванням їх до результатів педагогічних вимірювань при наявності чи відсутності апріорної інформації. При вивченні описових характеристик вибірки звернути увагу на відмінності цих характеристик при застосуванні різних шкал (номінальної, рангової, інтервальної та шкали відношень), що використовуються для опису досліджуваних ознак в педагогіці.

Перевірку статистичних гіпотез можна здійснювати на матеріалах психологічних та педагогічних досліджень та вимірювань. З рекомендаціями щодо вибору того чи іншого статистичного критерію в психолого-педагогічних дослідженнях можна ознайомитись в [3, 39-198] та [6, 187-193]. При вивченні

коефіцієнта кореляції доцільно використати ситуації з педагогічної практики, але наголосити, що кореляція дає уявлення лише про статистичний зв'язок, а не про причинно-наслідкову залежність між змінними. Наприклад, дослідити кореляцію між успішністю учнів з математики та іноземної мови. Якщо вона виявиться високою, пояснити, що це зумовлено не впливом однієї оцінки на іншу, а проявами дії деяких факторів (інтелекту, мотивації до навчання тощо). Регресійний аналіз доречно застосувати, наприклад, для дослідження залежності оцінки студентів на екзамені від успішності протягом семестру або залежності балів студента за одним тестом від балів за іншим тестом з однієї дисципліни. Застосування однофакторного дисперсійного аналізу можна показати на прикладі дослідження ефективності деякої педагогічної методики. Ознайомлення із двофакторним та мультифакторним дисперсійним аналізом планується у цьому курсі у межах часу, відведеного на самостійну та індивідуальну роботу. Поняття про методи класифікації та кластеризації теж доцільно дати на прикладах із психології чи педагогіки, обмежившись лінійним дискримінантним аналізом. Наприклад, за результатами навчання студента на першому курсі віднести його або до групи, яка успішно закінчить навчальний заклад, або до групи відрахованих. Планується на лекціях дати лише поняття про нелінійні класифікатори, метод найближчих сусідів, методи зниження розмірності. За новою програмою суттєво збільшується час на вивчення факторного аналізу, методи якого дозволяють виявити приховані (латентні) ознаки деякого складного явища, його внутрішню, часто приховану, структуру. Такими складними явищами є інтелект, особистість, мотивація тощо. Вони можуть характеризуватись великою кількістю спостережуваних змінних, а факторний аналіз дозволяє перейти до меншої кількості деяких узагальнених прихованих факторів, вплив яких визначається комбінацією кількох таких змінних. При цьому зменшується об'єм статистичних спостережень без втрати інформації про явище. Поява факторного аналізу є заслугою психологів Ч. Спірмена та Л. Терстоуна, а результатом його застосування у психології – факторні теорії особистості.

Після вивчення теорії ймовірностей та математичної статистики за оновленою програмою студенти матимуть необхідні знання та навички, щоб успішно засвоїти математичні методи у педагогічних дослідженнях та вимірюваннях. Вони матимуть уявлення про різні шкали для опису досліджуваних ознак, способи формування вибірки, її графічні та числові характеристики; здійснюватимуть статистичну оцінку параметрів генеральної сукупності; будуватимуть довірчі інтервали; перевірятимуть статистичні гіпотези; зможуть користуватися методами кореляційного, регресійного та дисперсійного аналізу; матимуть уявлення про дискримінантний та факторний аналіз.

Планується також оновити зміст деяких нормативних дисциплін, що вивчаються у попередніх семестрах. Зокрема, до дисциплін «Дискретна математика» та «Математична логіка і теорія алгоритмів» включаються розділи з теорії нечітких множин та нечіткої логіки, які будуть корисними при вивченні різних моделей тестування.

Метою дисципліни «Математико-статистичні методи в педагогічних вимірюваннях» є забезпечення фахової підготовки вчителів в галузі методики та практики педагогічних вимірювань. Опанувавши курс, майбутні вчителі будуть не лише свідомими неупередженими користувачами тестів, а зможуть самостійно планувати та проводити дослідження навчальної успішності учнів, кваліфіковано розробляти контрольньо-вимірювальні матеріали, якісно готувати учнів до Зовнішнього незалежного оцінювання, правильно інтерпретувати його результати, брати участь у роботі комісій з підготовки та оцінювання тестових завдань. Зміст даної дисципліни при підготовці вчителів математики з додатковою спеціалізацією «Освітні вимірювання» пропонується такий:

1. Вступ.

Поняття вимірювання у педагогічному контролі. Тест як засіб педагогічного вимірювання. Форми подання результатів вимірювання, геометрична інтерпретація. Матриця тестових відповідей, профілі відповідей. Джерела виникнення систематичних та випадкових помилок у вимірюваннях,

мінімізація помилок. Математико-статистичні пакети для обробки результатів педагогічного вимірювання.

2. Класична теорія тестування (*Classical Test Theory*).

Статистична обробка результатів у рамках класичної теорії тестування: мода, середнє вибіркове, дисперсія, середньо квадратичне відхилення та їх інтерпретація. Перевірка гіпотези про нормальний закон розподілу результатів тестування. Асиметрія та ексцес, їх інтерпретація та перевірка значущості. Коефіцієнт кореляції Пірсона, φ -коефіцієнт кореляції, коефіцієнт бісеріальної кореляції та їх інтерпретація.

Розрахунок основних математико-статистичних характеристик тестових завдань та тесту у рамках класичної теорії: складність, диференційна здатність завдань, правдоподібність дистракторів, гомогенність, ваговий коефіцієнт та коефіцієнт контингенції.

Основні постулати класичної теорії тестування. Істинний результат та стандартна помилка вимірювання. Індекс та коефіцієнт надійності. Фактори, що впливають на надійність. Процедури оцінювання надійності. Коефіцієнт α -Кронбаха. Надійність і довжина тесту. Довірчий інтервал для істинного результату.

3. Сучасна теорія тестування (*Item Response Theory*).

Латентні параметри складності завдання та рівня підготовленості тестованого. Функція успіху, характеристичні функції складності завдання та рівня підготовленості. Логістичні моделі: однопараметрична модель Г.Раша, дво- та трипараметрична моделі А.Бірнабаума. Оцінка параметрів функції успіху на основі спостережуваних балів. Метод моментів та метод максимальної правдоподібності. Аналіз точності отриманих оцінок та перевірка адекватності моделі Г.Раша за допомогою критерію χ^2 . Побудова характеристичних кривих. Інформаційні функції завдань та тесту. Конструювання тесту з використанням IRT.

4. Теорія узагальнення (*Generalizability Theory*).

GT як узагальнення класичної теорії тестів. G(*generalizability*)-дослідження та D(*decision*)-дослідження.

Врахування різних джерел помилок. Дисперсійний аналіз при обчисленні G-коефіцієнтів надійності. Стабільність тестових результатів, внутрішня узгодженість, надійність оцінок експертів.

5. Валідність тесту.

Поняття та види валідності. Критерії валідності. Статистичні процедури прогнозу та класифікації. Передбачення істинного результату на основі регресійної моделі. Дискримінантний аналіз для класифікації тестованих на дві і більше групи. Застосування факторного аналізу до виявлення кількості та природи факторів, що описують структуру кореляційних зв'язків між окремими тестами. Джерела підвищення валідності тесту.

6. Шкалювання результатів тестування.

Порядкові шкали. Метрична шкала. Перенесення латентних параметрів, отриманих за паралельними варіантами тесту, на єдину метричну шкалу. Перетворення єдиної метричної шкали на нормовану. Шкала станайнів та стенів. Шкала логітів. Остаточний результат тестованого.

Після вивчення дисципліни студенти повинні **знати**:

- види розподілів даних педагогічного вимірювання;
- методи і способи обчислення статистичних характеристик розподілу даних педагогічного вимірювання;
- методи перевірки статистичних гіпотез;
- основні постулати класичної теорії тестів;
- методи інтерпретації результатів математико-статистичної обробки даних педагогічного вимірювання у процесі створення тесту;
- характеристики тестових завдань та статистичні методи їх оцінки у рамках класичної теорії;
- процедури та способи оцінювання надійності тесту;
- основні математичні моделі сучасної теорії тестування;
- алгоритми і методи оцінки рівня підготовленості тестованих та характеристик тестових завдань у рамках сучасної теорії тестування;

- основні методи дисперсійного, кореляційного, дискримінантного та факторного аналізів у педагогічних вимірюваннях;
- основи теорії узагальнення;
- види, критерії та джерела підвищення валідності тесту;
- основні методи шкалювання результатів педагогічного вимірювання.

Студенти також повинні **вміти**:

- будувати матрицю та профілі тестових відповідей;
- застосовувати математико-статистичні пакети для обробки результатів педагогічного вимірювання;
- будувати оцінки основних параметрів статистичного розподілу та здійснювати перевірку їх значущості;
- інтерпретувати результати обробки даних тестування при конструюванні та застосуванні тесту;
- розраховувати основні статистичні характеристики завдань у рамках класичної та сучасної теорії;
- будувати характеристичні криві завдань та індивідуальні криві учасників тестування;
- будувати інформаційні криві завдань та тесту;
- досліджувати характеристики надійності тесту у рамках класичної та узагальненої теорії тестів;
- використовувати методи дисперсійного, кореляційного та факторного аналізу при конструюванні тесту;
- досліджувати показники валідності тесту;
- шкалювати результати педагогічних вимірювань.

Навчальний процес з дисципліни планується організувати у формі лекцій, практичних та лабораторних занять. На лабораторних заняттях студенти навчатимуться проводити необхідні розрахунки за допомогою програмних засобів. При вивченні дисципліни передбачається також індивідуальна та самостійна робота. Орієнтовний розподіл часу між видами роботи може бути таким:

№ з/п	Теми	Всього годин	З них:				
			Лекції	Практичні	Лабораторні	Індивідуаль-на робота	Самостійна робота
МОДУЛЬ I. Основи теорії тестів							
1.	Вступ	18	2	2		6	8
2.	Класична теорія тестування	28	6	2	4	8	8
МОДУЛЬ II. Сучасні теорії тестів							
3.	Сучасна теорія тестування	30	6	2	4	8	10
4.	Теорія узагальнення	18	2	2		6	8
МОДУЛЬ III. Валідність тесту та представлення результатів							
5.	Валідність тесту	34	6	2	4	10	12
6.	Шкалювання результатів тестування	16	4	2		4	6
Разом за курс		144	26	12	12	42	52

Література

1. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. – М.: Прогресс, 1976. – 494 с.
2. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО «Речь», 2003. – 350 с.
3. Чельшкова М.Б. Теория и практика конструирования педагогических тестов. – М.: Логос, 2002. – 432 с.
4. Linda Crocker, James Algina. Introduction to classical and modern test theory. – Wadsworth: Thomson Learning, 1986. – 528 p.
5. Ruth Ravid. Practical statistics for Educators. – Lanham: University Press of America, 2005. – 230 p.

МОЖЛИВОСТІ КОМП'ЮТЕРНОЇ ОБРОБКИ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ

Лісова Т.В.

Широке впровадження тестового контролю у педагогічну практику приводить до ситуацій, коли необхідно швидко та якісно проаналізувати отримані у результаті тестування дані, коректно їх проінтерпретувати та зробити об'єктивні висновки не лише про рівень успішності, а й про якість та надійність самого тесту. Зазвичай такий аналіз проводиться на основі класичної тестової теорії або на основі сучасної теорії тестів IRT (Item Response Theory). Обидві теорії на сьогодні мають добре розвинений математичний апарат, застосування якого полегшують спеціально розроблені комп'ютерні програми. Розрахунки за класичною теорією більш прості, їх можна проводити з використанням різних статистичних пакетів, а основні ідеї класичної теорії закладено у більшості тестових програм. Але, незважаючи на прозорість та зрозумілість висновків, отриманих за класичною теорією, вона має принципові недоліки, які не дозволяють вдосконалювати тест або проводити адаптивне тестування. Зокрема, тестові бали залежать від складності завдань, а складність завдань залежить від контингенту тестованих.

Перетворити набір тестових завдань на надійний та валідний тест зі стійкими характеристиками дозволяє сучасна теорія тестів IRT, яка дає потужний інструмент для вимірювання на інтервальній шкалі у спеціальних одиницях (логітах) латентних властивостей рівня підготовленості та складності завдань. Серед моделей IRT найчастіше використовуються одно-, дво- та трипараметричні логістичні моделі. Для оцінки параметрів цих моделей на основі даних тестування для певної вибірки використовують різні статистичні процедури, що вимагають великих об'ємів розрахунків. З кінця 70-х років минулого століття з'явилося багато комп'ютерних програм, у яких ці процедури було реалізовано.

Добре відомою у свій час для обчислень за методом

максимальної вірогідності на основі трипараметричної моделі була програма LOGIST (перша версія 1976 р.). Вона дозволяє одночасно оцінювати параметри для всіх завдань та отримувати оцінки латентної характеристики для всіх учасників тестування за сумісною процедурою максимальної вірогідності JML (*joint maximum likelihood procedure*). Недоліком такої процедури є відкрите питання про слухність отриманих оцінок параметрів завдань та необхідність мати великі вибірки (більше 1000 опитаних) для забезпечення більш точних оцінок. У програмі BILOG (1984 р.) для оцінки параметрів за трипараметричною моделлю використовується процедура маргінальної максимальної вірогідності MML (*marginal maximum likelihood procedure*), раніше відома, як безумовна процедура максимальної вірогідності. Оцінки, отримані за цією процедурою, є слухними. Програма BILOG та її вдосконалені версії широко використовується у країнах Північної Америки та Європи. Для однопараметричної моделі (або ж моделі Раша) початківцям рекомендується використовувати програму BICAL (1979 р.), оскільки її результати дуже прості для інтерпретації [1, ст. 474]. Розробники BICAL запропонували до процедури JML ввести додатковий множник, який дозволяє отримати майже слухні оцінки параметрів. Популярною серед користувачів є також програма WINSTEPS (Windows версія BIGSTEPS, 1991 р.), яка дозволяє калібрувати дихотомічні завдання за допомогою моделі Раша та політомічні завдання у рамках моделей Partial Credit та Rating Scale. Для оцінки параметрів тут також використовується процедура JML [6, ст. 131]. Популярності цій програмі додає те, що безкоштовно поширюється її академічна версія MINISTER, у якій можна аналізувати до 25 завдань на вибірці до 75 опитаних. Умовну процедуру максимальної вірогідності CML (*conditional maximum likelihood procedure*) для моделі Раша використали розробники програм RUMM (1990 р.) та WINMIRA (2001 р.). Приклад застосування програми RUMM у Росії можна знайти у монографії В.С. Кіма [3, ст.200]. До переліку популярних сьогодні програм можна також віднести MULTLOG (1991 р.), PARSCALE (1997 р.), продукти компанії ASC (Assessment System Corporation) та багато

інших. Користувачеві залишається лише вибрати потрібну модель, програмний засіб для її реалізації та правильно проінтерпретувати отримані результати обчислень.

Тут розглядаються можливості пакету ІТАР (Item and Test Analysis Package) компанії ASC та його застосування до аналізу 14 тестових завдань з теорії ймовірностей та математичної статистики у групі з 48 опитаних. Пакет містить шість програм, розроблених у 1995-1998 рр. та вдосконалених для Windows, які дозволяють здійснювати всебічний аналіз результатів тестування, анкетування,

```
14 o n 7
aaaaaabaaaaaaaa
444555555555555
YYYYYYYYYYYYY
stud001adaeacaaaaabe
stud002adaeaccbaacae
stud003adadadbbaaacbe
stud004adaaacaaaabaca
stud005adaaaebaaaaada
stud006adaaacaaaaaba
```

Рис.1. Фрагмент вхідного файлу

соціологічних чи психологічних опитувань у рамках класичної та сучасної тестової теорії.

Дані для аналізу у трьох програмах (ITEMAN, RASCAL та XCALIBRE) створюються як текстові ASCII файли і мають однаковий формат з деякими відмінностями, що залежать від типу тесту. На рис.1 наведено фрагмент вхідного файлу для дихотомічних завдань. У першому контрольному рядку через пропуск потрібно зазначити кількість завдань (до 750), символ для позначення пропущених відповідей (o), символ для позначення питань, на які опитуваний не встиг відповісти (n), кількість символів для ідентифікації опитаних (до 80). Другий рядок – ключ правильних відповідей на кожне завдання, третій – кількість альтернатив у кожному завданні (від 2 до 9). У четвертому рядку літерою Y позначають завдання, які включаються до аналізу, а N – ті, що не включаються. Далі наступним суцільним рядком – відповіді кожного опитаного, причому кількість опитаних для всіх програм необмежена.

Програма ITEMAN забезпечує традиційний аналіз тесту або опитування у рамках класичної теорії. Вона дозволяє аналізувати дані, отримані у дихотомічній та у багатопозиційних шкалах (типу шкали Лайкерта). Передбачена можливість аналізу питань з

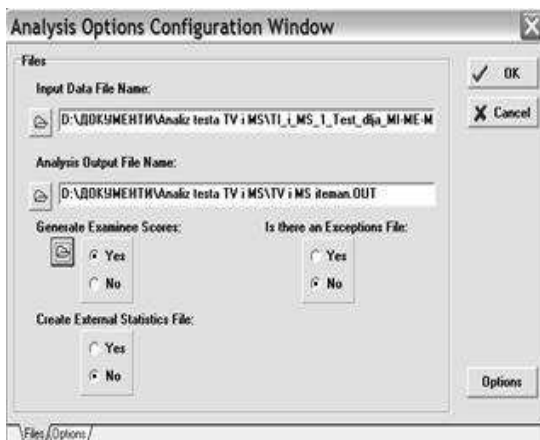


Рис.2. Вікно *Files* програми ITEMAN

кількома правильними відповідями. Для цього у вікні програми *Files* (рис.2), у якому визначаються всі робочі файли, потрібно вибрати кнопку *Yes* у полі *Exceptions Files* та сформувати додатковий файл для таких питань.

У всіх програмах передбачена можливість формування вихідного файлу у вигляді, зручному для експорту у інші статистичні пакети для подальшої обробки (*External Statistics File*). Параметри аналізу задаються у вікні *Options*. Тут є можливість вибрати вид кореляційних зв'язків, поділ опитуваних на слабку та сильну групи, аналіз тесту у підгрупах, уточнення параметрів завдань для коротких тестів. В ITEMAN, на відміну від інших програм пакету, не розрізняються пропущені питання та ті питання, на які опитуваний не встиг відповісти. Для аналізу питань, відповідь на які опитуваний пропустив, є три різні процедури. Вихідний текстовий файл містить загальні статистичні показники тесту (рис.3), кожного завдання та кожної альтернативи залежно від обраної опції аналізу (рис.4).

N of Items	14
N of Examinees	48
Mean	8.125
Variance	4.234
Std. Dev.	2.058
Skew	0.334
Kurtosis	-0.008
Minimum	4.000
Maximum	13.000
Median	8.000
Alpha	0.403
SEM	1.590
Mean P	0.580
Mean Item-Tot.	0.131
Mean Biserial	0.174
Max Score (Low)	7
N (Low Group)	18
Min Score (High)	9
N (High Group)	17

Number Correct	Freq- uency	Cum Freq	PR	PCT	
... No examinees below this score ...					
3	0	0	1	0	
4	2	2	4	4	####
5	3	5	10	6	#####
6	3	8	17	6	#####
7	10	18	38	21	#####>
8	13	31	65	27	#####>
9	8	39	81	17	#####>
10	2	41	85	4	####
11	2	43	90	4	####
12	4	47	98	8	#####
13	1	48	99	2	##
14	0	48	99	0	

-----+-----+-----+-----+-----+
5 10 15 20 25
Percentage of Examinees

Рис. 3. Фрагменти вихідного файлу в ITEMAN

Аналізуючи вихідний файл, можна зробити висновки, що група непогано впоралась із завданнями, більшість відповіла правильно на 7 - 10 питань із 14 запропонованих. Загальні статистичні показники тесту задовільні, однак аналіз характеристик кожного завдання дозволяє виявити ті з них, які не

Seq. No.	Scale -Item	Item Statistics			Alternative Statistics					
		Prop. Correct	Disc. Index	Point Biser.	Alt.	Prop. Total	Endorsing Low	High	Point Biser.	Key
13	0-13	.15	.07	-.17	A	.15	.11	.18	-.17	*
CHECK THE KEY										
a was specified, c works better										
					B	.33	.50	.18	-.49	
					C	.25	.17	.41	-.11	?
					D	.21	.17	.18	-.18	
					E	.06	.06	.06	-.13	
					other	.00	.00	.00		
14	0-14	.54	.49	.29	A	.54	.28	.76	.29	*
					B	.00	.00	.00		
					C	.02	.00	.06	-.01	
					D	.06	.00	.06	-.09	
					E	.38	.72	.12	-.68	
					other	.00	.00	.00		

Рис. 4. Фрагмент вихідного файлу в ITEMAN

коректно працюють у тесті і потребують вдосконалення.

Наприклад, завдання 13 (рис.4) має від'ємну кореляцію з усім тестом та низький показник дискримінуючої здатності. Слабкі студенти відповідали на це питання краще, ніж сильні

(можливо вгадували). Потрібно перевірити, чи не було помилки при введенні ключа, оскільки більшість опитаних обирала відповідь *c*) замість правильної *a*). Якщо незадовільні характеристики завдання не пояснюються механічними помилками, то таке завдання з тесту потрібно вилучити.

Програма RASCAL дозволяє аналізувати тест у рамках моделі Раша. Інколи її називають однопараметричною моделлю IRT. Хоча формально це так, але шлях розвитку моделі Раша та її ідеологія мають принципові відмінності від інших моделей. Г.Раш вважав, що якщо емпіричні дані не відповідають його теорії, то вони недостовірні і їх потрібно відкинути або вдосконалити. Введення ж дво- та трипараметричної моделі мало на меті вдосконалення саме теорії, щоб вона точніше описувала емпіричні дані. У RASCAL за ідеологією моделі Раша досліджується відповідність емпіричних даних моделі, чого немає у інших програмах даного пакету. Відповідно до основних припущень IRT ймовірність того, що опитаний правильно відповість на питання тесту, залежить від латентних (прихованих) параметрів рівня підготовленості *i*-ої особи θ_i та складності *j*-го завдання b_j :

$$P_{ij} = c_j + \frac{(1 - c_j)}{1 + e^{-Da_j(\theta_i - b_j)}},$$

де a_j - параметр дискримінуючої здатності завдання (у моделі Раша всі завдання мають однакову дискримінуючу здатність $a_j = 1$) c_j - параметр угадування ($c_j = 0$ для моделі Раша та двопараметричної моделі). Для оцінки латентних параметрів рівня складності завдань та рівня підготовленості опитаних у RASCAL використовується метод JML з деякими уточненнями для кращого узгодження з трипараметричною моделлю. Програма працює лише з дихотомічними даними, всі пропущені питання опрацьовуються як питання з неправильною відповіддю. Програма має два головних вікна **Files** та **Options**, у яких формуються усі необхідні робочі файли та визначаються опції аналізу. Є можливість обирати початок відліку на шкалі Раша: за середнім рівнем підготовки або за середньою складністю завдань. Функціональну залежність

можна вибрати у вигляді логістичної або нормальної кривої. Можна задати максимальну кількість ітерацій, внести поправку для невеликих вибірок, задати довільне лінійне перетворення

Sorted in Item Difficulty order					
Item	Difficulty	Std. Error	Chi Sq.	df	Scaled Diff
3	-3.331	0.906	2.683	7	70
1	-2.240	0.569	2.610	7	80
5	-1.694	0.467	3.318	7	85
11	-0.299	0.331	6.014	7	97
8	-0.100	0.323	5.115	7	99
9	-0.004	0.320	9.331	7	100
6	0.275	0.314	4.498	7	103
14	0.366	0.313	3.606	7	103
10	0.456	0.312	8.353	7	104
4	0.456	0.312	7.564	7	104
12	1.009	0.319	4.074	7	109
2	1.205	0.325	2.740	7	111
7	1.412	0.334	4.657	7	113
13	2.487	0.418	9.649	7	123

Рис.5. Фрагмент другої сторінки вихідного файлу RASCAL

шкали логітів. Вихідний текстовий файл, крім оцінок параметрів тесту (рис.5), містить графічне зображення шкали з відміченими значеннями латентних параметрів підготовленості та складності завдань (рис.6), графіки характеристичної та інформаційної функцій тесту. Оцінки рівня підготовленості кожного опитаного можна вивести у окремому файлі, який потрібно задати у полі *Examinee Scores* вікна *Files*.

На рис.5 крім оцінок складності завдань (другий стовпчик) є значення статистики χ^2 (четвертий стовпчик), яка вказує на відповідність емпіричних даних побудованій теоретичній моделі. Тут на рівні значущості 0,05 для 7 ступенів свободи можна стверджувати, що завдання не відповідатиме моделі, якщо значення χ^2 перевищуватиме 14,1. Як бачимо, жодне із завдань не потрапило у критичну область. За шкалою логітів на рис.6 можна визначити співвідношення між рівнем підготовленості опитаних та складністю запропонованих завдань. Можна зробити висновок, що тест майже збалансований, складність більшості завдань відповідає рівню підготовленості. Три найлегші завдання з тесту

складність. Можна вдосконалити тест, додаючи легкі завдання зі складністю у межах (-2) – (-1) логіти та більш складні завдання зі складністю 1 - 3 логіти.

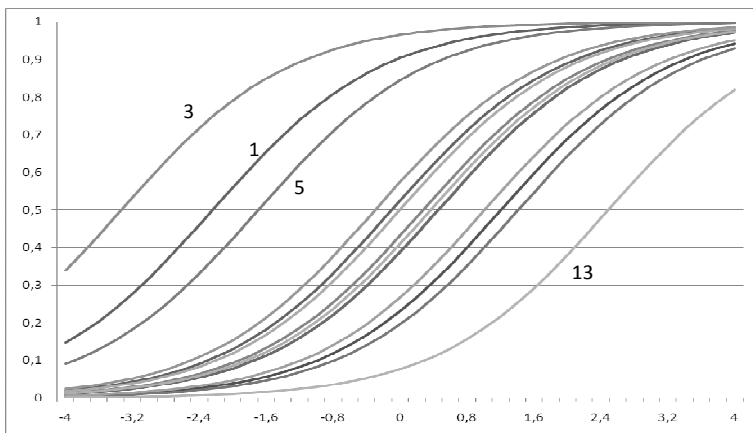


Рис. 7. Характеристичні криві завдань у моделі Раша

Проаналізувати результати тесту за дво- та трипараметричною моделями теорії IRT з урахуванням дискримінуючої здатності кожного завдання та угадування можна за допомогою програми XCALIBRE пакету ITPAR. Вона використовує маргінальний метод максимальної вірогідності (MML) для оцінки латентних параметрів обох моделей [7, ст. 48], що робить можливим її застосування для аналізу коротких тестів (менше 25 питань) та невеликих масивів опитаних (менше 1000 осіб). Програма працює лише з дихотомічними даними, які потрібно підготувати, як на рис.1. Для аналізу тут суттєво, чи тестований пропустив завдання, чи не встиг відповісти. Призначення вікон **Files** та **Options** тут таке ж, як у попередніх програмах. При визначенні опцій аналізу потрібно вибрати тип моделі, початкові розподіли для оцінок параметрів та кількість ітерацій. Якщо генерується файл з оцінками опитаних, то при запуску опції *Analyze* потрібно вибрати метод побудови оцінки латентного параметра рівня підготовленості (за замовчуванням це *Maximum-*

Likelihood).

Вихідний текстовий файл містить оцінки параметрів складності (*b*), дискримінуючої здатності (*a*) та угадування (*c*) для кожного завдання (рис.8), їх похибки, аналіз запитань та альтернатив, графіки характеристичної (рис.9) та інформаційної (рис.10) функцій тесту. Якщо обрано двопараметричну модель, то параметр $c = 0$.

З вихідного файлу для трипараметричної моделі (рис.8) бачимо, що завдання 3 було вилучено з аналізу як не інформативне, оскільки майже всі опитані дали вірну відповідь, крім однієї пропущеної. У стовпчику *Flg* для проблемних питань можлива поява міток: *P* – якщо значення оцінюваних параметрів виходять за межі моделі ($a < 0,3$, $b < -2,95$ або $b > 2,95$, $c > 0,4$); *K* – якщо деяка з альтернатив має кращу кореляцію з тестом, ніж

FINAL ITEM PARAMETER ESTIMATES										
Item	Lnk	Flg	a	b	c	Resid	PC	PBs	PBt	N
1			0.72	-2.41	0.26	0.35	0.94	0.19	0.35	48
2			0.74	1.86	0.25	0.45	0.36	0.45	0.25	48
3	--	Deleted	--	--	--	--	--	--	--	--
4			0.71	0.71	0.26	0.13	0.53	0.57	0.42	48
5			0.72	-1.90	0.26	0.13	0.90	0.36	0.45	48
6			0.69	1.05	0.32	0.61	0.57	0.33	0.12	48
7			0.75	1.78	0.22	0.39	0.31	0.55	0.43	48
8			0.72	-0.13	0.25	0.16	0.65	0.30	0.56	48
9			0.72	0.00	0.25	0.20	0.64	0.45	0.53	48
10			0.72	0.54	0.24	0.16	0.53	0.47	0.56	48
11			0.66	-0.14	0.28	0.49	0.69	0.15	0.29	48
12			0.71	2.41	0.30	0.38	0.40	0.14	-0.02	48
13			0.75	3.00	0.19	0.64	0.15	-0.00	-0.16	48
14			0.72	0.70	0.27	0.34	0.54	0.52	0.39	48

Рис.8. Фрагмент вихідного файлу в XCALIBRE

правильна відповідь (невірний ключ); *R* – якщо статистика відповідності моделі (стовпчик *Resid*) перевищує 2,0 [7, ст. 65]. У стовпчику *PC* маємо відсоток тих, хто правильно відповів на всі питання серед тих, хто дав вірну відповідь на дане питання.

Наступні стовпчики містять інформацію про кореляцію відповідей на кожне завдання з відповідями за весь тест (*PBs*) та з отриманими оцінками латентних параметрів (*PBt*). І у цій моделі завдання 13 є найскладнішим ($b = 3$), воно погано корелює з усім

тестом ($PBs = 0$), але не за рахунок угадування ($c = 0,19$). Шукаючи помилку у цьому завданні, потрібно врахувати, що об'єм вибірки надто малий для трипараметричної моделі і феномен 13-го завдання може пояснюватись зовсім іншими причинами. Аналізуючи характеристичні (рис.8) та інформаційні (рис.9) функції тесту, приходимо до висновку, що врахування параметра угадування значно зменшує роздільну здатність тесту та його інформативність.

Для слабкої групи опитаних результати, швидше за все, досягнуто за рахунок угадування. Тест найбільш інформативний саме для сильної групи, рівень підготовленості якої дорівнює або більше 1 логіта.

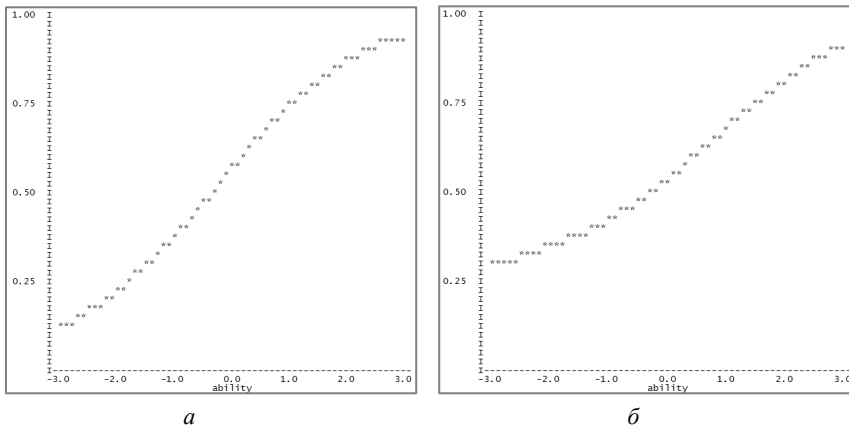


Рис.9. Характеристичні функції тесту для 2-параметричної (а) та 3-параметричної (б) моделі

Наступні три програми пакету ІТАР на основі попередньо знайдених статистичних характеристик тесту дозволяють оцінити його надійність (TESTINFO), валідність (TESTVAL) та порівняти оцінки рівня підготовленості, отримані за різними методиками (SCOREALL).

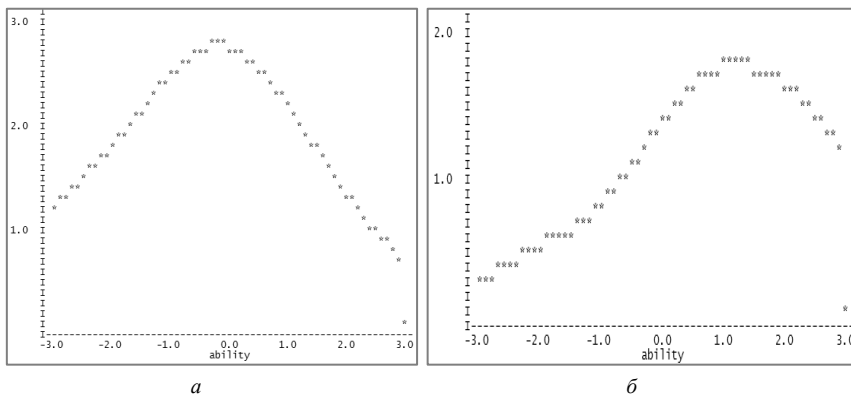


Рис.10. Інформаційні функції тесту для 2-параметричної (а) та 3-параметричної (б) моделі

Завдяки технічному оснащенню створювати та застосовувати тести у своїй роботі може кожен викладач, але недбало створені, неякісні тести тільки дискредитують технологію тестування. Як усі технології, що базуються на досягненнях науки, тестування вимагає серйозної підготовки тих, хто планує створювати та застосовувати педагогічні тести, передусім оволодіння строгими правилами та математичним апаратом, що застосовується в області педагогічних вимірювань. Мало просто створити тестові завдання, потрібно вміти на основі математико-статистичної обробки результатів тестування побачити об'єктивну картину рівня підготовленості учнів, намітити шляхи вдосконалення тесту для перетворення його на якісний інструмент педагогічного вимірювання. І на цьому шляху не можна нехтувати жодними засобами аналізу тестових завдань, навіть якщо відсутні серйозні програмні продукти. Так, на початковому етапі обробки результатів комп'ютерного тестування можна використовувати деякі характеристики завдань за класичною теорією, які пропонує модуль Moodle «Аналіз питання».

На прикладі підсумкового тесту з дисципліни «Теорія керування» для студентів фізико-математичного факультету дослідимо деякі класичні характеристики завдань, які пропонує

Moodle, та порівняємо їх із поглибленим дослідженням засобами IRT за допомогою програми WINSTEPS з метою вдосконалення тестових завдань та формування банку завдань за рівнем складності для проведення поточного та підсумкового контролю. Комп'ютерне тестування проводилось у середовищі Moodle. 38 студентів виконували 14 тестових завдань з випадковим порядком відповідей. Завдання полягали у виборі однієї або кількох правильних відповідей та у встановленні правильної відповідності. Кожне правильно виконане завдання оцінювалось в 1 бал, а неправильно – 0 балів. Тест розрахований на 30 хвилин. Оцінка за виконання тесту була складовою накопичених студентом балів і переводилась системою у шкалу від 0 до 20.

На початковому етапі обробки результатів тестування аналізуємо основні характеристики завдань у рамках класичної тестової теорії, які пропонує модуль Moodle «Аналіз питань»

Огляд Переоцінки Ручне оцінювання Аналіз питань									
Таблиця аналізу питань									
Сторінка: 1 2 (Далі)									
№ пит.	часткова оцінка	Кільк. відп.	% відповідей	% правильн. відповідей	Станд. пристр.	Дискр. індекс	Дискр. коеф.		
(1816)	(0.00)	37/38	(97%)	3 %	0.162	-0.07	-0.16		
	(1.00)	1/38	(3%)						
	(0.00)	0/38	(0%)						
	(0.00)	0/38	(0%)						
(1815)	(1.00)	18/38	(47%)	47 %	0.506	0.71	0.51		
	(0.00)	8/38	(21%)						
	(0.00)	9/38	(24%)						
	(0.00)	3/38	(8%)						
(1809)	(0.00)	0/38	(0%)	53 %	0.506	0.79	0.57		
	(1.00)	0/38	(0%)						
	(0.00)	0/38	(0%)						

Рис.11. Фрагмент таблиці аналізу питань

(рис.11). Для кожного питання можна бачити кількість студентів, які обирали ту чи іншу відповідь, відсоток правильних відповідей, обчислені стандартне відхилення, коефіцієнт дискримінативності та коефіцієнт бісеріальної кореляції.

Стандартне відхилення оцінює, наскільки різняться між собою відповіді різних студентів. Якби всі студенти, відповідаючи на деяке питання, набрали однакову кількість балів, то стандартне

відхилення дорівнювало б нулю. Стандартне відхилення обчислюється системою як квадратний корінь з суми квадратів відхилень оцінок від їх середнього значення, поділеної на кількість цих оцінок. Коефіцієнт дискримінативності характеризує якість тестового питання, його диференційну здатність відрізнити сильних студентів від слабких. Для розрахунку цього показника система за результатами відповідей на всі питання тесту поділяє студентів на три рівні групи: сильні, середні й слабкі студенти. Коефіцієнт дискримінативності для кожного питання обчислюється як доля правильних відповідей студентів сильної третини мінус доля правильних відповідей студентів слабкої третини. Він може набувати значення в діапазоні від +1 до -1. Наприклад, +1 означає, що на дане питання всі сильні студенти дали правильну, в усі слабкі – неправильну відповідь. Нульове значення коефіцієнта дискримінативності говорить про те, що на дане питання сильні і слабкі студенти відповіли однаково.

На питання, що має від'ємний коефіцієнт дискримінативності, слабкі студенти, які часто відповідають навмання, відповіли краще, ніж сильні студенти, схильні давати більш осмислені відповіді. Таке питання, швидше за все, містить помилку і потребує детального аналізу. Згідно з класичною тестовою теорією ефективним вважається питання, для якого коефіцієнт дискримінативності більший або дорівнює 0,4; від 0,3 до 0,39 – питання задовільне; від 0,2 до 0,29 – потребує аналізу на придатність у тесті; менше ніж 0,19 – вилучається або переробляється. Коефіцієнт бісеріальної кореляції теж є показником диференційної здатності завдання, який виражає ступінь статистичного зв'язку між результатами відповідей на дане завдання з індивідуальними балами всіх тестованих. У системі Moodle цей показник обчислюється за формулою $r = \frac{\sum_i \sum_j x_i y_j}{N s_x s_y}$, де $\sum_i \sum_j x_i y_j$ - сума добутоків відхилень набраних балів для даного питання і тесту в цілому, N - кількість відповідей на дане питання, s_x - стандартне відхилення набраних балів у цьому питанні, s_y - стандартне відхилення набраних балів для тесту в цілому. Цей показник також

набуває значень у діапазоні від -1 до +1. Так само, додатні значення свідчать, що на це питання краще відповіли сильні студенти, від'ємне – що на нього краще відповіли слабкі. Значення коефіцієнта бісеріальної кореляції, що дорівнює або більше за 0,3, є задовільним показником якості завдання.

Значення отриманих в Moodle показників усіх 14 завдань тесту з «Теорії керування» наведено у таблиці:

№ пит.	Відсоток прав.від.	Стандар. відхил.	Коеф. дискримін.	Коеф. бісер.кор.
1	26	0,45	0,57	0,59
2	87	0,34	0,79	0,56
3	50	0,28	0,38	0,14
4	67	0,34	0,71	0,62
5	74	0,45	0,71	0,35
6	98	0,11	0,62	0,20
7	53	0,51	0,79	0,58
8	58	0,50	0,14	-0,35
9	68	0,47	0,93	0,67
10	55	0,50	0,86	0,59
11	61	0,50	0,86	0,55
12	58	0,50	0,86	0,56
13	47	0,51	0,71	0,51
14	3	0,16	-0,07	-0,16

Питання 8 та 14 мають від'ємне значення коефіцієнта бісеріальної кореляції, а питання 3 та 6 - менше за 0,3. Всі інші питання мають задовільні значення усіх показників. Несподіваною виявилася реакція студентів на питання **14**, де потрібно вибрати одну правильну відповідь:

Для системи $\dot{x}(t) = A(t)x(t) + B(t)u(t)$ матриця

$$\{S_1(t), S_2(t), \dots, S_n(t)\}, \text{ де } S_1(t) = B(t), S_n(t) = A(t)S_{n-1}(t) - \frac{d}{dt}S_{n-1}(t)$$

називається матрицею:

- a.** імпульсних перехідних функцій;
- b.** фундаментальною;
- c.** керованості;
- d.** спостережуваності.

Всі студенти, за винятком одного, обрали відповідь (d) при правильній відповіді (c), оскільки вибирали відповідь за позначенням матриці буквою S , а не за способом її визначення. Студент, який дав правильну відповідь, витратив на виконання тесту 5 хвилин замість 30, тобто цю відповідь він міг вгадати. Дане питання можна залишити у тесті, якщо позначити матрицю керованості іншою буквою (відмінною від K , Φ та S).

У питанні 8 потрібно було також вказати одну правильну відповідь:

У задачі $\int_0^4 (x'^2 + x) dt \rightarrow \inf, |x'| \leq 1, x(4) = 0$ функція

Понтрягіна має вигляд:

- a.** $H = -tu + u^2$
- b.** $H = tu + u^2 + x$
- c.** $H = -tu - u^2$
- d.** $H = tu - u^2 - x$

Питання виявилось складним для сильних студентів, які намагалися не вгадувати правильну відповідь, а пройти усі необхідні кроки для побудови функції Понтрягіна: ввести керування, записати функцію Лагранжа та умови трансверсальності, розв'язати рівняння Ейлера та визначити множники Лагранжа. Слабкі ж студенти, не витрачаючи часу, ставили позначку навмання (деякі навіть вгадали). Таке громіздке питання недоцільно залишати у тесті з обмеженим часом виконання. Замість нього можна сформулювати кілька питань, які б контролювали вміння студентів на кожному кроці, необхідному для одержання виразу для функції Понтрягіна.

У завданні 3 потрібно вибрати усі можливі правильні відповіді (a, c, d):

При зведенні загальної задачі Больца до термінального виду з'являються:

- a.** нові диференціальні зв'язки
- b.** додаткові обмеження на керування
- c.** нові фазові координати
- d.** додаткові термінальні обмеження

Багато студентів вибрали не всі правильні відповіді, а лише деякі з них, певно, вважаючи, що правильних відповідей не може бути більше половини. Дане питання можна залишити у тесті, додавши ще варіанти відповідей. Аналізуючи питання б, приходимо до висновку, що його можна залишити у тесті у початковому вигляді як одне із простих питань і використовувати для перевірки залишкових знань.

Обробку результатів тестування у рамках сучасної теорії IRT можна провести за допомогою програми WINSTEPS (або її версія MINISTEP). Програма дозволяє отримати оцінки латентних параметрів рівня підготовленості θ та складності завдань β у рамках моделі Раша, головну ідею якої можна сформулювати таким чином: ймовірність P правильної відповіді на i -те питання тесту є функцією підготовленості тестованого і залежить від рівня підготовленості θ та складності цього завдання β_i . Для однопараметричної моделі Раша аналітичний вираз для даної функції:

$$P_i(\theta) = (1 + \exp(\theta - \beta_i))^{-1}.$$

Графік цієї залежності називається характеристичною кривою i -го завдання тесту (ICC). Аналогічно вважається, що ймовірність правильного виконання завдання j -м учасником тестування з рівнем підготовки θ_j залежить від складності завдання β і має вигляд $P_j(\beta) = (1 + \exp(\theta_j - \beta))^{-1}$. Одиницею інтервальної шкали Раша є логіт $d = \ln(P/(1-P)) = \theta - \beta$. Вимірювання у шкалі Раша залежать лише від співвідношення між рівнем підготовленості учасників тестування та складністю завдань, і не залежать від вибірки тестованих та набору завдань. Значення $d = 0$ відображає ситуацію, коли рівень підготовленості відповідає складності завдання, а ймовірність правильної відповіді на питання дорівнює $1/2$. Тобто значення $d = 0$ означає, що 50% тестованих

відповідають правильно на питання, а 50% – неправильно. А, наприклад, значення $-2,94$ відповідає «простому» питанню, на яке 95% тестованих відповідають правильно, а лише 5% - неправильно. Навпаки, значення $+2,94$ відповідає «складному» питанню, на яке лише 5% відповідають правильно.

Дані тестування з «Теорії керування», отримані в Moodle, конвертувались в Excel, а потім у програму MINISTEP. Результати конвертування вхідних даних у вимірювання Раша для всіх завдань тесту знаходяться у звіті ITEM STATISTICS (меню *Output Tables, 13. Item: measure*) (рис.12).

У першому стовпчику вказані номери завдань, у другому – кількість правильних відповідей, у третьому – кількість усіх відповідей. Результати вимірювання складності завдань у логітах (стовпчик MEASURE) наведені у порядку спадання (наприклад, завдання 14 майже на 3 логіти складніше, ніж завдання 1 і т.д.). У стовпчику MODEL S.E. наведена похибка вимірювання на основі моделі Раша, а у рядках MEAN та S.D. – середні значення та стандартні відхилення для значень у відповідних стовпчиках.

У стовпчиках INFIT та OUTFIT знаходяться параметри, що

ITEM STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
14	1	38	4.63	1.04	1.13	.4	4.21	1.8	-.12	.17	97.4	97.4	I14
1	10	38	1.69	.41	.68	-1.6	.63	-.9	.64	.42	92.1	78.2	I1
4	14	38	1.07	.38	.88	-.7	.75	-.9	.55	.45	73.7	72.9	I4
13	18	38	.52	.37	.85	-1.1	.73	-1.2	.59	.45	68.4	68.9	I13
7	20	38	.25	.37	.79	-1.6	.68	-1.5	.62	.45	81.6	68.1	I7
10	21	38	.12	.37	.89	-.8	.75	-1.1	.55	.45	63.2	68.6	I10
8	22	38	-.02	.37	1.99	5.6	3.20	5.9	-.45	.45	34.2	69.0	I8
12	22	38	-.02	.37	.96	-.3	.85	-.6	.50	.45	65.8	69.0	I12
11	23	38	-.15	.37	.95	-.3	.80	-.7	.50	.45	57.9	70.2	I11
9	26	38	-.39	.39	.74	-1.5	.60	-1.4	.64	.43	76.3	74.6	I9
3	27	38	-.74	.40	1.17	.9	1.58	1.5	.24	.43	73.7	76.5	I3
5	28	38	-.91	.41	1.06	.3	.85	-.3	.41	.42	71.1	78.2	I5
2	33	38	-1.95	.52	.71	-.8	.63	-.4	.54	.35	89.5	87.7	I2
6	37	38	-3.89	1.05	1.04	.3	.38	-.1	.24	.19	97.4	97.4	I6
MEAN	21.6	38.0	.00	.49	.99	-.1	1.19	.0			74.4	76.9	
S.D.	8.8	.0	1.82	.23	.31	1.8	1.07	1.9			16.4	9.9	

Рис.12. Характеристики завдань

характеризують відповідність даних моделі Раша. Значення MNSQ (*mean-square statistic*) характеризують рівень випадковості результатів або невідповідність даних моделі вимірювання. Найбільш очікувані значення MNSQ знаходяться поблизу 1. Великі значення MNSQ OUTFIT пов'язують з угадуванням відповідей, а великі MNSQ INFIT інтерпретуються як показник низької валідності завдань. Найбільш якісними і відповідними вважаються значення MNSQ у межах від 0,5 до 1,5. Значення більші за 1,5 вказують на невизначеність та «шум» у вхідних даних, значення менші за 0,5 теж небажані, бо свідчать про «інформаційну перевантаженість» питання. Аналіз починають із питань з високим значенням MNSQ.

Наприклад, для питання 8 маємо значення MNSQ INFIT 1,99, що дає підстави вважати його непридатним для цілей тестування, а високий показник OUTFIT 3,2 свідчить про намагання студентів вгадати відповідь. Питання 14 при задовільній валідності (1,13) має велике значення OUTFIT 4,21 завдяки ймовірному вгадуванню. Критичним є показник MNSQ і для питання 3. Це саме ті питання, що були виявлені у результаті аналізу за класичною теорією. У полі ZSTD наводяться

PERSON STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL			MODEL S.E.	INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		PERSON
	SCORE	COUNT	MEASURE		MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	085%	EXP%	
6	12	14	2.54	.98	1.01	.3	1.02	.5	.52	.54	92.9	91.3	P6
7	12	14	2.54	.98	1.01	.3	1.02	.5	.52	.54	92.9	91.3	P7
11	12	14	2.54	.98	1.01	.3	1.02	.5	.52	.54	92.9	91.3	P11
26	6	14	-.44	.63	1.07	.4	.84	.2	.47	.48	50.0	71.6	P26
27	6	14	-.44	.63	1.22	1.0	9.90	4.2	.17	.48	64.3	71.6	P27
36	6	14	-.44	.63	.93	-.2	.73	.0	.53	.48	64.3	71.6	P36
3	4	14	-1.30	.70	1.03	.2	.76	.2	.48	.47	71.4	79.8	P3
4	4	14	-1.30	.70	1.49	1.3	1.78	.9	.24	.47	71.4	79.8	P4
29	4	14	-1.30	.70	1.03	.2	.76	.2	.48	.47	71.4	79.8	P29
20	3	14	-1.85	.78	1.06	.3	.82	.2	.46	.47	78.6	84.5	P20
18	2	14	-2.57	.93	.96	.2	1.03	.4	.44	.45	92.9	89.9	P18
MEAN	7.9	14.0	.39	.70	1.00	.0	1.13	.3			74.4	76.9	
S.D.	2.5	.0	1.18	.12	.20	.7	1.47	.7			13.6	7.3	

Рис.13. Фрагмент характеристик учасників тестування

стандартизовані значення MNSQ. Прийнятними є значення від -2 до $+2$. І за цим показником завдання **8** повинно вилучатися із подальшого аналізу.

У стовпчику PT-MEA CORR. наводяться значення коефіцієнта кореляції. Як і раніше, привертають увагу питання **14**, **8**, **3** та **6** через свою неузгодженість із результатами тесту в цілому.

Аналогічний звіт PERSON STATISTICS можна отримати для всіх 38 учасників тестування (меню *Output Tables, 17. Person: measure*) (рис.13).

При всіх задовільних значеннях більшості показників насторожує значення 9,9 показника MNSQ OUTFIT для учасника під номером 27. Виявляється, що це саме той студент, якому вдалося за 5 хвилин дати кілька правильних відповідей, що майже напевне вказує на вгадування.

У MINISTEP можна отримати різні графічні звіти: характеристичні криві, інформаційні функції тощо. На рис.14 побудовані характеристичні криві усіх 14 завдань тесту, аналіз

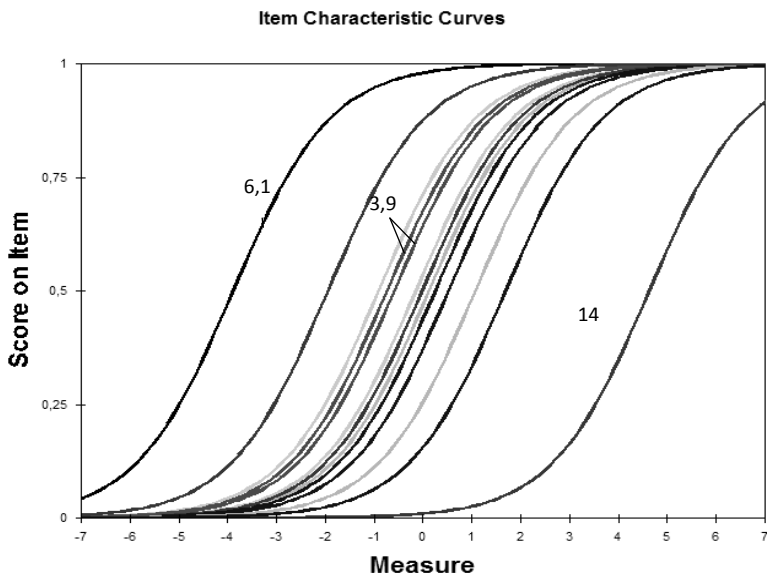


Рис.14. Характеристичні криві завдань у MINISTEP

взаємного розміщення яких допомагає вдосконалити тест як систему завдань зростаючої складності.

У даному тесті більшість завдань зосереджені в області середньої та меншої за середню складності. Потрібно додати складніші завдання, щоб рівномірно заповнити область від 0 до 5 логітів. Тим паче, що після коректування завдання 14 воно може виявитися не самим складним.

Характеристичні криві для завдань 6 та 12 накладаються, тому одне з них можна вилучити. Завдання, які мають дуже близькі характеристичні криві (наприклад, 3 та 9), потрібно аналізувати на їх доцільне включення до одного тесту. Можливо такі завдання можуть бути використані для паралельних тестів. Для ідеального тесту характеристичні криві повинні рівномірно заповнювати весь інтервал $(-5; +5)$.

Аналізувались графіки інформаційних функцій для всіх завдань та тесту в цілому (рис.15). Всі криві свідчать про задовільну інформативність завдань та тесту, крива на рис.15 має один чіткий максимум. Вдосконалити даний тест можна додаванням завдань з високою диференційною здатністю, щоб

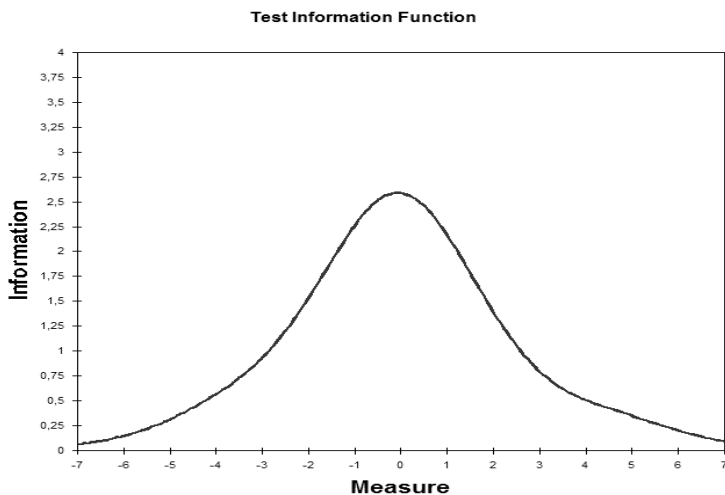


Рис.15. Інформаційна функція тесту

збільшити максимальне значення інформаційної функції та її крутизну.

Для демонстрації переваг IRT можна провести розрахунок складності завдань для групи сильних та слабких студентів (рис.16).

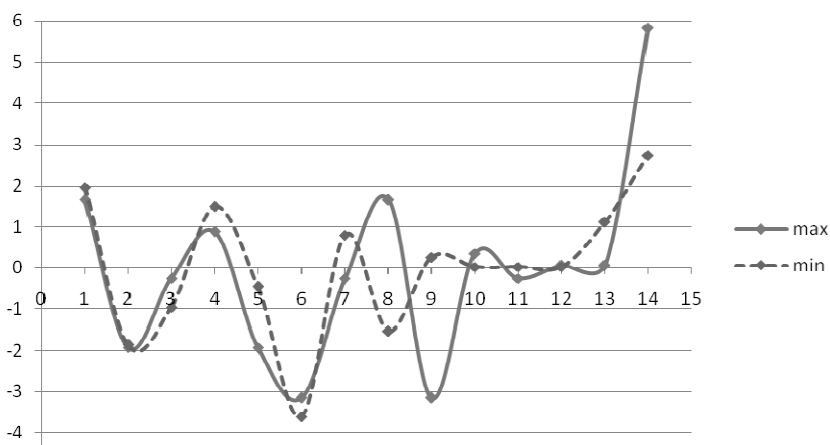


Рис.16. Порівняння складності завдань у двох групах

Теоретично складність завдань в обох групах повинна бути однаковою, однак, зважаючи на малий об'єм вибірки, результати порівняння можна визнати лише задовільними. Можливо, після коректування завдань 8, 9 та 14 даний тест зможе використовуватись для об'єктивного оцінювання знань студентів у групах з різним рівнем підготовленості.

Запропонований аналіз тесту у рамках класичної та сучасної теорій виявляє схожі закономірності та проблеми, хоча вимірювання за класичною схемою та у моделі Раша мають різну внутрішню логіку та різні обчислювальні процедури. Саме тому при оцінюванні надійності педагогічних вимірювань варто не замінювати, а доповнювати ці два підходи один одним.

Литература

1. Алгина Дж., Крокер Л. Введение в классическую и современную теорию тестов: учебник. – М.: Логос, 2010. – 668 с.
2. Дубина И.Н. Математические основы эмпирических социально-экономических исследований. – Барнаул: Изд-во Алт.ун-та, 2006. – 263 с.
3. Ким В.С. Тестирование учебных достижений. Монография. – Уссурийск: Издательство УГПИ, 2007. – 214 с.
4. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. – М.: Прометей, 2000. – 168 с.
5. Чельшкова М.Б. Теория и практика конструирования педагогических тестов. – М.: Логос, 2002. – 432 с.
6. Educational measurement: edited by Robert L. Brennan. – ACE/Praeger series on higher education, 2006. – 779 p.
7. User's manual for the Item and Test Analysis Package for Windows.– Assessment Systems Corporation, St.Paul MN, 2007.–94 p.
8. www.winsteps.com.
9. www.assess.com.

АНАЛІЗ МОДЕЛЕЙ ТА МЕТОДІВ ДІАГНОСТИКИ ЗНАНЬ

Малежик М.П., Сіткар Т.В.

Оцінювання знань в традиційному (вузькому) розумінні розглядається як визначення підсумкового рівня підготовки того, хто навчається у рамках деякої предметної галузі після проходження курсу навчання. Ключовим моментом цього твердження є нерозривність процесу контролю і процесу навчання. Це твердження правильне для традиційних форм навчання, альтернативних заочних і для нових форм, що розвиваються, таких як дистанційна освіта із застосуванням інформаційних технологій. Існують завдання оцінювання знань і поза учбовим процесом, які використовуються у виробничому середовищі. Основними з таких завдань є: атестація персоналу, проведення сертифікаційних іспитів, добір кандидатів на конкретні вакансії і т.п. Результатом оцінки знань, як правило, являється визначення деякої величини у рамках шкали оцінювання. Наслідки оцінювання можуть мати найрізноманітніші результати - від чисто морального ефекту, до визначення критичних висновків і ухвалення доленосного рішення.

У зв'язку з надзвичайною важливістю точності оцінювання знань необхідно підходити до цього явища як до процесу об'єктивного вимірювання і результати такого вимірювання опрацьовувати за стандартними математичними методами і супроводжувати стандартними характеристиками точності [6]. Увесь процес оцінювання знань повинен виконуватися у рамках науково-обґрунтованої методології, з практичною апробацією, що підтверджує правильність теорії.

Дослідження проблем оцінювання знань має кілька цілей:

максимізація об'єктивності оцінювання;

збільшення ефективності процесу оцінювання;

впровадження нових форм і способів оцінювання в сучасних (нетрадиційних) методичних системах навчання, атестації, сертифікації фахівців, і в інших сферах діяльності людей.

Тестування, як форма оцінювання знань, є одним з найбільш перспективних напрямів збільшення якості вирішення

поставлених завдань. Використання тестування в реальній діяльності дозволить підвищити детальність і точність оцінювання і, крім того, може являтися єдиною можливою формою перевірки знань у ряді ситуацій таких як:

перевірка знань висококваліфікованих фахівців, внаслідок дефіциту кваліфікованих контролюючих кадрів, складності організації екзаменаційної комісії, трудомісткості процесу контролю;

оцінювання знань в системах дистанційної освіти з використанням мережевих технологій Інтернет;

проведення масової атестації, сертифікації фахівців, які навчаються.

Використання відповідей, що вводяться у вільній текстовій формі, є найприроднішим і найбільш складним завданням при організації системи контролю знань. Використання подібних завдань має наступні переваги перед «традиційними» завданнями закритого типу :

збільшується точність і детальність оцінювання;

можна визначати правильність відповіді на поставлене питання з певною точністю;

виключається чинник випадкового вибору;

мінімізується можливість розсекречення.

Науковим напрямом, що лежить в основі моделювання і параметризації процесу тестування, є дві теорії: так звана «класична теорія тестування» і «сучасна теорія тестування», що дістала в іноземній літературі назву Item Response Theory (IRT). У основі теорії IRT лежать кілька математичних теорій, основні з яких: теорія ймовірності, математична статистика, дисперсійний і регресійний аналіз. При побудові тестуючих систем з відкритою формою відповіді у вигляді тексту, поданого природньою мовою (ПМ) використовуються: методи штучного інтелекту, методи лінгвістичного опрацювання текстів [2].

Отже, можна зробити висновок про актуальність робіт, спрямованих на побудову, дослідження і підтримку математичних моделей тестуючої системи заснованої на інтелектуальному опрацюванні відповідей випробовуваного, для дослідження усього

комплексу проблем автоматизації оцінювання знань.

При формуванні шкали оцінок досить часто зустрічається метод «проб і помилок». Тому реальні знання учня не отримують об'єктивного віддзеркалення і як наслідок - знижується стимулююча дія екзаменаційної оцінки на пізнавальну діяльність і якість навчального процесу в цілому.

У деяких моделях тестування оцінювання результатів здійснюється тільки за фактом правильності відповіді, тобто хід рішення в завданнях не перевіряється і не оцінюється. Такі, наприклад, закриті завдання з однозначною числовою відповіддю або бінарні тести [2].

Первинними відомостями при тестуванні знань є набраний бал випробовуваних або так званий первинний бал. Перевагою даної оцінки є її простота і наочність, Дійсно, чим більше завдань виконав випробований, тим вищий його бал.

Проте проблема полягає в тому, що первинний бал є не абсолютною, а відносною оцінкою. Він істотно залежить від важкості завдань тесту і на іншому тесті він може виявитися іншим, причому сама складності тесту у свою чергу визначається усім контингентом випробовуваних. Бажано мати об'єктивну оцінку рівня підготовленості випробовуваних, підтверджувану на різних тестах, що мають заздалегідь визначений рівень важкості завдань.

Другим істотним недоліком первинних балів є їх нелінійність відносно тих параметрів, які вони повинні характеризувати (рівень підготовленості). Зокрема, якщо тест складається з 100 завдань, то різниця в первинних балах $b_1 - b_2 = 86 - 82 = 4$ відповідає більшій відмінності в рівні підготовленості учасників, ніж та сама різниця для учасників тих, які мають, наприклад 23 і 19 балів. Порівнюючи первинні бали, необхідно розуміти, що первинні бали є лише індикатором підготовленості випробовуваних, а не її мірою.

Будь-які відомості для їх наступного застосування в завданнях тесту мають бути подані певним кількісним показником, розрахованим з використанням певної умовної одиниці.

Статистичне опрацювання результатів тестування на основі моделі Раша має важливі переваги, серед яких, необхідно відмітити наступні.

Модель Раша перетворює виміри, зроблені в дихотомічних і порядкових шкалах в лінійні виміри, в результаті якісні дані аналізуються за допомогою кількісних методів. Це дозволяє використовувати широкий спектр статистичних процедур.

Оцінка складності тестових завдань не залежить від вибірки випробовуваних, на яких була отримана, і оцінка рівня знань випробовуваних аналогічно не залежить від використовуваного набору тестових завдань. Пропуск даних для деяких комбінацій (випробовуваний ÷ тестове завдання) не є критичним. Крім того, модель Раша характеризується найменшим числом параметрів : один параметр рівня знань для кожного випробовуваного і тільки один параметр складності для кожного завдання.

Модель Раша спирається на поняття «складність завдання» і «рівень підготовленості». Так, одне завдання вважається складнішим, ніж інше, якщо ймовірність правильної відповіді на перше завдання менша, ніж на друге, незалежно від того, хто їх виконує. Аналогічно, підготовлений студент має велику ймовірність правильно відповісти на усі завдання, ніж менш підготовлений.

Завдяки простій структурі моделі існують зручні обчислювальні процедури для перевірки адекватності моделі : для усього набору тестових результатів, для кожного випробовуваного, для кожного завдання і для кожної конкретної відповіді.

Розглянемо модель Раша детальніше. Нехай тест складається з K різних завдань бінарного типу, (той хто тестується отримує 1, якщо відповів правильно, і 0 при неправильній відповіді) і його виконують N - студентів. В результаті виходить матриця відповідей A_{NK} що складається з N - рядків (i) і K стовбців (j).

$$A_{NK} = (a_{ij})$$

Число b_i , рівне сумі балів в i – рядку, називається первинним балом i - випробовуваного (воно дорівнює числу його правильних відповідей) :

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

При необхідності первинний бал можна виразити у відсотках (чи долях) таким чином $\frac{b_i}{K} 100\%$. Рівні підготовленості учасників А і В позначимо через S_A і S_B , а складність завдань через t (насправді усі завдання мають різний рівень складності t_K). У моделі Раша доводиться, що:

$$\frac{S_A}{S_B} = \frac{p_{At} q_{Bt}}{q_{At} p_{Bt}} \quad (1)$$

$$q_{At} = 1 - p_{At}$$

$$q_{Bt} = 1 - p_{Bt},$$

де p_{At} і p_{Bt} - ймовірність виконання завдання рівня складності t відповідно учасниками А і В, q_{At} і q_{Bt} - ймовірність невиконання завдання рівня складності t відповідно учасниками А і В. Із загальних міркувань вираз (1) повинен бути правильним для будь-якого рівня складності завдань і будь-якої пари учасників тестування. Нехай, якесь завдання має складність $t=1$ і необхідно порівняти складності двох завдань. У моделі Раша рівень складності визначається, як відношення ймовірності (q_{1t}) того, що деякий стандартний учасник випробування з одиничним рівнем підготовки ($S=1$) не виконає це завдання, до ймовірності (p_{1t}) його виконання :

$$t = \frac{1 - p_{Bt}}{p_{Bt}} .$$

Одиничний рівень підготовки і одинична складність завдання в моделі Раша пов'язані між собою. Використовуючи вираз:

$$\frac{S_A}{S_B} = \frac{p_{A1} q_{B1}}{q_{A1} p_{B1}} = \frac{p_{A1}}{(1 - p_{A1})} \cdot \frac{(1 - p_{B1})}{p_{B1}} ,$$

і припустивши, що рівень підготовленості саме учасника В є одиничним ($S_B=1$), отримаємо наступний вираз:

$$S_A = \frac{p_{At} q_{Bt}}{q_{At} p_{Bt}} = \frac{p_{At}}{(1 - p_{At})} \cdot \frac{(1 - p_{Bt})}{p_{Bt}} = \frac{p_{At}}{(1 - p_{At})} \cdot t \quad (2)$$

Рівняння (2) зв'язує рівень складності деякого завдання і рівень підготовленості деякого учасника з вірогідністю правильного виконання завдання і повинно бути правильним для завдань будь-якого рівня складності. Враховуючи спільність(загальність) отриманого рівняння (2), можна показати, що ймовірність $P(S, t)$, того, що учасник з рівнем підготовки S правильно виконає завдання складності t , виражається наступною формулою:

$$P(S, t) = \frac{S}{S + t} = \frac{1}{1 + \frac{t}{S}} \quad (3)$$

Ймовірність $P(S, t)$ дістала назву функції успіху. Як видно з виразу (3), функція успіху залежить тільки від відношення t до S , тому модель Раша називається однопараметричною.

Вводючи нові змінні:

$$\ln S = \theta , \quad S = \exp(\theta) ,$$

$$\ln t = \delta, \quad t = \exp(\delta)$$

вираз (3) можна переписати у виді:

$$P(\delta, \theta) = \frac{1}{1 + \exp(\delta - \theta)} \quad (4)$$

Формула (4) є основним рівнянням однопараметричної логістичної моделі Раша, одиниця виміру δ і θ називається логітом. При одному логіті ($\delta_0=1$ і $\theta_0=1$) ймовірність успіху $P(\delta, \theta)=0,5$, тобто ймовірність виконання стандартного завдання стандартним учасником має дорівнювати 0,5 (див. рис. 1).

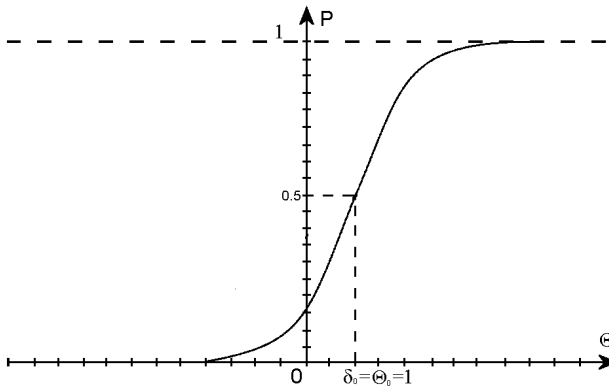


Рис. 1. Характеристична крива важкості завдання

Використовуючи модель Раша, можна зробити один дуже важливий висновок: чим вище рівень підготовки учасника, тим більше ймовірність виконання завдання будь-якого рівня складності. Варто відмітити, що параметри δ і θ називають латентними параметрами, оскільки вони не вимірюються безпосередньо в процесі тестування.

Функція успіху може бути отримана, виходячи з принципу максимуму інформації (H_o) про систему (мінімуму ентропії (I_o)) [4]. Для цього введемо наступні характеристики:

$$n_i = \frac{b_i}{K} = \frac{\sum_{j=1}^K a_{ij}}{K}$$

- середнє значення тестового балу учасника тестування за всією вибіркою завдань (K - число завдань в тесті), іншими словами середня успішність виконання усіх K завдань i -м випробовуваним;

$$n_j = \frac{c_j}{N} = \frac{\sum_{i=1}^N a_{ij}}{N}$$

- середнє значення балу завдання тесту за всією вибіркою випробовуваних (N - число учасників тестування), іншими словами середня успішність виконання j -м завдання усіма N випробовуваними.

Число b_i , рівне сумі балів в i -му рядку, називається первинним балом i -го випробовуваного (воно дорівнює числу його правильних відповідей) :

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

а число c_j , рівне сумі балів в k -му стовпці, називається первинним балом j -му завдання (воно дорівнює числу правильних відповідей на це завдання усіма випробовуваними) :

$$c_j = \sum_{i=1}^N a_{ij}.$$

Кількість різних станів системи (число способів розподілу 0 і 1 в матриці відповідей $A_{NK} = (a_{ij})$), при заданому значенні первинного балу j -го завдання c_j визначається числом комбінацій ($C_N^{c_j}$) по c_j з N :

$$C_N^{c_j} = \frac{N!}{c_j!(N-c_j)!},$$

а повне число станів системи W з урахуванням зміни j від 1 до K буде рівне:

$$W = \prod_{j=1}^K C_N^{c_j} = \prod_{j=1}^K \frac{N!}{c_j!(N-c_j)!}.$$

Інформаційна ентропія $I_e \propto -N \ln W$,

де

$$\ln W = \ln \left(\prod_{j=1}^K C_N^{c_j} \right) = \sum_{j=1}^K \ln \left\{ \frac{N!}{c_j!(N-c_j)!} \right\}.$$

Використовуючи формулу Стірлінга, можна отримати наступну формулу:

$$-\ln W = -\sum_{j=1}^K \ln \left\{ \frac{N!}{c_j!(N-c_j)!} \right\} = -\sum_{j=1}^K \left[N \ln N - c_j \ln c_j - (N-c_j) \ln(N-c_j) \right].$$

Щоб знайти розподіл, відповідний найбільшій статистичній вазі W , розглянемо комбінацію $-\delta(N \ln W)$, відповідну максимуму інформації (H_e) про систему (мінімуму ентропії (I_e)), :

$$-\delta(N \ln W) = -N \sum_{j=1}^K \left[-\frac{\ln c_j}{N} - \frac{1}{N} + \frac{\ln(N-c_j)}{N} + \frac{1-c_j}{N-c_j} \right] \delta c_j = -\sum_{j=1}^K \ln \frac{N-c_j}{c_j} \delta c_j = 0$$

чи:

$$\sum_{j=1}^K \ln \frac{N-c_j}{c_j} \delta c_j = 0.$$

Комбінації δc_j вибираються довільно, за винятком деякого їх числа рівного числу додаткових умов (множників Лагранжа). Усі комбінації δc_j можна розглядати як незалежні один від одного, а залежними від них величинами вважати множники Лагранжа. Вважатимемо одну з комбінацій $\delta c_j \neq 0$, а інші рівними 0. Тому до виразу

$$\sum_{j=1}^K \ln \frac{N - c_j}{c_j} \delta c_j = 0.$$

потрібно додати додаткові умови. В даному випадку є усього лише одна додаткова умова, яка зв'язує набраний індивідуальний первинний бал i -го випробуваного (b_i) з первинним балом j -го завдання, :

$$b_i = \sum_{j=1}^K \alpha_{ij} c_j,$$

де α_{ij} - множник, що визначає успішність виконання i -м випробуваним j -те завдання.

Індивідуальний бал i -го випробуваного є визначеним в результаті тестування, тому його варіація δc_j дорівнює 0.

$$\sum_{j=1}^K \alpha_{ij} \delta c_j = 0.$$

Таким чином:

$$\sum_{j=1}^K \left[\ln \frac{N - c_j}{c_j} + \alpha_{ij} \right] \delta c_j = 0.$$

З урахуванням того, що $\frac{N}{c_j} = \frac{1}{n_j}$ знаходимо:

$$n_j = \frac{1}{1 + \exp(\alpha_{ij})}.$$

Порівнюючи отриманий вираз з формулою:

$$P(\delta, \theta) = \frac{1}{1 + \exp(\delta - \theta)}$$

можна інтерпретувати n_j , як ймовірність успіху, тобто ймовірність виконання i -м учасником j -те завдання, а α_{ij} як $(\delta_j - \theta_i)$ різниця між складністю j -го завдання і рівнем підготовленості i -го учасника, виражену в логітах [8].

Проте, незважаючи на 40-річний досвід застосування цієї моделі в багатьох сферах, передусім в освіті і психології, досі тривають дискусії про істинну цінність і ефективність моделі Раша. Досі існують дві крайні точки зору на цю модель.

Найбільш переконані прибічники моделі Раша стверджують: чи «Можна зібрати або побудувати або сформулювати дані так, щоб вони відповідали визначенню виміру (моделі Раша)? Якщо ні, - то такі дані даремні».

Їх найбільш послідовні опоненти стверджують наступне: «Дані - це дані, а модель - це конструкція дослідника, яка схильна до помилок». Наприклад, при побудові регресії, не беручи до уваги ті або інші дані, можна отримати будь-яку залежність, але ми тим самим обмежуємо реальний світ даних. Таким чином, створюється штучна змінна, про яку обмежені.

Найбільш широко використовуваною моделлю діагностики знань нині є двохпараметрична модель Бірнбаума [4,5], що відноситься до теорії моделювання і параметризації педагогічних тестів або скорочено ТМППТ. У англomовній літературі ця теорія носить назву Item Response Theory (IRT). Принципова відмінність Item Response Theory від теорії Раша полягає в тому, що в даному випадку кожне тестове завдання розглядається, як самостійна структурна одиниця, параметри якої не залежать від параметрів інших завдань.

У основі усіх моделей Item Response Theory лежать так звані функції успіху, що мають заздалегідь відомий вигляд і що визначають залежність правильної відповіді на дане завдання від рівня підготовленості випробовуваного. Для оцінювання параметрів моделей Item Response Theory використовують спеціальні ітераційні методи, що є досить складними, проте ці моделі мають в порівнянні з моделлю Раша ряд переваг. Зокрема, для нанесення результатів на єдину шкалу за різними варіантами тесту немає необхідності вводити припущення про паралельність результатів. Більше того, можна звести воедино результати, отримані при використанні різних за складністю тестів. Моделі Item Response Theory менш чутливі до довжини тесту і рівня складності завдань. Слід зазначити, що дисперсія помилки оцінювання рівня підготовленості в моделі Раша дещо більше ніж в Item Response Theory.

Однією з основних і найбільш застосовних моделей Item Response Theory є так звана двохпараметрична модель Бірнбаума. Суть цієї теорії полягає в наступному [3]. Нехай параметр θ_i відображає рівень підготовленості i -го учасника випробувань, а параметри δ_j і d_j - деякі параметри j -го тестового завдання (θ_i і δ_j - будь-які числа, а d_j - тільки позитивні). Згідно моделі Бірнбаума ймовірність (p_{ij}) успішного виконання j -го завдання, i -м учасником визначається наступною формулою:

$$p_{ij} = \frac{1}{1 + \exp(-d_j(\theta_i - \delta_j))}.$$

Цей вираз показує, що ймовірність успіху p_{ij} залежить лише від d_j і $(\theta_i - \delta_j)$ тому модель Бірнбаума, дістала назву двохпараметричною. Ймовірність успіху p_{ij} як функція від θ_i при фіксованих значеннях $\delta_j = \delta_0$ і $d_j = d_0$ називається характеристичною функцією j -го завдання. На рис. 3 зображений

вид характеристичної кривої з параметрами $\delta_j = 0,3$ і $d_j = 1,2$. Дослідження характеристичних функцій показує, що:

$$\lim_{\theta_i \rightarrow -\infty} p(\theta_i, \delta_0, d_0) = 0,$$

$$\lim_{\theta_i \rightarrow +\infty} p(\theta_i, \delta_0, d_0) = 1.$$

Тобто із зменшенням рівня підготовленості учасника ймовірність правильно відповісти на завдання зменшується до 0, а при збільшенні рівня підготовленості збільшується до 1.

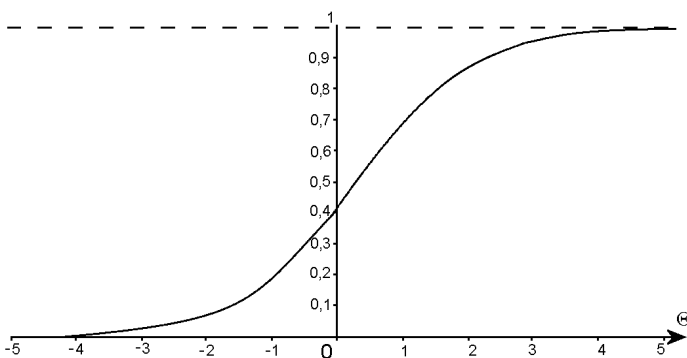


Рис. 2. Характеристична крива завдання з рівнем складності 0,3 і розрізняючою (диференціюючою) здатністю 1,2.

При збільшенні параметра δ_0 , що характеризує складність завдання, ймовірність успіху для одного і того рівня підготовленості зменшується. При невеликих значеннях параметра d_0 характеристична крива є пологою, отже, для випробовуваних з рівнем підготовленості $\theta_i < \delta_0$ і для учасників тестування з $\theta_i > \delta_0$ ймовірності правильної відповіді цього завдання приблизно рівні. Якщо значення параметра d_0 велике, то ймовірність успіху істотно відрізнятиметься, тому параметр d_j дістав назву коефіцієнта дискримінації. Цей коефіцієнт показує наскільки це завдання,

добре розрізняє учасників тестування з високим, і низьким рівнем підготовленості. На мал. 3 зображені характеристичні криві для завдань з однаковими параметрами складності $\delta = 1$, але з різними коефіцієнтами дискримінації $d_1 = 0,3$ і $d_2 = 2,8$. Характеристична крива, зображена суцільною лінією, відповідає завданню з коефіцієнтом дискримінації $d_1 = 0,3$, а пунктиром завданню з $d_2 = 2,8$. Розглянемо точку $\delta = 1$ на осі θ , з графіка видно, що для кривої з $d_1 = 0,3$ ймовірності виконання завдання учасниками з рівнями підготовленості $\theta < 1$ і з $\theta > 1$ розрізняються несуттєво. Для $\theta = 0$, відповідна ймовірність успіху дорівнює 0,43, а для $\theta = 2$, відповідна ймовірність дорівнює 0,57. Для характеристичної кривої завдання з $d_2 = 2,8$ ймовірність виконання завдання учасниками тестування з рівнями підготовленості $\theta < 1$ і з $\theta > 1$ відрізняються істотно. Для $\theta = 0$, відповідна ймовірність успіху дорівнює 0,06, а для $\theta = 2$, відповідна ймовірність дорівнює 0,94.[7]

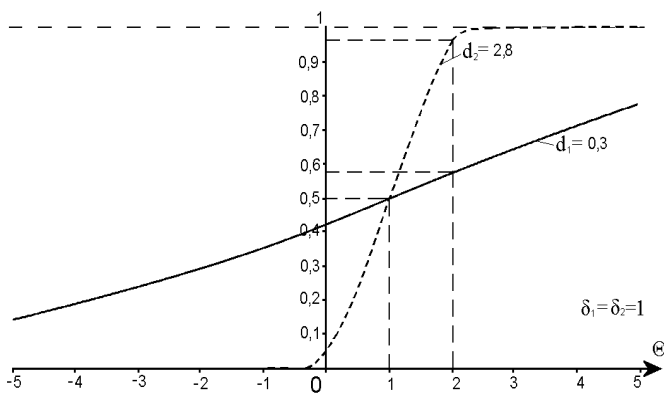


Рис. 3. Характеристичні криві завдання з однаковим рівнем складності $\delta = 1$ і з різними коефіцієнтами дискримінації d_j

Проаналізувавши моделі діагностики знань, ми прийшли до висновку, що серед усіх моделей найлегше реалізувати з

допомогою комп'ютера однопараметричну модель Раша. Якщо б модель Раша була б повністю адекватна результатам тестування, і були б відсутні помилки вимірів, то зміщення привели б до повного збігу значень, відповідних одному і тому ж значенню первинного балу для будь-яких варіантів тесту, що є на практиці маловірогідним.

Найбільш точною є Item Response Theory (двохпараметрична модель Бірнбаума), проте її досить важко реалізувати на практиці. Проте реалізація даної моделі дозволяє з максимальною точністю оцінювати знання та рівень підготовленості суб'єктів тестування. Крім того, у рамках двохпараметричної моделі первинні бали не є достатніми статистиками, тому випробовувані, які отримали однакові первинні бали в одному і тому варіанті тесту, можуть отримати різний остаточний бал.

Ситуація ще більш ускладнюється, якщо в процесі тестування використовувалися паралельні варіанти тесту. Для кожного з варіантів буде своя шкала оцінок рівнів підготовленості, рівнів складності завдань і їх диференціюючої здатності. І порівняння оцінок латентних параметрів з різних варіантів, без попереднього зведення до єдиної шкали є некоректним.

Але усі ці моделі не бездоганні.

Література

1. Аванесов, В.С. Методологические и теоретические основы тестового педагогического контроля: автореф. дис: д-ра пед. наук: 13.00.01 / В.С.Аванесов; -Петербургский гос. Уп-т.- СПб., 1994.
2. Аванесов, В.С. Научные основы тестового контроля знаний / В.С. Аванесов - М.: Иссл. центр, 1994. - 135 с.
3. Деревнина, АЛЮ. Системы тестирования в электронных учебниках / А.Ю. Деревнина, В.А. Семикнн // Информационные технологии. - 2002. - №5. - с. 39-44.
4. Интернет в гуманитарном образовании/ под ред. Е.С. Полат// Учеб. пособие для вузов-М.: Владос, 2001.-271 с.
5. Карпенко, Д.С. Автоматизированная система мониторинга эффективности усвоения знаний и качества тестовых заданий /

Д.С. Карпенко, О.М.Карпенко, Е.Н. Шлихунова // Инновации в образовании - 2001. - №2 - 69-86.

6. Самыловский, А.И. Государственные образовательные стандарты высшего профессионального образования как тестологическая основа внедрения образовательного тестирования / А.И. Самыловский // Вопросы тестирования в образовании, 2003, № 5.

7. Birnbaum, A. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F.M.Lord and M.R.Novick. Statistical Theories of Mental Test scores / A. Birnbaum - Reading, Mass.: Addison - Wesley. 1968.Сыр. 17-20.

8. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen, Denmark: Danish Institute for Educational Research, 1960.

Розділ 2. Організація навчального процесу та методика викладання

УДОСКОНАЛЕННЯ ПІДГОТОВКИ ФАХІВЦІВ З МОНІТОРИНГУ ЯКОСТІ ОСВІТИ

Сергієнко В.П., Войтович І.С.

Сучасні соціально-економічні та геополітичні зміни стали передумовою реформування національної системи освіти. Основною соціальною вимогою українського суспільства є потреба в якості освіти. Ураховуючи інтереси суспільства, Національна доктрина розвитку освіти визначає стратегічні напрями і наголошує на необхідності «підвищення якості освіти, оновленні її змісту та форм організації навчально-виховного процесу» [4]. Водночас держава бере на себе функції забезпечення моніторингу освітнього процесу шляхом регулярного збору інформації, об'єктивного аналізу та прийняття відповідних рішень [1]. Отже, основним засобом вивчення якості освіти є моніторинг, система інструментарію якого здатна оцінити ефективність освітнього процесу та передбачити подальші кроки до її підвищення.

На практиці нерідко використовуються окремі елементи моніторингу у формі контрольних робіт, екзаменів, інспекторських перевірок [3]. Однак вони недостатньо ефективні, оскільки такий моніторинг має епізодичний характер, залишаючи без уваги сам процес навчання та виховання. При цьому рідко використовуються діагностичні методики, за допомогою яких можна виявити причини недоліків у роботі колективу, зокрема виявити фактори,

що впливають на успішність. Комплексне впровадження педагогічного моніторингу дозволяє активізувати педагогічну діяльність, навчально-виховний процес, що піднімає професійний рівень педагогічних колективів, підвищує рівень якості освіти.

Моніторинг в освіті як предмет наукового дослідження у працях українських та зарубіжних учених розглядався переважно з позицій педагогічних наук як засіб удосконалення навчально-виховного процесу (Я. Болюбаш, Д. Гопкінз, К. Інгенкамп, В. Кальней, М. Курлехт, І. Лікарчук, О. Локшина, Т. Лукіна, О. Ляшенко, В. Максимова, В. Сергієнко, І. Тельнюк, С. Шишов та ін.). У вищих навчальних закладах підготовка фахівців з оцінювання якості освіти розпочалася порівняно недавно, тоді як основами моніторингу повинен володіти кожен педагог. Потребує включення цих питань і у програми курсів підвищення кваліфікації педагогічних кадрів.

Для реалізації цього завдання нами підготовлено до впровадження навчальний курс «Моніторинг якості освіти», метою якого є формування компетентностей педагогічних працівників з питань організації та проведення моніторингу освітніх послуг навчальних закладів, ознайомлення з методиками та методами педагогічного оцінювання, необхідними умовами організації та проведення моніторингу якості освіти, ознайомлення із сучасними програмами та результатами національних і міжнародних порівняльних досліджень якості освіти. Завдання курсу: 1) познайомити студентів (слухачів курсів підвищення кваліфікації педагогічних кадрів) з основами моніторингу якості освіти, 2) сформувані вміння педагогічного оцінювання; 3) навчити здійснювати моніторинг якості освіти.

Після вивчення дисципліни студент (слухач курсів підвищення кваліфікації педагогічних кадрів) повинен *знати*: науково – понятійний апарат педагогічного оцінювання, види моніторингу якості освіти, нормативні документи, які регламентують проведення моніторингу якості освіти, сучасний стан та розвиток моніторингу якості освіти; *уміти* застосовувати на практиці сучасні види моніторингу якості освіти.

У перший модуль „**Науково-теоретичні основи педагогічних вимірювань і моніторингу якості освіти**” ми

включили теми: 1. Теоретичні основи педагогічних вимірювань. 2. Міжнародні та українські порівняльні моніторингові дослідження. 3. Класифікація навчальних цілей. Вимоги та критерії оцінювання навчальних досягнень; 4. Об'єкти, інструментарій, мета і завдання моніторингу якості освіти. Другий модуль „**Організація моніторингу якості освіти**” передбачає вивчення тем: 5. Нормативно-правове забезпечення моніторингових досліджень в освіті; 6. Обробка та інтерпретація результатів моніторингу якості освіти.

Скептики стверджують, що «моніторинг – це науково–педагогічне чи психологічне дослідження і ним не потрібно займатися кожному вчителю», однак хочемо відзначити, що кількісні виміри можуть здійснюватись будь-ким, а якісні виміри – експертом у даній галузі. Тому одним із важливих етапів моніторингу є відповідна інтерпретація результатів дослідження фахівцями (переведення їх на мову педагогіки, психології, управління) та підготовка відповідних рекомендацій.

Важливою умовою ефективності моніторингу є відповідність його процедур вимогам культури оцінки, яку характеризують [2]:

- наявність чітких критеріїв оцінки;
- розроблені процедури оцінки;
- наявність кваліфікованих експертів;
- розроблені форми фіксації інформації в ході моніторингу;
- розроблені часові характеристики оцінки;
- зв'язок оцінки з системою прийняття рішень.

Моніторинг опосередковано та безпосередньо впливає на якість освіти. Опосередкований вплив пов'язаний з тим, що моніторинг забезпечує управління інформацією, необхідною для прийняття рішень, а безпосередній вплив здійснюють вже ці рішення. Безпосередній вплив забезпечують самі процедури моніторингу. Так, наприклад, якщо студентами систематично проводиться оцінка організації навчального процесу, діяльності викладачів, то сам факт здійснення моніторингу забезпечує управлінський вплив. Викладачі починають корегувати свою діяльність відповідно до критеріїв, за якими вони оцінюються, як

це реалізовано у Херсонському державному університеті [5] та інших ВНЗ.

Таким чином, на основі результатів експериментального дослідження та теоретичного аналізу ми вважаємо, що наукова організація моніторингу якості освіти дає можливість приймати відповідні управлінські рішення та прогнозувати навчально-виховні ситуації на наступний навчальний рік; оперативно втручатися і вносити відповідні корективи до педагогічного процесу; конкретно планувати роботу з відповідної проблеми з учителями та учнями; створювати умови для порівняння власної оцінки діяльності педагогічного колективу з незалежною оцінкою.

Тому впровадження курсу «Моніторинг якості освіти» є важливим етапом удосконалення системи підготовки та перепідготовки педагогічних кадрів в спектрі сучасних тенденцій розвитку освіти.

Література

1. Лукіна Т. О. Моніторинг якості освіти: теорія і практика. – К.: Вид. дім «Шкільний світ», 2006. – 128 с.
2. Моніторинг якості освіти: світові досягнення та українські перспективи / [за заг. ред. О.І.Локшиної] - К.: К.І.С., 2004. - 128 с.
3. Педагогічне оцінювання і тестування. Правила, стандарти, відповідність. [наукове видання] / Я.Я. Болубаш, І.С. Булах, М.Р. Мруга, І.В. Філончук.– К.: Майстер-клас, 2007.– 272 с.
4. Указ Президента України від 17 квітні 2002 р. № 347 «Про Національну доктрину розвитку освіти». – Режим доступу: <http://zakon.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=347%2F2002>
5. Управління інформаційними технологіями вищих навчальних закладів: [метод. посіб.] / О.В. Співаковський, Я.Б. Федорова, О.О. Глущенко, Н.А. Кудас. — [2-ге вид., доповн. і переробл.] — Херсон: Айлант, 2007. — 299 с.

ВИКЛАДАННЯ ТЕМИ «ВАЛІДНІСТЬ» В ЗАГАЛЬНОМУ КУРСІ ОСВІТНІХ ВИМІРЮВАНЬ: КОМПЛЕКСНИЙ ПІДХІД

Ковальчук Ю.О.

Поняття валідності та валідації інструментів вимірювання є одним з чи не найскладніших для розуміння та застосування в теорії освітніх вимірювань. Необхідність валідації нових тестів є наслідком самої природи тестування, яка передбачає, з одного боку, наявність певних попередніх теоретичних конструкцій, а з іншого боку, – відшукання часткових свідчень правдоподібності цих конструкцій, подальшого узагальнення свідчень, а також правильної інтерпретації та застосування результатів тестування.

У змісті поняття валідності, починаючи з 20-х років минулого століття, акцент поступово зміщувався від концентрації навколо ідеї відповідності результатів тестування зовнішньому критерію (критеріальна валідність), до необхідності оцінювання насамперед *правильності інтерпретації та використання* результатів тестування. На цьому шляху теорія валідації пройшла через період панування так званої «троїстої» моделі валідності (змістова, критеріальна та конструктивна типи валідності), до виділення конструктивної валідності як загальної, що охоплює також поняття змістової та критеріальної валідності, і далі – до трактування валідності як аргументу в рамках загальної теорії валідації М. Кейна.

Модель Кейна валідності як аргументу запропонована ним ще у 1982 році, але тоді ця теорія не набула помітного поширення. Однак саме Кейну було довірено написати розділ «Валідація» для четвертого видання фундаментального збірника «Educational Measurement» [1], який видано у 2006 році під егідою АСЕ та NCME. Підхід Кейна є спробою надати розробникам та користувачам тестів єдину методологію валідації, і це є добрим аргументом для покладання цієї методології в основу викладу відповідного розділу в навчальних курсах з теорії вимірювань в освіті. Далі ми намагатимемося прослідкувати логіку розвитку поняття валідності та валідації стосовно інструментів освітніх

вимірювань, стисло викласти сучасний підхід до цієї проблеми, узагальнений М. Кейном, та запропонувати методологію викладання розділу «Валідність» у начальних курсах з основ освітніх вимірювань.

Еволюція поняття валідності в педагогічному тестуванні.

Валідність в науці – це відповідність емпіричних досліджень тій меті, заради якої вони проводяться.

В психометрії тест називається *валідним*, якщо він адекватно вимірює саме ту рису чи конструкт, для вимірювання якої він був створений.

Укладачі та користувачі тестів часто обмежуються лише перевіркою часткових свідчень валідності. У деяких випадках це цілком виправданий підхід. Наприклад, якщо вчитель в кінці уроку пропонує учням невеликий тест на засвоєння щойно пройденого матеріалу, достатньо, крім визначення правила нарахування балів за окремі завдання та правильної процедури проведення тестування, забезпечити також змістову валідність цього тесту. В інших випадках, за наявності валідного критерію (наприклад, іншого тесту, який використовувався з тією ж метою), цілком достатньо перевірити, «за рівних інших умов», чи достатньо високою є кореляція результатів тестування з результатами застосування критерію. Але для виготовлення стандартизованого інструменту вимірювання, для якого передбачається масове повторне використання, і результати якого можуть інтерпретуватися й застосовуватися у різних контекстах, дослідження валідності має бути комплексним, тобто воно повинне охоплювати усі аспекти валідності.

Для ясності розглянемо значно простіший, з точки зору валідизації, вид вимірювання – фізичне вимірювання, наприклад, вимірювання температури тіла людини. Очевидно, що, перш за все, нам потрібен якісний інструмент – градусник. Чи можемо ми замість градусника використати кімнатний термометр? Якщо так, то з якими відмінностями й обмеженнями? Якість градусника можна визначати по-різному. Наприклад, ми можемо порівняти його показання з показаннями іншого градусника, якість якого не викликає у нас сумнівів («критеріальна» валідність), або

перевірити показання нашого градусника на достатньо великій групі цілком здорових осіб. Крім того, ми повинні правильно «читати» шкалу градусника. Чи виконане градування шкали за Цельсієм, чи за Фаренгейтом? Нехай ми упевнилися у добрій якості градусника і вміємо зчитувати його показання. Чи є це достатніми для того, щоб отримати правильні результати його конкретного використання? Очевидно, ні, оскільки неправильно могла бути процедура вимірювання, наприклад, замість необхідних семи хвилин, пацієнт тримав градусник дві хвилини, або тримав градусник не під пахвою, а у кишені пальто. Якщо градусник має добру якість і процедура вимірювання була правильною, то виникає наступна проблема – як слід інтерпретувати отриманий результат? Що означає, наприклад, що градусник показує температуру 37 градусів за Цельсієм? Яка температура вважається нормальною, і чи є ця норма застосовною також і для конкретного пацієнта? Зауважимо, що інтерпретація отриманого результату вимірювання температури тіла залежить також і від мети, з якою проводилося вимірювання. Наприклад, якщо лікар хоче перевірити, як подіяв на пацієнта, у якого напередодні була висока температура, призначений йому лікувальний засіб, він може інтерпретувати температуру в 37 градусів як позитивний результат. Навпаки, якщо пацієнт звернувся до лікаря вперше, це ж саме показання повинне трактуватися лікарем як «підвищена температура». У цьому випадку лікар може інтерпретувати показання градусника як симптом можливої хвороби. Хвороба на цій початковій стадії трактується лікарем як комплекс відповідних симптомів, серед яких температура тіла відіграє певну визначену наперед роль: для одних хвороб це важливий показник, для інших – менш важливий, або й зовсім не важливий. Нарешті, на основі отриманого результату вимірювання лікар повинен прийняти певне рішення. Чи слід лікувати пацієнта? Якщо так, то за допомогою яких ліків чи процедур? Амбулаторно чи стаціонарно? Можна заперечувати відповідність останніх запитань темі валідності вимірювання. Але, з практичної точки зору, аспект прийняття рішень на основі результатів вимірювання вигідно включити до загального дослідження валідності вимірювання, інакше нам доведеться

розглядати окремо «валідність прийняття рішень». А тепер задамося останнім запитанням: якщо ми забезпечили перевірку усіх перелічених свідчень валідності вимірювання температури тіла, чи можемо ми, нарешті, стверджувати, що вимірювання є цілком валідним? На жаль, ствердну відповідь ми можемо дати лише з тією чи іншою мірою впевненості, щоправда, ніколи не повною. Можна наводити скільки завгодно питань, які випали з нашого розгляду. Наприклад, чи був збитий ртутний стовпчик до достатньо низького рівня перед початком вимірювання температури?

Розглянутий приклад свідчить про необхідність комплексної валідизації навіть такого простого, у порівнянні з психометричним, фізичного вимірювання. Але для більшості фізичних вимірювань згадані вище проблеми й методи їх вирішення є цілком очевидними, тому й термін «валідність» для них зазвичай не використовується.

Психометричні вимірювання, до яких ми відносимо й педагогічні, відрізняються від фізичних кардинально. Головними відмінностями психометричних вимірювань є латентність вимірюваних величин, і те, що отримані результати є практично у кожному конкретному випадку частинними й потребують ряду узагальнень.

Розробник чи користувач тесту ніколи не може із стовідсотковою упевненістю сказати, що саме вимірюється даним тестом. Якщо ми хочемо визначити вагу власного тіла, у нас не виникає сумнівів, що для цього слід скористатися вагами, а не, скажімо, лінійкою. З іншого боку, коли у нас в руках опиняється інструмент на зразок ваг, ми можемо впевнено стверджувати, що може і чого не може вимірювати цей інструмент. З тестами все виглядає куди складніше. Наприклад, тест на *рівень інтелекту*, за умови значного скорочення часу на його виконання, може перетворитися на тест *здатності до концентрації розумових зусиль*, і користувач тесту може такого перетворення не помітити. У тесті досягнень з історії можуть зустрічатися настільки великі уривки тексту, що тест вимірюватиме радше швидкість читання, ніж власне знання історії.

Джерел поганої валідності психологічних вимірювань є багато, вони пов'язані з самою природою тестування. Основною причиною можливого погіршення валідності є вибірковий метод – звуження цільової популяції (людей, проявів їх поведінки) до множини її окремих представників, з наступним узагальненням отриманих результатів на всю популяцію.

В стандартній процедурі тестування подібне звуження відбувається тричі (рис. 1):

1. З області поведінки, яка досліджується (target domain) обираються для тестування лише деякі прояви – ті, які в принципі можуть бути перевірені тестом. Вони складають популяцію, або генеральну сукупність, тестових завдань (не слід її плутати з поняттям банку завдань).

2. Для тесту з популяції тестових завдань виконується вибірка завдань.

3. Тест отримує свої характеристики після апробації лише на вибірці з цільової популяції осіб, тобто тих осіб, для яких тест призначений.

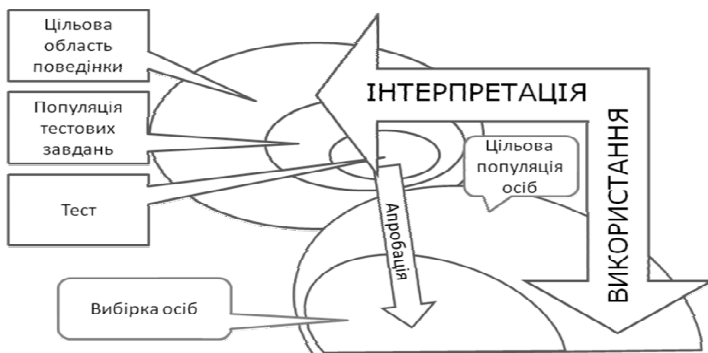


Рис. 1. Співвідношення множин суб'єктів та об'єктів тестування.

Але цим не обмежується коло джерел можливої «поганої» валідності тесту. Результати одного і того ж тесту, отримані на одній і тій же вибірці з цільової популяції суб'єктів, можна по-

різному *інтерпретувати*. Неправильна інтерпретація результатів тестування перекреслює всі попередні зусилля, направлені на забезпечення валідності тесту. З іншого боку, теоретично можна припустити, що для деякого новоствореного тесту може існувати така, хай і невідома, інтерпретація його результатів, яка зробить тестування валідним, і тоді постає завдання відшукати цю правильну інтерпретацію.

Навіть після правильної інтерпретації результатів тестування можуть бути прийняті неправильні *рішення*. Наприклад, якщо дитина не пройшла тест на готовність відвідувати дитячий садок, і було прийнято рішення відкласти прийом до садка до наступного року, це рішення може бути як правильним, так і неправильним, в залежності від того, з яких саме причин дитина не склала тест успішно. Не всі дослідники погоджуються з тим, що подібні проблеми якимось стосуються валідності тестування. Але, включаючи проблеми правильності прийнятих на основі тестування рішень до кола проблем валідизації, ми вчинимо розумно, бо тим самим ми не залишимо цих проблем поза увагою і зменшимо ризик прийняття неправильних рішень.

Щойно сказане стосується також і можливих *соціальних наслідків* тестування. Тестування, особливо широкомасштабне, доленосне для його учасників, часто супроводжується збуреннями в суспільстві, породжуючи напругу, скандали, міфи, непорозуміння, політичні спекуляції. Було б правильним, якби розробники і користувачі тестів зважали на можливі суспільні ефекти і включали дослідження соціальних наслідків тестування до кола питань валідизації вимірювань.

Говорити «валідність тесту» некоректно у тому розумінні, що потрібно уточнювати, як будуть проінтерпретовані і використані результати тестування, які наслідки вони будуть мати. Тому більш правильно говорити про *валідність як про відповідність передбачуваної інтерпретації та використання результатів тестування тій меті, заради якої створено тест*.

Традиційно виділяють кілька видів валідності. З 20-х по 50 роки минулого століття в теорії валідизації панівну роль відіграло поняття критеріальної валідності. *Критеріальна валідність* визначає ступінь відповідності результатів тестування

зовнішнім, тобто таким, які не стосуються тесту, критеріям. Наприклад, критерієм для тесту особи на здатність до навчання в університеті може бути середній бал за 1 курс, отримані цією особою після вступу в університет. Зрозуміло, що сам критерій повинен бути валідним. Критеріальна валідність визначається чисельно як коефіцієнт кореляції між тестом і критерієм, обчисленим для репрезентативної вибірки осіб, і називається у цьому випадку *коефіцієнтом валідності*. З точки зору відмінностей за часовою ознакою, критеріальну валідність поділяють на поточну (конкурентну) та прогностичну. *Поточна* валідність отримується при порівнянні результатів тестування з уже відомими на момент тестування результатами. Якщо критерієм є інший тест, то слід обґрунтувати використання нового тесту. Наприклад, слід показати, що новий тест у порівнянні з критерієм є коротшим, зручнішим, чи має якісь інші переваги. Якщо кореляція між новим тестом і критерієм є дуже високою, це може ставити під сумнів необхідність використання нового тесту, адже він не даватиме нової у порівнянні з критерієм інформації. Якщо ж кореляція між тестом і валідним критерієм є надто низькою, то тест не може бути визнаним валідним.

Прогностичну валідність розглядають у випадках, коли тест призначений для передбачення рівня успішності особи в певному виді діяльності в майбутньому. Такими є тести здібностей та тести відбору. У цих випадках коефіцієнт валідності обчислюється як кореляція між результатами тестування групи осіб і критерієм, який виражається в оцінці реальної діяльності осіб (див. приклад про порівняння результатів вступного тесту з оцінками за 1 курс). Дослідження прогностичної валідності вимагає багато часу, адже після тестування групи осіб потрібно дочекатися того моменту, коли ця група осіб достатньо проявить себе в обраному виді діяльності настільки, щоб за результатами цієї діяльності отримати валідні критеріальні оцінки. Якщо обставини не дозволяють досліджувати валідність тесту таким чином, то можна запропонувати тест групі осіб, яка вже задіяна у даному виді діяльності і для якої вже існують критеріальні оцінки, після чого обчислити кореляцію між результатами тестування і критерієм.

Валідність тесту досягнень прийнято досліджувати шляхом порівняння його змісту із змістом тієї області, для оцінки якої він призначений. У цьому випадку говорять про *змістову валідність* тесту. На відміну від критеріальної валідності, змістова валідність визначається не чисельно, а у вигляді суджень експертів. Змістова валідність повинна забезпечуватися вже на початкових етапах створення тесту шляхом ретельного аналізу цільової області і складання специфікації тесту та описання окремих тестових завдань у такий спосіб, щоб зміст цільової області був представлений у тесті достатньо повно і пропорційно. Ця робота тільки на перший погляд може здатися простою, адже в поняття успішності засвоєння цільової області, наприклад, певної навчальної дисципліни, повинні включатися і необхідні рівні когнітивних процесів. Наприклад, тест, у якому переважають завдання на знання фактів, не дасть змоги повно виявити рівень розуміння цих фактів та їх взаємозв'язків, здатність особи до самостійного критичного аналізу, оцінювання та застосування набутих знань, а ці якості зазвичай входять до переліку педагогічних цілей, які ставляться при викладанні навчальних дисциплін, отже, складають зміст цільової області.

Розробка тестів особистості привела до появи поняття *конструктної валідності*. Для подібних тестів часто не існує прийнятних критеріїв, і неможливо однозначно визначити зміст цільової області поведінки.

Конструктна валідність виникла з потреби вимірювати *теоретичні конструкти*. Класичний приклад конструкту – здібність до чогось.

Поняття конструктної валідності ввели Кронбах і Міл. В статті [2], яка вважається класичною, вони розглядали цей вид валідності як альтернативу критеріальній та змістовій валідності. Конструктна валідність, за Кронбахом і Мілом, повинна була застосовуватися, коли «тест інтерпретується як міра деякого атрибуту чи якості, які не визначені операціонально». Тобто коли не існує ні прийнятного зовнішнього критерію, ні змістового описання даного атрибуту чи якості.

Щоправда, додають Кронбах і Міл, практично для кожного тесту існує потреба визначення, які психологічні конструкти у ньому задіяні.

Спочатку конструктна валідність розглядалася як окремий випадок валідності. Так, в [3] розглядається чотири види валідності як такі, що є пов'язаними з чотирма видами інтерпретації:

- прогностична і конкурентна (різновиди критеріальної валідності);
- змістова валідність;
- конструктна валідність.

Ці види валідності становили разом так звану «троїсту модель валідності». Однак в кінці 1970-х намітилося 2 тренди в подальшому розвитку теорії:

1. Стійкий інтерес до чіткого визначення, які саме види свідчень потрібні для тих чи інших інтерпретацій та використання тестів.
2. Визнання необхідності створення єдиної концепції валідності.

Базою для єдиної концепції валідності стала конструктна валідність. В кінці 1980-х років Мессік розробив розширену модель конструктної валідності як основу (framework) єдиного поняття валідності. В [4] Мессік визначає валідність як інтегральне оціночне судження про ступінь підтримки емпіричними свідченнями і теоретичними міркуваннями адекватності прийнятності висновків і дій, заснованих на тестових балах чи інших видах оцінок.

В [1] М. Кейн виділив три аспекти, у яких конструктна модель вийшла за межі теоретико-залежного контексту, в якому вона була спочатку запропонована:

1. Між 1955 і 1989 роками основний наголос змістився від валідизації тестів до розробки і валідизації пропозицій щодо інтерпретації та використання тестових балів.
2. Конструктна модель валідності потребує більшою мірою теоретичного дослідження, ніж просто емпіричних свідчень.

3. Фокусування конструктивної валідності на теорії веде до можливості і потреби оспорювати запропоновану інтерпретацію і розробляти альтернативні інтерпретації.

Валідність як аргумент: підхід М. Кейна [1]. У викладі Кейна, *аргумент валідності* (validity argument) є інструментом загальної оцінки інтерпретації та використання тестових балів, що передбачаються для даного тесту. Головною метою при цьому є відшукування ясних та взаємно узгоджених свідчень «за» або «проти» запропонованих інтерпретацій чи застосувань, і, якщо можливо, свідчень для альтернативних інтерпретацій/застосувань. З цією метою спочатку розробляється *інтерпретативний аргумент* (interpretative argument), який слугує своєрідною канвою для вироблення аргументу валідності. Інтерпретативний аргумент складається з ряду припущень та декларативних висновків.

Підхід Кейна до валідації можна сформулювати як послідовність кроків, що може ітеративно повторюватися:

1. Пропонується інтерпретація тестових балів в термінах інтерпретативного аргументу.

2. Створюється попередня версія аргументу валідності шляхом відшукування усіх доступних свідчень правдоподібності інтерпретативного аргументу.

3. Детально оцінюється справедливість припущень та висновків.

4. Якщо потрібно, переформулюються інтерпретативний аргумент та аргумент валідності, після чого повторюється крок 3. Так відбувається доти, доки не свідчення правдивості задекларованих висновків не стануть достатніми для їх визнання або відхилення.

Цей процес нагадує процес створення теорій у природничих науках. Вже з розглянутого вище прикладу фізичного вимірювання можна зробити такі важливі висновки: 1) дослідження валідності вимірювання має бути комплексним; 2) навіть комплексне дослідження не дає повної гарантії валідності. Подібну ситуацію математик може характеризувати як таку, у якій є лише багато необхідних умов для доведення деякого твердження, і жодної – достатньої. Для фахівців гуманітарної

сфери подібна ситуація є цілком звичною. Таким чином, можна розглядати комплексну валідацію як міні-теорію, засновану на системі практичної аргументації. Розглянемо загальні принципи побудови такої теорії. За С. Тулміним [5], аргументація – це переважно процес верифікації вже існуючих ідей; існує шість взаємопов’язаних компонентів процесу аргументації:

1. *Твердження (claim)*. Твердження повинне бути завершеним. Наприклад, якщо хтось намагається переконати, що він є громадянином Великобританії, то його твердженням буде «я громадянин Великобританії»

2. *Свідчення (evidence)*. Це факт, на який посилаються, як на підставу для твердження. Наприклад, особа з попереднього прикладу може підтримати своє висловлювання іншими даними «я народився на Бермудських островах».

3. *Підстава (warrant)*. Висловлювання, що дозволяє перейти від свідчення до твердження. Для того щоб перейти від свідчення «я народився на Бермудських островах» до твердження «я громадянин Великобританії», особа повинна використовувати підстави для усунення розриву між твердженням і свідченням, вказавши, що «людина, народжена на Бермудських островах, юридично може бути громадянином Великобританії».

4. *Підтримка (backing)*. Доповнення, спрямоване на підтвердження висловлювання, вираженого в підставі. Підтримка має бути використана, коли підстава сама по собі не є достатньо переконливою для опонента.

5. *Спростування/контраргумент (rebuttal)*. Висловлювання, що вказує на обмеження, які можуть застосовуватися. Прикладом контраргументу є: «Людина, що народилася на Бермудських островах, може легально бути громадянином Великобританії, тільки якщо вона не зрадила Великобританії і не є шпигуном іншої країни».

6. *Визначник (qualifier)*. Слова та фрази, що виражають ступінь впевненості автора у його твердженні. Це такі слова і фрази, як «ймовірно», «можливо», «неможливо», «безумовно», «ймовірно» або «завжди». Твердження «Я безумовно є громадянином Великобританії» несе в собі набагато більшу

ступінь впевненості, ніж твердження «Я імовірно є громадянином Великобританії».

Перші три елементи розглядаються як основні, тоді як потреба у трьох останніх виникає не завжди.

Крім загальної філософії науки, на теорію Кейна вплинули також різні методології, поширені в психометрії, зокрема, висновки узагальненої теорії тестування (Generalizability Theory), а також методологія оцінювання програм (Program Evaluation).

Кейн виділяє чотири широкі категорії інтерпретативних аргументів:

1. Інтерпретація для рис особистості (traits). Сюди відноситься наприклад, випадок вимірювання навчальних досягнень учня з певного предмету.

2. Інтерпретація, заснована на теорії. Такого виду інтерпретації вимагають, наприклад, результати тесту здібностей.

3. Якісна інтерпретація. Стосується області якісного (на відміну від кількісного) оцінювання і підходить для валідизації, зокрема, оцінювання учнів під час занять у класі.

4. Процедури прийняття рішень. Прикладом може бути інтерпретативний аргумент для Програми відповідальності NCLB (No Child Left Behind Act) в США.

Інтерпретаційний аргумент для вимірювання риси.

Розглянемо детальніше процес побудови інтерпретативного аргументу для випадку, коли тест призначений для вимірювання проявів деякої риси. *Риса* (trait) – схильність індивідуума поводитися або діяти певним чином у відповідь на деякий стимул або завдання, за певного набору умов. Під це визначення підходить поняття навчальних досягнень, отже, сюди відноситься найпоширеніший у педагогічному тестуванні вид тестування – тестування досягнень.

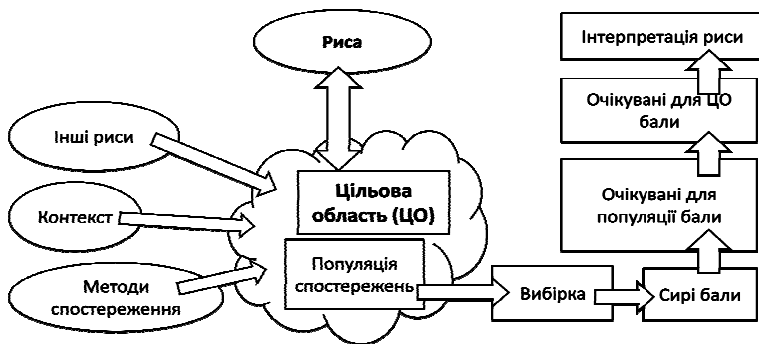


Рис. 2. Процедура вимірювання й інтерпретативний аргумент.

Риса асоціюється з поняттям *цільової області* можливих спостережень, і очікувані бали особи, які представляють цю особу по відношенню до цільової області, є *цільовими балами*.

Цільова область може бути дуже широкою, як при деяких визначеннях інтелекту, більш вузькою, як при визначенні рівня успішності десятикласника з алгебри, або зовсім вузькою, наприклад, навички з виконання деякого завдання.

На рис. 2 схематично зображено взаємозв'язок між процедурою вимірювання і структурою інтерпретативного аргументу у випадку вимірювання риси.

Нехай потрібно дослідити валідність предметного тесту для випадку, коли метою вимірювання є з'ясування рівня успішності (в іншій термінології – рівня навчальних досягнень) особи з даного предмету. Порядок узагальнень інтерпретації тестових балів у цьому випадку є таким (в дужках пропонуються короткі назви відповідних рівнів узагальнення):

1. Від спостережених відповідей – до тестових балів (*скоринг*).
2. Від тестових балів – до балів за ту частину предмету, яку представляв тест (*генералізація*).
3. Від балів з частини предмету, яку представляє тест – до балів з усього предмету (*екстраполяція*).
4. Від балів з предмету – до словесного описання рівня успішності (*імплікація*).

Розглянемо можливі твердження інтерпретаційного аргументу й методи отримання свідчень валідності для кожного з цих рівнів.

Скоринг. Під цим терміном розумітимемо нарахування тестових балів за заданими укладачем правилами. На даному рівні інтерпретаційний аргумент може містити такі твердження:

1. Схема нарахування тестових балів є прийнятною.
2. Схема нарахування тестових балів використовувалась правильно.
3. Нарахування тестових балів було неупередженим.
4. Нараховані бали узгоджується з обраною моделлю шкалування.

Аргумент валідизації для цих тверджень повинен будуватися вже на початку створення тесту. Методи забезпечення валідності для цього етапу мають якісний характер і, здебільшого, форму експертних висновків. Найкраще було б, якби дві або й більше груп експертів виробляли незалежно одна від одної власні схеми оцінювання. Ці різні схеми слід проаналізувати, порівнюючи бали, отримані для тесту за цими схемами. Важливо також правильно організувати та контролювати роботу оцінювачів, якщо така робота диктується видами тестових завдань (наприклад, тест містить есе).

Практично завжди до отриманих «сирих» тестових балів застосовується той чи інший метод шкалування. Наприклад, в Україні тести ЗНО протягом останніх років шкалювалися за методом еквіпроцентильної нормалізації. Доцільність використання цього методу є предметом постійних дискусій. Справедливо вказується головний його недолік – неможливість інтерпретувати тестові бали як дійсний рівень успішності (що знає й уміє випускник, і чого він не знає і не вміє), оскільки у випадку застосування цього методу оцінювання є нормо-орієнтованим. Тим не менше, для порівняння результатів тестування з предмету у різних сесіях ЗНО і за різні роки, цей метод є чи не єдиною правильним (крім, хіба що, застосування методів теорії IRT), з огляду на непорівнюваність самих тестів. Зауважимо, що метод еквіпроцентилів рекомендований Європейською Комісією для зарахування наявних оцінок студента на новому місці навчання

при його переході до іншого університету [6]. З точки зору забезпечення валідності вимірювання, шкалування не повинне суперечити тій кінцевій меті, заради якої проводиться тестування. Так, при застосуванні методу еквіпроцентильної нормалізації у ЗНО важливим є відповідність його головному принципу – той випускник, який отримав більше «сирих» балів у порівнянні з іншим випускником, також повинен отримати більше балів і після проведення процедури шкалування.

Генералізація. Практично завжди тестуванням може бути охоплена не вся область поведінки суб'єкта, яка є предметом вимірювання. Це стосується й предметних тестів. Наприклад, якщо тест з іноземної мови проводиться у письмовій формі, неможливо перевірити правильність вимови учня. З іншого боку, тестові завдання завжди є лише вибіркою з усіх можливих тестових завдань, які разом складають деяку генеральну сукупність завдань (за термінологією математичної статистики). Саме ця генеральна сукупність тестових завдань й репрезентує ту частину поведінки особи у цільовій області (у нашому випадку – проявів рівня успішності з предмету), яка покривається тестуванням. Не слід плутати генеральну сукупність тестових завдань з банком тестових завдань. Доцільно вважати, що генеральна сукупність містить нескінченно багато завдань. Адже легко погодитися з тим фактом, що тести двох різних незалежних укладачів, будучи вибірками з генеральної сукупності, практично ніколи не містять однакових завдань.

При просуванні інтерпретації тестових балів від конкретного тесту до генеральної сукупності тестових завдань важливо отримати свідчення на користь таких тверджень (інтерпретаційний аргумент):

1. Завдання тесту складають репрезентативну вибірку з генеральної сукупності тестових завдань.

2. Дана вибірка тестових завдань є достатньо великою, щоб контролювати випадкову похибку вимірювання.

Аргумент валідності на цьому рівні будується методами математичної статистики. Зокрема, контроль похибок вибірок здійснюється методами теорії надійності та узагальненої теорії тестування (Generalizability Theory).

Екстраполяція. На рівні екстраполяції інтерпретація тестових балів переноситься на всю область, яка оцінюється (у нашому випадку – це рівень успішності з даного предмету, наприклад, з математики чи з історії України).

Це досить складний і відповідальний етап у дослідженні валідності. Інтерпретаційний аргумент для цього етапу може містити наступні твердження:

1. Тестові бали є релевантною мірою рівня успішності з даного предмету.

2. Систематичні похибки вимірювання не перешкоджають екстраполяції.

Методи збору свідчень валідності на даному етапі можна умовно поділити на аналітичні й емпіричні.

До аналітичних методів належить, зокрема, перевірка широти покриття тестуванням усього різноманіття проявів рівня успішності з предмету. Зауважимо, що істотно збільшити широту покриття можна, замінивши паперове тестування комп'ютерним. Наприклад, при тестуванні з іноземної мови комп'ютерна форма дозволяє відносно легко організувати аудіювання. Перевірка широти покриття тестом успішності засвоєння предмету передбачає, крім перевірки змістової частини предметної області, також і контроль процесів мислення учня. Слід намагатися порівняти ті процеси мислення, які учень використовує під час виконання тестових завдань, з тими процесами мислення, які він демонструє в тій частині володіння предметом, яка не підлягає тестуванню. Важливу роль на цьому етапі відіграють факти і методи когнітивної психології. Зокрема, для виявлення процесів мислення, застосовуваних учнем під час вирішення ним конкретних завдань, використовується метод документування міркувань уголос.

Певну роль відіграє забезпечення так званої очевидної валідності (*face validity*) тесту, особливо тоді, коли тест (як у випадку ЗНО) є тестом високої відповідальності.

Окремої уваги заслуговує дослідження негативного впливу стандартизації вимірювання. Заходи стандартизації, маючи на меті зменшення випадкової похибки вимірювання, є одночасно джерелом систематичних похибок, що є результатом сильного

звуження усього розмаїття умов і способів, у які проявляється рівень успішності учня з предмету, до жорстких, максимально однакових для всіх змісту і процедури тестування.

До емпіричних методів екстраполяції тестових балів слід віднести порівняння результатів тестування з критерієм (*критеріальна валідність*); з результатами інших видів оцінювання (*конвергентна валідність*); з оцінками дивергентних рис (*дискримінантна валідність*). Зауважимо, що для дослідження критеріальної валідності потрібен валідизований критерій, а такий є у розпорядженні дослідника дуже рідко, тому те, що часто декларується як критеріальна валідність, повинне бути віднесене швидше до конвергентної валідності. Аналіз матриці «багато рис – багато методів (Multitrait-Multimethod)», яка складається з коефіцієнтів кореляції дивергентних рис, кожна з яких виміряна кількома різними способами, є потужним методом дослідження дискримінантної валідності.

Часом може стати у нагоді використання висновків *попередніх досліджень валідності*, якщо нова ситуація (нові умови, нова популяція учнів, новий тест) мало відрізняється від попередньої.

Імплікація. Якщо на попередніх рівнях валідизації валідність тесту перевіряється для тестових балів, то на рівні імплікації предметом дослідження є вже словесна інтерпретація цих балів. Тут потрібно перевірити, чи є прийнятною смисловою інтерпретація результатів вимірювання риси або теоретичного конструкту, і чи узгоджуються властивості отриманих результатів з висновками, асоційованими з визначенням риси чи конструкту. Вербальна інтерпретація тестових балів повинна забезпечуватися вже на стадії розробки тесту. Пізніше, за наявності емпіричних даних, перевіряється узгодження смислової інтерпретації результатів вимірювання з тими висновками щодо співвідношення з іншими змінними, які закладаються в концепції даної риси чи конструкту.

Інтерпретаційний аргумент для вимірювання теоретичного конструкту. Якщо предметний тест використовується для вимірювання теоретичного конструкту, яким є, наприклад, здатність особи до навчання, між рівнями

екстраполяції ті імплікації у дослідженні валідності тесту з'являється ще один рівень: *теоретична інтерпретація*. Ріса (рівень успішності з дисципліни), яка вимірюється, у цьому випадку відіграє роль одного з багатьох можливих *індикаторів конструктору*. Теорія, яка спочатку існує лише «в голові» дослідника, передбачає певне співвідношення між індикаторами, а також між даним конструктором та іншими конструкторами, і результати тестування повинні або підтверджувати цю теорію, або спростовувати її.

Конструктна валідність досліджується насамперед методами кореляційного аналізу та моделювання структурними рівняннями (Structural Equation Modeling). Останній зараз все частіше використовується і претендує на універсальність. Що стосується самої теорії, яка будується для даного конструктору, то вирішальну роль тут відіграють досягнення когнітивної психології, методи моделювання когнітивних процесів.

Підсумовуючи, підкреслимо важливість комплексного дослідження валідності тестів, зокрема, предметних, і, як наслідок, необхідність логічного впорядкування цього дослідження. Оскільки комплексна валідизація вимірювання є різновидом системи практичної аргументації, то підхід, який базується на побудові інтерпретаційного аргументу й відповідного йому аргументу валідності, є запорукою її повноти та високої якості.

Структура розділу «Валідність» в навчальних курсах з основ теорії освітніх вимірювань. Як бачимо, за час існування теорії освітніх вимірювань поняття валідності та валідизації тестів істотно трансформувалися. Це вимагає від викладача основ теорії освітніх вимірювань дотримуватися принципу історизму при побудові схеми викладання навчального матеріалу. За традицією, яка склалася протягом останніх десятиліть, більшість літературних джерел подають проблему валідизації освітніх вимірювань у відповідності до «троїстої моделі». Очевидним недоліком такого підходу є фрагментарність, яка створює враження, ніби окремі види валідності (змістова, критеріальна, конструктна) є самодостатніми, ніяк не зв'язаними між собою, причому далеко не завжди зрозуміло, з чого повинен розпочинатися процес валідизації конкретного вимірювання, чим він має закінчуватися, і,

головне, чи достатньо зібраних в процесі валідизації свідчень для того, щоб назвати даний тест валідним.

З іншого боку, конкретні види валідності тісно пов'язані з характерними для кожного з них методами та інструментами (наприклад, індекси критеріальної валідності).

Таким чином, розділ курсу, присвячений проблемі валідності, повинен:

1) розпочинатися з огляду історії розвитку понять валідності та процесу валідизації інструменту вимірювання;

2) містити детальний виклад окремих видів валідності;

3) розглядати валідизацію як інтегральний процес побудови інтерпретативного аргументу та аргументу валідизації (в термінології М. Кейна, яка, однак, не є достатньо вдалою в українському перекладі і вимагає більш адекватного представлення);

4) «Прив'язати» окремі види валідності до інтегрального процесу валідизації.

Дуже важливим є відшукування конкретних прикладів повної валідизації тестів, виконаних за схемами, запропонованими М. Кейном (і їх стислий виклад, оскільки подібні приклади можуть мати об'єм у десятки і сотні сторінок тексту).

Література

1. Educational Measurement / sponsored jointly by National Council on Measurement in Education and American Council on Education; edited by Robert L. Brennan. – 4th ed. p. cm. – (ACE/Praeger series on higher education). – 779 p. – ISBN 0-275-98125-8.

2. Cronbach, L. and Meehl, P. (1955). Construct validity in psychological tests, *Psychological Bulletin*, 52, 4, 281-302.

3. American Psychological Association, American Educational Research Association, & National Council on Measurements in Education. Standards for educational and psychological tests and manuals. – Washington, DC: American Psychological Association, 1974.

4. Messick, S. (1989). Validity. In R.L. Linn (Ed.), Educational Measurement (3rd ed., pp. 13-103.) New York: American Council on Education and Macmillan.

5. Toulmin S.E. The Uses of Argument (1958). – 2nd edition – 2003: ISBN 0–521–53483–6.

6. ECTS Users' Guide. – Luxembourg: Office for Official Publications of the European Communities, 2009.

КЛЮЧОВІ ПІДХОДИ ДО ВИКЛАДАННЯ ТЕМИ «НАДІЙНІСТЬ» У ЗАГАЛЬНОМУ КУРСІ ОСВІТНІХ ВИМІРЮВАНЬ

Ковальчук Ю.О.

Загальний навчальний курс освітніх вимірювань є стрижневим у програмах як спеціалізації бакалаврів з освітніх вимірювань, так і магістратури за спеціальністю *Освітні вимірювання* в НДУ імені Миколи Гоголя. У свою чергу, тема «надійність» є однією з ключових тем цього курсу. У літературі [1-3], яка заслуговує на роль базової при вивченні цієї теми, пропонуються дещо різні підходи та акценти щодо її висвітлення. Вивчення цих та інших джерел дозволяє висловити наступні зауваження стосовно ролі, структури та змісту даної теми.

1. Хоча у більшості навчальних посібників вивчення основ надійності тестів передує темі їх валідності, є сенс у тому, щоб викладати надійність вже після валідності, а останню – якомога раніше, оскільки сучасна концепція валідності є фактично всеохоплюючою і включає у себе також і поняття надійності. Надійність у порівнянні з валідністю є досить вузьким, «технічним» терміном.

2. Корисно при початковому уведенні поняття надійності використовувати для ілюстрації приклади, які стосуються фізичних вимірювань. Ця рекомендація стосується й інших концепцій і тем освітніх вимірювань. Фізичним вимірюванням

притаманний значно нижчий рівень абстрактності і вони значно доступніші для розуміння. Не лишнім буде також розглянути, як використовується слово «надійність» у розмовній мові і показати спільне у цьому відношенні з теоретичним поняттям надійності тесту.

3. Слід дати порівняльну характеристику понять валідності та надійності, оскільки обидва терміни використовуються часто разом, і обидва використовують статистичне поняття кореляції. При цьому слід акцентувати як на відмінності цільових областей цих двох фундаментальних понять психометрії, так і на тому, що їх об'єднує.

4. Викладання теми «Надійність» не може бути відірваним від положень класичної моделі тестової оцінки. Ця модель, яка має окрему цінність в системі освітніх вимірювань, слід викладати на достатньо високому і строгому теоретичному рівні, незважаючи на деяку її математичну складність.

5. З останнього впливає необхідність достатньо високої попередньої математико-статистичної підготовки студентів. Тому перед вивченням основ класичної теорії тестування студентам важливо пригадати такі ключові поняття теорії ймовірностей та математичної статистики, як випадкова величина та її розподіл, дисперсія і стандартне відхилення, стандартна похибка, коваріація, кореляція.

6. Не варто розглядати різні види надійності (ретестову, надійність еквівалентних частин тесту тощо) як окремі поняття. Слід наголосити, що існує єдине поняття коефіцієнта надійності, а усі так звані види надійності є лише емпіричними оцінками цього абстрактного коефіцієнта, залежними від способу організації дослідження тесту на надійність.

7. При вивченні різних видів надійності слід акцентувати на відповідних джерелах похибок вимірювання. Варто подати відповідну інформацію у вигляді таблиці.

8. При обговоренні різних джерел похибок слід викласти основи узагальненої теорії тестів (*Generalizability Theory*) як засобу

дослідження впливу різних джерел похибок на точність вимірювання.

9. Разом з тим доцільно показати, як можна визначати частку дисперсії спостережених оцінок, зумовлену впливом того чи іншого джерела похибок, порівнюючи між собою різні види оцінок коефіцієнта надійності.

Розглянемо ту частину матеріалу теми «Надійність», яку можна вважати опорною. Зауважимо, що тут не розглядаються такі важливі теми, ознайомлення з якими є бажаним у загальному курсі освітніх вимірювань, як надійність батарей тестів з визначеними ваговими коефіцієнтами субтестів, та узагальнена теорія тестів.

1. Загальне поняття надійності та похибки вимірювання

Поняття надійності, разом з поняттям валідності, є фундаментальною характеристикою тесту, без якої тестування не може вважатися вимірюванням. Разом з тим, у порівнянні з валідністю, надійність є більш технічною характеристикою, яка стосується насамперед проблеми точності вимірювання. У повсякденній мові ми називаємо надійним помічника або надійним другом людину, на яку можна покласти у певній складній ситуації, тобто людину, дії якої у визначених умовах є цілком прогнозованими. Подібно до цього, тест вважається надійним, якщо його багаторазове використання у схожих умовах приводить до схожих же результатів. Ця цілком зрозуміла з практичної точки зору вимога, однак, не може вважатися строгим визначенням надійності, оскільки поняття схожості можна надто вільно трактувати: схожість може бути більшою або меншою, сильнішою або слабшою. Проте ми не можемо замінити тут слово «схожі» на «однакові», оскільки після такої заміни практична перевірка виконання цієї вимоги стає неможливою.

Справді, навіть якщо один і той же тест пред'являти одній і тій же групі екзаменованих, ми змушені будемо робити це, як мінімум, у різні моменти часу. Але ж психічні процеси, які невпинно протікають у мозку людини, наявність пам'яті, тренуваності, здатності до навчання призводять до того, що результати першого тестування неодмінно впливатимуть на

результати другого, і ми отримаємо вже для двох сеансів тестування дві різні, хоча, можливо, й схожі ситуації.

Навіть для простих фізичних вимірювань спостерігається щось подібне. Якщо, приміром, зважити з великою точністю кілька разів дерев'яний брусок, то результати різних зважувань будуть дещо відмінними. Іншими словами, вимірюванню властива певна *похибка*. Важливим завданням є ідентифікація джерел похибки вимірювання. Якщо між зважуваннями бруска будуть достатньо великі проміжки часу, і вологість повітря буде при різних зважуваннях різною, істотним джерелом похибки вимірювання буде здатність дерева змінювати свою вологість, і, як наслідок, вагу. Інші джерела похибок, такі, як ретельність зчитування показів інструменту зважування, при цьому можуть бути такими, що їх середній результат буде близьким до нуля внаслідок взаємної компенсації похибок з від'ємними і додатними значеннями.

Різні схеми організації повторного психометричного вимірювання неодмінно містять різні істотні джерела похибок, які призводять до того, що результати вимірювання щоразу, взагалі кажучи, будуть дещо відмінними. Так, якщо групі осіб двічі пред'являється один і той же тест, таким джерелом мінливості похибки є мінливість часових інтервалів між першим і другим сеансом тестування; якщо групі осіб пред'являються у різні моменти часу різні форми одного і того ж тесту (їх називають паралельними), то з'являється додаткове джерело похибки – відмінність у змісті завдань паралельних форм; якщо результати одиничного тестування оцінюються групою експертів незалежно один від одного, то істотним джерелом похибки є відмінність у вподобаннях та критеріях оцінювання між експертами. Крім того, оскільки тест складається з окремих завдань, узгодженість у результатах між окремими завданнями та завданнями і тестом в цілому теж є предметом надійності.

Відповідно до того, яка схема повторного використання тесту застосовується, можна говорити про різні види надійності. Надійність при цьому будемо шукати у числовому вираженні. Але теоретично більш правильний підхід полягає у тому, що для даного тесту існує деяка ідеальна величина, яка називається

коефіцієнтом надійності, а різні схеми практичного дослідження надійності дають різні *оцінки* цього коефіцієнта. Для уведення поняття коефіцієнта надійності далі розглянемо так звану класичну модель тестової оцінки. Цю модель, разом із деякими припущеннями, та висновками, які з них випливають, називають класичною теорією тестування (*CTT – Classical Test Theory*). Пізніше познайомимося з так званою узагальненою теорією (*Generalizability Theory*), яка дозволяє в окремих випадках ідентифікувати вплив різних джерел похибок вимірювання.

2. Класична модель тестової оцінки

Розробка моделі істинної оцінки тестованого була розпочата ще Чарльзом Спірменом (*Charles Spearman*), а потім продовжена різними дослідниками.

Центральним положенням класичної теорії тестування є твердження про те, що спостережена тестова оцінка X_{pf} , яку отримав екзаменований p в результаті виконання ним форми f даного тесту, є сумою двох складових – істинної оцінки екзаменованого T_p та похибки вимірювання E_{pf} :

$$X_{pf} = T_p + E_{pf}.$$

Істинна оцінка особи T_p відповідає її рівню вираженості вимірюваної якості і є незмінною для різних форм тесту. Форми тесту тут вважаються *строго паралельними*, тобто вони задовольняють наступним чотирьом вимогам.

1. Вони мають ідентичні специфікації.
2. Розподіли спостережених оцінок при пред'явленні різних форм різним однаковим за об'ємом репрезентативним вибіркам з популяції екзаменованих є однаковими:

$$F(X_f) = F(X_g) = F(X_h) = \dots$$

3. Результати пред'явлення цим вибіркам екзаменованих будь-яких двох форм мають однакову коваріацію:

$$\sigma_{X_f X_g} = \sigma_{X_f X_h} = \sigma_{X_h X_g} = \dots$$

4. Якщо Z – деяка міра тієї ж самої або іншої якості осіб, коваріація результатів пред'явлення різних форм з Z є однаковою:

$$\sigma_{X_f Z} = \sigma_{X_g Z} = \sigma_{X_h Z} = \dots$$

Наступне припущення полягає у тому, що якщо экзаменований складає повторно різні строго паралельні форми тесту за умови, що попереднє тестування жодним чином не впливає на наступне (повне стирання з пам'яті), середнє значення похибок вимірювання, за умови наближення сеансів тестування до безмежності, прямує до нуля:

$$E_f(E_{pf}) = 0.$$

(тут E_f означає математичне очікування по множині строго паралельних форм тесту).

Ще одне припущення полягає у тому, що якщо будь-яка строго паралельна форма тесту пред'являється групі экзаменованих, очікуване середнє похибок вимірювання наближається до нуля при прямуванні кількості экзаменованих до нескінченності:

$$E_p(E_{pf}) = 0.$$

(тут E_p означає математичне очікування по множині усіх экзаменованих).

З цих припущень випливає, що коваріація між істинними балами та похибками вимірювання для будь-якої з паралельних форм тесту дорівнює нулю, і коваріація між похибками вимірювання для будь-яких двох паралельних форм тесту теж дорівнює нулю. Іншими словами, між істинними балами і похибкою вимірювання при адмініструванні однієї форми тесту, а

також між похибками вимірювання при адмініструванні різних форм тесту існує лінійна незалежність. З іншого боку, істинні оцінки та похибки вимірювань як компоненти спостережених оцінок корелюють з ними.

З того, що істинні оцінки екзаменованих і похибки вимірювання некорельовані, випливає той ключовий факт, що для окремої форми тесту дисперсія спостережених оцінок є сумою дисперсій істинних оцінок і похибок вимірювання:

$$D(X_f) = D(T) + D(E_f).$$

Також з припущень класичної моделі випливає, що коваріація між спостереженими оцінками, отриманими за паралельні форми тесту, дорівнює дисперсії істинних оцінок:

$$\sigma_{X_f X_g} = \sigma_T^2.$$

Поділивши на добуток середньоквадратичних відхилень спостережених балів за відповідними формами тесту, отримаємо коефіцієнт кореляції:

$$\rho_{X_f X_g} = \frac{\sigma_{X_f X_g}}{\sigma_{X_f} \sigma_{X_g}} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}.$$

Цю величину назвемо коефіцієнтом надійності. Таким чином, *коефіцієнт надійності* – це коефіцієнт кореляції між оцінками за гіпотетичні строги паралельні форми тесту.

Ще один важливий факт випливає також з наведених рівностей: коефіцієнт надійності дорівнює квадрату коефіцієнта кореляції між істинними та спостереженими оцінками при однократному тестуванні. Як зазначалося раніше, квадрат коефіцієнта кореляції між двома змінними визначає частку дисперсії, яку одна змінна привносить у дисперсію іншої змінної. Таким чином, величина

$$1 - \rho_{X_f X_g} = 1 - \rho_{XT}^2$$

визначає частку загальної дисперсії спостережених оцінок, зумовлену дисперсією похибки вимірювання.

Зауважимо, що припущення класичної теорії тестування не вимагають, щоб для двох різних осіб, яким пред'являлися повторно різні паралельні форми тесту, мінливість спостережених балів була однаковою. Теорія і практика вимірювань свідчать, що мінливість оцінок за різні форми тесту з завданнями, виміряними за дихотомічною шкалою, є меншою для осіб з дуже великими або дуже малими істинними оцінками, ніж для осіб з істинними оцінками, близькими до центру розподілу. Стандартне відхилення σ_{T_p} для осіб з рівнем T_p називається *умовною стандартною похибкою вимірювання*. В свою чергу, *безумовна стандартна похибка вимірювання* визначається як корінь квадратний з середньої очікуваної по всій групі екзаменованих дисперсії похибки вимірювання. Це означає, що стандартна похибка вимірювання може бути визначеною лише відносно певного розподілу істинних оцінок. Для вибірок з різних популяцій екзаменованих ця величина, взагалі кажучи, буде різною. Тому називати її безумовною можна лише з урахуванням цієї обставини.

Між безумовною стандартною похибкою вимірювання, коефіцієнтом надійності та дисперсією спостережених оцінок існують такі співвідношення:

$$\begin{aligned}\sigma_E &= \sqrt{\sigma_X^2 (1 - \rho_{X_f X_g})}; \\ \rho_{X_f X_g} &= 1 - \frac{\sigma_E^2}{\sigma_X^2}; \\ \sigma_X^2 &= \frac{\sigma_E^2}{1 - \rho_{X_f X_g}}.\end{aligned}$$

Таким чином, при наявній дисперсії спостережених оцінок, надійність тесту можна трактувати як через термін коефіцієнта

надійності, так і за допомогою поняття стандартної похибки вимірювання.

3. Відмінність між поняттями надійності та валідності

Як ми вже зазначали, надійність тесту стосується точності вимірювання. Натомість, валідність пов'язана з самою природою атрибутів, які вимірюються. Не валідний тест може виявитися цілком надійним. Припустимо, що групі екзаменованих з математики помилково пред'являвся тест з мови. І внутрішня узгодженість тестових завдань, і повторне пред'явлення цього тесту можуть вказувати на високу надійність, тоді як змістова валідність його, очевидно, є абсолютно неприйнятною. З іншого боку, не надійний тест ніколи не може вважатися валідним, тому в схему комплексного дослідження валідності тесту входить також і дослідження його надійності.

Надійність тесту стосується лише ситуацій, у яких інструменти вимірювання з подібною якістю застосовуються у подібний спосіб. Валідність же має справу також із співставленням результатів даного вимірювання з результатами альтернативних вимірювань тієї самої риси чи конструкту, або й навіть з результатами вимірювання інших якостей (дискримінантна валідність). Кореляція між двома альтернативними тестами з мови може розглядатися як оцінка надійності тесту – у тій мірі, у якій альтернативні тести можуть вважатися паралельними формами одного тесту. Повторне тестування за допомогою одного і того ж тесту однієї і тієї ж групи осіб дасть іншу оцінку коефіцієнта надійності. Але кореляція між оцінками за два цілком різні тести, скажімо, розробленими різними тестовими компаніями, вже не може розглядатися як оцінка коефіцієнта надійності, тобто різні тести, призначені для вимірювання нехай навіть і одного й того ж конструкту, не можуть розглядатися як повторне використання одного інструменту вимірювання.

Разом з тим, надійність, як і валідність, стосується не стільки самого тесту, скільки очікуваної інтерпретації та використання його результатів.

Розглянемо тепер різні схеми дослідження надійності тесту, які дозволяють знайти різні оцінки коефіцієнта надійності.

3. Випадок повторного тестування

Ретестова надійність. Найбільш очевидний метод оцінювання надійності вимірювання – його повторне застосування до однієї і тієї ж групи осіб. У цьому випадку обчислюється коефіцієнт кореляції між двома отриманими вибірками результатів. Оскільки обидва рази використовується той самий тест, не виникає проблем із забезпеченням строгої паралельності форм тесту.

Але виконати два вимірювання для одних і тих же суб'єктів одночасно неможливо. Тому потрібно визначити, яким має бути проміжок часу між двома тестуваннями. Саме тривалість цього проміжку є тим додатковим джерелом похибки вимірювання, який впливає на точність вимірювання і на показники надійності тесту. В свою чергу, на вибір проміжку часу між тестуваннями сильно впливає те, яка саме якість вимірюється. Деякі психологічні конструкти є цілком стійкі у часовому вимірі, тому для них вплив часового інтервалу між тестуваннями є неістотним. Для інших якостей характерна помірна мінливість у часі. Наприклад, коефіцієнт інтелекту (IQ) для дітей мало змінюється протягом такого періоду, як дошкільний, і два тестування IQ протягом цього періоду можуть показати високу надійність вимірювання. Якщо ж одне тестування провести у дошкільному віці, а інше – у старшому шкільному, або у дорослому віці, слід очікувати набагато більш слабкої кореляції результатів. Взагалі кажучи, при повторному вимірюванні фактично будь-якої якості кореляція результатів із зростанням інтервалу часу між вимірюваннями зменшується.

Нарешті, ретестовий метод зовсім не підходить для оцінки надійності вимірювання навчальних досягнень. Припустимо, що повторно за допомогою одного і того ж тесту вимірюється рівень навчальних досягнень з математики. Під час першого виконання тесту учні мимоволі запам'ятовують відповіді на завдання, у правильності розв'язання яких вони упевнені. При повторному виконанні тесту вони швидко справляються з цими завданнями і

мають більше часу для роботи над нерозв'язаними раніше завданнями. Крім того, під час перерви між тестуваннями їх мозок мимоволі, свідомо чи підсвідомо, працював над нерозв'язаними завданнями, і це також збільшує шанси справитися з цими завданнями. Збільшення ж часу між двома тестуваннями, хоча й послаблює ефект запам'ятовування, породжує нові проблеми, адже у цьому випадку учні або продовжували вивчати математику і розвивати свій рівень логіко-математичного мислення, або, навпаки, забували пройдений матеріал і втрачали набуті знання й уміння. Перше призведе до завищеної оцінки істинного значення коефіцієнта надійності вимірювання, друге – до заниженої. В обох випадках істинний рівень навчальних досягнень учнів змінювався.

Оскільки інтервал часу між двома вимірюваннями обирається користувачами тесту довільно, можна формально говорити про вплив вибірки часових інтервалів як джерела похибки при оцінюванні коефіцієнта надійності.

В усіх випадках зменшення часу між двома тестуваннями до нуля, тобто повторне тестування без перерви, породжує принаймні ще одну проблему – вплив втомлюваності тестованих.

Використання паралельних форм тесту. Нехай тепер у двох сеансах тестування екзаменованим пред'являлися різні (але паралельні) форми тесту. Для тестів навчальних досягнень вплив вибірки часових інтервалів, особливо коротких, можна вважати у цьому випадку більш слабким. Але цей вплив усе ж залишається, оскільки учні запам'ятовують не тільки конкретні розв'язки завдань з попереднього тестування, але й загальні методи та підходи до їх розв'язання, а вони мають бути спільними для обох форм тесту. Таким чином, паралельні форми тесту в аспекті принципів виконання завдань можуть виявитися однією і тією ж формою. Хоча використання паралельних форм тесту поширене на практиці, оскільки нівелює такі небажані дії, як пряме запам'ятовування відповідей чи списування, слід пам'ятати, що ефект «натаскування» учнів на одній з форм істотно впливає на успішність виконання ними іншої форми. Щоправда, якщо тренуваність усіх тестованих між двома сеансами тестування

зросте більш-менш рівномірно, це не вплине на величину коефіцієнта надійності.

При повторному тестуванні за допомогою паралельних форм тесту з'являється принципово нове, у порівнянні з ретестовим методом, джерело похибки оцінки коефіцієнта надійності – *вибірка змісту завдань*. Слід розуміти, що поняття строго паралельних форм тесту, сформульоване вище, є ідеальним, на практиці мало досяжним. Незважаючи на вимогу однакових специфікацій для відповідних завдань у двох паралельних формах тесту, їх зміст все ж є відмінним. Наприклад, один і той же учень може знати, скільки буде 6 помножити на 8, і не знати, скільки буде 6 помножити на 7, хоча формально можна вважати ці два завдання паралельними.

Таким чином, при визначенні оцінки коефіцієнта надійності методом повторного використання паралельних форм тесту існують два істотні джерела похибки – вибірка часових інтервалів і вибірка змісту завдань.

4. Випадок одноразового тестування

Дослідження надійності вимірювання методами повторного тестування є затратним з точки зору людських та матеріальних ресурсів, а у деяких випадках – взагалі неприйнятним. Оскільки тест складається з багатьох окремих завдань, з'являється ідея штучного поділу тесту на частини, які вважалися б більшою чи меншою мірою паралельними формами тесту. У цьому випадку для дослідження надійності достатньо провести однократне тестування вибірки з цільової популяції осіб. Оцінку коефіцієнта надійності, отриману в такий спосіб, називають *оцінкою внутрішньої узгодженості тесту*.

В усіх випадках оцінювання коефіцієнта надійності за допомогою результатів одноразового тестування тест ділиться на частини і обчислюється кореляція між результатами виконання цих частин репрезентативною вибіркою осіб.

Оскільки обрані частини тесту не можуть бути строго паралельними формами, вводиться поняття *еквівалентних частин*. Це поняття означає послаблення вимог порівняно з поняттям

строкої паралельності. Пригадаємо, що поняття строгої паралельності форм тесту включає чотири вимоги: однаковість специфікацій завдань, ідентичність розподілів спостережених оцінок, однаковість коваріації між усіма парами форм, однаковість коваріації між кожною з форм та результатами іншого вимірювання. Були запропоновані різні означення еквівалентності форм, кожне з яких тією чи іншою мірою послаблює ці вимоги.

Першим таким означенням було уведене Лордом та Новіком (Lord & Novick, 1968) поняття *τ-еквівалентності*. Це поняття залишає у силі вимогу незмінності істинної оцінки екзаменованого по множині всіх форм, але не вимагає, щоб вимірювання за дома формами виконувалося з однаковою точністю – дисперсія похибки може бути для різних форм різною. Прикладом є дві форми, які відрізняються лише кількістю завдань. Для таких форм виконуються лише дві останні з чотирьох вимог до строго паралельних форм. Розподіли спостережених оцінок за цими формами мають однакові очікувані середні значення, але можуть мати різні очікувані дисперсії.

Близьким до поняття *τ-еквівалентності* є запропоноване тими ж авторами поняття *істотної τ-еквівалентності*. Це поняття допускає, що для всіх екзаменованих існує константа, на яку відрізняються їх істинні бали, отримані за різними двома формами. Оскільки значення коваріації не чутливе до зміни значень однієї з вибірок на одну й ту ж константу, то й у цьому випадку залишаються виконаними останні дві з чотирьох вимог до строго паралельних форм. При цьому розподіли спостережених оцінок за цими формами можуть мати різні очікувані середні значення і різні очікувані дисперсії.

Однорідні форми (Jöreskog, 1971) – ще одне поняття, яке означає послаблення вимог паралельності. У доповнення до припущень істотної *τ-еквівалентності*, це поняття допускає також існування константи-множника, на який відрізняються істинні оцінки, отримані за двома формами:

$$T_{pf} = b_{fg}T_{pg} + C_{fg},$$

де T_{pf} – істинна оцінка особи p за виконання форми f , T_{pg} – істинна оцінка особи p за виконання форми g , b_{fg} та C_{fg} – відповідні константи. Іншими словами, для однорідних форм тесту допускається лінійна залежність між істинними балами екзаменованих.

Зауважимо, що однорідність означає також істотну τ -еквівалентність, яка, в свою чергу, включає у себе звичайну τ -еквівалентність.

Формули Спірмена-Брауна та Рюлона-Гуттмана. У 1910 році Спірмен та Браун запропонували формулу для оцінки коефіцієнта надійності за двома строго паралельними половинами тесту X_1 і X_2 :

$$\rho_{XX'} = \frac{2\rho_{X_1X_2}}{1 + \rho_{X_1X_2}}.$$

Тут і далі $\rho_{XX'}$ означає відповідну оцінку коефіцієнта надійності. Формула Спірмена-Брауна враховує той факт, що кореляція оцінюється для тесту, удвічі довшого від його половин.

Альтернативна формула Рюлона-Гуттмана, вже для істотно τ -еквівалентних форм, має вигляд:

$$\rho_{XX'} = 1 - \frac{\sigma_{X_1 - X_2}^2}{\sigma_X^2}.$$

Тут в чисельнику дробу стоїть дисперсія різниць між спостереженими балами за відповідні половини тесту, у знаменнику – дисперсія балів за весь тест. Для строго паралельних форм формули Спірмена-Брауна та Рюлона-Гуттмана дають однаковий результат. У інших випадках формула Спірмена-Брауна дає дещо більше значення, ніж формула Рюлона-Гуттмана. Наприклад, якщо дисперсії частин дорівнюють 6 і 8 відповідно, а кореляція між ними 0,7, то за формулою Рюлона-Гуттмана отримаємо оцінку 0,819, а за формулою Спірмена-Брауна – 0,824. Припущення про істотну τ -еквівалентність зазвичай є більш

прийнятним, ніж про строгу паралельність, але у цьому випадку використання формули Спірмена-Брауна не є достатньо обґрунтованим. Для однорідних форм не існує строгої формули оцінки коефіцієнта надійності. Такі формули існують лише при додаткових припущеннях. Так, припустивши, що дисперсії істинних оцінок і похибки вимірювання є такими, які б одержувалися лише внаслідок простої зміни довжин частин тесту, для цих части можна обчислити ефективні довжини:

$$\lambda_1 = \frac{\sigma_{X_1}^2 + \sigma_{X_1 X_2}}{\sigma_X^2},$$

$$\lambda_2 = 1 - \lambda_1.$$

Тоді справедлива формула Ангофа-Фелдта:

$$\rho_{XX'} = \frac{4\sigma_{X_1 X_2}}{\sigma_X^2 - \frac{(\sigma_{X_1}^2 - \sigma_{X_2}^2)^2}{\sigma_X^2}}.$$

Рекомендується для випадку, коли відношення дисперсій спостережених балів за частини тесту (більшої до меншої) не перевищує 1,15, використовувати просту в обчисленні формулу Спірмена-Брауна, хоча формула Рюлона-Гуттмана є більш прийнятною, а для випадку, коли це відношення знаходиться у межах між 1,15 та 1,30, використовувати формулу Рюлона-Гуттмана. Якщо ж це відношення є більшим від 1,30, слід використовувати формулу Ангофа-Фелдта.

Методи поділу тесту на дві частини. При визначенні, які завдання до якої частини тесту слід віднести для дослідження внутрішньої узгодженості, потрібно керуватися двома основними принципами. По-перше, слід добиватися максимальної паралельності частин. Зазвичай тест складається з завдань, розташованих у порядку зростання їх труднощі. У цьому випадку буває достатнім простий поділ тесту за принципом: завдання з

парними номерами відносяться до однієї частини, завдання з непарними номерами – до іншої частини.

Галіксен (Gulliksen, 1950) описує наступний метод поділу тесту. Завдання зображуються точками на площині відповідно до їх труднощі та коефіцієнта кореляції між завданням та тестом в цілому. Завдання, близькі між собою візуально, групуються у пари чи більші кластери. Далі всередині кожного кластеру завдання випадковим чином розподіляються до частин тесту. Цю процедуру за потреби можна виконувати окремо для певних частин цільової області вимірювання або різних форматів тестових завдань, щоб максимально забезпечити паралельність як у аспекті контенту, так і у статистичних властивостях.

По-друге, якщо у тесті є групи завдань, об'єднані спільним змістом, кожену групу слід всю відносити до однієї з частин. Наприклад, якщо тест на ефективність читання складається з кількох текстів, після яких ідуть групи завдань, що стосуються цих текстів, то віднесення завдань однієї групи до різних частин тесту може призвести до штучного завищення кореляції, якщо припустити, що різні групи завдань відносяться до різних частин цільової області вимірювання. Якщо всі завдання груп віднести до однієї, тієї чи іншої частини тесту, це може збільшити дисперсію частин, але не коваріацію між частинами, що дозволяє інтерпретувати вплив вибірки змісту як джерело похибки вимірювання.

Поділ тесту більше ніж на дві частини. Формули К'юдера-Річардсона та альфа Кронбаха. У більшості випадків для дослідження внутрішньої узгодженості тесту краще ділити тест більше ніж на дві частини. Якщо у тесті немає зв'язаних однаковим змістом груп завдань, то тест бажано ділити на максимальну кількість частин – по одному завданню у частині.

Потрібно враховувати, що чим більш однорідною є цільова область вимірювання, тим більшою внутрішньою узгодженістю повинен володіти тест. Наприклад, для тесту, який перевіряє лише уміння учнів множити числа, внутрішня узгодженість має бути більшою, ніж для тесту на всі арифметичні операції.

Для тесту з дихотомічними відповідями на завдання К'юдеру та Річардсону належить кілька формул оцінки коефіцієнта внутрішньої узгодженості, з яких найбільш часто використовується так звана формула KR-20:

$$\rho_{XX'} = \frac{n}{n-1} \cdot \frac{\sigma_X^2 - \sum_{i=1}^n p_i q_i}{\sigma_X^2},$$

де n – кількість завдань у тесті, σ_X^2 – дисперсія оцінок за тест, p_i і q_i – частки тих, хто справився і, відповідно, не справився з i -тим завданням.

Можна математично довести, що оцінка коефіцієнта надійності, отримана за формулою KR-20, дорівнює середньому оцінок, отриманих при поділі тесту на дві частини усіма можливими способами. Оскільки при розщепленні тесту на дві частини обирають такий поділ, який забезпечував би максимальну паралельність частин, то формула KR-20 дає дещо нижчий результат, ніж формули для поділу тесту на дві частини, описані вище. Таким чином, різниця між значеннями, знайденими за цими формулами, та формулою KR-20, є показником неоднорідності тесту.

Формула KR-20 у наведеному нами вигляді не підходить для тестів з політомічними відповідями (тобто такими відповідями, які можуть вважатися частково правильними). Для цього випадку існує більш універсальна формула, яку прийнято називати формулою *альфа Кронбаха*:

$$\rho_{XX'} = \frac{n}{n-1} \cdot \frac{\sigma_X^2 - \sum_{i=1}^n \sigma_i^2}{\sigma_X^2},$$

де σ_i^2 – дисперсія оцінок, отриманих за i -те завдання тесту.

Саме альфа Кронбаха як показник надійності обчислюється та публікується для тестів зовнішнього незалежного оцінювання в Україні.

5. Надійність оцінювача

В деяких тестах, таких як тести креативності чи проєктивні особистісні тести, а також у тестах навчальних досягнень з завданнями з розгорнутою відповіддю, велику роль відіграє суб'єктивізм оцінювача. Надійність оцінювача можна визначити, організувавши оцінювання двома незалежними фахівцями. Між двома наборами оцінок, виставлених цими фахівцями за тест, обчислюється звичайний коефіцієнт кореляції. Джерелом цієї похибки оцінки коефіцієнта надійності у цьому випадку є вибірковість оцінювачів.

6. Загальний огляд оцінок коефіцієнта надійності

Різні методи оцінки коефіцієнта надійності можна класифікувати у відповідності до кількості необхідних сеансів тестування та форм тесту. Подамо цю класифікацію у вигляді таблиці:

Таблиця 1.

Кількість тестувань	Кількість форм тесту	
	Одна	Дві
Одне	1) Метод розщеплення на еквівалентні половини 2) Формула К'юдера-Річардсона	4) Метод паралельних форм (безпосередній)
Два	3) Ретестовий метод	5) Метод паралельних форм (з часовим інтервалом)

Будь-яку оцінку коефіцієнта надійності можна інтерпретувати у частках дисперсії, спричиненої різними джерелами. Так, величина оцінки 0,85 означає, що 85% дисперсії результатів тестування спричинені мінливістю вимірюваної якості у цільовій популяції осіб, а решта 15% - дисперсією похибок.

Подамо у вигляді таблиці зв'язок різних оцінок коефіцієнта надійності з джерелами похибок:

Таблиця 2.

Вид оцінки коефіцієнта надійності	Джерела дисперсії похибок
1) Ретестовий	Часова вибіркковість
2) Паралельних форм (безпосередній)	Вибірковість змісту
3) Паралельних форм (з часовим інтервалом)	Часова вибіркковість плюс вибіркковість змісту
4) Еквівалентних половин тесту	Вибірковість змісту
5) KR-20 та альфа Кронбаха	Вибірковість змісту плюс неоднорідність змісту
6) Оцінювача	Відмінність між оцінювачами

Покажемо, як спеціально підібраний план дослідження надійності допомагає оцінити вплив різних джерел похибки вимірювання (Анастасі, Урбіна).

Нехай 100 учнів проходили тестування на креативність двічі з інтервалом у два місяці за допомогою паралельних форм тесту. Нехай оцінка коефіцієнта надійності за методом паралельних форм (з часовим інтервалом) складала 0,7. Нехай також метод еквівалентних половин за формулою Спірмена-Брауна дав для обох форм оцінку 0,8; у оцінюванні був задіяний додатковий експерт, який оцінював відібрану навмання половину учнівських робіт, і надійність оцінювача складала 0,92. Тоді вплив часового інтервалу та вибіркковості змісту дають $1 - 0,7 = 0,3$. Тут 1 означає 100% дисперсії. З іншого боку, вплив лише вибіркковості змісту дорівнює

$1 - 0,8 = 0,2$. Звідси отримуємо що $0,3 - 0,2 = 0,1$ – вплив вибіркковості часу між тестуваннями. Вплив заміни оцінювача дорівнює

$1 - 0,92 = 0,08$.

Тоді сумарна оцінка дисперсії похибок дорівнює

$$0,2 + 0,1 + 0,08 = 0,38.$$

Звідси істинна дисперсія, зумовлена відмінностями у рівні креативності учнів, дорівнює $1 - 0,38 = 0,62$.

Загалом дослідження впливу різних джерел похибки вимірювання є предметом розгляду спеціальної теорії, яку в українському перекладі можна назвати узагальненою теорією (точніше, теорією узагальнюваності – англ. Generalizability Theory), яка заснована на методах спеціального розділу математичної статистики дисперсійного аналізу.

Література

1. Анастаси А., Урбина С. Психологическое тестирование. – 7-е изд. – СПб.: Питер, 2007. – 688 с.: ил. – (Серия «Мастера психологии»). – ISBN 978-5-272-00106-1.

2. Educational Measurement. – 4th edition /Edited by Robert L. Brennan. – ACE, 2006.

3. Linda Crocker, James Algina. Introduction to classical and modern test theory. – Wadsworth: Thomson Learning, 1986. – 528 p.

ВИКЛАДАННЯ ПОНЯТТЯ НАДІЙНОСТІ ТЕСТУ, ЯК ОСНОВНОЇ ЙОГО СТАТИСТИЧНОЇ ХАРАКТЕРИСТИКИ

Паращук С.Д.

Під час апробації тесту необхідно оцінювати його за різними показниками. Ці показники стосуються як самого тесту в цілому, так і окремих складових його завдань. Зазвичай для отримання необхідних оцінок використовуються математико-статистичні методи, які становлять основу професійної підготовки фахівця з освітніх вимірювань.

Одним із важливих показників тесту є його надійність. Цей показник у певній мірі характеризує якість тесту. Перевірка тесту на надійність є обґрунтування якості тестових вимірювань.

Формули для обчислення надійності тесту вперше були запропоновані в рамках класичної теорії тестів. Основоположником цієї теорії є відомий британський психолог, автор факторного аналізу Чарльз Едвард Спірмен (1863-1945).

Вивчення поняття надійності майбутніми фахівцями з освітніх вимірювань не повинно обмежуватися знайомством лише з набором необхідних формул для обчислення надійності, а й повинно передбачати формальне обґрунтування цих формул і розуміння їх застосування. Оскільки підготовка таких фахівців на Україні розпочата недавно, то відбувається етап становлення нових навчальних дисциплін для відповідної спеціальності. При цьому уточнюється зміст дисциплін, розробляється їх методичне забезпечення. Метою даної статті є визначення змісту мінімального необхідного навчального матеріалу, який повинен вивчатися фахівцями з освітніх вимірювань і який дає уявлення про проблеми визначення надійності тесту.

При відборі навчального матеріалу використовувалися джерела [1-5], які можуть використовуватися при відборі навчального матеріалу. Підручник [1] містить цілий розділ присвячений ґрунтовному аналізу методів визначення надійності. Загалом для вивчення пропонуються такі питання:

1. Основні положення класичної теорії тестування.
2. Коефіцієнт надійності.
3. Дисперсія складених тестових оцінок.
4. Надійність складених тестових оцінок.
5. Процедури оцінювання надійності тесту. Формули для обчислення коефіцієнта надійності.
6. Фактори, що впливають на коефіцієнти надійності.
7. Наближене обчислення істинних оцінок.

Коротко охарактеризуємо зміст цих питань, зосереджуючи увагу на обов'язкових положеннях, що рекомендуються до вивчення.

1. Основні положення класичної теорії тестування.

Вивчення надійності тесту доцільно розпочати зі знайомства з основами класичної теорії тестів, оскільки саме в рамках цієї теорії вперше було запропоновані підходи до обчислення

надійності. Отримані формули дозволяли оцінювати надійність нормативно-орієтованих тестів.

Класична теорія тестів базується на класичній моделі істинної оцінки, яку запропонував Спірмен. Суть цієї моделі полягає в тому, що будь яка отримана тестова оцінка X може розглядатися як композиція двох гіпотетичних оцінок – істинної оцінки T і випадкової похибки вимірювання E . Ця композиція записується у вигляді суми $X = T + E$.

Істинна оцінка або істинний бал випробовуваного трактується як середнє спостережуваних оцінок, отриманих по нескінченному числу повторюваних тестувань при використанні одного і того ж тесту. Також можна вважати, що істинний бал випробовуваного отримується в результаті тестування по гіпотетичній генеральній сукупності завдань нескінченного тесту. З математичної точки зору істинна оцінка є математичне сподівання $M(X)$ випадкової величини X .

Похибка вимірювання відображає степінь відхилення спостережуваного балу від істинного балу випробовуваного. Причини виникнення похибок вимірювання можуть бути різні. Зазвичай виділяють два типи помилок: систематичні та випадкові. До систематичних відносять помилки, що породжуються недостатньою якістю тесту.

У класичній теорії тестів використовується поняття «паралельні форми тесту». Дві форми тесту вважаються паралельними, якщо вони розроблені на основі одної специфікації, складаються із однакової кількості завдань попарно рівної трудності зі співпадаючими характеристиками. Для кожної вибірки випробовуваних істинні бали випробовуваних на обох формах повинні бути однаковими. Також для спостережуваних балів випробовуваних форми повинні мати однакові розподіли середніх, дисперсії. Коваріації результатів тестування повинні співпадати для різних пар форм.

Загалом класична теорія тестів ґрунтується на наступних п'яти основних положеннях.

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

$$X = T + E, \tag{1}$$

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

$$T = M(X). \quad (2)$$

Істинна оцінка є статистичним поняттям, що базується на очікуваному значенні, отриманому в даному процесі вимірювання.

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

$$M(E) = 0.$$

Це положення ґрунтується на ідеї тестування нескінченної сукупності випробовуваних за допомогою даного набору паралельних форм тесту.

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

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

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

$$\rho_{E_1 E_2} = 0.$$

Серед наведених п'яти положень перші два є означеннями. Решту три можна отримати логічними міркуваннями із означень. За наявності достатнього часу доцільно навести ці міркування та проілюструвати наочними матеріалами. 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.

Зауважимо також, що в навчальному посібнику Челишкової [2] наводяться інші постулати класичної теорії тестів.

2. Коефіцієнт надійності. У цьому питанні потрібно обґрунтувати необхідність поняття надійності тесту та дати його визначення. Також доцільно подати виведення формули для коефіцієнта надійності.

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*. [2]

У класичній теорії тестів надійність обчислюють за допомогою коефіцієнта кореляції за результатами випробування по двох паралельних формах. У підручнику [1] вводиться спочатку поняття показника надійності

$$\rho_{XT} = \frac{\sigma_T}{\sigma_X}, \quad (3)$$

як коефіцієнта кореляції між істинними оцінками (Т) і оцінками (Х), що спостерігаються за результатами множини тестувань, які повторюються. Цей показник рівний відношенню стандартного відхилення істинних оцінок до стандартного відхилення оцінок, що спостерігаються.

Коефіцієнт надійності означається як кореляція між оцінками по паралельних тестових формах. Наведемо два варіанти його виведення.

Позначимо відхилення оцінок і-того випробовуваного по двох паралельних тестах як $x_1 = X_{i1} - \bar{X}_1$ і $x_2 = X_{i2} - \bar{X}_2$. За класичною моделлю істинної оцінки $X_{i1} = T_{i1} + E_{i1}$. Для генеральної сукупності випробуваних цю рівність у відхиленнях можна переписати так $X_{i1} - \bar{X}_1 = (T_{i1} - \bar{T}_1) + (E_{i1} - \bar{E}_1)$ або коротко $x_1 = t_1 + e_1$. У наведених формулах використані позначення з індексами для $\bar{X}, \bar{T}, \bar{E}$ - відповідних середніх величин. Аналогічно, для другого паралельного тесту отримуємо теж рівність у відхиленнях $x_2 = t_2 + e_2$.

Формула кореляції між оцінками X_{i1} і X_{i2} , що спостерігаються по двох паралельних тестах, має вигляд

$$\rho_{X_1 X_2} = \frac{\sum_{i=1}^N (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)}{N \cdot \sigma_{X_1} \cdot \sigma_{X_2}} \text{ або через відхилення}$$

$$\rho_{X_1 X_2} = \frac{\sum_{i=1}^N X_1 \cdot X_2}{N \cdot \sigma_{X_1} \cdot \sigma_{X_2}}.$$

Її можна записати так

$$\rho_{X_1 X_2} = \frac{\sum_{i=1}^N (t_1 + e_1) \cdot (t_2 + e_2)}{N \cdot \sigma_{X_1} \cdot \sigma_{X_2}}.$$

Якщо виконати алгебричні перетворення, то вираз набуде вигляду

$$\rho_{X_1 X_2} = \frac{\sum_{i=1}^N t_1 \cdot t_2}{N \sigma_{X_1} \sigma_{X_2}} + \frac{\sum_{i=1}^N t_1 \cdot e_2}{N \sigma_{X_1} \sigma_{X_2}} + \frac{\sum_{i=1}^N t_2 \cdot e_1}{N \sigma_{X_1} \sigma_{X_2}} + \frac{\sum_{i=1}^N e_1 \cdot e_2}{N \sigma_{X_1} \sigma_{X_2}}.$$

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

$$\rho_{X_1 X_2} = \frac{\sum_{i=1}^N t_1 \cdot t_2}{N \sigma_{X_1} \sigma_{X_2}} = \frac{\sum_{i=1}^N t_1^2}{N \sigma_{X_1}^2}$$

або

$$\rho_{X_1 X_2} = \frac{\sigma_T^2}{\sigma_X^2}. \quad (3)$$

Таким чином, *коефіцієнт надійності* може бути математично отриманий як *відношення дисперсії істинної оцінки до дисперсії оцінки, яка спостерігається* (або є квадратом показника надійності).

Важливо зауважити, що розглянутий коефіцієнт надійності для ряду тестових оцінок є просто теоретичним поняттям. Його можна було б отримати, якби насправді існували паралельні тести. Окрім того, невідомі істинні оцінки випробовуваних, а значить неможливо обчислити дисперсію істинної оцінки.

Іншу формулу для коефіцієнта надійності можна отримати із важливої рівності для дисперсій

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2,$$

яка виконується в класичній теорії.

З неї легко отримати співвідношення $\frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}$ або з урахування (3)

$$\rho_{XX'} = 1 - \frac{\sigma_E^2}{\sigma_X^2}. \quad (4)$$

Формула (4) є іншим виразом для коефіцієнта надійності. За нею теж не можна обчислити коефіцієнт надійності, бо неможливо визначити похибки істинних оцінок. З цієї формули випливає, що коефіцієнт надійності не може бути більшим за 1. В ідеалі він рівний 1 тоді, коли дисперсія похибки рівна нулю або дисперсії істинної та спостережуваної оцінок – рівні.

З методологічної точки зору доцільно подати перший варіант виведення коефіцієнта надійності, тому що саме він характеризує коефіцієнт як кореляцію між тестовими балами по двох паралельних формах.

3. Дисперсія складених тестових оцінок. Зазвичай підсумкова оцінка за тест обчислюється як сума оцінок за кожне із завдань тесту. Окрім того, тест може складатися із субтестів і також загальна оцінка за тест буде складатися із оцінок субтестів.

Це приводить до необхідності введення поняття складеної тестової оцінки.

Припустимо, що є два набори оцінок X_{i1} та X_{i2} за два завдання для одної у тої ж генеральної сукупності випробовуваних. Для кожного випробовуваного складена оцінка по цих завданнях матиме вигляд $C_i = X_{i1} + X_{i2}$. Середнє значення складених оцінок дорівнюватиме $\bar{C} = \bar{X}_1 + \bar{X}_2$, а відхилення оцінок - $C_i - \bar{C} = (X_{i1} - \bar{X}_1) + (X_{i2} - \bar{X}_2)$, які коротко позначимо як $c_i = x_{i1} + x_{i2}$. Тоді дисперсія складених оцінок матиме вигляд

$$\sigma_C^2 = \frac{\sum_{i=1}^N c_i^2}{N} = \frac{\sum_{i=1}^N (x_{i1} + x_{i2})^2}{N} = \frac{\sum_{i=1}^N x_{i1}^2}{N} + \frac{\sum_{i=1}^N x_{i2}^2}{N} + \frac{\sum_{i=1}^N 2 \cdot x_{i1} \cdot x_{i2}}{N}.$$

За означенням $\frac{\sum_{i=1}^N x_{i1}^2}{N} = \sigma_1^2$, $\frac{\sum_{i=1}^N x_{i2}^2}{N} = \sigma_2^2$ - це дисперсії. Останній доданок перетворюється до виду

$$\frac{\sum_{i=1}^N 2 \cdot x_{i1} \cdot x_{i2}}{N} = \frac{2 \sum_{i=1}^N x_{i1} \cdot x_{i2}}{N} \cdot \frac{\sigma_1 \cdot \sigma_2}{\sigma_1 \cdot \sigma_2} = \frac{2 \sum_{i=1}^N x_{i1} \cdot x_{i2}}{N \cdot \sigma_1 \cdot \sigma_2} \cdot (\sigma_1 \cdot \sigma_2).$$

Оскільки $\frac{\sum_{i=1}^N x_{i1} x_{i2}}{N \sigma_1 \sigma_2}$ є кореляцією ρ_{12} , то $\frac{\sum_{i=1}^N 2 x_{i1} x_{i2}}{N} = 2 \rho_{12} \sigma_1 \sigma_2$.

Остаточно отримуємо формулу

$$\sigma_C^2 = \sigma_1^2 + \sigma_2^2 + 2\rho_{12}\sigma_1\sigma_2 \quad (5)$$

дисперсії складеної оцінки. Ця дисперсія не рівна сумі дисперсій окремих оцінок. Тому загальна дисперсія тесту перевищує суму дисперсій завдань тесту. Формулу (5) можна узагальнити на довільну скінченну кількість завдань. Вона використовується при виведенні надійності складеної оцінки.

4. Надійність складених тестових оцінок. При вивченні цього питання потрібно розглянути дві відомі формули для оцінки

надійності складених тестових оцінок – формулу Спірмена-Брауна та коефіцієнт альфа Кронбаха. Доцільно вивести ці формули та обговорити їх призначення.

Існують два методи визначення надійності складеної оцінки в термінах статистичних властивостей її компонентів.

1. Прогностична формула Спірмена-Брауна

$$\rho_{CC'} = \frac{k\rho_{ii'}}{1 + (k-1)\rho_{ii'}} \quad (6)$$

обчислює надійність складеної оцінки С як функцію від окремих компонентів. У цій формулі величина С визначає складену загальну оцінку по k строго паралельних тестових формах, $\rho_{ii'}$ - коефіцієнт надійності одної із паралельних форм (i-тої). Формула Спірмена-Брауна дозволяє оцінити надійність складеної оцінки по паралельних тестах у випадку, коли відома надійність одного з тестів.

2. У реальних ситуаціях тестування не можна бути впевненим, що всі форми строго паралельні. Тоді можна отримати нижню оцінку для коефіцієнта надійності складеної оцінки

$$\rho_{CC'} \geq \frac{k}{k-1} \left(\frac{\sum_{i \neq j} \sigma_{ij}}{\sigma_C^2} \right) \text{ або } \rho_{CC'} \geq \frac{k}{k-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_C^2} \right).$$

Вираз

$$\frac{k}{k-1} \left(\frac{\sum_{i \neq j} \sigma_{ij}}{\sigma_C^2} \right) \quad (7)$$

називається коефіцієнтом альфа Кронбаха. Якщо тест komponують із непаралельних субтестів, то можна оцінити нижню межу коефіцієнта надійності тесту 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.

Наведемо виведення формули Спірмена-Брауна. Нехай задано k паралельних форм тесту. Позначимо через X_{ij} тестові бали i -того випробовуваного по j -тій формі. Тоді складена оцінка i -того випробовуваного матиме вигляд

$$C_{iX} = X_{i1} + X_{i2} + \dots + X_{ik}.$$

За узагальненою формулою (5) дисперсія спостережуваної складеної оцінки дорівнює

$$\sigma_{CX}^2 = \sum_{j=1}^k \sigma_{iX}^2 + \sum_{i \neq j} \rho_{ij} \sigma_{iX} \sigma_{jX}. \quad (8)$$

У цю формулу ми добавили всюди індекс X , щоб підкреслити, що дисперсія обчислюється для спостережуваних балів випробовуваних. Сума $\sum_{i \neq j} \rho_{ij} \sigma_{iX} \sigma_{jX}$ складається із $k(k-1)$

коваріаційних доданків, індекси i, j позначають номери паралельних форм тесту. Оскільки тести паралельні, то всі кореляції ρ_{ij} рівні між собою, а також рівні між собою σ_{iX} .

Рівність (8) набуде вигляду $\sigma_{CX}^2 = k\sigma_{iX}^2 + k(k-1)\rho_{ij}\sigma_{iX}^2$ або $\sigma_{CX}^2 = k\sigma_{iX}^2(1 + (k-1)\rho_{ij})$. Коефіцієнт кореляції ρ_{ij} можна вважати коефіцієнтом надійності $\rho_{ii'}$ i -того тесту тому, що індекси i, j є номерами паралельних тестів. Формула (8) отримує такий остаточний вигляд

$$\sigma_{CX}^2 = k\sigma_{iX}^2(1 + (k-1)\rho_{ii'}). \quad (9)$$

Обчислимо тепер дисперсію складеної істинної оцінки. Складена істинна оцінка i -того випробовуваного має вигляд $C_{iT} = T_{i1} + T_{i2} + \dots + T_{ik}$. Істинні оцінки цього випробовуваного по всіх паралельних формах рівні. Тому $C_{iT} = kT_{ij}$, де T_{ij} - це істинна оцінка i -того випробовуваного по j -му паралельному тесту. Тоді дисперсія складеної істинної оцінки дорівнює

$$\sigma_{CT}^2 = \frac{\sum_{i=1}^N (k(T_{ij} - \bar{T}))^2}{N} = \frac{k^2 \sum_{i=1}^N (T_{ij} - \bar{T})^2}{N} = k^2 \sigma_{jT}^2,$$

де σ_{jT}^2 - дисперсія істинної оцінки по j -му паралельному тесту.

Остаточно отримуємо рівність

$$\sigma_{CT}^2 = k^2 \sigma_{iT}^2. \quad (10)$$

Тепер за формулою (3) обчислюємо надійність складеної оцінки

$$\rho_{CC'} = \frac{\sigma_{CT}^2}{\sigma_{CX}^2} = \frac{k^2 \sigma_{iT}^2}{k \sigma_{iX}^2 (1 + (k-1) \rho_{iI'})}.$$

Оскільки відношення $\sigma_{iT}^2 / \sigma_{iX}^2$ дорівнює коефіцієнту надійності $\rho_{iI'}$ i -того паралельного тесту, то остаточно отримуємо формулу (6).

5. Процедури оцінювання надійності тесту. Формули для обчислення коефіцієнта надійності. У даному питанні вивчаються практичні методи обчислення коефіцієнта надійності тесту. На практиці ми маємо справу зі скінченною вибіркою випробовуваних. З методичної точки зору доцільно використовувати різні позначення для теоретичних і практичних величин. Тому реально обчислюваний коефіцієнт надійності прийнято позначати як r_n , а теоретичний ми познали як $\rho_{XX'}$.

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 тестових форм. Друга група передбачає однократне тестування.

Методи двократного тестування.

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

$$(r_{XY})_{\text{пер}} = \frac{N \sum_{i=1}^N X_i Y_i - \left(\sum_{i=1}^N X_i \right) \left(\sum_{i=1}^N Y_i \right)}{\sqrt{\left(N \sum_{i=1}^N X_i^2 - \left(\sum_{i=1}^N X_i \right)^2 \right) \left(N \sum_{i=1}^N Y_i^2 - \left(\sum_{i=1}^N Y_i \right)^2 \right)}}, \quad (11)$$

де $(r_{XY})_{\text{пер}}$ - коефіцієнт надійності тесту за ре тестовим методом; X_i - індивідуальний бал i -го випробуваного в першому тестуванні; Y_i - індивідуальний бал i -го випробуваного в другому тестуванні. Такий 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 (11) і він називається *коефіцієнтом еквівалентності*. Чим вищий коефіцієнт еквівалентності, тим з більшою впевненістю можна розраховувати на те, що оцінки по різних формах можуть використовуватися як взаємно замінювані.

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 тесту випробовуваними. 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. За формулою (11) обчислюють коефіцієнт надійності $(r_{XY})_{\text{розщ}}$, 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)

$$r_n = \frac{2(r_{XY})_{\text{розщ}}}{1 + (r_{XY})_{\text{розщ}}}.$$

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

$$\hat{\alpha} = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_X^2} \right),$$

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.

6. 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

$$r'_n = \frac{kr_n}{1 + (k - 1)r_n},$$

де k – число разів, на яке збільшується довжина тесту; r_n – коефіцієнт надійності тесту до збільшення його довжини; r'_n – 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_n та задавши число k , можна обчислити очікувану надійність r'_n . З іншої сторони, маючи показники r_n та r'_n , можна обчислити число k .

7. Похибка вимірювання надійності. За формулою (4) коефіцієнт надійності в статистичних характеристиках набуває виду $r_n = 1 - \frac{S_E^2}{S_X^2}$. Виконавши перетворення, отримаємо

$$S_E = S_X \sqrt{1 - r_n}. \quad (12)$$

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.

7. Наближене обчислення істинних оцінок. Методи регресійного аналізу дозволяють прогнозувати оцінки істинних балів випробовуваних по розподілу спостережуваних балів і коефіцієнту надійності тесту. Регресійне рівняння має вигляд $T_i = \bar{X} + r_n (X_i - \bar{X})$ і воно може використовуватися для наближеного обчислення істинної оцінки випробовуваних.

Висновок. Запропонований до вивчення перелік питань загалом дає уявлення про складність проблеми визначення надійності тесту. Подальше вивчення надійності можна продовжити в рамках сучасної теорії тестів IRT для відповідних параметричних моделей, де використовується аналогічні методи, що розглянуті вище. Окрім того, доцільно ознайомити майбутніх

фахівців з підходами до визначення надійності критеріально-орієнтованих тестів, яка ґрунтується на теорії генералізації.

Література

1. Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов.- М.: Логос, 2010.- 668 с.
2. Чельшкова Н. Б. Теория и практика конструирования педагогических тестов : учебное пособие / Н. Б. Чельшкова. – М. : Логос, 2002. – 432 с.
3. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов.- М.: Прометей, 2000.- 168 с.
4. Ким В.С. Тестирование учебных достижений. Монография.- Усурийск: Изд-во УГПИ, 2007. -214 с.
5. Вимірювання в освіті: Підручник / За редакцією О.В.Авраменко. – Кіровоград: Лисенко В.Ф.,2011. – 360 с.

МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ ПРАКТИЧНИХ ЗАНЯТЬ ЗА ЗМІСТОВИМ МОДУЛЕМ «НАДІЙНІСТЬ ТЕСТУ»

Янчукова Н.В., Авраменко О.В.

Вступ. Обрахунок надійності тестової оцінки посідає основну позицію в практичному створенні тестових завдань. Надійність, як величина, вказує на те, що учасник тестування з плином певного проміжку часу може хоча б приблизно відтворити свій результат і, що його отриманий бал по тесту максимально відповідає оцінці реальних знань з деякої області. Огляд поняття надійності тесту з різних літературних джерел представлено нижче.

Надійність тесту – це характеристика методики тестування, що показує точність вимірів і стабільність результатів до впливу сторонніх випадкових факторів. Це означає, що невелика зміна умов тестування і стану учасників тестування повинно привести

до невеликих змін кінцевого результату. Надійність тесту пов'язана з поняттям стандартної помилки, чим вище надійність, тим менше стандартна помилка вимірів. Іноді помилка вимірів трактується як статистична величина, що показує степінь відхилення отриманого бала від істинного бала учасника.

Під надійністю (реліабільністю) вимірів розуміється степінь надійності, або точності, з якою може бути виміряний той чи інший конкретний признак. Надійність теста характеризує відтворюваність його результатів. Надійність характеризується коефіцієнтом надійності. Коефіцієнт надійності – це кореляційний коефіцієнт, що показує степінь збігу результатів тестування, що здійснюється в однакових умовах одним тестом.

З точки зору Л.Крокер і Дж.Алгини [4] надійність тесту розглядається, як гарантія того, що результати можуть бути хоча б приблизно відтворені при повторному тестуванні тих же самих тестованих в подібних умовах. В практичних умовах надійність – це ступінь, в якому відхилення оцінок тестованих залишаються відносно стабільними при повторному представленні тестованим того ж самого тесту чи його паралельних форм.

Поняття надійності і валідності педагогічного тесту надзвичайно важливі, оскільки саме вони характеризують тест інструмент для вимірювання. Тест з невідомими надійністю і валідністю непридатний для вимірів. [3]

Надійність та методи її обчислення розглядаються в багатьох роботах [1], [2], [7], [8], [12]-[14], де показаний математичний апарат обчислення коефіцієнтів надійності та представлено його обґрунтування.

Метою даної статті є висвітлення елементів математичного обґрунтування поняття надійності в теорії освітніх вимірювань та класифікація основних типів завдань по темі «Надійність тесту», а також вироблення рекомендацій до використання при підготовці магістрів спеціальності 8.18010022 «Освітні вимірювання».

1. Надійність тесту та методи її визначення. Наведемо основні теоретичні позиції змістового модуля «Надійність тесту» [3], [4]-[6], [9]-[11].

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 визначення коефіцієнта надійності полягає у виділенні «факторної дисперсії», що породжується в даному випадку впливом завдань та учасників тестування та «залишкової дисперсії», обумовленої випадковими причинами.

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, $b^* = b + \Delta$.

Обчислимо величини $S_{\text{загал}} = \sum_{i=1}^n \sum_{j=1}^k b_{ij}^2 - \frac{b^2}{nk}$,
 $S_{\text{учасн}} = \frac{1}{k} \sum_{i=1}^n b_i^2 - \frac{b^2}{nk}$, $S_{\text{завд}} = \frac{1}{n} \sum_{j=1}^k b_j^2 - \frac{b^2}{nk}$, $S_{\text{ост}} = S_{\text{загал}} - S_{\text{завд}} - S_{\text{учасн}}$,
де $b_i = \sum_{j=1}^k b_{ij}$, $b_j = \sum_{i=1}^n b_{ij}$, $b_{ij} = \sum_{i=1}^n b_{ij} \sum_{j=1}^k b_{ij}$.

Оцінімо дисперсії $D_{\text{учасн}} = \frac{S_{\text{учасн}}}{n-1}$,

$$D_{\text{завд}} = \frac{S_{\text{завд}}}{k-1}, D_{\text{ост}} = \frac{S_{\text{ост}}}{(n-1)(k-1)}.$$

Значення $\sqrt{D_{\text{ост}}}$ дає *середню квадратичну помилку вимірів*.

• Якщо усі завдання *однакові* для усіх учасників, то дисперсія помилки $D_e\{b^*\} = D_{\text{учасн}} - D_{\text{ост}}$, а коефіцієнт надійності $r = \frac{D_{\text{учасн}} - D_{\text{ост}}}{D_{\text{учасн}}}$.

- Якщо учасникам пропонувалися різні завдання, то

$$S_e = S_{\text{загал}} - S_{\text{учасн}}; \quad \hat{D}_e = \frac{S_e}{n(k-1)} = 5,58; \quad \hat{y}\{\Delta\} = \sqrt{\hat{D}_e}, \quad \text{а}$$

$$\text{коефіцієнт надійності } \hat{r} = \frac{D_{\text{учасн}} - \hat{D}_e}{D_{\text{учасн}}}.$$

Середнє значення коефіцієнта надійності при всіляких розщеплюваннях тесту на дві половини, може бути обчислене

$$\text{формулою Кьюдера-Річардсона} \quad r = \frac{k}{k-1} \cdot \frac{D\{b\} - \sum_{j=1}^k p_j q_j}{D\{b\}}, \quad \text{при}$$

цьому само розщеплювання практично не потрібно. Формула Кьюдера-Річардсона застосовна, якщо елементи a_{ij} матриці відповідей А можуть набувати тільки двох значень 0 та 1. Існує її узагальнення у вигляді так званого коефіцієнта альфа Кронбаха.

Для підвищення надійності можна збільшити кількість завдань, 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.

Кореляційна методика оцінювання надійності полягає у вимірюванні коефіцієнту кореляції між результатами двократного тестування (або розщеплення). *Коефіцієнтом надійності* виступає коефіцієнт кореляції.

Для встановлення відповідності між варіантами тесту або двох частин розщепленого тесту аналізуються матриці відповідей, елементи яких можна інтерпретувати, як реалізації дихотомної випадкової величини. *Коефіцієнт дихотомної кореляції*

$$\hat{\rho} = \frac{\overline{xy} - \bar{x}\bar{y}}{(\hat{p}_X \hat{q}_X \hat{p}_Y \hat{q}_Y)^{1/2}} = \frac{\hat{p}_{XY} - \hat{p}_X \hat{p}_Y}{(\hat{p}_X \hat{q}_X \hat{p}_Y \hat{q}_Y)^{1/2}}.$$

Для перевірки гіпотези про відсутність кореляції між даними двох тестів у припущенні нормального розподілу використовують критерій Стьюдента.

Коефіцієнт надійності r повного тесту пов'язаний з коефіцієнтом r' надійності в результаті розщеплення тесту

$$r = \frac{2r'}{1+r'}, \quad \text{що є частковим випадком формули Спірмена – Брауна}$$

$$r = \frac{lr'}{1+(l-1)r'}, \quad \text{де } l - \text{відношення числа завдань тесту до}$$

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 $r\{x\}$ і $r\{y\}$, тут випадкові величини x та y – результати повторного тестування або розщеплення тесту; $d_i = r\{x_i\} - r\{y_i\}$.

Рангова кореляція оцінюється також коефіцієнтом Кендалла ϕ , для чого треба підрахувати в стовпці ранг y_i кількість R_i рангів, що перевищують $r\{y_i\}$ та сума $R = \sum_{i=1}^{n-1} R_i$. Вибірковий коефіцієнт рангової кореляції Кендалла і його дисперсія знаходяться по формулах

$$\hat{r} = \frac{4R}{n(n-1)} - 1, \quad D\{\hat{r}\} = \frac{2(2n+5)}{9n(n-1)}.$$

При рівні значущості α та конкуруючій гіпотезі $\phi \neq 0$ критичне значення $\phi_{\text{крит.}} = U_{\text{крит.}} \sqrt{\frac{2(2n+5)}{9n(n-1)}}$, де $U_{\text{крит.}}$ визначається умовою $\Phi(U_{\text{крит.}}) = \frac{1-\alpha}{2}$, функція Лапласа $\Phi(u)$; n - об'єм вибірки. Якщо $|\hat{\phi}| < \phi_{\text{крит.}}$, то немає підстав для відмови від нульової гіпотези, і рангова кореляція визнається незначущою. Якщо $|\hat{\phi}| > \phi_{\text{крит.}}$, то зв'язок слід вважати значущим.

2. Аналіз основних типів практичних завдань. Для модуля «Надійність тесту» дисципліни «Класичні тестові моделі», що викладається майбутнім магістрам спеціальності 8.18010022 «Освітні вимірювання» створено посібник практичних завдань, укладачі якого спирались на такі відомі підручники та посібники [4], [10], [11]. Головною метою створення посібника є показати, що всі висновки, які випливають з теоретичних питань можуть бути математично обґрунтовані або доведені, навчити правильній інтерпретації отриманих статистичних даних та надати основний теоретичний матеріал, математичний апарат та розв'язані теоретико-практичні типові задачі для навчання майбутніх

фахівців з освітніх вимірювань різнобічній оцінці розроблених тестів, або тих, що існують.

Серед практичних завдань можна виділити три основні типи теоретичні завдання, математичні завдання на аналіз тесту, математичні завдання на виведення співвідношень.

Застосування практичних *завдань теоретичного типу* дозволяє систематизувати пройдений теоретичний матеріал, розвивати логічне мислення і розуміння ситуацій, що можуть виникати в процесі створення завдань та під час тестування і впливати, негативно чи позитивно, на результати. В даному випадку будь-який вплив відхиляє отримані результати від істинних оцінок, тому вважається негативним фактором. Наприклад, вплив систематичних і випадкових помилок на результати тестування, або зміна коефіцієнту надійності в залежності від того якій репрезентативній вибірці він запропонований, або від зміни часу виділеного на виконання тесту.

Наведемо два приклади теоретичних завдань.

Завдання 1. Для кожної ситуації вкажіть тип оцінки надійності, що найбільш підходить наступним умовам.

A. 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?

B. Шкільний вчитель хоче перевірити на учнях свого класу надійність підсумкового тесту по біології, що містить завдання з множинним вибором.

C. 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.

D. 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.

Завдання 2. Порівняйте відмінності в методах поділу навпіл і методах підрахунку коваріації завдань для оцінювання надійності по відношенню вимог до даних і поясніть чому вони дають різний результат.

Метою використання таких завдань є закріплення знань теоретичного матеріалу з розділу «Процедури для оцінювання надійності», а також понятійного апарату та можливості аналізувати і робити висновки із поданих ситуацій. Передбачається володіння такими основними поняттями:

- процедури, що потребують двох пред'явлень тесту (метод взаємозамінюваних форм, що потребує конструювання двох подібних форм теста і пред'явлення обох форм одній і тій же групі тестованих, метод тест-ретест (ретестовий метод), що потребує пред'явлення однієї форми тесту групі через певний проміжок часу та комбінація двох методів – тест-ретестовий метод з застосуванням взаємозамінюваних форм (в даному випадку процедура оцінки надійності включає представлення першої форми тесту групі тестованих, а через деякий проміжок часу другої форми));

- процедури, що потребують одноразового пред'явлення тесту (метою використання таких процедур є визначення узгодженості між собою результатів одних і тих же тестованих по всім завданням тесту);

- процедури, за якими розділяють тест на взаємозамінювані форми (метод оцінки кореляції (тобто оцінюється кореляція між двома формами тестових завдань і за її величиною формують висновок про паралельність даних тестових форм), метод в основі якого лежить використання передбачуваних тестових оцінок (розробник тесту повинен виявити джерела помилки вимірів, які

вносять найбільше перешкод в корисну інтерпретацію оцінки, і планувати дослідження надійності таким чином, щоб ефекти впливу цих помилок могли бути оцінені у вимірюваннях));

• помилки, що впливають на кінцевий результат: випадкові (помилки, що впливають на оцінку тестованого через різні непередбачувані фактори) та систематичні (помилки, що постійно впливають на оцінку індивіда через деякі специфічні характеристики тестованого чи тесту). Перший тип зменшує як відтворюваність, так і практичну повноцінність тестових оцінок, другий тип являється причиною зниження точності тестових оцінок, тобто скорочує їх практичну корисність.

Другий тип завдань на аналіз надійності тесту є **завдання математичного типу**. У них задача ставиться так, що спочатку потрібно розрахувати певні величини, а потім інтерпретувати їх в залежності від ситуації. Наприклад, розрахунок всіх представлених в теоретичній частині коефіцієнтів надійності і порівняльний аналіз отриманих результатів, вибір оптимального значення. Нижче наведемо характерні завдання такого типу.

Завдання 3. Початковий тест містить 40 тестових завдань і має коефіцієнт надійності (розрахований за формулою KR20) рівний 0,85. Розробник тесту розглядає можливість зменшення кількості тестових завдань на 5. Оцініть як це вплине на надійність тесту.

Завдання 4. Тест, що складається з 10 завдань був пред'явлений 9-ти учням. Матриця, що об'єднує результати по завданням і учням представлена в таблиці 1. 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.

B. Знайти кореляцію між субтестами.

C. Яка оцінка надійності для тесту повної довжини?

D. Сказати, користуючись результатами візуального аналізу, який із субтестів, на вашу думку буде мати більш високу оцінку надійності, обґрунтувати відповідь.

E. Який з методів поділу тесту забезпечить найбільше наближення до оцінки коефіцієнта точності? Обґрунтувати відповідь.

F. Яка величина коефіцієнта надійності за формулами KR20 і KR21.

Пояснити чому відрізняються отримані величини коефіцієнтів за формулами KR20 і KR21.

Таблиця 1.

Учень	Завдання										заг. оцінка
	1	2	3	4	5	6	7	8	9	10	
A	1	0	1	1	1	0	0	0	1	0	5
B	1	1	1	1	1	0	0	0	0	0	5
C	0	1	1	1	1	1	1	0	1	0	7
D	0	0	1	0	0	0	1	0	0	0	2
E	1	0	0	1	1	1	0	0	1	0	5
F	1	1	1	1	1	0	1	1	1	0	8
G	1	0	0	1	1	1	1	0	0	1	6
H	0	0	0	1	0	0	1	0	0	0	2
I	1	1	1	0	0	1	1	0	0	0	5
	6	4	6	7	6	4	6	1	4	1	

Завдання 5. Експерт цікавиться питанням, що більше впливає на рівень уваги студентів вищого навчального закладу: стать чи вік? Розглядаються студенти різних факультетів і форм навчання. Результаті приведені в таблиці2.

Використавши двохфакторний дисперсійний аналіз даних, визначити коефіцієнт генералізації, що вказує на систематичний вплив одного із факторів (стать, вік) на рівень уваги студента. Проаналізувати отримані результати. Зробити висновки.

Таблиця 2.

Студенти					
Чоловіки			Жінки		
від 18 до 25 років	від 26 до 35	старші 35	від 18 до 25 років	від 26 до 35	старші 35
16	15	13	17	15	12
17	16	14	15	17	10
15	13	13	16	14	10
16	14	10	16	14	9

Завдання 6. Наступна таблиця має дані за двома гіпотетичними субтестами стандартизованої батареї тестів навчальних досягнень по використанню різних типів програмного забезпечення, кожен з яких має дві взаємозамінювані форми. Потрібно дослідити дані, наведені в таблиці 3, і відповісти на питання.

Таблиця 3.

Характеристики	Використання комерційного програмного забезпечення	Використання безкоштовного програмного забезпечення
Форма А		
Середнє	30	28
Стандартне відхилення	8	12
Число завдань	50	55
KR20	0,83	0,85
Форма В		
Середнє	32	28
Стандартне відхилення	6	12
Число завдань	50	55
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?

В. Нехай оцінка студента X — 20 балів по математичному тесту обрахунків за формою A , 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?

С. По формі B математичного тесту обрахунків студент Y отримав оцінку 10 балів, а студент Z — 29 балів. Які значення їх можливих істинних оцінок?

В даних завданнях, в першу чергу, перевіряється знання студентом математичних формул, що відповідають теоретичному матеріалу, та вміння аналізувати обраховані результати. В приведених вище прикладах розглядаються різні види коефіцієнтів надійності (стабільності, еквівалентності, внутрішньої узгодженості), що вибираються в залежності від основних джерел помилок (зміна тестованих з плином часу, вибіркова перевірка змісту від форми до форми, вибіркова перевірка змісту чи виявлення некоректних завдань), та проводиться їх обрахунок і аналіз доцільності використання в даній ситуації.

Завдання такого типу дають можливість підкріплювати свої твердження відповідними математичними показниками, що є важливим фактором для студентів з нематематичною базовою освітою.

До *математичних завдань на виведення співвідношень* віднесемо задачі на доведення тверджень чи виведення формул. Такі завдання необхідні для розвитку і розширення математично — статистичної бази студентів. Нижче наведені характерні завдання названого типу.

Завдання 7. *Вказівка по тестуванню для індивідуального використання має оцінки надійності (метод поділу тесту), що розраховані за формулою: $\hat{c}_{XX} = \frac{4\hat{c}_{AB}\hat{y}_A\hat{y}_B}{y_X^2}$ де A, B — оцінки по половинам тесту. Довести, що дана формула еквівалентна формулі для коефіцієнта Рюлона..*

Завдання 8. *Покажіть, що коваріація між отриманими по двом тестам оцінкам еквівалентна коваріації між їх істинними*

оцінками, тобто докажіть, що $U_{X_1 X_2} = U_{T_1 T_2}$.

Завдання 9. На основі формули KR20 виведіть формулу для коефіцієнта надійності KR21.

Наведені вище приклади орієнтовані на якісне знання студентів математичного апарату обрахунку надійності, формул і означень елементарної статистики та класичної моделі істинної оцінки. Вони не є типовими і вимагають від магістранта логічного мислення, знання основних формул і методів їх виведення, аналізу отриманих результатів.

Кінцевою метою застосування всіх трьох типів завдань з одного боку виробити у магістранта навички практичного аналізу тесту на надійність та закріпити теоретичні знання цього матеріалу, а з іншого боку - виявити ступінь засвоєння їм поданий теоретичний матеріал.

Отже, тут проведено аналіз теоретичних підходів до поняття надійності тесту та проведено класифікацію відповідних практичних завдань. Серед практичних завдань виділено три основні типи — теоретичні завдання, математичні завдання на аналіз тесту, математичні завдання на виведення співвідношень. Наведені характерні завдання кожного із трьох названих типів. Відмічено, що завдання першого типу направлені на покращення розуміння понятійного апарату та можливості аналізувати і робити висновки із реальних ситуацій. Завдання другого типу дають можливість підкріплювати свої твердження відповідними математичними показниками, що є важливим фактором для магістрантів з нематематичною базовою освітою. Завдання третього типу вимагають від магістранта логічного мислення, знання основних формул і методів їх виведення, аналізу отриманих результатів.

Література

1. Гарвілл Л.М. Стандартна похибка вимірювання (навчальний модуль Національної Ради США з оцінювання у сфері освіти) // Вісник. Тестування і моніторинг в освіті. – 2010. - №3. –

с.40-48

2. Горох В.П., Соколов О.Ю. Порівняльний аналіз шкалювання результатів ЗНО різними методами // Вісник. Тестування і моніторинг в освіті. – 2010. - №12. – с.22-28

3. Ким В.С. Тестирование учебных достижений. Монография. – Усурийск: Издательство УГПИ, 2007. – 214 с.

4. Крокер Л., Алгина Дж. Введение в классическую и современную теорию тестов: Учебник. – Логос, 2010. – 668 с.

5. Майоров А.Н. Теория и практика создания тестов для системы образования. (Как выбирать, создавать и использовать тесты для целей образований.). – М., Интеллект-центр, 2001. – 286 с.

6. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов . – Москва, 2000. – 168с.

7. Раков С.А. Теоретичні засади шкалювання результатів ЗНО методом еквіпроцентної нормалізації // Вісник. Тестування і моніторинг в освіті. – 2010. - №12. – с.11-22

8. Фрарі Р.Б. Обчислення балів за формулою для тестів множинного вибору (коригування вгадування) // Вісник. Тестування і моніторинг в освіті. – 2008. - №9. – с.32-40

9. Чельшкова М.Б. Теория и практика конструирования педагогических тестов: Учеб. пособие. – М. Логос, 2001. – 432 с.

10. Baker F.B. The Basics of Item Response Theory.- Portsmouth NH: Heinemann Educational Books, 1985.- 131 pp.

11. Hambleton R.K., Swaminathan H., Rogers H. J. Fundamentals of Item Response Theory. Newbury Park, CA: Sage, 1991.

12. Ragosa D.R., Willet J.B. Demonstrating the reliability of the difference score in the measurement of change. Journal of Educational Measurement , 20, 335-343.

13. Stanley J.C. Reliability. In R. Thorndike, (Ed.), Educational Measurement, (2-nd ed., pp. 356-442). (1971) Washington, D.C: American Council on Education.

14. Traub R. E., Rowley G.L. Understanding reliability. Educational measurement: Issues and practice, 1991, 10(1), 37-45.

КОНСТРУЮВАННЯ ТЕСТІВ: НОВА НАВЧАЛЬНА ДИСЦИПЛІНА ДЛЯ СТУДЕНТІВ ВИЩИХ ПЕДАГОГІЧНИХ НАВЧАЛЬНИХ ЗАКЛАДІВ

Кухар Л.О., Сергієнко В.П.

Україна визначила напрямки розвитку освіти і науки, здійснює вдосконалення освітньої діяльності в контексті європейських вимог, зробила конкретні кроки для приєднання до Болонського процесу.

Однією з найважливіших сфер розвитку євроінтеграції є сфера вищої освіти, а одним із найважливіших показників є рівень знань студентів з навчальних дисциплін. Найоб'єктивнішим засобом оцінювання рівня знань сьогодні вважають тести, які дозволяють неупереджено оцінити навчальні досягнення студентів.

В Україні проведені дослідження, результатом яких є рекомендації щодо організації тестування знань студентів в умовах кредитно-модульної системи навчання, проте недостатньо дослідженими залишаються питання аналізу впливу систем тестування на якість знань, класифікації тестів і особливостей їх складання з метою підвищення ефективності організації навчання студентів.

У зв'язку із реформуванням вищої освіти в Україні виникла необхідність у впровадженні та вивченні нових навчальних курсів, дисциплін, адаптації навчальних програм відповідно результатів навчання [1]

Вагоме місце у процесі навчання та самостійної підготовки студента займає тестування. Та, на жаль, не всі матеріали які використовуються мають належне оформлення та перевіряють саме те що потрібно. Випускник вищого педагогічного навчального закладу, прийшовши до школи на роботу зустрічається із наступною проблемою: як скласти тест таким чином, щоб він був об'єктивним, надійним, перевіряв уміння, знання, навички, акцентував увагу учителя на проблемах у вивченні матеріалу та ін.

В рамках проекту за програмою Європейського Союзу TEMPUS-IV «Освітні вимірювання, адаптовані до стандартів ЄС» на базі Інституту інформатики НПУ імені М.П. Драгоманова розроблені навчальний план та курси бакалаврської підготовки за напрямом «6.040302 Інформатика*», з додатковою спеціалізацією «Освітні вимірювання».

Однією із дисциплін, що входять до циклу «Освітні вимірювання» є «Конструювання тестів». Вивчення цієї дисципліни сприяє підготовці вчителя інформатики і в галузі теорії та практики педагогічних вимірювань з використанням тестів, передбачає вивчення питань прикладного напрямлення: розрізнати контрольний матеріал за цілями використання, інтерпретувати результати тестування, розуміти основні характеристики якості контрольного матеріалу і можливості використання комп'ютерних технологій при створенні, проведенні і опрацюванні результатів тестування.

Одними із провідних питань, які розглядаються в курсі «Конструювання тестів» є наступні: якість освіти; використання тестів як інструменту педагогічних вимірювань; переваги тестування перед іншими формами контролю знань, умінь та навичок; про науку тестологію, її основні напрямки розвитку; класифікацію тестів; проблеми, пов'язані зі складанням тестових завдань; види тестового контролю; цілі процесу навчання, таксономія Блума; стандартизація тесту; формати тестових завдань; вимоги до завдань в тестовій формі; опрацювання та подання результатів тестування; шкалювання та подання результатів тестування; методика і технологія тестування та ін.

Процес створення в Європі єдиного освітнього простору супроводжується розробленням єдиних критеріїв і стандартів у цій галузі, де якість вищої освіти є основою створення цього процесу. Сьогодні саме освіта є одним із найважливіших чинників у формуванні майбутнього нашої країни. Виокремлення проблеми якості освіти зумовлене об'єктивними причинами: по-перше, від якості людських ресурсів залежить рівень розвитку країни та її конкурентоспроможності на економічних ринках; по-друге, якість освіти набуває все більшого значення в забезпеченні

конкурентоспроможності випускників вищої школи на ринку праці. [7]

Сучасне розуміння якості освіти не є достатнім для виходу України до міжнародних освітніх програм, тому якість освіти має розглядатися як: ступінь відповідності реальних результатів освіти ринковій кон'юктурі; показник матеріально-технічної і ресурсної забезпеченості освітнього процесу; комплексний показник чинників престижності й економічної ефективності освіти; показник досконалості змісту, технологій і системи оцінювання досягнень; показник інвестиційної принадності освіти тощо.

Нагальною потребою є введення рейтингової системи оцінювання як вищого навчального закладу, так і навчальних курсів. Доцільним є також поступовий розвиток рейтингового експертного оцінювання від рейтингу вищого навчального закладу до рейтингу освітніх програм.

Ефективним інструментом вимірювання знань студентів є тест. Тест – це інструмент, що складається з вивіреної системи тестових завдань, стандартизованої процедури проведення, заздалегідь спроектованої технології опрацювання і аналізу результатів, призначений для вимірювання якостей і властивостей особи, зміна яких можлива в процесі систематичного навчання [18].

Застосування тестування як методу вимірювання у різних сферах сприяло появі та використанню великої кількості різноманітних тестів, що зумовило необхідність їхньої класифікації (за рівнем уніфікації, за рівнем впровадження, за статусом використання, за співвідношенням із нормами або критеріями, за видом тестового контролю).

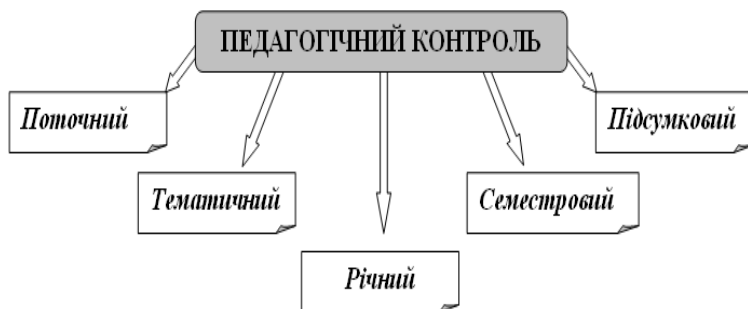
Відмінна риса тесту – наявність вимірювання, функція якого полягає в тому, щоб давати кількісні і якісні відомості про прогрес навчання, недоробки і прогноз успішності.

Тестовий контроль відрізняється від інших методів контролю тим, що він є спеціально підготовленим контрольним набором завдань, який дозволяє надійно і адекватно кількісно оцінити знання студентів за допомогою статистичних методів. Об'єктивність досягається шляхом стандартизації процедури проведення, стандартизації і перевірки показників якості завдань і

тесту в цілому. Педагогічний контроль поділяється на кілька видів залежно від обсягу навчального матеріалу, який підлягає контролю, та часу протягом якого він вивчався.

За змістом навчального матеріалу контроль поділяється на три види:

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



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

Тест як інструмент вимірювання використовується в більшості країн світу. Тестологія як теорія та практика тестування існує більше 120 років, і за цей час накопичено великий досвід використання тестів в різних сферах людської діяльності, включаючи і освіту.

Тестологія (від англ. test – проба, грецьк. logos – знання)– міждисциплінарна наука про створення якісних та науково обґрунтованих вимірювальних діагностичних методик.

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

Теорію тестових педагогічних вимірювань називають коротко IRT - Item Response Theory. У цій теорії не ставляться і не

розв'язуються фундаментальні проблеми валідності і надійності тесту: тест там заздалегідь вважається надійним і валідним. Обчислення IRT зводяться до отримання оцінок параметрів складності завдання і до вимірювання рівня ЗУН випробуваних.

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

В Україні поки що не склалося уявлення про педагогічну тестологію як окрему науку або наукову галузь:

- ✓ немає посади тестолога в реєстрах відділів кадрів;
- ✓ педагоги не сприймають тестології, оскільки у ній занадто багато математики;
- ✓ програмісти вважають, що тестологія - пройдений етап.

Ключовим поняттям тестології є поняття «педагогічний тест».

Тестом називають спеціально розроблений комплекс завдань для перевірки будь-яких характеристик, наприклад рівня засвоєння студентом тієї чи іншої теми, готовності до навчання тощо.

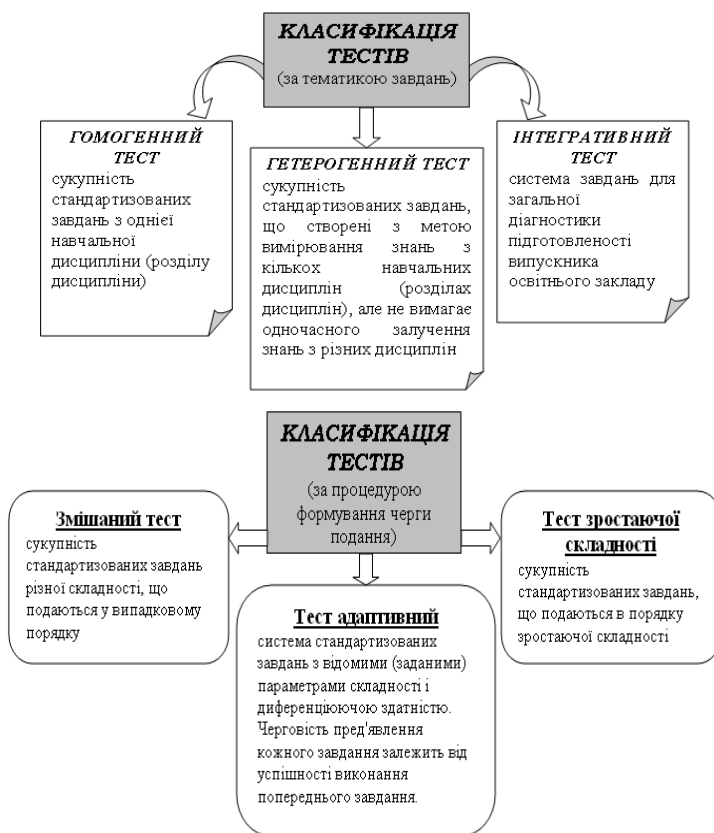
Так В.С. Аванесов дає наступне визначення поняттю «педагогічний тест»: **«Педагогічним тестом** називається система завдань специфічної форми, певного змісту, складності, що зростає – система, яка створюється з метою об'єктивної оцінки структури і вимірювання рівня підготовленості тих, хто навчається» [8].

Іноді під поняттям «тест» розуміють засіб педагогічного вимірювання, або стандартизований метод дослідження, призначений для кількісного та якісного оцінювання характеристик людини.

Застосування тестування як методу вимірювання у різних сферах сприяло появі та використанню великої кількості різноманітних тестів. Це зумовило необхідність їхньої класифікації. Її здійснюють за різними критеріями та принципами, наприклад:

- ✓ за рівнем уніфікації (тести стандартизовані, нестандартні);
- ✓ за рівнем впровадження (загальнодержавні, навчального закладу, вчительські);
- ✓ за статусом використання (обов'язкові, пілотні, дослідницькі);
- ✓ за співвідношенням із нормами або критеріями (тести досягнень, тести порівняння або тести відбирання);
- ✓ за видом тестового завдання (тести із закритими завданнями, тести з відкритими завданнями).

Крім поданих вище, можна представити ще такі класифікації тестів:



Одним з визначальних компонентів процесу навчання є його мета. *Мета навчання* – це ідеальне мисленнєве передбачення кінцевого результату процесу навчання, це те, до чого прагнуть педагог та учні.

Перехід від загального уявлення про результати навчання до конкретнішого передбачає чіткий опис того, що учень може досягти в результаті навчання. Загальний прийом конкретизації цілей – використання дієслів, які вказують на конкретну дію: аналізувати, синтезувати, інтерпретувати, оцінювати, розуміти, перетворювати, застосовувати, висловлювати, створювати тощо.

Автором однієї з перших схем педагогічних цілей був американський вчений Б. Блум. Ним випущена в світ перша частина «таксономії» (1956 р.), де описуються цілі пізнавальної (когнітивної) області.

Блум окреслив шість рівнів освітніх цілей, які використовуються для визначення розвитку в учнів навичок мислення високого рівня (аналіз, синтез, оцінювання).

Структура тесту залежить від мети тестової перевірки знань. Правильно розроблені тестові завдання перевіряють всю пізнавальну сферу, яка містить (відповідно до складності): знання, розуміння, застосування, аналіз, синтез, оцінювання.

- I. **Знання** - це здатність запам'ятовувати факти, принципи, процеси.
- II. **Розуміння** - здатність учнів розуміти матеріал, який вивчають.
- III. **Застосування** - здатність використовувати засвоєний матеріал у нових ситуаціях.
- IV. **Аналіз** - здатність структурувати навчальний матеріал таким чином, що стає зрозумілою загальна організаційна структура.
- V. **Синтез** - здатність поєднати окремі частини знань для отримання цілого, що набуває нової якості.
- VI. **Оцінювання** - здатність судити про цінність даного навчального матеріалу в рамках поставленої мети.

Існує багато різних думок щодо форми подання тестових завдань у тесті. Хтось надає перевагу тестовим завданням у прямій

формі, дехто рекомендує використовувати тестові запитання у питальній формі.

Вибір способу формулювання тестового завдання, у формі твердження або у формі запитання повинен визначитися двома обставинами: по-перше, необхідно вибирати ту форму, яка буде максимально зрозумілою для випробуваних, і, по-друге, одна з форм завдань вимагає неодмінно формулювання питання у формі твердження – це завдання альтернативних відповідей, оскільки вони вимагають відповіді так чи ні, які краще підходять до тверджень.

Вид тестового контролю визначається *цілями тестування* – здобуття відомостей про рівень знань студентів на певному етапі навчання.

Кожен освічений користувач тесту має розуміти, що таке тестові норми і як ними користуватися.

Тест як засіб вимірювання називають стандартизованим тестом. [15]

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

Стандартизований тест складається із чотирьох складових:

- ✓ специфікацій;
- ✓ системи завдань;
- ✓ інструкції щодо проведення тестування;
- ✓ технології перевірки, обробки і аналізу результатів тестування.

Технологія створення тестів навчальних досягнень складається із багатьох етапів. Ця процедура називається стандартизацією тесту і передбачає такі види діяльності:

- ✓ складання технічного завдання на підготовку тесту (специфікації);
- ✓ формування банку тестових завдань;
- ✓ аналіз доступності тестових завдань і відбраковування неякісних завдань;

- ✓ аналіз дистракторів та їхнє коригування;
- ✓ формування робочого варіанта тесту;
- ✓ формування репрезентативної вибірки для експериментальної перевірки цього варіанта;
- ✓ знаходження коефіцієнтів надійності за різними методиками і різних видів надійності;
- ✓ експертне дослідження змістової валідності;
- ✓ розроблення кількох варіантів тесту та перевірка їхньої еквівалентності; розроблення нормативів тесту;
- ✓ створення інструктивно-методичного забезпечення тестування.

Під час створення тестових завдань фахівці дотримуються певних правил. [4][8][9]. Наведемо деякі з них:

1. Кожне тестове завдання має оцінювати досягнення важливої навчальної мети.
2. Кожне тестове завдання має перевіряти відповідний рівень засвоєння знань.
3. Умова повинна містити чітко сформульоване і зрозуміле для екзаменованого завдання.
4. Варіанти відповідей мусять бути вірогідними (правдоподібними).
5. Усі дистрактори мають бути вірогідними (правдоподібними).
6. Інформація, що міститься в одному тесті, не повинна давати відповіді на інше завдання.

В сучасній тестології (Кляйн К., Анастасі А., Раш Д., Аванесов В.С., Челишкова М.Б. та ін.) виділяють основні 4 типи завдань в тестовій формі: [4], [8], [10], [15]

ФОРМАТ А – тестові завдання з *однією найкращою відповіддю*

Структура цього тестового завдання наступна: умова (завдання), запитання за цією умовою та чотири – п'ять варіантів відповідей, один з яких є найбільш правильним. Неправильні відповіді можуть бути не абсолютно неправильними, але вони все ж таки менш правильні, ніж еталонна відповідь. Форматом А при відповідній зміні умови і запитання можна перевірити з першого по третій рівні когнітивного домену: знання, розуміння і застосування знань.

Принципи підбору відповідей:

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

ФОРМАТ Х – тестові завдання *множинного вибору* «правильно – неправильно»

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

До основних елементів завдань з вибором кількох правильних відповідей відносяться: інструкція для тих хто тестується. Вона пишеться так: «Обведіть номери всіх правильних відповідей». Змістовна основа зазвичай співпадає з іншими формами і видами завдань. Гарний спосіб покращення змісту завдань – використання фасету. Не рекомендується формулювати завдання у вигляді заперечення, у випадку, коли запитують про те, що не використовується, не відноситься, не робиться та ін..

Відповіді на завдання обов'язково відносяться до одного роду чи виду термінів. Загальна кількість відповідей має бути в діапазоні від 5 до 14. Особливістю завдань з вибором кількох правильних відповідей є те, що в них потрібно визначати не лише правильні відповіді, але й оцінювати повноту своєї відповіді. В завданнях цього формату найчастіше перевіряються лише засвоєння першого рівня когнітивного домену – знання фактів та концепцій, тобто декларативних, а не процедурних знань.

ФОРМАТ N – тестові завдання з *кількома найкращими відповідями*

У тестових завданнях такого формату екзаменованій має вибрати певну кількість (наприклад 2, 3 чи 4) відповідей із

запропонованого списку варіантів. Перелік варіантів відповідей зазвичай досить довгий (до 30 варіантів).

Формат **N** доцільно використовувати у ситуаціях, в яких описуються висновки, дослідження, рекомендації, в яких доречно дати більше однієї відповіді.

За діагностичними властивостями, тестове завдання формату **N** діагностує знання, розуміння і застосування.

ФОРМАТ В – тестові завдання *логічних пар* або *визначення відповідності*.

Тестові завдання формату **В** належать до категорії логічних пар, що складається з 3-5 визначень або цифрових значень, які є варіантами відповідей та переліку слів чи фраз, що містять завдання. Варіанти відповідей позначено буквами, а слова чи фрази, що містять завдання, - пронумеровано. Екзаменованій має дібрати до кожного пронумерованого завдання один найбільш відповідний варіант відповіді, позначений буквою.

Кожен варіант відповіді може використовуватись більше одного разу або не використовуватись взагалі. За наявністю або відсутністю варіантів відповіді виділяють тести *закритої* та *відкритої* форм.

Закриті тестові завдання перевіряють повноту засвоєння та рівень сформованості змістової лінії навчальної програми.

Види закритих тестових завдань:

1. Тестові завдання з вибором однієї правильної відповіді
2. Тестові завдання з вибором кількох правильних відповідей
3. Тестові завдання на встановлення відповідності

Види відкритих тестових завдань:

1. Тестові завдання з пропусками
2. Тестові завдання на доповнення
3. Завдання з короткою відповіддю
4. Завдання з розгорнутою відповіддю

Вимоги до завдань в тестовій формі

В умовах кредитно – модульної системи організації навчального процесу для об'єктивності оцінювання знань

студентів, крім вибору форми представлення відповідей, важливим є забезпечення змісту запитань, який дозволить вважати тест валідним.

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

Основні правила розроблення якісних тестів:

- ✓ Тест повинен мати необхідний і достатній рівень складності.
Складність — міра розумових зусиль, потрібних для вибору відповіді. Часто цей параметр називається вагою і визначається як частка правильних або неправильних відповідей при відповіді на завдання в групі тестованих.
- ✓ Тест повинен бути об'єктивним і надійним.
Надійність — міра правильності і адекватності відображення тестом рівня знань учнів. Надійний тест забезпечує рівні права кожній групі учнів і кожному учню в групі.
- ✓ Тест повинен бути стійким і таким, що може бути співвіднесений певній шкалі.
Стійкість тесту — міра збереження надійності і валідності при перенесенні тесту в інше, аналогічне середовище, міра рівнозначності і однорідності тестів для різних груп учнів.
Шкалювання тесту — здатність тесту відображати результати навчальних досягнень в деякій системі (шкалі) оцінок або балів.
- ✓ Тест повинен бути репрезентативним.
Репрезентативність тесту (або бази тестів) — міра повноти охоплення завданнями навчального матеріалу, програми, відображення тестами різних рівнів навчання.
- ✓ Тест повинен бути значущим і дискримінантним.
Значущість тесту — міра необхідності, актуальності включення в тест ключових знань.

Дискримінантність тесту – міра диференціації тих хто тестується щодо максимального або мінімального рівня навчання.

- ✓ Тест повинен бути достовірним, науковим, несуперечливим.
Достовірність, науковість, несуперечність тесту – міра істинності тесту, узгодженості, відповідності сучасному стану науки і технології, методиці навчання. Необхідно забезпечити узгодженість завдань тесту як між собою (внутрішня узгодженість), так і з іншими тестами (зовнішня узгодженість).
- ✓ Тестове завдання повинно бути сформульовано ясно і чітко, не повинно допускати двозначного тлумачення змісту і сприяти формулюванню правильної відповіді.
- ✓ Кожне завдання тесту повинне бути функціонально завершеним, тобто перевіряти конкретне знання, уміння або навик.

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

- ✓ Відповіді повинні містити не більше 2 – 3 ключових слів по умові запитання.

Бажано будувати відповіді однакової форми і довжини. Кількісні відповіді бажано упорядковувати за зростанням або спаданням, а якщо відповіді легко обчислити, то першою не повинна бути вказана правильна відповідь.

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

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

- ✓ Тест повинен складатися з мінімальної кількості завдань, достатніх для повного визначення рівня навчальних досягнень.

Рекомендується уникати непотрібного дублювання знань, що перевіряються, умінь і навичок, а також передбачити всі необхідні.

Для створення якісних тестів, необхідним є використання тестових завдань різних когнітивних рівнів (знань, розуміння, застосування, аналізу, синтезу, оцінювання).

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

Основною метою вивчення даної дисципліни є прагнення дати студентам уявлення про оцінювання якості освіти на сучасному рівні, навчити їх розробляти тестові завдання та тести, опрацьовувати результати тестувань та оцінювати якість тестових завдань та тесту в цілому.

Завданнями даної дисципліни є розкриття важливості оцінювання освітніх досягнень студентів з позиції сучасних вітчизняних і міжнародних пріоритетів в освіті; навчити створенню і адекватному використанню сучасних контрольних вимірювальних матеріалів для тестування і вміння інтерпретувати отримані при тестуванні результати; розкрити можливості інструментальних програмних систем для розроблення тестових завдань для психолого-педагогічної діагностики

Підготовка кваліфікованих учителів, а також підвищення кваліфікації працівників вищих навчальних закладів в галузі теорії і практики педагогічного тестування є одним із найважливіших завдань, тому введення дисципліни «Конструювання тестів» є важливою та необхідною складовою процесу реформування освіти в умовах кредитно-модульної системи організації навчального процесу. Цілком очевидно є перспектива викладання цієї дисципліни й для студентів інших спеціальностей вищих педагогічних навчальних закладів, що буде передумовою формування професіоналізму в оцінюванні ефективності освітнього процесу в контексті незалежного оцінювання.

Література

1. Болонський процес: головні принципи входження в Європейський простір вищої освіти / Упорядники Журавський В.С., Згуровський М.З.. - К.: ІВЦ "Видавництво "Політехніка", 2003. - 200 с.
2. Аванесов В. С. Теоретические основы разработки заданий в тестовой форме: учеб. пособие / В. С. Аванесов. – М. : Изд-во МГТА, 1995. – 95 с.
3. Аванесов В. С. Математические модели педагогического измерения / В. С. Аванесов. – М. : Исслед. центр, 1994. – 26 с.
4. Булах І.Є., Мруга М.Р. Створюємо якісний тест. Навчальний посібник. / К. : Майстер-клас, – 2006 –160 с.
5. Основи педагогічного оцінювання. Частина 1. Теорія. Навчально-методичні та інформаційно-довідкові матеріали для педагогічних працівників. / К.: Майстер-клас, – 2005 –96 с.
6. Основи педагогічного оцінювання. Частина 2. Практика. Навчально-методичні та інформаційно-довідкові матеріали для педагогічних працівників. / К.: Майстер-клас, – 2005 –56 с.
7. Болюбаш Я.Я., Булах І.Є., Мруга М.Р., Філончук І.В. Педагогічне оцінювання і тестування. Правила. Стандарти. Відповідальність. Наукове видання. / К.: Майстер-клас, – 2007 – 272 с.
8. Аванесов В.С. Композиция тестовых заданий / В. С. Аванесов. – М. :Адепт, 1998. - 216 с.
9. Гулюкина Н. А. Педагогический тест: этапы и особенности конструирования и использования / Н. А. Гулюкина, С. В. Клишина. – Новосибирск : Изд-во НГТУ, 2006. – 151 с.
10. Чельшкова Н. Б. Теория и практика конструирования педагогических тестов : учебное пособие / Н. Б. Чельшкова. – М. : Логос, 2002. – 432 с.
11. Аванесов В. С. Научные проблемы тестового контроля знаний / В. С. Аванесов. – М. : Исслед. центр, 1994. – 135 с.
12. Аванесов В. С. Основы научной организации педагогического контроля в высшей школе / В. С. Аванесов. – М. : Исслед. центр, 1989. – 168 с.
13. Гласс Дж. Статистические модели в педагогике и психологии / Дж. Гласс, Дж. Стэнли. – М. : Прогресс, 1976. – 496 с.

14. Ингекамп К. Педагогическая диагностика / К. Ингекамп. – М. : Педагогика, 1991. – 240 с.

15. Клайн П. Справочное руководство по конструированию тестов : введение в психометрическое проектирование : пер. с англ. / П. Клайн ; под ред Л. Ф. Бурлачука. – Киев : ЛТД, 1994. – 288 с.

16. Майоров А. Н. Тесты школьных достижений: конструирование, проведение, использование / А. Н. Майоров. – СПб. : Образование и культура, 1996. – 304 с.

17. 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.

18. Фігурська Л. В. До проблеми забезпечення ефективності тестових технологій як засобу контролю професійної підготовки майбутніх спеціалістів //Професійні компетенції та компетентності вчителя (Матеріали регіонального науково-практичного семінару). – Тернопіль: Вид-во ТНПУ ім.В.Гнатюка, 2006. – 188 с.

ТЕОРІЯ І ПРАКТИКА КОНСТРУЮВАННЯ ЗАВДАНЬ В ТЕСТОВІЙ ФОРМІ

Кухар Л.О.

Однією із сфер євроінтеграції України є вища освіта, а одним із найважливіших показників європейської якості освіти – рівень навченості студентів. Визначення рівня навченості студентів полягає в їх оцінюванні за допомогою певного інструментарію. Нині одним із ефективних, технологічних та об'єктивних засобів оцінювання студентів є тести - як своєрідний інструментарій тестування.

Застосування тестування як методу вимірювання рівня навченості зумовило появу науково-педагогічних досліджень, спрямованих на розробку відповідної теорії, різних видів тестів, технології тестування та опрацювання і інтерпретації отриманих результатів. Нині налічується досить велика кількість різних тестів.

Результатом досліджень, проведених в Україні, стали рекомендації щодо проведення тестування студентів в умовах кредитно-модульної системи навчання, проте недостатньо дослідженими залишаються питання аналізу впливу тестового контролю на якість знань студентів, класифікації тестових завдань, особливостей їх конструювання та стандартизації з метою підвищення ефективності навчального процесу у вищій школі.

Тестування займає вагомe місце у процесі навчання та самостійної роботи студента. Та, нажаль, не всі матеріали які використовуються для тестового контролю мають належне оформлення та перевіряють саме те що потрібно.

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

Відмінна риса тесту – можливість вимірювання у кількісній і якісній формі, що дозволяє встановити динаміку якості навчання та виконати його діагностичний аналіз.

Тестовий контроль відрізняється від інших методів контролю тим, що він є спеціально підготовленим контрольним набором завдань, який дозволяє за допомогою статистичних методів визначити наявність та рівень змін.

Структура тесту залежить від мети тестової перевірки знань. Правильно розроблені завдання в тестовій формі перевіряють всю пізнавальну сферу. Відповідно до оновленої таксономії Блума доцільно розглядати окремо знання і пізнавальні процеси (запам'ятовування, розуміння, використання, аналіз, оцінювання, створення).

З практичного досвіду конструювання тестів можна запропонувати таку етапність послідовність конструювання тестів.

1. Визначення мети тестування.
2. Добір змісту навчального матеріалу.
3. Проектування матриці тесту.
4. Формування структури банку тестових завдань.
5. Конструювання тесту відповідно до рівнів пізнавальної діяльності.

6. Проведення тестування.
7. Оцінювання та аналіз результатів тесту.

У результаті розвитку тестології та практичного застосування тестів відбувся логічний перехід від тестових завдань до «завдань у тестовій формі». Якщо **тестове завдання** – це складова одиниця тесту (кожне завдання є невід'ємною його частиною, тому видалення хоча б одного завдання неминуче призведе до виникнення прогалин в оцінюванні) як засобу контролю, що відповідає не тільки формальним, але й певним статистичним вимогам (частка неправильних відповідей в кожному завданні; співвідношення балів за завданнями та балів за весь тест). То **завдання в тестовій формі** – це новий педагогічний засіб, який відповідає таким вимогам:

- 1) відповідність меті (кожне тестове завдання має оцінювати досягнення важливої освітньої цілі, слід уникати перевірки тривіальних або надмірно вузькоспеціальних знань);
- 2) лаконічність (досягається шляхом ретельного добору слів, символів, графіків, які дозволяють досягнути максимального розуміння змісту завдання);
- 3) логічна форма висловлювання;
- 4) умова має містити чітко сформульоване завдання, яке фокусується на одній проблемі.
- 5) варіанти відповідей (дистрактори) мають бути гомогенними (однорідними).
- 6) усі дистрактори мають бути вірогідними (правдоподібними).
- 7) відомості, що містяться в одному тестовому завданні, не повинні давати відповідь на інше тестове завдання.
- 8) не рекомендується використовувати як правильну відповідь чи дистрактор фрази "все з вищевказаного", "нічого з вищевказаного".
- 9) не рекомендується використовувати як правильну відповідь чи дистрактор фрази "жоден варіант відповіді неправильний", "немає правильної відповіді", "усі відповіді правильні", "інколи", "ймовірно" тощо;
- 10) при формулюванні умови необхідно уникати наступних підказок:

- ✓ граматична невідповідність між умовою та варіантами відповідей;
- ✓ повторення у правильній відповіді слів з умови;
- ✓ використання прикладів з підручника чи лекції як тестових завдань;
- ✓ найдовша правильна відповідь;
- ✓ найдетальніша правильна відповідь;
- ✓ дистрактори, що виключають один одного;
- ✓ ситуації, коли одне тестове завдання є підказкою для іншого.

- 11) використання однакових правил оцінювання відповідей;
- 12) наявність однакової інструкції для всіх тестованих;
- 13) адекватність інструкції формі та змісту завдання;
- 14) визначений рівень складності;

та дозволяє організувати повноцінний навчальний процес, за якого формуються не тільки знання, уміння, навички, а й відбувається розвиток критичного мислення та рефлексивність особистості.

Існує багато думок щодо класифікації тестових завдань, зокрема, виділяють тестові завдання закритої та відкритої форм.

У тестових завданнях відкритої форми варіанти відповідей не надаються ані особі, яка тестується, ані особі, яка їх перевіряє, тобто це тестові завдання з вільною формою відповіді. Їх поділяють на:

- завдання з пропусками,
- завдання на доповнення,
- завдання з короткою відповіддю,
- завдання у формі структурованого есе.

В сучасній тестології (К. Кляйн, Анастасі А., Раш Д., Аванесов В.С., Челишкова М.Б. та ін.) виділяють основні 4 види закритих завдань в тестовій формі подання:

- тестові завдання з однією правильною відповіддю;
- тестові завдання з вибором кількох правильних відповідей;
- тестові завдання на встановлення правильної послідовності;
- тестові завдання на встановлення відповідності.

Якість тесту прийнято визначати за кількома властивостями.

- ✓ Тест повинен мати необхідний і достатній рівень складності.

Складність — міра розумових зусиль, потрібних для вибору відповіді. Часто цей параметр називається вагою і визначається як частка правильних або неправильних відповідей при відповіді на завдання в групі тестованих.

✓ Тест повинен бути валідним.

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

✓ Тест повинен бути надійним.

Надійність — міра правильності і адекватності відображення рівня знань студентів. Надійний тест забезпечує рівні права кожній групі студентів і кожному студенту в групі.

✓ Тест повинен бути об'єктивним.

Об'єктивність досягається шляхом стандартизації процедури проведення тестування, стандартизації і перевірки показників якості завдань і тесту в цілому.

Стандартизація тесту – сукупність експериментальних, методичних і статистичних процедур, що забезпечують створення суворо фіксованих компонентів тесту (інструкції, набору завдань, методу опрацювання результатів і підрахунку балів, способу інтерпретації).

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

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

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

Отже, із розвитком тестології відбувся логічний перехід від тестових завдань, як засобу контролю рівня навченості студентів, до завдань у тестовій формі, що, передусім, дають можливість застосувати тести ще й для навчання, використовувати більшу кількість матеріалу. Використання завдань в тестовій формі дозволяє студенту самостійно виявляти прогалини у власних знаннях та вживати заходів щодо їх усунення.

Наявність якісно розроблених завдань в тестовій формі для формування тесту є невід'ємною умовою ефективності тестування.

Використання навчального потенціалу завдань в тестовій формі є одним із пріоритетних напрямів реалізації принципу єдності та зв'язку навчання і контролю.

СТАНОВЛЕННЯ І РОЗВИТОК СИСТЕМ ОЦІНЮВАННЯ І МОНІТОРИНГУ ЯКОСТІ ОСВІТИ В УКРАЇНІ

Сліпухіна І. А.

Забезпечення якості (Quality Assurance)¹ вищої освіти не є суто європейською проблемою. В усьому світі зростає увага до якості та стандартів, що пояснюється швидким розвитком вищої освіти і водночас підвищенням вартості освітніх послуг для держави і людей. Європа має намір здійснити свою мету стати найдинамічнішою та інтелектуальною економікою у світі (Лісабонська Стратегія). Нові ініціативи і вимоги, що виникають як в Європі, так і поза її межами, з огляду на інтернаціоналізацію вищої освіти, потребують відповідної реакції. Формування системи забезпечення якості, заснованої на єдиних стандартах і рекомендаціях, свідчить про виникнення справжнього європейського виміру у сфері забезпечення якості, що має посилити привабливість Європейського простору вищої освіти [1].

Встановлено, що забезпечення якості вищої освіти є багатоплановим і включає: наявність необхідних ресурсів (кадрових, фінансових, матеріальних, інформаційних, наукових, навчально-методичних тощо); організацію навчального процесу, яка найадекватніше відповідає сучасним тенденціям розвитку національної та світової економіки та освіти; контроль освітньої діяльності ВНЗ та якості підготовки фахівців на всіх етапах навчання та на всіх рівнях, зокрема на рівні ВНЗ, державному та міжнародному (європейському) рівнях. Аналіз Європейських стандартів забезпечення якості вищої освіти дозволив виділити

¹ Термін "Quality Assurance" включає процеси оцінювання, акредитації та аудиту

основні вимоги і напрямки забезпечення якості освіти та механізми їх реалізації в Україні.

Забезпечення якості освіти при формуванні зони європейської освіти є однією з головних умов довіри, мобільності, мотивації студентів, сумісності й привабливості європейської вищої освіти.

В Європі історично сформувались «англійська» модель якості освіти, яка базується на внутрішньому самооцінюванні академічної спільноти, і «французька» модель, що ґрунтується на зовнішньому оцінюванні діяльності ВНЗ з погляду його відповідальності перед суспільством. В Європі поки що відсутня єдина система інституційного оцінювання діяльності освітніх установ, аналогічна системі акредитації в США. Проте в кожній країні існують свої підходи до забезпечення й оцінювання якості вищої освіти [2].

Європейська мережа гарантії якості у вищій освіті (ENQA) була утворена у 2000 році. Основою її створення стали пілотний проект «Європейський пілотний проект з оцінювання якості вищої освіти» (1994-1995 рр.), матеріали Рекомендацій Єврокомісії (98/561/ЕС від 24 вересня 1998 р.) з європейського співробітництва у питаннях гарантії якості у вищій освіті й текст Болонської декларації 1999 року [3].

Призначення ESG – надати допомогу і визначити орієнтири як для вищих навчальних закладів при розробленні своїх власних систем забезпечення якості, так і для агенцій, які здійснюють незалежні перевірки. Крім того, ці стандарти і рекомендації роблять свій внесок у створення спільної основи для діяльності як закладів освіти, так і агенцій із забезпечення якості.

4 березня 2008 року в Брюсселі (Бельгія) було створено Європейський реєстр забезпечення якості (EQAR), засновником якого стала група E4. Головною умовою включення до реєстру є відповідність ESG та ряд інших критеріїв, визначених у доповіді групи E4 на конференції у Лондоні 2007 року [3]. Метою діяльності EQAR є забезпечення прозорої і доступної інформації про надійні агенції із забезпечення якості освіти, що працюють в Європі, а також сприяння мобільності студентів та підвищення довіри до вищих навчальних закладів. За даними інвентаризації

виконання групою країн-учасниць вимог Болонського процесу на 2007 рік [4]. Україна отримала за індикаторами забезпечення якості освіти оцінку 3,5 за п'ятибальною шкалою при загальній оцінці країн-учасниць за цим показником – 4,1.

Приєднавшись до Болонського процесу, Україна взяла на себе зобов'язання проводити роботу з приведення якості національної освіти відповідно до європейських стандартів.

Проблема оцінювання якості освіти є наразі однією найактуальніших для усієї системи освіти України. Загальна риса системних змін в освіті – направленість на забезпечення впровадження в життя системи менеджменту якості освіти, як системи, яка направлена на досягнення результатів в навчанні відповідно до вимоги сучасного стану розвитку суспільства. Ми є свідками новітніх процесів в існуванні і функціонуванні цивілізації – тотальної інформатизації, глобалізації, технократизації. В інформаційному суспільстві, у неперервній системі технологічних інновацій якість освіти є основним аргументом людського розвитку, яка забезпечить такий рівень життєвої і професійної компетентності людини, який задовольняв би її прагнення до самовдосконалення і саморозвитку і, як наслідок, потреби суспільства в освічених і висококультурних громадянах.

Національна стратегія розвитку освіти в Україні на 2012-2021 роки констатує той факт, що наразі освіта є пріоритетним напрямом державної політики України. Держава виходить з того, що освіта – це стратегічний ресурс соціально-економічного, культурного і духовного розвитку суспільства, поліпшення добробуту людей, забезпечення національних інтересів, зміцнення міжнародного авторитету й іміджу нашої держави, створення умов для самореалізації кожної особистості.

Модернізація структури, змісту й організації освіти на засадах компетентнісного підходу, переорієнтації на цілі сталого розвитку; забезпечення доступності та неперервності освіти упродовж життя; розвиток наукової та інноваційної діяльності в освіті, підвищення якості освіти на інноваційній основі; розвиток наукової та інноваційної діяльності в освіті; забезпечення національного моніторингу системи освіти; створення сучасної

матеріально-технічної бази системи освіти; інтегрування національної системи освіти у європейський і світовий освітній простір – визначальна частина напрямків розвитку освіти в Україні на 2012-2021 роки.

З метою прискорення процесу реформування освітньої галузі на виконання Програми економічних реформ на 2010-2014 роки «Заможне суспільство, конкурентоспроможна економіка, ефективна держава» на урядовому рівні прийнято низку надзвичайно важливих для освіти державних цільових програм серед яких: підвищення якості шкільної природничо-математичної освіти на період до 2015 року; розвитку професійно-технічної освіти на 2011-2015 роки; впровадження у навчально-виховний процес загальноосвітніх навчальних закладів інформаційно-комунікаційних технологій «Сто відсотків».

Продовжується робота щодо реалізації державних програм, спрямованих на інформатизацію та комп'ютеризацію загальноосвітніх, професійно-технічних та вищих навчальних закладів, упровадження інформаційних та комунікаційних технологій в освіті і науці; забезпечення загальноосвітніх, професійно-технічних і вищих навчальних закладів сучасними технічними засобами навчання природничо-математичних і технологічних дисциплін; державних цільових програм роботи з обдарованою молоддю; реалізацію принципів Болонського процесу у вищій освіті, подальше впровадження кредитно-модульної системи навчання тощо.

Водночас глибина освітніх реформ, якість і ефективність роботи навчальних закладів і установ системи національної освіти не можуть повністю задовольнити сучасні потреби особистості і суспільства.

Нинішній рівень освіти в Україні не дозволяє їй повною мірою виконувати функцію ключового ресурсу соціально-економічного розвитку країни і підвищення добробуту громадян. Залишається низькою престижність освіти і науки у суспільстві.

Інтеграція України у світовий освітній простір вимагає постійного вдосконалення національної системи освіти, пошуку ефективних шляхів підвищення якості освітніх послуг, апробації та впровадження інноваційних педагогічних систем, реального

забезпечення рівного доступу всіх її громадян до якісної освіти, можливостей і свободи вибору в освіті, модернізації змісту освіти і організації її адекватно світовим тенденціям і вимогам ринку праці, забезпечення безперервності освіти та навчання протягом усього життя, розвиток державно-громадської моделі управління.

З огляду на визначені пріоритети найважливішим для держави є виховання людини інноваційного типу мислення та культури, проектування акмеологічного освітнього простору з урахуванням інноваційного розвитку освіти, запитів особистості, потреб суспільства і держави. Якісна освіта є необхідною умовою забезпечення сталого розвитку суспільства, консолідації усіх його інституцій, гуманізації суспільно-економічних відносин, формування нових життєвих орієнтирів особистості.

Серед найближчих перспективних завдань, які потребують негайного виконання декларуються такі:

- оновлення цілей і змісту освіти на основі компетентнісного підходу та особистісної орієнтації, урахування світового досвіду та принципів сталого розвитку; перехід від процесної до результатної, компетентнісної парадигми освіти;
- забезпечення системного підвищення якості освіти на інноваційній основі; створення сучасного психолого-педагогічного та науково-методичного супроводу навчально-виховного процесу;
- розвиток системи державно-громадського управління освітою, забезпечення об'єктивного оцінювання якості освітніх послуг.

Термін *«якість освіти»* є інтегральною характеристикою системи освіти, який відображає ступінь відповідності реальних освітніх результатів, які досягаються у процесі навчання, нормативним вимогам, соціальним і особистісним очікуванням. Дослідження різноманітних інформаційних джерел вказує на існування широкого спектру критеріїв, за якими визначається якість освіти. Найчастіше, оцінюючи якість освіти, фахівці мають на увазі рівень освітніх досягнень учнів, продуктивність освітнього процесу, науково-педагогічний потенціал освітнього

закладу. Якість освіти в контексті програми ЮНЕСКО «Освіта для всіх» включає серед іншого:

- *здорових учнів*, які отримують хороше харчування, готові брати участь у освітньому процесі за підтримки сім'ї і суспільства;
- *освітнє середовище* – здорове, безпечне, забезпечене необхідними ресурсами і сприятливими умовами;
- *зміст освіти*, який відображено у відповідних навчальних програмах і матеріалах для оволодіння базовими знаннями, вміннями і навичками, а також знаннями в таких галузях як гендер, здоров'я, харчування, подолання хвороб ВІЛ/СНІД, світове співіснування;
- *навчальний процес*, спрямований на розвиток і формування особистості;
- *результати*, які включають знання, вміння і особистісну позицію, і пов'язані з національними завданнями в галузі освіти і з позитивною участю у суспільному житті [4].

Поняття *якості вищої освіти* на думку більшості дослідників – це збалансована відповідність вищої освіти (як освітнього середовища, процесу, результату) різноманітним цілям, вимогам, потребам, умовам, нормам (стандартам).

Наразі за даними соціологічних досліджень (дані Інституту Горшеніна, м. Київ) щодо якості вітчизняної освіти, майже половина українців (45,2 %) задовільно оцінила якість освіти, що надається сучасною середньою школою. Чверть опитаних (24,9 %) оцінили його як «хороший», кожен п'ятий (19,4 %) дав оцінку «погано» і лише 5,8 % опитаних оцінили якість середньої освіти в Україні на «відмінно». Вагаються з відповіддю на це питання 4,7 % респондентів [7].

Основою європейських домовленостей у рамках Болонського процесу, до якого приєдналася Україна, є саме якість освіти. Єдине розуміння якості освіти є необхідною умовою створення єдиного освітнього простору, який включає в себе підвищення мобільності студентів, викладачів, випускників, зближення позицій європейських країн у економічній, соціальній і культурній сферах. Розроблення стратегії і тактики управління якістю освіти наразі є темою і метою досліджень державного

рівня. Згідно з Болонською угодою, відповідальність за якість освіти покладена насамперед на кожен окремий навчальний заклад, й у такий спосіб забезпечується можливість перевірки якості системи навчання на національному і міжнародному рівнях. Система акредитації освітніх закладів та освітніх програм складає основу забезпечення якості освіти. Якість освіти забезпечується на трьох рівнях:

- 1) на рівні держави – акредитація;
- 2) на рівні суспільства – зовнішня експертиза;
- 3) на рівні освітньої установи – зовнішня і внутрішня експертизи.

До основних завдань системи оцінювання якості освіти відносяться:

1. Оцінювання рівня освітніх досягнень учнів освітніх установ для їх підсумкової атестації та добору на наступний щабель навчання.

2. Оцінка якості освіти на різних ступенях навчання в рамках моніторингових досліджень якості освіти (федеральних і міжнародних).

3. Формування системи вимірників для різних користувачів, що дозволяє ефективно реалізовувати основні цілі системи оцінювання якості освіти.

Для здійснення стабільного розвитку і нового якісного прориву в національній системі вищої освіти планується зокрема:

- завершення розроблення та прийняття Верховною Радою України Законів України « Про вищу освіту» (нова редакція) та «Про національну рамку кваліфікацій»;

- розроблення стандартів вищої освіти нового покоління, зорієнтованих на компетентнісний підхід в освіті, узгоджених із новою структурою освітньо-кваліфікаційних (освітньо-наукового) рівнів вищої освіти;

- подальше удосконалення процедур і технології зовнішнього незалежного оцінювання як передумови забезпечення рівного доступу до навчання у вищій школі.

Визначальним показником потужності вищого навчального закладу, критерієм спроможності надавати освітні та наукові

послуги відповідних кваліфікаційних рівнів вищої професійної освіти, які відповідають вимогам державних освітніх стандартів МОНМС України, надавати наукові послуги, які відповідають вимогам споживачів (замовників) – підприємств та організацій профільних галузей та інших зацікавлених сторін є функціонування в навчальному закладі *системи менеджменту якості* (СМЯ), яка констатує політику і цілі університету для постійного покращення і підвищення задоволеності споживачів.

СМЯ використовує такі методичні засоби: засоби отримання даних; засоби надання даних; методи статистичного опрацювання даних; теорію загального менеджменту; теорію мотивацій і міжособистісних стосунків; економічні розрахунки; системний аналіз надання освітніх послуг; керування за допомогою управління.

Основними складовими СМЯ в освіті є:

1) констатування потреб і очікувань споживачів та інших зацікавлених сторін в галузі змісту та якості освіти;

2) наявність політики і цілей навчального закладу, які задовольнятимуть вище вказані потреби;

3) визначення процесів, які необхідні для досягнення визначеної, і відповідальних осіб;

4) визначення необхідних ресурсів і забезпечення ними відповідальних за процеси при досягненні мети;

5) розроблення і впровадження методів для вимірювання ефективності і результативності кожного процесу на підставі ключових показників якості;

6) визначення механізмів, необхідних для попередження невідповідностей і усунення їх причин. Реалізація цих механізмів в процесах СМЯ;

7) розроблення і застосування процесу для постійного покращення усієї СМЯ.

Одним з піонерів з впровадження в життя СМЯ освіти серед вітчизняних університетів є Національний авіаційний університет, керівництво якого систематично координує діяльність структурних підрозділів відповідно до вимог Міжнародного стандарту ISO 9001:2008, державного стандарту ДСТУ ISO 9001:2009, що спрямоване на подальше підвищення якості

освітньої та наукової діяльності, що перевіряється і констатується аудитами Бюро Верітас Сертифікейшн Україна.

Таким чином можна говорити, що якість підготовки фахівців в умовах сучасного ВНЗ визначається рядом зовнішніх і внутрішніх чинників, серед яких:

- вимоги Болонської угоди в якості освіти;
- атестаційні та акредитаційні вимоги;
- конкуренція ВНЗ на ринку освітніх послуг і на ринку праці;
- вимоги споживачів до освітніх послуг;
- застосування систем менеджменту якості в галузі освіти;
- політика держави щодо широкого впровадження ринкових механізмів в освіті;
- зростаючі вимоги персоналу ВНЗ до якості організації освітнього процесу.

Модернізація системи управління освітою ґрунтується на засадах інноваційних стратегій відповідно з принципами сталого розвитку, створення сучасних систем освітніх проектів та *моніторингу*²; розвитку моделі державно-громадського управління, посилення ролі та взаємодії усіх суб'єктів освітньої політики, у якій особистість, суспільство й держава стають рівноправними суб'єктами і партнерами.

Для ефективного управління, прийняття обґрунтованих рішень з управління якістю необхідно володіти надійною і достовірною інформацією про хід освітнього процесу, діяльності системи управління тощо. Отримання такої інформації можливе при здійсненні моніторингу.

Освітній моніторинг – форма організації, збирання, зберігання, обробки та розповсюдження інформації про діяльність педагогічної системи, що забезпечує безперервне стеження за її станом і прогнозуванням її розвитку.

² Моніторинг – процес спостереження і реєстрації даних про певний об'єкт на часових інтервалах, що неперервно прилягають один до одного, протягом яких значення даних суттєво не змінюються.

Оцінювання якості освіти з огляду на динаміку зміни її результатів – дуже продуктивний спосіб, що вимагає систематичного моніторингу результатів. Таким чином, фундаментальною основою механізму управління якістю освіти виступає психолого-педагогічний, медичний, соціальний моніторинг, а сутністю процесу управління розвитком якості освіти є рефлексивний підхід як у розвитку самого об'єкта (освітнього процесу), так і управління ним.

Механізм керування якістю освіти пов'язаний з його структурними елементами: збирання інформації від потенційних соціальних замовників → формування соціального замовлення → визначення місії вищої школи → співвідношення обраного варіанта з наявними можливостями → вибір типу управління (на процес або на результат) → визначення параметрів оцінювання результатів освіти → діагностика особистості учнів → прогнозування результатів освіти учнів → відповідність якості освіти поставленій меті.

Даний механізм відповідає науково-теоретичним положенням у розробленні визначення якості і заснований на відповідних підходах і принципах. Механізм управління якістю універсальний, він дозволяє визначити якість освіти на всіх рівнях освітнього процесу від учня (студента) до адміністрації освітньої установи.

Місія сучасного ВНЗ полягає в побудові освітнього простору як простору життєвого самовизначення людини у всій різноманітності його проявів у сучасній культурній практиці, що в свою чергу визначає новий етап розвитку освітнього закладу, спрямований на: підготовку фахівців нового типу; всебічну підтримку розвитку вітчизняної освіти в різних регіонах країни; на створення нових переваг університетської освіти та підвищення її конкурентоспроможності.

Параметри міжнародних оцінок якості освіти, зокрема, відомих міжнародних досліджень PISA, TIMSS, CIVIC, LINGUA, PIRLS відображають такі характеристики якості результатів освіти, як компетентність, мотивація до неперервної освіти і

професійного росту, які складають характеристику сучасного культурного типу особистості [8].

Отже, не даремно одним із головних напрямів модернізації освіти сучасного періоду виступає перш за все, корінна зміна якості підготовки конкурентоздатних фахівців, які в сучасних умовах є головним критерієм оцінювання діяльності навчальних закладів, що включає в себе:

- 1) подальше підвищення якості освіти;
- 2) розроблення і введення державного стандарту нового покоління;
- 3) упровадження незалежного оцінювання якості освіти;
- 4) підготовку компетентних, практико-орієнтованих фахівців, що відповідають потребам ринку праці.

На необхідність виконання вище зазначених завдань вказано в Указі Президента України від 30 вересня 2010 р. № 926 «Про заходи щодо забезпечення пріоритетного розвитку освіти в Україні» [5] і відповідній Постанові Кабінету міністрів України від 14 грудня 2011 р. № 1283 «Про затвердження Порядку проведення моніторингу якості освіти» [6].

Відповідно до вказаних документів моніторинг проводиться шляхом: збирання та проведення аналізу інформації про стан системи освіти; підготовки статистичної та аналітичної інформації про якість освіти. Він проводиться на локальному, регіональному та загальнодержавному рівні. Об'єктами моніторингу є: інформація про учасників освітнього процесу, зокрема про стан їх здоров'я, соціальний захист, умови життя та навчання, готовність до провадження певного виду діяльності, рівень задоволення потреби в освітніх послугах; процеси, що відбуваються у системі освіти, та характеристики її стану; результати навчальної діяльності; навчально-методичне, матеріально-технічне, нормативно-правове, кадрове забезпечення освітнього процесу. Моніторинг проводить державна наукова установа "Інститут інноваційних технологій і змісту освіти" МОНмолодьспорту, установи та організації, що належать до сфери управління.

Методами проведення моніторингу є:

- 1) опитування різних груп респондентів;
- 2) тестування;

- 3) збирання статистичних даних про стан системи освіти за встановленою виконавцем формою;
- 4) вивчення документів навчальних закладів, органів управління освітою.

Моніторинг проводиться в п'ять етапів.

На першому етапі визначається мета, завдання, строки, процедура, методи проведення моніторингу та критерії оцінювання.

На другому етапі готуються інструктивно-методичні матеріали щодо проведення моніторингу.

На третьому етапі проводиться моніторинг.

На четвертому етапі проводиться аналіз результатів моніторингу, готується статистична та аналітична інформація про якість освіти.

На п'ятому етапі оприлюднюється зазначена інформація.

МОНМС України враховує результати моніторингу під час формування та реалізації державної політики в галузі освіти.

Однак в процесі виконання вище вказаних завдань виникли деякі труднощі і протиріччя. Так, директор Центру тестових технологій і моніторингу якості освіти Ігор Лікарчук вважає, що споживачі освітніх послуг не можуть дізнатися про якість освіти через відсутність незалежної системи моніторингу. Про це він повідомив 8 лютого 2012 року під час проведення прес-конференції, що була присвячена презентації аналітичної доповіді «Про стан моніторингу якості освіти в Україні» [9].

За даними авторів звіту, Україна є єдиною країною Європейського простору вищої освіти, де немає незалежної недержавної агенції з оцінювання якості освіти, незважаючи на те, що наша держава є урядовим членом Європейського реєстру забезпечення якості освіти.

Наразі в Україні одні й ті ж органи управління освітою виконують як управління галуззю так і функції контролю якості освіти. Як наслідок, замовники та споживачі освітніх послуг – учні, студенти та їх батьки, роботодавці, незалежні експерти, громадські організації не мають можливості дізнатися, якою ж є реальна якість української освіти.

Саме через це більшість реформаторських заходів у системі освіти не досягають мети, ведуть до неефективного використання бюджетних і приватних коштів, знецінюють сутність і зміст освітньої діяльності.

Є думка, що нормативні документи, прийняті в листопаді 2011 року – січні 2012 року, не можуть забезпечити створення системи моніторингу освіти. Зокрема, це стосується положення про національну систему рейтингового оцінювання діяльності вищих навчальних закладів, що затверджене Міністерством освіти і науки, молоді та спорту України.

«Ці документи орієнтовані на отримання формальних показників, проведення безсистемних ситуативних контрольних процедур і посилення централізації в управлінні освітньою галуззю», - підкреслив Ігор Лікарчук.

Автори аналітичної доповіді щодо моніторингу якості освіти в Україні також повідомили, що в Україні немає визначеної й адаптованої з освітніми цілями системи державних освітніх стандартів, аналіз виконання яких міг би давати підстави для висновків про якість функціонування освіти.

«У державі не розроблено та не впроваджено національної системи освітніх індикаторів, яка б могла стати важливим джерелом інформації про функціонування галузі та забезпечила порівняння її результатів із аналогічними даними інших країн світу, зокрема тих, що входять до складу Організації економічного співробітництва та розвитку (ОЕСР)».

Отже, імплементація європейської системи забезпечення якості освіти в Україні передбачає здійснення таких заходів:

1. Розроблення рекомендацій щодо запровадження національних Стандартів та для забезпечення якості в Європейському просторі вищої освіти.
2. Створення системи акредитацій та Національного освітнього акредитаційного центру.
3. Підвищення міжнародного рівня участі України у зовнішній системі забезпечення якості.
4. Розроблення та запровадження правових механізмів участі роботодавців в процедурі контролю та суспільної відповідальності за якість вищої професійної освіти.

5. Підвищення рівня участі студентів у національній системі забезпечення якості.
6. Запровадження постійно діючого моніторингу якості вищої освіти з урахуванням світового, європейського та національного досвіду.
7. Удосконалення рейтингової системи оцінювання результатів діяльності вищих навчальних закладів як інструментів управління якістю вищої освіти.
8. Запровадження і розвиток практики забезпечення доступності і прозорості інформації з питань забезпечення якості вищої освіти на всіх рівнях, а саме: окремого ВНЗ, національному та міжнародному рівнях [1].

Література

1. Система забезпечення якості вищої освіти у Болонському процесі та механізми її імплементації в Україні / М. Карпенко <http://old.niss.gov.ua/Monitor/juni08/16.htm>
2. Report to the London conference of ministers on a European Register of Quality Assurance Agencies // www.enqa.eu/pubs.lasso
3. Bologna Process Stocktaking London 2007 // www.dfes.gov.uk/londonbologna/
4. Джанетт Кобли, Миске Уитт и др. Определение понятия качества в образовании /Джанетт Кобли, Миске Уитт и др. Доклад ЮНИСЕФ. – Нью-Йорк, 2000. – 125 с.
5. Національна стратегія розвитку освіти в Україні на 2012-2021 роки/ <http://www.mon.gov.ua/images/files/news/12/05/4455.pdf>
6. Про затвердження Порядку проведення моніторингу якості освіти Постанова КМУ №1283 від 14.12.11 року/ <http://osvita.ua/legislation/other/25860/>
7. Якість освіти в Україні потрібно наближати до Європейських стандартів/ А.Ярошенко <http://kiev.spravedlivo.org.ua/2011/10/alla-yaroshenko-yakist-osviti-v-ukraini-potribno-nablizhati-do-yevropejskix-standartiv/>
8. Якість освіти у контексті вимог сучасності/ Архипова С.П./ http://www.nbuv.gov.ua/portal/soc_gum/vchu/N135/N135p011-014.pdf
9. <http://osvita.ua/vnz/news/27399/>

**ЛОГІКА ПОБУДОВИ ПРОГРАМИ ВСТУПНОГО
ФАХОВОГО ВИПРОБУВАННЯ НА СПЕЦІАЛЬНОСТІ
8.18010022 ОСВІТНІ ВИМІРЮВАННЯ У КДПУ ім.
В.ВИННИЧЕНКА**

Ріжняк Р.Я.

Вступ. Інтеграція України у Європейський простір вищої освіти висуває необхідність встановлення національних параметрів проведення вступних випробувань при зарахуванні на навчання на освітньо-кваліфікаційний рівень магістр. У правилах прийому до Кіровоградського державного педагогічного університету імені Володимира Винниченка [1] зазначається, що для конкурсного відбору осіб на здобуття освітньо-кваліфікаційного рівня магістр за спеціальністю 8.18010022 Освітні вимірювання на основі базової та повної вищої освіти конкурсний бал обчислюється як сума результату фахового випробування з основ технологій освітньої діяльності, вступного екзамену з іноземної мови, середньозваженого балу додатку до диплому освітньо-кваліфікаційного рівня бакалавра та середнього балу результатів підсумкової державної атестації на освітньо-кваліфікаційному рівні бакалавр. Спеціальність 8.18010022 Освітні вимірювання належить до специфічних категорій, а тому участь у конкурсі на здобуття рівня магістр за вказаною спеціальністю можуть брати випускники різних бакалаврських спеціальностей. Фаховий вступний іспит “Основи технологій освітньої діяльності” має передбачати визначення рівня засвоєння навчальних програм (що розкривають структуру цього іспиту) для різних природничо-математичних, технічних напрямів підготовки та ряду гуманітарних спеціальностей, розроблених згідно національних стандартів вивчення технологій навчальної діяльності у вищій школі, що узгоджені з міжнародними стандартами у світлі рекомендацій Болонського процесу.

В даній статті автори мають на меті виходячи із вимог стандартів вищої освіти (Освітньо-професійна програма та Освітньо-кваліфікаційна характеристика спеціальності 8.18010022 Освітні вимірювання), правил прийому [1] та особливостей

профорієнтаційної роботи з випускниками ОКР бакалавр різних спеціальностей розкрити логіку та організаційні передумови створення програми вступного випробування для організації конкурсного відбору осіб на здобуття освітньо-кваліфікаційного рівня магістр за спеціальністю 8.18010022 Освітні вимірювання. З поставленої мети логічно випливають задачі дослідження: 1) визначити складові компоненти змісту програми фахового вступного випробування з основ технологій освітньої діяльності для відбору претендентів на навчання за спеціальністю 8.18010022 Освітні вимірювання; 2) визначити алгоритм підготовки вступної екзаменаційної документації з врахуванням специфіки підготовки абітурієнтів різних природничо-математичних, технічних напрямів підготовки та ряду гуманітарних спеціальностей.

Таким чином, об'єктом дослідження даної роботи є процес відбору осіб на здобуття освітньо-кваліфікаційного рівня магістр за спеціальністю 8.18010022 Освітні вимірювання, а його предметом – організаційні передумови та логіка проведення фахового вступного випробування з основ технологій освітньої діяльності для відбору претендентів на навчання за цією спеціальністю.

Складові компоненти змісту програми фахового вступного випробування з основ технологій освітньої діяльності та логіка проведення фахового випробування. Вступний іспит з основ технологій освітньої діяльності до магістратури за спеціальністю 8.18010022 Освітні вимірювання проводиться з метою: а) організувати якісний та фаховий процес відбору осіб на здобуття освітньо-кваліфікаційного рівня магістр за цією спеціальністю; б) врахувати специфіку підготовки абітурієнтів (ОКР бакалавр) різних природничо-математичних, технічних напрямів підготовки та ряду гуманітарних спеціальностей. Тому вступний іспит з основ технологій освітньої діяльності має базуватися на відповідних навчальних програмах бакалаврату, призначених для вищих навчальних закладів, які здійснюють підготовку фахівців з різних спеціальностей. Укладачами програми іспиту враховані рекомендації Міністерства освіти і науки, молоді та спорту України [2] щодо організації фахових випробувань, а саме визначення рівнів володіння

освітніми технологіями, компетенцій суб'єкта випробування, можливих контекстів користування діловою мовою, науковою термінологією, орієнтування у процедурах складання програм та контролю результатів навчання. Програми є загальними за характером і можуть використовуватись для різних спеціальностей та спеціалізацій професійної підготовки студентів. Програми складені у відповідності до вимог кредитно-модульної системи організації навчального процесу.

Вступний іспит з основ технологій освітньої діяльності до магістратури за спеціальністю 8.18010022 Освітні вимірювання визначає рівень володіння претендентами на навчання освітніми технологіями у контекстах:

- *Володіння інформаційно-комунікаційними технологіями* як складовою частиною освітніх технологій взагалі – поняття про апаратну частину ПК та про їх програмне забезпечення, уміння працювати з операційною системою, володіння текстовим редактором, електронними таблицями, системами управління базами даних, презентаціями, системами автоматичної обробки документів, системами автоматичної конвертації документів в електронний вигляд, зі скануванням зображень та їх сегментацією, з процесами розпізнавання тексту, з системами автоматичного перекладу та перевірки правопису, з інформаційними технологіями Internet;
- *Володіння діловою українською мовою* – широта лексичного запасу, рівень володіння базовою українською граматикою, навички української розмовної мови та культури усного спілкування; володіння сталими виразами та кліше, характерними для ділової української мови; рівень формування вміння працювати з особовими та інформаційними документами;
- *Володіння основами організації наукових досліджень* – знання і уміння щодо наукових досліджень, що є системоутворюючим фактором організації та проведення майбутнім магістром наукових досліджень у сфері предметної діяльності, а саме знання загальнотеоретичних положень наукових досліджень: методологічні принципи наукових досліджень і доцільність їх

застосування в дослідницькій роботі; методи наукових досліджень та їх застосування в наукових дослідженнях; загальнотеоретична структура та підходи до проведення науково-дослідної роботи;

- *Володіння термінологією та закономірностями філософії* – розуміння світогляду як духовно-практичного способу освоєння світу, соціальної спрямованості філософського знання, духовно-творчої природи філософії, поняття про структуру процесу пізнання, про форми та методи пізнання, про проблему істини в теорії пізнання, тлумачення логіки як науки про засоби та форми мислення, поняття як форми мислення, судження та його логічної структури, використання законів логічного мислення, засвоєння співвідношення законів формальної та діалектичної логіки, поняття про умовиводи, їх роль та структуру, про типологію умовиводів, про доведення, їх роль та структуру;
- *Володіння фактами та причинно-наслідковими зв'язками з історії України* – знання та розуміння складного процесу формування та розвитку багатомільйонного українського народу, його діяльності в соціально-економічній, духовній, політичній і державній сферах з давніх-давен до сьогодення, розгляд історії України в тісному взаємозв'язку з глобальними історичними процесами, з історією її найближчих сусідів, з якими у різні часи українці перебували у складі різних держав;
- *Володіння основами економічних знань* – трактування економіки, як сфери суспільного життя; розуміння змісту господарських відносин на певній території; поняття про виробничі ресурси, їх класифікацію, про виробництво як процес суспільної праці, про еволюцію виробництва, розуміння суті та структури економічної системи, знання основних типів економічних систем: ринкова економіка «вільної конкуренції», адміністративно-командна, змішана, традиційна системи; уявлення про сутність, передумови формування та функції ринку, про принципи ринкової економіки, про становлення і еволюцію грошових відносин, про підприємництво, його суть, принципи і форми функціонування, про циклічні коливання

економічного розвитку;

- *Володіння елементами теорії ймовірностей та математичної статистики* – поняття про елементи комбінаторики, про основні поняття теорії ймовірностей: стохастичний експеримент (випадковий дослід), випадкова подія, ймовірність, про статистичне та класичне означення ймовірності, основні терміни та задачі математичної статистики, про можливі варіанти дослідження вибірки.

Загальною метою програми з основ технологій освітньої діяльності є визначення рівня формування у студентів професійних освітньо-технологічних компетенцій, що сприятиме їхньому ефективному функціонуванню у культурному розмаїтті навчального та професійного середовищ. Тематика розділів програми вступного іспиту розроблена з урахуванням специфіки педагогічного університету і є уніфікованою програмою, що дає можливість охопити основну проблематику основ технологій освітньої діяльності і здійснити моніторинг студентів різних фахових напрямків. Таким чином, ми можемо визначити розділи програми вступного фахового випробування для організації конкурсного відбору осіб на здобуття освітньо-кваліфікаційного рівня магістр за спеціальністю 8.18010022 Освітні вимірювання так:

1. Інформатика та інформаційно-комунікаційні технології.
2. Ділова українська мова.
3. Основи наукових досліджень.
4. Філософія.
5. Історія України.
6. Економіка.
7. Теорія ймовірностей та математична статистика.

Перелічені розділи програми налічують по 10 питань, за виключенням розділу «Інформатика та інформаційно-комунікаційні технології», до складу якого входить 30 питань. Збільшення «ваги» цього розділу вступного випробування пов'язане з тим, що інформаційно-комунікаційні технології на сьогоднішній день вносять суттєві корективи в усі компоненти методичної системи навчання. Доцільність та ефективність використання ІКТ в навчальному процесі вищого навчального

закладу нерозривно пов'язані з поняттям інформаційної культури – системи компетентностей, необхідних для використання ІКТ у професійній діяльності. Саме це і стало визначальним у розподіленні переваг при формуванні білетів вступного випробування. Таким чином питання з розділу «Інформатика та інформаційно-комунікаційні технології» наявне у всіх білетах випробування. Інші два питання білету обираються абітурієнтом за принципом вільного вибору двох з шести розділів, що лишилися. Наведемо приклад одного з білетів випробування.

Білет № 1.

1. Оформлення слайдів комп'ютерної презентації за допомогою кольорів, фону, шаблонів (з розділу «Інформатика та інформаційно-комунікаційні технології»).
2. Виникнення українського козацтва. Запорізька Січ: устрій, побут (з розділу «Історія України»).
3. Види науково-дослідної роботи студентів (з розділу «Основи наукових досліджень»).

Висновки. На даний момент у КДПУ ім. В.Винниченка створена достатньо продумана збалансована система організації фахових вступних випробувань до магістратури з освітніх вимірювань, яка проходитиме літом 2012 року перше випробування в умовах реальної вступної компанії. Розроблені програми компонентів вступного випробування є взаємоузгодженими та враховують специфіку підготовки абітурієнтів різних природничо-математичних, технічних напрямів підготовки та ряду гуманітарних спеціальностей, так як базується на відповідних навчальних програмах бакалаврату, призначених для вищих навчальних закладів, які здійснюють підготовку фахівців з різних.

В Україні система підготовки фахівців освітніх вимірювань на різних освітньо-кваліфікаційних рівнях знаходиться на етапі становлення. У трьох ВНЗ, учасниках Проекту TEMPUS [3], розроблені внутрішні стандарти підготовки магістрів освітніх вимірювань, розробляється державний стандарт магістерської програми. Аналіз стандарту підготовки магістрів, експертне оцінювання змісту навчальних дисциплін у системі підготовки фахівців, створення лінійної структурно-логічної схеми

підготовки, кількісний аналіз внеску окремих дисциплін у основні компоненти фахової підготовки, просторове моделювання структури підготовки [4], а також дослідження особливостей профорієнтаційної роботи з випускниками ОКР бакалавр різних спеціальностей дали можливість виважено та обґрунтовано підійти до створення програм вступних випробувань для вступу на ОКР магістр спеціальності 8.18010022 Освітні вимірювання.

Література

1. Правила прийому до Кіровоградського державного педагогічного університету імені Володимира Винниченка у 2012 році. – Кіровоград, 2011. – 49 с.

2. Про затвердження Умов прийому до вищих навчальних закладів України в 2012 році // Наказ МОНмолодьспорт №1179 від 12.10.11 року.

3. Авраменко О.В., Ковальчук Ю.О., Сергієнко В.П., Сільвестров Д.С. Проект «Освітні вимірювання, адаптовані до стандартів ЄС» за програмою Європейського Союзу TEMPUS // Вісник ТІМО.– 2009.– № 9.– С. 44-47.

4. Авраменко О.В., Ріжняк Р.Я. Передумови створення різнорівневих програм підготовки фахівців з освітніх вимірювань у КДПУ ім. В.Винниченка // Наукові записки НДУ ім. М. Гоголя. Психолого-педагогічні науки.- 2011.- № 10.- С. 182-195.

АСИСТЕНТСЬКА ПРАКТИКА МАГІСТРАТУРИ 8.18010022 «ОСВІТНІ ВИМІРЮВАННЯ»

Лутченко Л.І.

Асистентська практика магістрів спеціальності 8.18010022. «Освітні вимірювання» (галузь знань 1801 Специфічні категорії) є логічним продовженням теоретичного навчання магістра і завершальним етапом його професійно-педагогічної підготовки. Вона проводиться в умовах, максимально наближених до реальних умов майбутньої професійної діяльності у навчальних закладах II-IV рівнів акредитації.

Мета практики – підготовка до цілісного виконання функцій фахівця з освітніх вимірювань, викладача дисциплін з освітніх вимірювань, до проведення системи навчально-виховної роботи з студентами, формування у студентів-практикантів досвіду педагогічної та оціночної діяльності, виховання морально-етичних якостей викладача вищого навчального закладу, індивідуального творчого стилю педагогічної діяльності, потреби в самоосвіті.

Особливістю асистентської практики є те, що крім педагогічної складової (магістранти вникають у коло реальних проблем професійної праці викладача, вивчають зміст і обсяг його роботи), практиканти повинні виконати ряд завдань освітньої складової програми практики, що стосуються моніторингу якості навчальних досягнень та рівня підготовки студентів, створення банку тестових завдань, їх калібрування.

Асистентська практика магістрів організовується деканатом фізико-математичного факультету спільно з кафедрою прикладної математики, статистики та економіки (ПМСЕ) на основі наказу по університету, де затверджуються факультетський керівник практикою, методисти з фаху та бази практики.

Перед початком асистентської практики проводиться настановна конференція, на якій магістрантам повідомляють мету практики, пояснюють її завдання та зміст, а також порядок проходження. По закінченню практики проводиться підсумкова конференція за результатами всіх видів робіт, виконаних практикантами.

В обов'язки факультетського керівника входить координація методичного керівництва з боку фахових методистів, які здійснюють систематичний контроль за проходженням практики студентами. По закінченню практики факультетський керівник за результатами обговорення практики на засіданні кафедри ПМСЕ готує письмовий звіт деканові фізико-математичного факультету, який заслуховується на вченій раді факультету.

Керівництво практикою покладається на факультетського керівника та завідувача кафедрою ПМСЕ, які планують розподіл магістрів-практикантів між викладачами-методистами, а також контролюють процес проходження практики студентами.

Загальне керівництво науково-педагогічною практикою здійснює декан факультету.

Оскільки в магістратуру з освітніх вимірювань вступають студенти, які навчалися на педагогічних спеціальностях, різних за фахом: «Математика», «Фізика», «Технологічна освіта», «Географія», «Біологія», «Історія» та ін., то виникає питання, яким чином магістранти можуть проявити здобутий потенціал з фахової підготовки. Вирішення цієї проблеми ми бачимо у тісному співробітництві з відділом моніторингу якості освіти нашого вузу, факультетом довузівської та післядипломної освіти, який організовує навчання старшокласників для підвищення якості їх підготовки (з предметів природничо-математичного, філологічного та суспільно-гуманітарного циклу) на вечірніх та очно-заочних курсах, зокрема, учні готуються до зовнішнього незалежного оцінювання з обов'язкових та обраних дисциплін, та Кіровоградським обласним інститутом післядипломної педагогічної освіти імені Василя Сухомлинського (у т.ч. з центром сучасних технологій оцінювання якості освіти).

Основні завдання та організаційні питання практики висвітлюються під час настановної конференції, на якій обов'язкова присутність як практикантів, так і їх фахових методистів.

Завдання практики. Магістранти під час практики повинні виконати такі завдання:

- протягом перших 2-3 днів практики скласти індивідуальний план роботи та ознайомитися з групою студентів (учнів), з якими вони будуть працювати, з документацією, що веде й оформлює викладач;
- протягом практики вести щоденник психолого-педагогічних спостережень;
- розробити конспект лекції дисципліни з освітніх вимірювань та конспекти 2-3 практичних (семінарських, лабораторних) занять, узгодити їх з фаховим методистом практики, після чого методист затверджує конспекти і дає дозвіл на проведення заняття, або його частини;
- провести 2-3 практичних (семінарських, лабораторних) заняття та зробити їх самоаналіз;

- відвідати 3-4 заняття студентів-практикантів, взяти участь у їх обговоренні та аналізі;
- провести позакласне заняття, пов'язане з підготовкою студентів (учнів) до тестування;
- сконструювати тематичні завдання у тестовій формі з закритими відповідями, провести тестування слухачів курсів або студентів, здійснити математико-статистичну обробку результатів тестування та їх аналіз;
- написати статистичну довідку за отриманими результатами проведеного тестування;
- написати звіт про проходження практики.

Заняття, проведені студентами оцінюються з позицій:

- вміння самостійно підготуватися та провести заняття;
- досягнення мети заняття;
- вміння творчо підходити до організації та проведення занять, використання інноваційних технологій навчання;
- вміння володіти аудиторією;
- техніка мови;
- вміння зробити підсумок заняття.

Присутність керівника практики на кожному занятті, що проводить студент-практикант є обов'язковою, як і подальший їх аналіз. Корисно, щоб заняття обговорювалися та аналізувалися при участі групи студентів-практикантів, тому одним із завдань практики є відвідування практикантом заняття, що проводить інший практикант, та його аналіз. Аналіз занять здійснюється за єдиною схемою, яка об'єднує загально-дидактичні, психологічні та методичні вимоги до його проведення. Дуже важлива в процесі аналізу правильна оцінка студентом своєї роботи, вміння його самостійно визначати позитивне та побачити недоліки в реалізації розробленого заняття.

Тестування, проведене студентами-практикантами, оцінюється з урахуванням: організаційно-технологічного забезпечення, проведення тестування, структурування тестових завдань, якості математико-статистичної обробки результатів тестування, їх аналізу, якості написання статистичної довідки, *вироблення методичних рекомендацій.

Звітна документація. Після закінчення педагогічної практики, магістранти повинні здати на кафедру прикладної математики, статистики та економіки таку документацію:

1. Індивідуальний план роботи студента-практиканта.
2. Щоденник психолого-педагогічних спостережень.
3. Текст розробленої тематичної лекції та розгорнуті плани-конспекти практичних (семінарських, лабораторних) занять, проведених у вузі з підписом та оцінкою методиста з фаху.
4. Самоаналіз проведених занять.
5. Конспект позакласного заходу.
6. Сконструйовані завдання у тестовій формі, їх математико-статистичну обробку, аналіз та зроблені висновки.
7. Статистичну довідку, написану за результатами виконаного тестування.
8. Звіт про проходження практики.
9. Витяг із засідання кафедри (фахової циклової комісії) вузу II-IV рівня акредитації з аналізом результатів проходження науково-педагогічної практики магістрантом та указанням оцінки у вигляді цифрового балу (у випадку, коли магістрант проходив практику не в КДПУ ім. Володимира Винниченка).

Характеристику на магістранта складає фаховий методист, який перевіряє здану практикантом документацію та виставляє загальну оцінку з асистентської практики, враховуючи:

- 1) оцінку якості проведених занять,
- 2) оцінку організації і проведення тестування та аналізу його результатів,
- 3) оцінку проведеної позааудиторної навчально-виховної роботи з студентами (учнями),
- 4) оцінку за оформлення документації,
- 5) оцінку за ставлення до практики.

Підсумки практики. Підведення підсумків практики проводиться на підсумковій конференції. Доцільно її провести у вигляді диспуту, круглого столу тощо. Це звичайно вимагає великої майстерності як від викладача, так і від студентів-практикантів, тому корисно використовувати певний досвід, який мають студенти з педагогічних практик бакалаврату. Попереднє ознайомлення з проблемами, з якими зустрілися магістри під час

практики, дасть змогу сформулювати тему диспуту, дібрати літературу, і під час підсумкової конференції спрямувати обговорення у потрібне русло.

Таке підведення підсумків практики викликає значний інтерес у магістрантів та стає ще однією формою їх професійної фахової підготовки.

Бажано, щоб результати практики знайшли відображення в дипломних магістерських роботах, наукових статтях, тезах конференцій.

Досвід проведення асистентської практики 2012 року показав, що доцільним було б перенесення терміну практики після теоретичного навчання II семестру, у якому магістранти вивчають когнітивну психологію та методику викладання спеціалізованих дисциплін з «Освітніх вимірювань». Тоді програму практики можна збагатити цікавими й змістовними завданнями з психології, крім того вивчення методики викладання спеціалізованих курсів дасть можливість магістрантам проводити заняття на більш високому методичному рівні.

Загалом, проведення асистентської практики сприяє формуванню у студентів-практикантів досвіду педагогічної та оціночної діяльності викладача вищого навчального закладу, індивідуального творчого стилю його педагогічної діяльності, розвитку організаторських здібностей та становленню як фахівця у галузі освітніх вимірювань.

ДЕРЖАВНА АТЕСТАЦІЯ В МАГІСТРАТУРІ 8.18010022 ОСВІТНІ ВИМІРЮВАННЯ

Авраменко О.В.

Державна атестація в магістратурі 8.18010022 Освітні вимірювання складається з двох етапів: державного екзамену з освітніх вимірювань та магістерської роботи. Державна атестація (3 кредити) проводиться протягом двох тижнів після закінчення 10-місячної програми підготовки.

Дисципліни, які виносяться у 2011-2012 н.р. на державний екзамен з освітніх вимірювань, можна умовно розділити на дві групи: *«Теорія вимірювань в освіті»* та *«Практична реалізація вимірювань в освіті»*.

Наведемо список питань державного екзамену:

I. «Теорія вимірювань в освіті»:

«Основи педагогічного оцінювання»

1. Загальнопедагогічні вимоги, що покладено в основу оцінювання навчальних досягнень учнів. Рівні навчальних досягнень учнів, якісна характеристика рівнів і загальні критерії їх оцінювання.
2. Історія, сучасний стан та тенденції розвитку систем оцінювання в освіті. Принципи і функції педагогічного оцінювання.
3. Зовнішнє незалежне оцінювання: зміст і процедура. Організаційно-технологічне забезпечення зовнішнього незалежного оцінювання.
4. Основні види і методи педагогічного оцінювання.

«Вимірювання в освіті»

1. Педагогічне вимірювання. Латентні змінні. Компоненти процесу педагогічних вимірювань. Інструмент вимірювання.
2. Типи вимірювання. Основні характеристики нормо-орієнтованого та критеріально-орієнтованого вимірювання. Відмінності між нормо-орієнтованими і критеріально-орієнтованими тестами.
3. Класифікація навчальних цілей. Таксономія Блума.
4. Поняття завдання у тестовій формі, тестового завдання, тесту. Вимоги до завдань у тестовій формі. Форми тестових завдань. Переваги та недоліки різних форм тестових завдань.
5. Завдання з вибором однієї та декількох правильних відповідей. Принципи формулювання та оцінювання.
6. Завдання на встановлення відповідності та правильної послідовності. Завдання відкритої форми з короткою та розгорнутою відповіддю. Принципи формулювання та оцінювання.
7. Поняття моніторингу. Комплекс моніторингових показників (індикаторів). Поняття якості освіти, її структура. Основні групи показників якості освіти.
8. Поняття моніторингу якості освіти. Моніторинг як кваліметричний інструмент. Етапи проведення моніторингового дослідження.

«Класичні тестові моделі»

1. Основні положення класичної теорії тестування. Паралельні та еквівалентні тести.
2. Кореляційна матриця. Коефіцієнт кореляції Пірсона, φ - коефіцієнт кореляції, коефіцієнт бісеріальної кореляції та їх інтерпретація.
3. Розрахунок основних математико-статистичних характеристик тестових завдань та тесту у рамках класичної теорії.
4. Поняття надійності тесту в класичній теорії. Коефіцієнти надійності Спірмана – Брауна та α -Кронбаха та їх порівняння.
5. Процедури оцінювання надійності тесту. Формули для обчислення коефіцієнта надійності.
6. Поняття валідності тесту. Змістова, критеріально-орієнтована, конструктна валідності та процедури для їх оцінювань.

«Математично-статистичні методи в освітніх вимірюваннях»

1. Знаходження законів розподілу випадкових величин на основі дослідних даних педагогічних вимірювань.
2. Вибірковий метод Оцінка результатів вибіркового спостереження у педагогії та психології, розповсюдження їх на генеральну сукупність. Мала вибірка.
3. Загальні методи вивчення зв'язку соціально-педагогічних явищ і процесів. Непараметричні показники зв'язку. Рангові коефіцієнти зв'язку.

«Моделі та методи IRT»

1. Основні припущення IRT. Однопараметрична модель Раша. Латентні параметри. Логіти. Характеристичні криві.
2. Дво- та три параметричні моделі Бірнбаума: недоліки та переваги. Порівняння моделей Бірнбаума з моделлю Раша.
3. Матриця відповідей. Первинні бали: обчислення, статистичні характеристики, властивості.
4. Обчислення оцінок латентних параметрів на основі емпіричних даних. Стандартні помилки вимірювання. Обчислення стійких оцінок латентних параметрів методом найбільшої правдоподібності.
5. Побудова характеристичних кривих для досягнень учасників тестування (PCC) та для складності тестових завдань (ICC) на основі емпіричних даних. Два підходи теорії параметризації тестів.

Перевірка гіпотези про відповідність емпіричних даних моделі Раша

6. Інформаційні функції тестового завдання та системи тестових завдань у різних моделях тестування
7. Роздільна здатність тесту (РСТ) Оцінка дискримінаційної здатності тесту. Аналіз дистракторів
8. Надійність. Середня квадратична помилка вимірів. Методи обчислення коефіцієнта надійності. Ретестова надійність
9. Валідність: класифікація, математичне тлумачення. Міра валідності завдань тесту. Метод підвищення контрастності груп
10. Класифікація шкал: кількісні та якісні шкали. Порядкові, метричні, нормовані та номінальні шкали у педагогічних вимірюваннях.
11. Порядкові шкали у педагогічному тестуванні. Відсоткова шкала. Процентильна шкала. Нормалізовані шкали типу $N(\mu; \sigma^2)$. Шкали нормалізованих процентилів. Нормалізовані шкали з постійним кроком. Ітераційна процедура для обчислення виставкових балів
12. Метричні шкали. Перенесення латентних параметрів на єдину метричну шкалу. Перетворення єдиної метричної шкали в нормовану. Остаточний бал учасників тестування

II. «Практична реалізація вимірювань в освіті»:

«Конструювання тестів»

1. Класифікації педагогічних тестів.
2. Основні етапи конструювання педагогічного тесту.
3. Планування змісту тесту та загальні принципи відбору.
4. Методичне оснащення тесту та його складові. Склад інструкції для особи, яка проводить тест, та склад інструкції для випробовуваного.
5. Системи завдань в тестовій формі. Ланцюгові завдання. Приклади.

«Моніторинг якості освіти»

1. Поняття якості освіти, її структура. Основні групи показників якості освіти.
2. Поняття моніторингу якості освіти. Комплекс моніторингових показників (індикаторів). Міжнародні системи критеріїв і показників якості шкільної освіти.
3. Моніторинг як кваліметричний інструмент. Етапи проведення

моніторингового дослідження.

4. Міжнародні моніторингові порівняльні дослідження якості освіти. Організатори міжнародних досліджень якості освіти (TIMSS, PISA, PIRLS, SITES).

«Комп'ютерні технології в тестуванні»

1. Види, типи, етапи, вимоги, характеристики комп'ютерного тестування.
2. Інформаційна безпека процедур масового комп'ютерного тестування.
3. Сервіси Web 2.0, та їх використання в освіті та освітніх вимірюваннях.
4. Програмні засоби для локального тестування, загальні характеристики, приклади використання.
5. Мережеве програмне забезпечення для проведення тестування. Система myTestX.
6. Мережеве програмне забезпечення для проведення тестування. Веб орієнтовані системи. Система x-TLS.
7. Загальна характеристика VLE Moodle, створення курсів.
8. Середовище Moodle, банк тестових завдань, створення тесту.
9. Аналіз тестових завдань та засоби оцінювання в Moodle.
10. Комп'ютерне адаптивне тестування. Види адаптації, алгоритми формування тесту

«Методика викладання освітніх вимірювань»

1. Розвиток освітніх вимірювань як окремої галузі. Складові освітніх вимірювань та зв'язок з іншими галузями. Проблеми освітніх вимірювань.
2. Методична система навчання освітніх вимірювань.
3. Методи досліджень у галузі освітніх вимірювань.
4. Вимоги до професійної відповідальності при оцінюванні. Стандарти освітнього та психологічного тестування. Кодекс справедливого тестування в освіті.
5. Стандарти компетентності вчителів з оцінювання учнів. Ключові компетенції фахівця з освітніх вимірювань. Освітньо-кваліфікаційна характеристика фахівця з освітніх вимірювань.
6. Методика навчання математико-статистичних основ освітніх вимірювань. Методика проведення лекційних та практичних занять з дисциплін математико-статистичного циклу.
7. Методика навчання розробки та використання тестів.

Використання засобів дистанційного навчання. Методика використання комп'ютерних технологій навчання і тестування.

- Методика навчання основ освітніх вимірювань та моніторингу в галузі освіти. Методика проведення лекційних та семінарських занять організаційно-педагогічного циклу.

Структура екзаменаційного білету передбачає перше питання з групи дисциплін «Теорія вимірювань в освіті», друге питання з групи дисциплін «Практична реалізація вимірювань в освіті», а також третє питання практичного характеру. Практичні завдання є достатньо різноманітними.

Приклад 1. Знайти вибіркові коефіцієнти кореляції Спірмена, перевірити його на значимість:

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

Приклад 2. 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.

№ 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.

Приклад 3. 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.

№з	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

Приклад 4. 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 тестам. 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, що середня помилка вимірів кожної оцінки повинна дорівнювати 0?

Приклад 5. Побудувати характеристичні криві для досягнень учасників тестування (РСС) з рівнями підготовки: (-1) логіт, 0 логітів, $+1$ логіт. Побудувати інформаційні функції для цих учасників та інформаційну функцію групи учасників.

Приклад 6. 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	Дистрактори			
	1	2	3	4
Емпірична частота вибору дистрактора n	130	175	120	175

Перевірити гіпотезу про рівномірний розподіл частот вибору дистракторів з рівнем значущості $\alpha = 0,05$.

Тематика магістерських робіт затверджується на початку навчального року, вона повинна відповідати стандарту спеціальності. Кожна робота повинна містити наукову новизну та мати практичне значення. Наведемо деякі теми дипломних робіт з метою та основними завданнями:

Тема «Статистична перевірка деяких гіпотез тестування»

Мета: аналіз можливостей різних статистичних методів щодо визначення відповідності імовірнісної моделі тестування експериментальним даним та застосування статистичних критеріїв до перевірки гіпотез щодо результатів педагогічного експерименту

Завдання:

- огляд гіпотез тестування та статистичних методів для їх перевірки;
- підготовка та проведення тестування, первинна обробка його результатів;
- формування основних та альтернативних гіпотез для поглибленого аналізу результатів педагогічного експерименту: порівняння емпіричної та теоретичної імовірності успіху, перевірка рівномірності розподілу дистракторів, перевірка значущості розбіжності різних результатів тестування на метричній шкалі;
- вибір статистичних критеріїв для перевірки гіпотез, застосування обраних критеріїв;
- педагогічне тлумачення статистичних результатів щодо якості тесту та значущості його результатів та рекомендації для вдосконалення тесту;
- теоретичні висновки щодо застосування статистичних критеріїв до перевірки гіпотез тестування.

Тема «Дослідження проблеми взаємозв'язку між дисперсіями латентних параметрів IRT»

Мета: виявлення взаємозв'язку між дисперсіями латентних параметрів IRT за результатами статистичного аналізу результатів тестування, аналіз тестування варіантами тесту різної складності та результатів проведення одного тесту серед учасників різного рівня підготовки

Завдання:

- огляд особливостей та виявлення проблем сучасної теорії тестування;

- вивчення математичного обґрунтування проблеми взаємозв'язку між дисперсіями латентних параметрів IRT;
- педагогічна інтерпретація проблеми взаємозв'язку між дисперсіями латентних параметрів IRT;
- підготовка, проведення тестування варіантами різного рівня складності, дисперсійний аналіз його результатів;
- підготовка, проведення тестування одним варіантом тесту серед учасників різного рівня підготовки, дисперсійний аналіз його результатів;
- педагогічне тлумачення статистичних результатів щодо взаємозв'язку між дисперсіями латентних параметрів IRT та вироблення рекомендацій щодо врахування проблеми моделі тестування.

Тема «Аналіз методів перевірки паралельності варіантів тесту»

Мета: порівняльний аналіз методів перевірки паралельності варіантів тесту на основі результатів застосування різних методів до результатів тестування.

Завдання:

- висвітлення поняття паралельності варіантів тесту та його застосування у педагогічних вимірюваннях;
- огляд статистичних методів перевірки паралельності варіантів тесту;
- підготовка та проведення тестування різними варіантами одного тесту;
- перевірка гіпотези про рівність рівнів складності однойменних завдань різних варіантів одного тесту;
- перевірка гіпотези про однорідність одного завдання у різних варіантах тесту;
- перевірка гіпотези про однорідність кожного завдання у всіх варіантах тесту;
- педагогічне тлумачення результатів перевірки статистичних гіпотез та вироблення рекомендацій щодо паралельності варіантів тесту.

Тема «Дослідження методів перенесення латентних параметрів на єдину метричну шкалу»

Мета: перенесення латентних параметрів на єдину метричну шкалу методом перекриттів варіантів та методом достатніх статистик та порівняння можливостей названих методів.

Завдання:

- огляд методів шкалювання результатів тестування та їх застосування у педагогічних вимірюваннях;
- підготовка та проведення тестування з варіантами, які мають перекриття;
- використання перекриття варіантів тесту для перенесення латентних параметрів на єдину метричну шкалу;
- обчислення достатніх статистик для теоретичного обґрунтування та практичного здійснення перенесення латентних параметрів на єдину метричну шкалу;
- формулювання висновків про переваги та недоліки розглянутих методів перенесення латентних параметрів на єдину метричну шкалу.

Тема: «Педагогічне тестування з математики у процесі підготовки до ЗНО»

Мета: на основі вивчення та аналізу наукової психолого-педагогічної математико-статистичної та методичної літератури, практичного досвіду вчителів з даної проблеми, визначити місце й обсяг та розробити зміст педагогічного тестування учнів з алгебри й початків аналізу у процесі підготовки до ЗНО з позицій сучасних вимог

Завдання:

- Проаналізувати психолого-педагогічну, математико-статистичну та методичну літературу, практичний досвід вчителів з питань організації педагогічного тестування учнів з математики у процесі підготовки до ЗНО; на цій основі здійснити аналіз структурних елементів проблеми дослідження та намітити шляхи її розв'язання.
- Визначити місце, зміст та обсяг педагогічного тестування учнів з математики у системі занять на уроках та в

позаурочний час.

- Розробити систему тестових завдань з алгебри й початків аналізу та провести їх апробацію.
- Забезпечити добір ефективних математико-статистичних методів для обробки результатів тестування, калібрування тестових завдань та підвищення їх надійності й валідності.
- Зробити висновки та розробити методичні рекомендації для вчителів щодо організації та проведення тестування з математики.

Тема: «Рейтингова система оцінювання знань учнів професійного ліцею»

Мета: на основі вивчення та аналізу наукової психолого-педагогічної математико-статистичної та методичної літератури, практичного досвіду викладачів з даної проблеми, розробити рейтингову систему оцінювання знань учнів з курсу «Виробниче навчання професії «Кухар».

Завдання:

- Проаналізувати психолого-педагогічну, математико-статистичну та методичну літературу, практичний досвід вчителів з питань запровадження рейтингової системи оцінювання знань учнів; на цій основі здійснити аналіз структурних елементів рейтингової системи навчання.
- Визначити принципи і функції рейтингової технології навчання у професійному ліцеї.
- Розробити рейтингову систему оцінювання знань ліцеїстів з курсу «Виробниче навчання професії «Кухар».
- Експериментально перевірити розроблену систему оцінювання знань студентів з курсу «Виробниче навчання професії «Кухар» у Кіровоградському професійному ліцеї імені О.С. Єгорова.
- Розкрити перспективи застосування системи рейтингового контролю в освітній практиці, зробити висновки.

Магістерська робота спеціальності 8.18010022 Освітні вимірювання має засвідчити, що випускник є компетентним

фахівцем у галузі вимірювань в освіті, вміє використовувати надбані теоретичні знання та навички роботи з сучасними технологіями для прийняття рішень, володіє методикою наукового дослідження, узагальнює та аналізує наукові джерела, статистичний матеріал, працює з нормативно-правовими актами.

Методичне забезпечення державної атестації підготовлене за Проектом 145029-TEMPUS-2008-SE-JPCR «Освітні вимірювання, адаптовані до стандартів ЄС». Розроблено декілька навчальних посібників, практикуми, навчально-методичні комплекси тощо. У 2011 році видано підручник «Вимірювання в освіті» за редакцією Авраменко О.В., який широко відображає основні напрями теорії та практики освітніх вимірювань, оскільки він планувався як допомога випускникам магістратури спеціальності «Освітні вимірювання» при підготовці до державної атестації. При підготовці матеріалів були опрацьовані різноманітні наукові та науково-методичні джерела, списки яких подано наприкінці кожного розділу підручника.

У підготовці до друку підручника «Вимірювання в освіті» 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;
- Парашук 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.

Перша державна атестація магістрів освітніх вимірювань відбудеться у травні 2012 року. Це буде перевірка не тільки досягнень випускників магістратури, але і перевірка майстерності професорсько-викладацького складу, задіяного у розробці програми їх підготовки. Сподіваємося, що і магістранти, і викладачі успішно пройдуть останній етап виконання програми Проєкту 145029-TEMPUS-2008-SE-JPCR «Освітні вимірювання, адаптовані до стандартів ЄС».

Не має сумніву, що перші магістри освітніх вимірювань зроблять вагомий внесок у піднятті на професійний рівень процесу тестування в освіті в Україні.

МЕТОДИКА НАВЧАННЯ ОСВІТНІХ ВИМІРЮВАНЬ ЯК ДИСЦИПЛІНА ПРОГРАМИ ПІДГОТОВКИ МАГІСТРІВ

Лупан І.В.

Курсом методики навчання завершується підготовка магістрів спеціальності «Освітні вимірювання». Досвід його викладання у 2011-2012 навчальному році для магістрантів Кіровоградського державного педагогічного університету імені Володимира Винниченка дозволяє викрити проблеми, які закономірно виникають при викладанні нового курсу, а також специфічні, притаманні лише курсу методики навчання освітніх вимірювань.

Цікаво, що проблеми методики підготовки фахівців у галузі освітніх вимірювань турбували науковців відтоді, як тестування та оцінювання в освіті набуло масових масштабів, про що свідчать, зокрема, публікації У.Дж.Фіндлі та А.Р.Гулліксона [1, 2].

На сьогодні основною проблемою є те, що формування галузі знань «Освітні вимірювання» знаходиться у стані активного становлення, остаточно не визначено її об'єкт та предмет, відповідно формується методична система навчання освітніх вимірювань. 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 підготовленому за проектом «Освітні вимірювання, адаптовані до стандартів ЄС» підручнику авторами зроблено спробу внести ясність у ці питання. Отже об'єктом освітніх вимірювань на нашу думку є «знання, уміння, навички, компетенції, інші кваліфікаційні характеристики, які підлягають оцінюванню в освітній діяльності та в галузі сертифікації та ліцензування професійної діяльності» [3, с. 310]. А предметом – «теоретичні засади, методи і засоби оцінювання та вимірювання, доцільні для прийняття рішень в галузі освіти, ліцензування та сертифікації професійної діяльності» [Там само].

Другою проблемою є становлення галузевої термінології, починаючи з основних: як з'ясувалося, в англomовній літературі використовують два різних терміна «assessment» та «evaluation», які українською перекладаються як «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 сучасної 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 вищого навчального закладу 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.

Третя проблема методики навчання пов'язана з тим, що спеціальність «Освітні вимірювання» відноситься до галузі знань «Специфічні категорії», це означає, що випускник кожної спеціальності може вступити до магістратури з даної спеціальності. При цьому фахівець освітніх вимірювань за короткий час навчання 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 [4, 5].

Основним напрямком у підготовці магістра є напрям 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».

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», – а також «Основи педагогічного оцінювання», «Ділове спілкування іноземною мовою», «Моніторинг якості освіти», «Педагогіка вищої школи» та «Класичні тестові моделі».

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.

Все це відображено у освітньо-кваліфікаційній характеристиці та в освітньо-професійній програмі підготовки магістра [6].

Магістрантам, які отримують також і кваліфікацію викладача з освітніх вимірювань необхідно поєднувати методики навчання дисциплін базового фаху, отриманого на освітньому рівні «бакалавр», педагогіки та психології, з методиками формування математичних та інформатичних компетентностей, необхідних кваліфікованому фахівцю з освітніх вимірювань.

Ще однією проблемою курсу методики є те, що саме в рамках даного курсу розглядаються питання професійної відповідальності при тестуванні та оцінюванні. Існує багаторічний досвід зарубіжних країн, у першу чергу Сполучених Штатів, з питань регулювання етичних, технічних та юридичних аспектів у галузі освітніх вимірювань [7]. Однак в Україні вони також знаходяться у стадії становлення [8].

Література

1. Warren G. Findley. When and how to teach educational measurements. – JSTOR Journal of Educational Measurement, Vol 1, No 1 (Jun, 1964), pp 87-89. – [Ел. ресурс] Article first published online: 12.09.2005. – Реж. дост.: <http://onlinelibrary.wiley.com/doi/10.1111/j.1745-3984.1964.tb00160.x/abstract>

2. Arlen R. Gullickson. Teacher Education and Teacher-Perceived Needs in Educational Measurement and Evaluation. – Journal of Educational Measurement. – Vol. 23, No. 4 (Winter, 1986), pp. 347-354.

3. Авраменко О.В., Лупан І.В., Ріжняк Р.Я. Освітні вимірювання: стандарти, форми та засоби навчання. //Вимірювання в освіті: Підручник / За редакцією О.В.Авраменко. – Кіровоград: Лисенко В.Ф., 2011. – С. 309-359.

4. 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.

5. 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.

6. Авраменко О.В., Ріжняк Р.Я. Передумови створення різнорівневих програм підготовки фахівців з освітніх вимірювань у КДПУ ім. В.Винниченка. // Наукові записки НДУ ім. М.Гоголя. Психолого-педагогічні науки. – 2011. – №10. – С.182-195.

7. 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.

8. 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.nbuv.gov.ua/e-journals/ITZN/em8/emg.html>. – Заголовок з екрана.

Розділ 3. Комп'ютерні технології в тестуванні

ПРО ДЕЯКІ АСПЕКТИ ВИКЛАДАННЯ ДИСЦИПЛІНИ «КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ ТЕСТУВАННЯ»: WEB-ОРІЄНТОВАНІ СИСТЕМИ ТЕСТУВАННЯ

Котяк В.В.

Вивчення практики комп'ютерного тестування у навчальному процесі ВНЗ призводить до усвідомлення необхідності наукового узагальнення накопиченого досвіду та розробки ефективних шляхів організації та проведення комп'ютерного тестування, адекватного цілям і змісту професійної підготовки майбутніх бакалаврів та магістрів. Процес змін в системі освіти, заданий Болонською декларацією, включає в себе, зокрема, введення нових механізмів і процедур забезпечення якості освіти.

Оскільки успішність навчання багато в чому зумовлюється оперативністю і достовірністю інформації про навчальні досягнення, вирішення завдання вдосконалення процесів управління якістю вищої професійної освіти передбачає підвищення ефективності контролю і оцінки якості підготовки студентів вищих навчальних закладів на кожному з рівнів.

Цим вимогам відповідає комп'ютерне тестування результатів навчання, що забезпечує отримання об'єктивної, достовірної інформації у сфері якості освіти.

Структура навчальної програми дисципліни «Комп'ютерні технології тестування»

Дисципліна «Комп'ютерні технології у тестуванні» є складовою програми підготовки магістрів в галузі освітніх вимірювань.

Метою її викладання є забезпечення фахової підготовки магістрів в галузі теорії та практики педагогічних вимірювань з використанням комп'ютерних технологій тестування. Опанувавши курс, майбутні фахівці ознайомляться з організаційними та методичними аспектами використання комп'ютерних систем призначених для проведення тестування, отримають уміння та навички роботи з широким колом програмних продуктів, призначених для проведення тестування, вмітимуть використовувати віртуальні освітні середовища та матимуть ґрунтовні знання у сфері сучасних комп'ютерних технологій у тестуванні.

Передбачається, що попередньо студенти отримали ґрунтовну психолого-педагогічну підготовку та мають достатній рівень комп'ютерної грамотності, прослухавши такі нормативні курси, як «Інформатика», «Педагогіка», «Загальна психологія» тощо.

Після вивчення дисципліни студенти повинні знати:

- інформаційні та методологічні складові безпеки комп'ютерного тестування;
- стан та перспективи розвитку систем тестування;
- види систем для проведенні комп'ютерного тестування;
- основні онлайн сервіси для організації та проведення тестування;
- засоби обробки результатів «паперового» тестування;
- склад та будову найбільш популярних пропріетарних та вільно розповсюджуваних систем комп'ютерного тестування;
- склад та призначення віртуальних освітніх середовищ;
- формати експорту-імпорту тестових завдань;
- основи адаптивного тестування.

Студенти також повинні вміти:

- організувати комп'ютерне тестування;
- використовувати інтернет-сервіси для проведення тестування;
- опрацьовувати результати бланкового тестування засобами ABBYY TestReader, ABBYY FlexiCapture, або аналогічними програмними засобами;
- створювати різні типи тестових питань в найбільш розповсюджених середовищах для комп'ютерного тестування;
- використовувати широкоживані програмні продукти для проведенні комп'ютерного тестування;
- експортувати та імпортувати тестові завдання;
- експортувати та аналізувати результати тестів;
- проводити аналіз результатів тестування.

Пропонуємо наступну структуру навчального матеріалу:

Змістовий модуль I

Організаційні та методичні особливості комп'ютерного тестування.

Освітні інформаційні технології

Сфери застосування ОІТ. Мета і завдання впровадження ОІТ в системі освіти

Типологія програмних засобів учбового призначення. Дидактичні вимоги до програмно-педагогічних засобів. Методичні вимоги до програмно-педагогічних засобів.

Недоліки ОІТ і способи їх компенсації.

Тенденції ролей учасників і засобів учбового процесу в умовах впровадження ОІТ.

Комп'ютерні технології атестації

Педагогічне і психологічне тестування.

Бланкове тестування. Десять недосконалостей.

Переваги комп'ютерного тестування.

Загальні достоїнства і недоліки тестових технологій.

Види, типи, етапи, вимоги, характеристики комп'ютерного тестування.

Комп'ютерні технології навчання

Переваги та недоліки комп'ютерного навчання.

Якість комп'ютерного навчання. Проблема автоматичної активізації процесів запам'ятовування.

Інформаційна безпека процедур масового комп'ютерного тестування

Десять основних вимог інформаційної безпеки тестувань і забезпечення їх виконання

Шляхи забезпечення інформаційної безпеки масових комп'ютерних тестувань. Канали просочування конфіденційної інформації. Забезпечення секретності БТЗ.

Стимули збереження конфіденційності тестових матеріалів в США та Росії.

Забезпечення конфіденційності тесту в масштабі регіону і країни. Десять кроків безпечної технології тестування. Десять кроків безпечної технології обробки результатів і публікації. Десять напрямів зниження загальної собівартості і собівартості засобів безпеки. Десять відомих в практиці тестування способів фальсифікації

Змістовий модуль II

Комп'ютерні технології у тестуванні

Сервіси Web 2.0 та їх використання в освіті

Загальна характеристика та класифікація сервісів Web 2.0, їх використання в освіті та тестуванні.

Варіанти сервісів для розгляду:

Соціальні сховища закладинок

<http://www.bobrdobr.ru>

<http://Del.icio.us>

Соціальні сховище мультимедіа даних

<http://picasaweb.google.com>

<http://www.panoramio.com>

<http://www.teachertube.com>

<http://www.youtube.com/education>

<http://audacity.sourceforge.net>

Мережеві щоденники (блоги)

<http://Livejournal.ru>

<http://www.blogger.com>

<http://Liveinternet.ru>

Соціальні геосервіси

<http://maps.google.com>

<http://maps.yandex.ru>

<http://wikimapia.org>

Wiki-сервіси

<http://www.wikipedia.org>

<http://www.wiktionary.org>

<http://www.wikibooks.org>

<http://www.wikiversity.org>

<http://commons.wikimedia.org>

<http://www.eduwiki.uran.net.ua>

Соціальні сервіси для спільної роботи з документами

<http://www.slideshare.net>

<https://docs.google.com>

<https://www.zoho.com/docs>

Карти знань

<http://bubbl.us>

<http://www.graphviz.org>

Огляд онлайн сервісів тестування.

Онлайн сертифікація.

Вільні онлайн системи створення тестових завдань та проведення тестування.

Варіанти програмних продуктів для розгляду:

Quizful - <http://quizful.net>,

Мастер-тест - <http://master-test.net>,

Специалист - <http://specialist.ru>,

Classmarker - <http://www.classmarker.com>,

ProProfs - <http://www.proprofs.com>,

Огляд вільнорозповсюджуваних систем комп'ютерного тестування.

Локальні та мережеві системи, їх складові та принципи функціонування.

Локальні, мережеві та веб-орієнтовані системи тестування. Порівняння вільно розповсюджуваних та пропрієтарних систем. Варіанти програмних продуктів для розгляду:

Hot Potatoes - <http://hotpot.uvic.ca/>,

Айрен - <http://www.irenproject.ru/>,

qedoc - <http://www.qedoc.org>,

iTest - <http://itest.sourceforge.net/documentation/index.shtml>,

NetTest - <http://kpolyakov.narod.ru/prog/nettest.htm>,

MyTest - <http://mytest.klyaksa.net/>,

ADTester - <http://www.adtester.org/>,

OpenTEST - <http://opentest.com.ua/>

TCEXAM - <http://www.tcexam.org/>,

Concerto: R-Based Online Adaptive Testing Platform - <http://www.psychometrics.cam.ac.uk/page/300/concerto-testing-platform.htm>

x-TLS - <http://xtls.org.ua/>,

LimeSurvey - <http://www.limesurvey.org/>.

Огляд пропрієтарних систем комп'ютерного тестування.

Основні складові систем, монолітна та модульна організація систем комп'ютерного тестування, їх використання в локальній та глобальній мережі. Порівняльний аналіз вибраних систем. Захист та надійність збереження даних тестування.

Варіанти програмних продуктів для розгляду:

SunRav TestOfficePro - www.sunrav.ru/testofficepro.html,

TestBox Pro - <http://www.testbox.softvea.ru/products/testbox.html>,

FastTEST Web - <http://www.fasttestweb.com/>,

та інші.

Використання системи підтримки освіти Moodle для організації комп'ютерного тестування.

Структура та склад Moodle. Організація банку тестових питань, типи питань (базові та додаткові), види тестів. Підготовка завдань для комп'ютерного тестування.

Змістовий модуль III

Обробка результатів тестування

Засоби оцінювання та аналіз результатів тестування в середовищі Moodle.

Журнал в LMS Moodle. Базові засоби аналізу тесту. Середньоквадратичне відхилення, індекс дискримінації, коефіцієнт дискримінації. Експорт результатів тестування.

Обробка результатів «паперових» тестів, та засоби підготовки таких тестів до автоматичної обробки.

ABBYU TestReader, ABBYY FlexiCapture. Їх склад, призначення, функціонал та використання.

Комп'ютерне адаптивне тестування.

Коротка історія адаптивного тестування. Комп'ютерне адаптивне тестування: психометрична теорія і комп'ютерні алгоритми. Побудова банку питань. Представлення питань тестування та тестування досвіду тестованих. Звіт про результати. Переваги комп'ютерного адаптивного тестування.

При викладанні курсу, особливу увагу слід звертати на практичну складову, а саме на оволодіння практичними навичками роботи в різноманітних програмних середовищах для організацій та проведення тестування. Цим питанням присвячений другий модуль курсу, на який передбачено найбільшу кількість годин. Нижче зупинимося, більш докладно, на найбільш важливих, на нашу думку, темах навчальної програми.

Онлайн сервіси тестування. Сервіс Мастер-тест

Незважаючи на велике розмаїття сервісів для онлайн тестування, якісних продуктів досить мало.

В якості творчого завдання, при вивченні теми, може бути порівняльний аналіз наявних онлайн сервісів. Особливу увагу слід приділити сервісу Мастер-тест і включити його, як обов'язковий елемент порівняння. Незважаючи на те, що протягом останнього

року змін в сервісі не вносилося і ніяких модифікацій не відбувалося, сервіс працює в повному обсязі і його закриття не передбачається.

Мастер-Тест - це безкоштовний Інтернет сервіс, який дозволяє створювати тести. Ви можете створювати як онлайн тести так і скачати і проходити тест без підключення до інтернету. І для цього не потрібно встановлювати на комп'ютер додаткові програми.

Мастер-Тест - освітній сервіс. На сторінках якого немає інформації, яка буде відволікати від проходження тесту. Основна ідея сервісу - проводити інтерактивне тестування знань студентів та учнів.

Сервіс орієнтується, в першу чергу, на потреби освітян.

Основні можливості сервісу Майстер-Тест

1) Створення онлайн тестів

- з великим набором видів питань,
- з необмеженою кількістю питань і варіантів відповідей в онлайн тесті,
- із застосуванням сучасних інтернет-технологій, які роблять процес створення онлайн тестів простішим і швидшим,
- з можливістю додавання в онлайн тести зображень, музичних файлів, відео, і т.д.,
- з можливістю вказівки джерел інформації, що б студент міг за результатами онлайн тестування визначити, що йому варто вивчити або який навчальний матеріал повторити,
- з можливістю вказівки різної кількості балів (від 1 до 10) для кожного питання в онлайн тесті,
- з різними варіантами оцінювання результату онлайн тестування.

2) Тестування студентів з наступними можливостями:

- перегляд докладного результату здачі тесту,
- зазначення часу на здачу тесту,
- зазначення дати, після якої студенту буде надана можливість здати онлайн тест,
- зазначення дати, до якої можна буде здати онлайн тест.

3) Публікація онлайн тестів в каталозі на сайті Майстер-Тест

Програмні засоби для локального тестування. Система Hot Potatoes

Одним з найпростіших програмних комплексів для створення тестових завдань є Hot Potatoes, що і зумовило його вибір, як першого програмного продукту, що розглядається в межах курсу. Повноцінної системою тестування вона не може вважатися, так як не містить ніяких засобів збору та аналізу результатів виконання тестових завдань.

Hot Potatoes – інструментальна програма-оболонка, що надає викладачам можливість самостійно створювати інтерактивні завдання, тести для контролю і самоконтролю учнів без знання мов програмування і без залучення фахівців в області програмування.

За допомогою програми можна створити 10 типів тестових завдань з різних дисциплін з використанням текстової, графічної, аудіо- та відеоінформації.

Особливістю цієї програми є те, що створені завдання зберігаються в стандартному форматі веб-сторінки: для їх використання тестованим необхідний лише веб-браузер.

Тестованим не потрібна програма Hot Potatoes: вона потрібна тільки викладачам для створення і редагування тестових завдань.

Програма широко використовується у всьому світі для створення завдань при вивченні будь-яких дисциплін.

Програма розроблялася в 1997-2003 р.р. Центром інформаційних технологій в гуманітарній освіті Університету Вікторії, Канада (<http://web.uvic.ca/hcmc>).

Автори програми: Stewart Arneil, Martin Holmes, Hilary Street. Martin Holmes автор останньої, 6-ї версії програми, випущеної в жовтні 2003 року.

Скачати Hot Potatoes можна з офіційного сайту <http://hotpot.uvic.ca>.

До складу Hot Potatoes входять 5 блоків для складання завдань і тестів різних видів. Кожен блок може бути використаний

як самостійна програма (відповідні ярлики розташовані у папці з програмою).

Назви блоків та типи завдань		Піктограма	Розширення імені файлу
JQuiz	Вікторина – завдання з множинним вибором відповіді (4 типи завдань)		*.jqz
JCloze	Заповнення пропусків		*.jcl
JMatch	Встановлення відповідності (3 типи завдань)		*.jcw
JCross	Кросворд		*.jmx
JMix	Відновлення послідовності		*.jmt
The Masher	Інструменти		*.jms

Всі вправи 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.

Програма містить також додатковий блок The Masher (Інструменти), що дозволяє об'єднувати створені завдання та інші навчальні матеріали до тематичних блоків, уроків та навчальних курсів.

Незалежно від того, яка програма використовується, створення завдання або тесту складається з наступних етапів:

- введення текстових даних, питань і відповідей;
- настройка конфігурації створюваного завдання або тесту;
- збереження проекту для подальшої зміни;
- збереження завдання у форматі HTML-сторінки.

Створені завдання і тести можна вивести на друк, а також об'єднати в блоки і уроки.

Інтерфейс всіх п'яти програм однотипний. Основні дії можна виконувати через головне меню і кнопки панелей інструментів. Зручніше використовувати саме панелі інструментів.

Ще одна особливість даного програмного продукту – це інтеграція завдань в найбільш популярне на сьогоднішній день віртуальне освітнє середовище Moodle. Тому особливу увагу слід звернути на завдання, аналогів яких немає в Moodle. Одним з таких тестових завдань є кросворд.

Кроссворд (JCross)

Програма JCross призначена для швидкого складання кросвордів. Цей вид завдань дуже популярний в середній школі.

Для складання кросворду треба запустити програму HotPotatoes, а потім вибрати програму JCross з меню Potatoes або на екрані HotPotatoes клацнути на відповідній картоплі.

Для розміщення слів у 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 *Автоматическое создание* в панелі інструментів або вибрати пункт меню *Управление сеткой/Автоматическая сетка*.

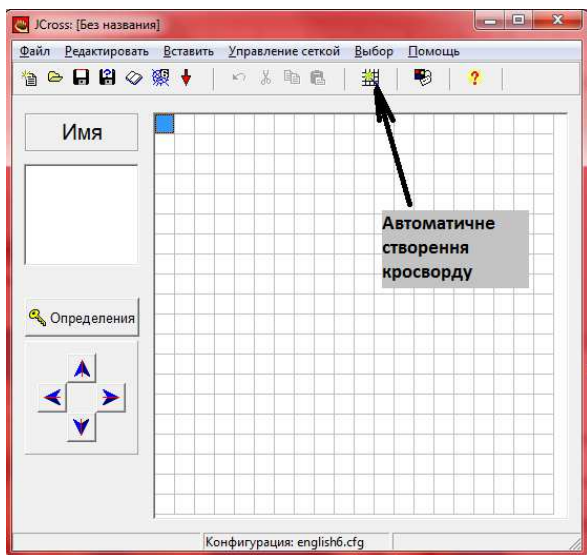


Рис. 1. Створення завдання типу JCross

При автоматичному розміщенні слів всі слова вводяться списком в поле введення слів у діалозі представленому на рис. 2, при цьому всі слова вводяться в окремі рядки, а потім програма підбирає варіанти їх розміщення. У полі **Максимальний розмір** вказується максимальна кількість клітинок у кросворді. Клацання на кнопці **Создать сетку** призведе до створення кросворду.

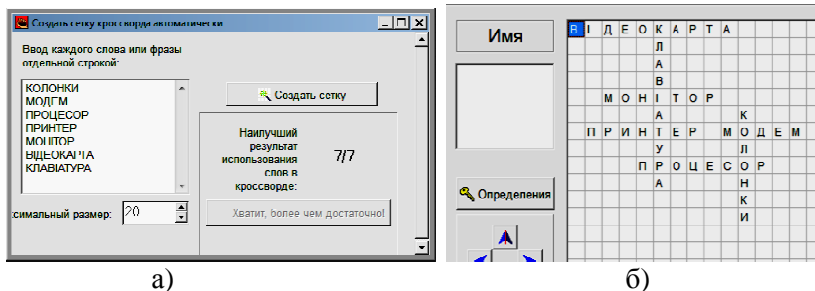


Рис. 2. Введення слів для кросворду завдання JCross

Наступний крок полягає в тому, що треба ввести визначення до слів. Ці визначення описують зазначені слова. Для введення визначень треба натиснути кнопку **Определения**. З'явиться діалог **Определение** (рис. 3).

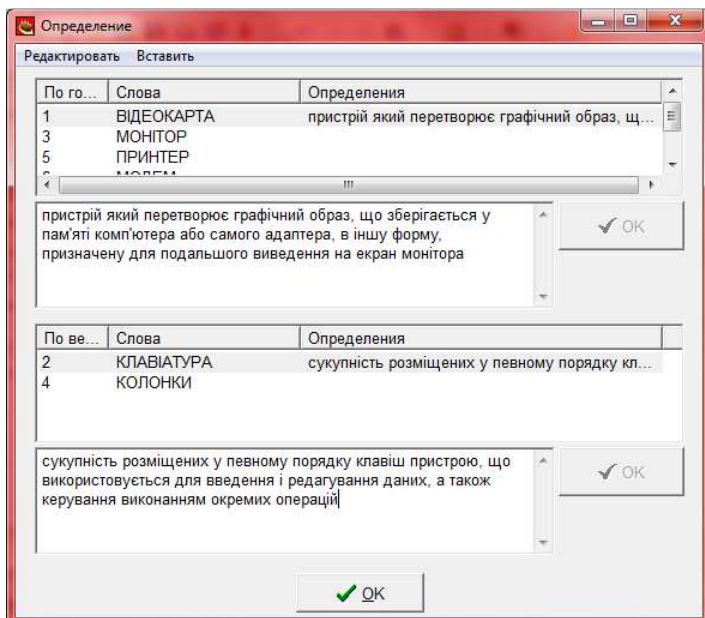


Рис. 3. Введення визначень слів кросворду

У верхній частині екрану розташовані слова, які в кросворді розміщені по горизонталі, у нижній частині – по вертикалі. Для введення визначення треба виділити потрібне слово у списку, в полі під списком ввести визначення і натиснути кнопку ОК. Послідовність введення визначень не має значення. Після введення визначень можна вносити будь-які зміни у сітку - при переміщенні слів програма буде зберігати відповідні визначення.

Визначення до кожного слова можуть з'являтися всі відразу, як в кросвордах, що публікуються в газетах і журналах, або ж по

одному визначенню. Для вибору цього режиму відображення використовується прапорець Показувати всі визначення внизу кросворду. А потім треба зберегти проект і опублікувати кросворд у вигляді html-сторінки.

Мережеве програмне забезпечення для проведення тестування

Найбільший клас програмних продуктів для тестування – мережеві програмні комплекси. Такі програмні комплекси, як правило побудовані з використанням клієнт-серверної архітектури.

iTest

iTest – класичний приклад клієнт-серверної архітектури. Клієнтська частина містить досить значний функціонал, що полягає не тільки в простому демонструванні завдань та записі відповідей, а й функції вибору завдань реалізовані в клієнтській частині. Хоча правила вибору визначаються на сервері. Основний недолік – невеликий набір типів тестових завдань.

ADTester

Програми пакета:

Конструктор тестів - програма для створення тестів. Дозволяє створювати і редагувати тести будь-якої складності.

Тестер - програма для проведення тестування. Має максимально простий інтерфейс і безліч параметрів проведення тестування, дозволяють підлаштувати її саме під ваші цілі.

Адмін панель - програма адміністрування. Дозволяє робити маніпуляції з користувачами та групами користувачів, будувати матриці правильності та проводити аналіз результатів тестування.

Можливості пакета:

1. Необмежена кількість питань у тесті та кол-во варіантів відповіді.

2. П'ять типів питань:

- одиночний вибір;
- множинний вибір;
- введення відповіді з клавіатури;
- відповідність;

- порядок.

3. Можливо встановити режим тестування: контрольний і навчаючий. У режимі навчання при неправильному відповіді можна подивитися підказку з даного питання, результати тестування не враховуються. У режимі контролю ведеться статистика проходження тесту.

4. Кожне питання має свій "вагу" (ціну в балах).

5. Кожне питання може супроводжуватися підказкою для тестованого

6. Точне відтворення зовнішнього вигляду тестів (розмір шрифтів, їх колір, назва, стиль ...).

7. Питання можуть супроводжуватися: зображеннями, html і RTF документами, таблицями, діаграмами і різними OLE документами.

8. Контроль відображення статистики користувачеві (кол-во питань у тесті, номер поточного питання, індикатори правильності відповіді на кожне питання, що залишилося час).

9. Можна як дозволити, так і заборонити показ результатів тестування для користувача.

10. Редагування тесту може бути захищене паролем (зашифрованим методом стійкою криптографії).

11. Тестування може бути обмежене за часом.

Розвиток програмного пакету на даний час припинено, останні роки оновлень не має, і адмін панель та конструктор тестів на сайті розробника не доступні, хоча і є можливість скачати комплекс на багатьох інших сайтах.

MyTest

Найбільшу зацікавленість у студентів прививченні дисципліни викликає система MyTestX. Дана система поєднує в собі простоту використання та досить значний функціонал.

MyTest це набір програмних засобів – програма тестування, редактор тестів і журнал результатів – для створення тестових завдань, проведення комп'ютерного тестування, збору та аналізу результатів, виставлення оцінок за вказаною в тесті шкалою. Програма легка і зручна у використанні.

Програма MyTest X працює з дев'ятьма типами завдань: одиничний вибір, множинний вибір, встановлення порядку проходження, встановлення відповідності, вказівка істинності чи хибності тверджень, ручне введення числа, ручне введення тексту, вибір місця на зображенні, перестановка букв. У тесті можна використовувати будь-яку кількість будь-яких типів завдань. У завданнях з вибором відповіді (одиничний, множинний вибір, вказівка порядку, вказівка істинності) можна використовувати до 10 (включно) варіантів відповіді.

Модуль тестування (MyTestStudent) є "плеєром тестів". Він дозволяє відкрити або отримати по мережі файл з тестом і пройти тестування. Хід тестування, сигналізація про помилки, спосіб виведення результату тестування залежать від параметрів тесту, заданих в редакторі.

Для створення тестів призначений дуже зручний редактор тестів (MyTestEditor) з дружнім інтерфейсом. За допомогою редактора можна створити новий тест або змінити існуючий. В редакторі настроюється процес тестування: порядок завдань і варіантів, обмеження часу, шкала оцінювання та багато іншого.

Журнал тестування (MyTestServer) дозволяє організувати тестування більш зручним чином. За допомогою нього можна роздавати файли з тестами по мережі, отримувати результати з усіх комп'ютерів тестованих і аналізувати їх у зручному вигляді.

Також є додатковий модуль MyTestBuilder для створення автономних тестів у форматі exe (встановлюється окремо).

За допомогою програм MyTest можна організувати як локальне так і мережеве тестування.

Більш детальну інформацію про програму, та останню версію програми можна знайти за адресою <http://mytest.klyaksa.net> – розділ інформаційно-освітнього порталу Клякс@.net, присвячений цій програмі.

Всі права на програму MyTest належать її автору: Башлакову Олександрові Сергійовичу, м.Унеча, Брянської обл.

Програма MyTest поширюється безкоштовно (Freeware). Будь-який освітній заклад, вчитель і учень можуть безкоштовно використовувати програму на основі ліцензійної угоди без будь-яких грошових відрахувань. Для отримання більш докладної інформації про права використання та поширення програми дивіться ліцензійну угоду.



Рис. 4. Вікно «О программе» MyTest X

Освоєння програмного продукту не викликає у студентів труднощів незалежно від рівня їх комп'ютерної грамотності. Створення тестових завдань просте та інтуїтивно зрозуміле. Зупинемося безпосередньо тільки на створенні тесту та налаштуванні його параметрів.

Створення та налаштування тестів в MyTest

Тест – це система тестових завдань, параметрів і правил їх застосування.

Завдання, параметри тестування, звуки та зображення до завдань для кожного окремого тесту – все зберігається в одному файлі тесту. Ніяких баз даних, ніяких зайвих файлів – один тест – один файл. Файл з тестом зашифрований і стиснутий.

Команда меню **Інформація про тест...** містить зведені дані про тест.

Команда меню **Імпорт** дозволяє імпортувати завдання з іншого тесту.

Команда меню **Експорт** призначена для підготовки тесту до друку на папері (рис. 5).

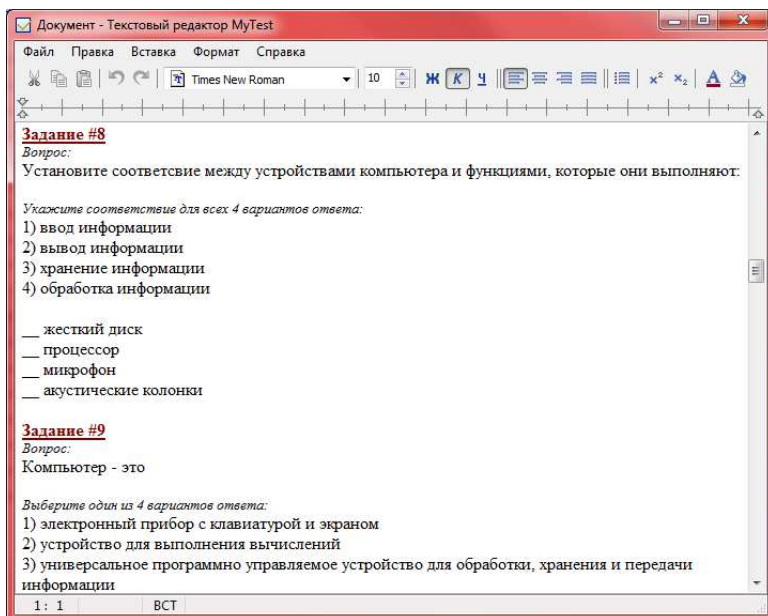


Рис. 5. Приклад експорту (друку) завдань в MyTest

Параметри тесту і правила застосування

Кожен тест може містити:

1. Заголовок та опис тесту, які містять: заголовок тесту, автора (упорядника) тесту, електронну пошту автора та опис тесту.
2. Інструкцію тестованому – повідомлення, яке буде показано тестованому на початку тесту. Пізніше це повідомлення можна подивитися по команді Параметри тесту / Інструкція тестованому. Інструкція може містити форматований текст, малюнки і формули.
3. Замітки – це текст, який не видно тестованому. Читати його або змінити можна тільки відкривши тест в редакторі. У замітках можна зберігати будь-яку інформацію. Наприклад, що і в якому завданні хочеться змінити, якісь зауваження по тесту і т.д. Замітки можуть містити форматований текст, малюнки і формули.

4. Оцінювання. Оцінка учня обчислюється за балами. Рівень оцінки у відсотках (для того щоб не міняти при видаленні/додаванні завдань) задається в редакторі тестів. Система оцінювання може бути будь-якою – від двобальної (залік/незалік) до 100-бальної. Для оцінки можна вказати назву – тоді виводиться не цифра, а назва (корисно для залік/незалік). У редакторі для зручності є шаблони оцінювання, Ви можете використовувати їх, а можете задати свої критерії оцінювання.

У редакторі Ви можете вказати зараховувати тільки 100% вірні відповіді або враховувати неповні відповіді (наприклад, в завданнях на множинний вибір відзначені не всі або навпаки зайві варіанти).

5. Порядок питань: звичайний або випадковий. Визначає, в якому порядку тестованому буде запропоновано завдання. Для випадкового порядку перемішування завдань відбувається при кожному запуску тесту.

6. Порядок варіантів: звичайний або випадковий. Визначає в якому порядку тестованому буде запропоновано варіанти відповідей. Якщо включений випадковий порядок, то у різних учнів буде різний порядок варіантів відповіді.

7. Формулювання: випадкове, основне, друге і т.д. Завдання тесту може містити кілька формулювань питання (від 1 до 5). Якщо формулювання не вказано, то виводиться основне.

8. Теми (групи завдань)

Теми, або групи, завдань призначені для зручного поділу завдань на групи і обмеження кількості завдань для учня. Ви можете вказати, скільки саме завдань з якої теми буде задано. При обмеженні кількості завдань вони вибираються випадковим чином. Кількість тем у тесті не обмежена. Змінити тему завдання можна в будь-який момент (рис. 6).

9. Обмеження за часом. Ви можете обмежити не тільки час обдумування кожного завдання, але й час усього тесту. Обмеження для завдання вказується при редагуванні завдань, обмеження для всього тесту в меню *Параметри тестування*.

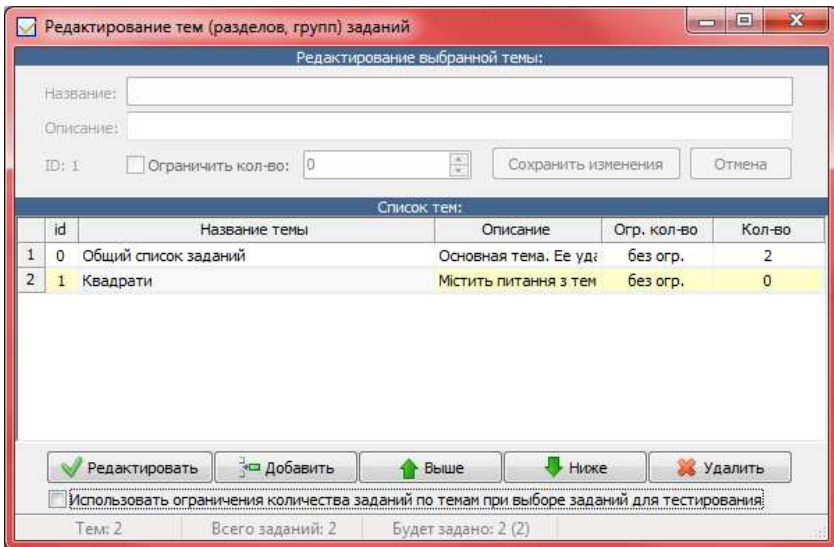


Рис. 6. Редагування тем в MyTest

10. Обмеження на кількість запусків. Ви можете вказати максимальну кількість запусків тесту до перезавантаження комп'ютера. Тобто якщо вкажіть 3, то тест можна буде запустити (почати тестування) 3 рази. Щоб ще запустити потрібно перезавантажити комп'ютер.

11. Паролі. Ваші тести надійно захищені. Для зручності Ви можете задати пароль на відкриття тесту для тестування, відкриття тесту для редагування, на запуск тестування, для збереження захищених результатів.

12. Режими тестування.

У програмі передбачено декілька незалежних один від одного режимів тестування: Навчальний - при помилці тестований отримує відповідне повідомлення і може подивитися пояснення завдання (якщо воно Вами задане). Помилкові і вірні варіанти підсвічуються різними кольорами. Штрафний – при помилці у тестованого будуть відніматися бали. У цьому режимі додатково з'являється кнопка «Я не знаю ...». К्लецання по ній присвоює за

завдання нуль балів. Цей режим працює по-різному в залежності від налаштувань оцінювання (зараховувати 100% вірні відповіді чи ні). Вільний – дозволяє переміщуватися по завданнях в будь-якому порядку. У будь-який момент можна повернутися до завдання, якщо на нього ще не дана відповідь. Додатково з'являється кнопка "Пропустити" – перехід до наступного завдання без відповіді (пізніше можна повернутися) і список завдань (у якому по-різному відзначені переглянуті і виконані завдання), що дозволяє перейти до будь-якого невиконання завдання. Монопольний – вікно програми займає весь екран і його неможливо згорнути.

13. Параметри тестування і виведення результатів Ви можете дозволити / заборонити

- показувати результат тестованому;
- зберігати результат в текстовий файл;
- зберігати результат в захищений файл;
- відправляти результати по мережі модулю Журнал;
- відправляти проміжні результати по мережі модулю Журнал для вкладки Монітор;
- відправляти результати по електронній пошті;
- показувати детальний звіт тестованому по закінченню тесту (виводиться текст всіх заданих завдань і вказівка вірна чи ні відповідь);
- встановити мінімальний відсоток завдань для виведення результату (використовується для захисту відповідей методом підбору);
- показувати / підсвічувати правильну відповідь у навчальному режимі.

Всі параметри тесту в одному вікні (рис. 7) можна змінити за допомогою команди **Параметри тесту / Панель параметрів тесту...**(F2).

Журнал тестування MyTestServer

Журнал (сервер) – модуль програми MyTest, що дозволяє централізовано приймати і обробляти результати тестування, роздавати тести за допомогою комп'ютерної мережі. Для відправки

та отримання результатів, відправки файлів з тестами використовується протокол Інтернету TCP/IP.

Переваги використання Журналу:

- тестування проходить більш спокійно. Вчителю немає необхідності переміщатися по класу для отримання результатів;
- швидка підготовка до тестування, захист тестів від несанкціонованого доступу. Учителю не потрібно копіювати файли з тестами на комп'ютери учнів, тести можна "роздати" прямо з Журналу;
- учитель отримує дуже детальну інформацію про результати тестування, а не тільки одну оцінку;
- учитель може налаштувати вигляд таблиці результатів під себе (сховати можна будь-які колонки);
- учитель може швидко проаналізувати результати (наприклад, які питання викликали найбільші труднощі) і відразу ж провести роботу над помилками. Є можливість перегляду графіку тестування і швидкого переходу в Редактор тестів до вибраного завданням;
- упорядкувати результати по будь-якому критерію;
- учитель може зберегти результати у файл. Цей файл потім можна відкрити і обробити пізніше;
- експортувати результати в текстовий файл (для подальшого відкриття в електронних таблицях) або HTML-файл;
- вчителю зручно виставляти оцінки за тест тощо.

Для використання журналу необхідно щоб:

1. Учитель запустив модуль журналу програми MyTest.
2. Учні запустили програму MyTest і почали тестування. Тести можна відкрити з свого комп'ютера, мережевої папки або отримати по мережі від вчителя.
3. Як тільки учень розпочинає тестування (вибрав тест і почав його проходити) учитель отримує повідомлення про це.
4. Як тільки учень закінчує тестування вчитель отримує детальну інформацію про його результати.

5. Отримані результати вчитель може проаналізувати і/або зберегти в файл.

Для зручності вікно Журналу може згортатися в трей (відображатися позначкою в правому нижньому куті біля годинника). При отриманні нових результатів може бути показано спливаюче віконце (у лівому нижньому куті) з короткою інформацією про результати.

Вікно журналу (рис. 7) містить кілька вкладок: Результати, Монитор, Сообщения, Лог, Раздать тест, Раздать список.

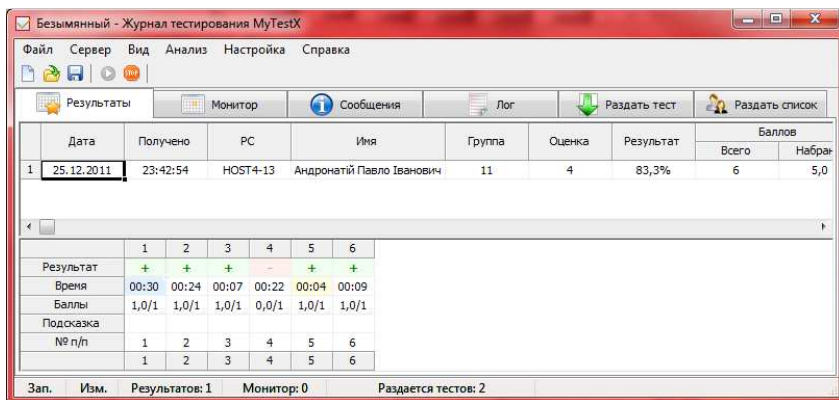


Рис. 7. Вікно журналу MyTest, вкладка Результаты

На вкладці Результати відображається детальна інформація про результати проходження тесту: дата отримання результату, час отримання результату, мережеве ім'я комп'ютера, ім'я учня (яке він вводить при запуску тесту), група (клас) – додаткове поле до імені, оцінений, результат у відсотках, балів всього, балів набрано, завдань всього, завдань задано, завдань правильно, завдань помилково, завдань пропущено, завдань підказок, завдань результативність ($100 \cdot \text{Кількість правильно} / \text{Кількість завдань з відповіддю}$), час початку тесту та завершення тесту (за часом на комп'ютері учня), час всього тесту (не залежно від часу на комп'ютері учня, захищено від змін системного часу), тема тесту, повне ім'я файлу з тестом, CRC файлу з тестом (якщо учень

підмінив або сплутав файл з тестом Ви відразу це виявите по іншим значенням), UID файлу з тестом (при збереженні тесту кожному присвоюється свій унікальний ідентифікатор (якщо учень підмінив або сплутав файл з тестом Ви відразу це виявите по іншим значенням), режим закінчення тесту та інше.

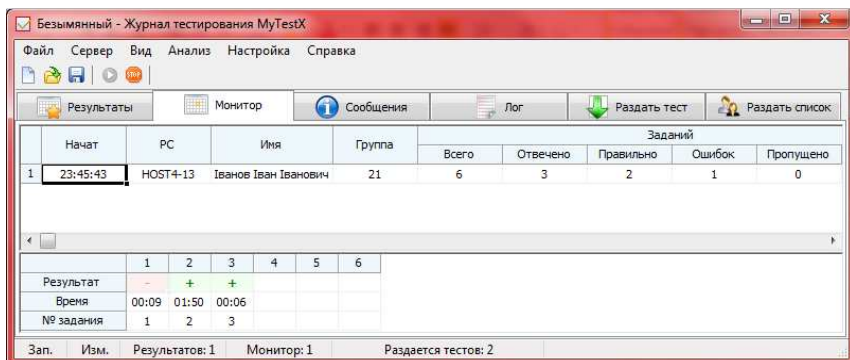


Рис. 8. Вікно журналу MyTest, вкладка Монитор

На вкладці Монитор виводиться інформація про процес тестування. На рис. 8 видно список комп'ютерів (і учні) що виконують тест (і який тест) і скільки завдань виконав учень (+ скільки з них вірно чи помилково, % результативності), скільки балів (максимум, можливо було б набрати за ці відповіді, набрано, % попереднього результату).

На вкладці Раздать тест можна вказати які файли з тестами будуть отримувати тестовані (рис. 9) за командою Файл/Получить по сети (або автоматично при відповідній настройці). Ця можливість економить час учителя – немає необхідності копіювати файл на комп'ютери учнів. Ще одна перевага – файл з тестом не зберігається на компютерах тестованих, у них немає можливості відкрити його для підглядання, тому не потрібно задавати паролі.

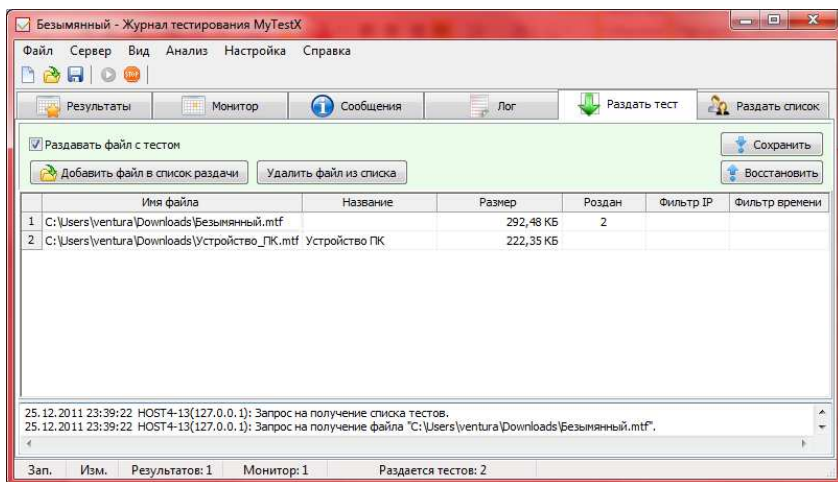


Рис. 9. Вікно журналу MyTest, вкладка Раздать тест

Якщо в список для роздачі тестів додано тільки один файл, то учень при запиті тесту (Файл/Получить по сети) відкриє цей файл. Якщо ж в список додано декілька файлів, то учень при запиті тесту побачить вікно, де пропонується вибрати який саме файл він бажає отримати. Після вибору файлу з цього списку він буде завантажений.

Якщо Журнал запущений, але не вибраний файл для роздачі, то учень при команді Файл/Получить по сети отримає повідомлення про відмову в отриманні тесту.

На вкладці Раздать список можна вказати файл зі списком користувачів (тестованих). Якщо в модулі тестування зроблені відповідні настройки, то користувач зможе отримати цей список і вибрати себе зі списку, а не вводити вручну.

Аналіз результатів

Допомогу при аналізі результатів тестування можуть надати діаграми та зведені таблиці з меню Анализ. Всі пункти меню показані на рис. 10.

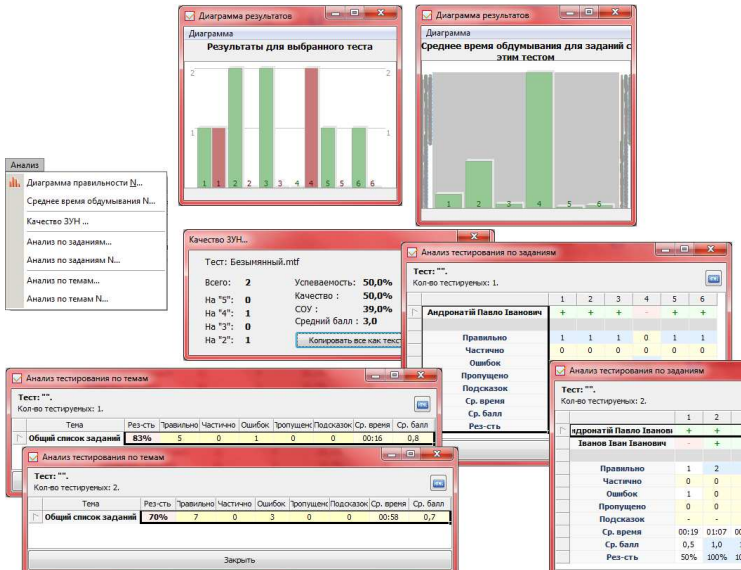


Рис. 10. Аналіз тестів в MyTest

Для аналізу результатів доступні дві діаграми – гістограми: діаграма правильності та середнього часу обдумування. На діаграмі правильності кожному завдання відповідають дві колонки (зелена – правильно, червона – помилково). На діаграмі середнього часу обмірковування – кожному завдання відповідає один стовпчик.

Якщо запустити Редактор тестів MyTestX і в ньому відкрити тест для якого побудована діаграма, то клацання по стовпцю діаграми з одночасним натисканням клавіші Alt, переведе Редактор на передній план і в ньому буде обрано відповідне завдання. Ця можливість дозволяється зручно проаналізувати помилки і розібрати завдання з учнями.

Аналіз тестування за темами дозволяє дізнатися результативність по кожній темі (групі завдань) тесту. Кожен рядок таблиці відповідає окремій темі в тесті. Таким чином можна з'ясувати, наприклад, завдання яких тем викликають найбільші

труднощі в учнів і скорегувати процес навчання. Природно, даний аналіз має сенс, якщо Ви використовуєте розподіл завдань за темами.

Обидва зазначених вище види аналізу можна проводити як для одного окремого учня, так і для всіх учнів, що проходили той же тест.

Веб орієнтовані системи тестування

Завдяки бурхливому розвитку мережі Інтернет, найбільш перспективний напрямком у сучасних систем тестування є веб-орієнтовані системи, тому розгляду таких систем слід приділяти значну увагу.

ТАО

ТАО® - <http://www.iao.lu/html/> - платформа з відкритим вихідним кодом, що вільно розповсюджується. Вона дозволяє адмініструвати, створювати та проводити мультимедійні тести.

Система підтримує широке коло типів тестів: стандартні, психометричні, адаптивні тести

ТАО використовується у відомому широкомасштабному дослідженні PISA (Programme for International Student Assessment). ТАО створена і підтримується університетом Люксембурга.

OpenTEST 2.0

В Харківському національному університеті радіоелектроніки розроблена і на протязі декількох років в ряді навчальних закладів України знаходиться в експлуатації система комп'ютерного тестування знань студентів OpenTEST.

Основною особливістю системи OpenTEST 2.0 є її спрямованість на забезпечення тестувань учнів з максимально суворою звітністю. Областю застосування можуть бути різноманітні підсумкові тестування, заліки, іспити, кваліфікаційні тести і будь-які інші види контролю знань учнів в яких головну роль відіграє максимально об'єктивна оцінка знань.

Також особливістю комп'ютерної системи тестування OpenTEST 2.0 є низькі вимоги до апаратних і програмних ресурсів,

що дозволяє проводити тестування навіть у погано оснащених комп'ютерних класах.

В системі OpenTEST 2.0 приділено багато уваги проблемам безпеки при проведенні тестувань. Так як при контрольних тестуваннях кінцева оцінка відіграє величезну роль, об'єктивність її виставлення повинна бути максимальна, а також всі можливі варіанти фальсифікації результатів повинні бути виключені.

- Захист від злому паролів до всіх модулів системи методом перебору (тимчасове блокування акаунта, блокування IP адреси).
- Унікальний алгоритм динамічної зміни хеш-ідентифікатора клієнта, який не дозволяє одночасну роботу під одним акаунтом з декількох комп'ютерів і запобігає будь-яким несанкціонованим спробам використання даних на проміжних ланках мережі.
- Використання захищеного протоколу передачі даних SSL.
- Ведення докладного структурованого логу подій.
- Інтерфейс для проведення гнучкого аналізу логу подій за вказаними критеріями.
- Щоденний автоматичний аналіз логів із формуванням звіту з безпеки.
- Шифрування паролів на стороні клієнта засобами JavaScript не дозволяють перехопити чийсь пароль при роботі з системою OpenTEST 2.0 навіть при відмові від використання протоколу SSL.

Модульна структура OpenTEST 2.0 забезпечує легкість розширення функціональності без необхідності внесення змін в існуючі модулі.

Основні модулі OpenTEST 2.0:

- модуль «Тестування»;
- модуль «Управління користувачами»;
- модуль «Управління тестами»;
- модуль «Управління тестуванням»;
- модуль «Результати тестування»;
- модуль «Адміністрування».

HTML-код всіх модулів системи OpenTEST 2.0 оптимізований для коректної роботи в усіх сучасних браузерах (IE, Mozilla, Opera, Safari). Система OpenTEST 2.0 є повністю мультимовною, всі дані передаються в браузер в універсальній кодуванні UTF-8, що дозволяє підтримувати навіть такі мови як китайська, арабська, відображати різноманітні спеціальні символи і т.п.

При проектуванні модулів системи OpenTEST 2.0 основну увагу було приділено високій ергономічності системи а також орієнтації на роботу з великим потоком користувачів. Система дозволяє проводити тестування одночасно більше 1000 користувачів. Продуктивність системи забезпечується високо оптимізованим кодом і всесвітньо визнаними засобами побудови web-додатків. Система OpenTEST 2.0 функціонує на основі зв'язки web-сервер Apache + PHP + MySQL. Це найбільш поширені технології, на яких побудовано більше половини всіх провідних ресурсів в мережі Internet. Всі ці технології є платформенно незалежними, що дозволяє інстальювати систему OpenTEST 2.0 практично на будь-який комп'ютер.

Обмеження у використанні системи лише для контролюючих функцій та занадто складна система безпеки, значно ускладнюють засвоєння системи студентами.

TSExam

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 всесторонній моніторинг системи та усунення її недоліків, таким чином з часом досягається значно більша стабільність та надійність програмног 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, 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, перегляду списку онлайн користувачів та перегляду статистики роботи користувача (рис. 11).

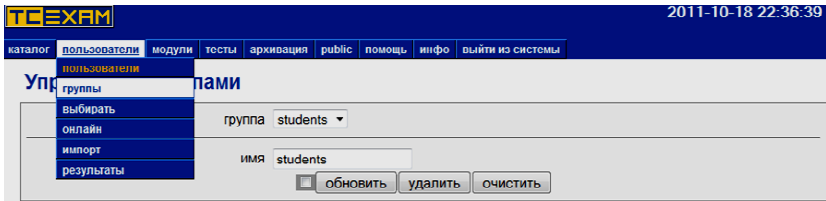


Рис. 11. Розділ Користувачі TSEХam

Створення груп передбачає можливість надання права проходження тесту визначеній групі і у визначений час.

У системі можлива робота тільки зареєстрованих користувачів (рис. 12). Також передбачено розподіл прав користувачів на різні рівні, що дає можливість надавати різним користувачам/групам доступу до різних компонентів системи.

TSEХam використовує потужну систему захисту, що базується як на куки так і на перевірці IP адреси для запобігання несанкціонованого доступу.

У системі TSEХam 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 перегляду 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.

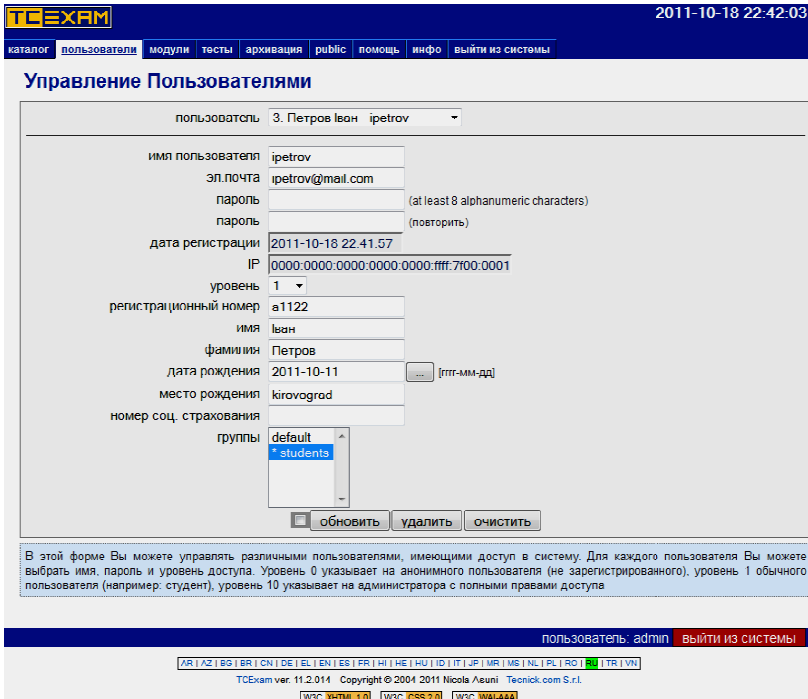


Рис. 12. Управління параметрами користувача TCEXAM Модулі

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 питань з відкритою формою, 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.

Системи комп'ютерного адаптивного тестування.

Не зважаючи на досить суттєве обмеження в годинах на вивчення даного курсу, необхідно приділити увагу системам комп'ютерного адаптивного тестування.

На сьогоднішній день існують вільнорозповсюджені системи тестування в яких в більшій чи меншій мірі реалізовано адаптивність.

В першу чергу слід ознайомитися з системою Психометричного центру університету Кембріджа – Concerto.

Concerto

Concerto – веб-орієнтована платформа адаптивного тестування для створення та проведення мультимедійних динамічних тестів.

Вона поєднує в собі гнучкість HTML презентації з обчислювальною потужністю мови R, а також безпеку і продуктивність бази даних MySQL. Це абсолютно безкоштовна система для комерційних і освітніх цілей з відкритим вихідним кодом.

Система проста у використанні, гнучка, модульна, побудована на основі потужної мови R.

R — мова програмування і програмне середовище для статистичних обчислень, аналізу та представлення даних в графічному вигляді. Розробка R відбувалась під істотним впливом двох існуючих мов програмування: в основному мови програмування S з семантикою успадкованою від Scheme. Мова програмування R названа в честь перших літер прізвищ її засновників Роса Іхаки (Ross Ihaka) та Роберта Джентлмена (Robert Gentleman) працівників Оуклендського Університету, в Новій Зеландії. Незважаючи на деякі принципові відмінності, більшість програм, написаних на мові програмування S запусकाються в середовищі R.

R розповсюджується безкоштовно за ліцензією GNU General Public License у вигляді вільно доступного вихідного коду або відкомпільованих бінарних версій більшості операційних систем: Linux, FreeBSD, Microsoft Windows, Mac OS X, Solaris. R використовує текстовий інтерфейс, однак існують різного роду графічні інтерфейси користувача (див. Графічні Редактори Скриптів та IDE).

R містить широкі можливості для проведення статистичних аналізів, включаючи лінійну і нелінійну регресію, класичні статистичні тести, аналіз часових рядів (серій), кластерний аналіз і багато іншого. R легко розбудовується завдяки використанню додаткових функцій і пакетів доступних на сайті Comprehensive R Archive Network (CRAN). Більша частина стандартних функцій R, написана на мові R, однак існує можливість підключати код написаний на C, C++, або Fortran. Також за допомогою програмного коду на C або Java можна безпосередньо маніпулювати R об'єктами.

Існує досить докладна інструкція по створення тестових завдань в системі Concert - <http://code.google.com/p/concerto-platform/wiki/Concerto3SimpleTest>

OSCATS

OSCATS бібліотека, що використовує IRT, модель когнітивної діагностики та адаптивні алгоритми вибору. Бібліотека написана використовуючи об'єктно-орієнтовану мову C з

використанням GObject, та перенесена на мови Python, Perl, PHP, та Java.

OSCATS – вільнорозповсюджувана система з відкритим вихідним кодом. Розповсюджується у вигляді готового програмного комплексу для ОС Windows та у вигляді вихідного коду мовою C.

Повний перелік можливостей та підтримуваних алгоритмів можна прочитати за адресою:

<http://code.google.com/p/oscats/wiki/Features>.

ЛАБОРАТОРНІ РОБОТИ З КУРСУ «ВИБІРКОВІ ОБСТЕЖЕННЯ У ПСИХОЛОГІЇ, СОЦІОЛОГІЇ ТА ПЕДАГОГІЦІ»

Лупан І.В.

Якісне проведення освітніх вимірювань, як і будь-яке статистичне дослідження, неможливе без правильної побудови досліджуваної вибірки. Саме тому до навчального плану спеціальності «Освітні вимірювання» включено дисципліну «Вибіркові обстеження у педагогіці, соціології та психології». Теоретичний матеріал для даного курсу, основний зміст та тренувальні вправи визначено у посібниках [1, 2, 3] викладачів Київського національного університету імені Тараса Шевченка, які брали участь у проєкті Tempus-Tasis «Статистичні аспекти в економіці», та у перекладених російською мовою посібниках зарубіжних авторів [4, 5]. Однак вказані посібники призначені у першу чергу для математиків і потребують серйозної попередньої підготовки. Крім того у роботах різних авторів використовується різна символіка для позначення основних понять, тому її опанування потребує додаткових часових затрат. Необхідно створити посібник, який би відповідно до теоретичних засад

вибіркових обстежень забезпечував майбутніх дослідників у галузі освітніх вимірювань корисними алгоритмами та інструментами вибіркового відбору і оцінювання параметрів досліджуваної популяції за отриманими вибірковими. У такому сенсі важливою складовою посібника має стати практична частина, яка б містила задачі та завдання для лабораторних робіт.

Основним змістом лабораторних робіт має стати дослідження студентами ефективності різних вибірових планів на матеріалі реальних суцільних обстежень. Такі роботи дозволяють порівняти параметри сформованих пізніше вибірок з параметрами генеральної сукупності та оцінити ефективність обраної схеми побудови вибірки, а також з'ясувати її переваги та недоліки, порівняно з іншими, а також сформулювати у студентів уявлення про те, що вибіркоче обстеження дозволяє лише оцінити (з більшою чи меншою точністю) істинні значення досліджуваних параметрів.

Розглянемо для прикладу можливі результати дослідження простої випадкової та систематичної вибірки.

Просту випадкову вибірку зручно будувати за допомогою інструменту «Вибірка» з пакета аналізу MS Excel. Щоправда, таким чином будується вибірка з повтореннями, однак при досить великому обсязі генеральної сукупності та порівняно невеликому обсязі досліджуваної вибірки імовірність вибору повторних значень надзвичайно мала. Метою лабораторної роботи у даному випадку є дослідження кількості простих випадкових вибірок, які дозволяють «правильно» оцінити генеральні параметри, зокрема, генеральне середнє та сумарне значення. Як приклад генеральної сукупності використаємо множину міст України, а як досліджувану ознаку – відомості про населеність міст. За даними, наведеними у [6], середня населеність 454 міст України та загальна кількість міського населення становлять відповідно 62662 та 28448620 осіб.

Зазвичай для оцінки генеральних параметрів будують одну-дві випадкові вибірки. Однак для оцінки «правильності» слід побудувати їх якомога більше, наприклад, 20-50. У такому разі дійсно вдасться співставити отримані результати з обраним рівнем значущості α (при достатньо великій кількості вибірок α визначає

долю вибірок, за якими досліджувані параметри будуть обчислені неправильно) [7]. Обсяг побудованих вибірок може бути довільний, але однаковий для всіх.

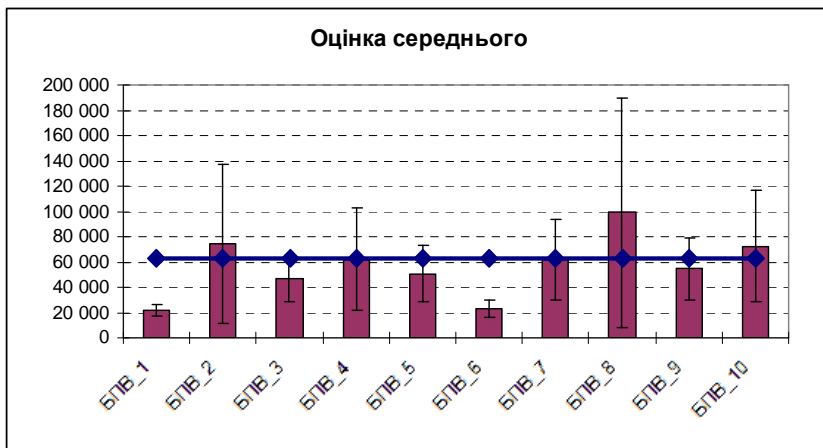


Рис. 1.

На рис. 1 горизонтальною лінією зображено істинне середнє значення, стовпцями – середні, обчислені за даними десяти вибірок по 20 одиниць у кожній. «Вусики» на стовпцях показують межі довірчих інтервалів. Як видно на рисунку, у двох випадках з десяти справжнє середнє не потрапляє до довірчого інтервалу, що відповідає прийнятому рівню значущості ($\alpha=0,2$). Результати оцінювання сумарного значення будуть аналогічні.

Наступне завдання лабораторної роботи полягає в обчисленні обсягу, необхідного для задоволення допустимої похибки обчислень. Як правило за першою вибіркою, якщо доведено її репрезентативність, обчислюють параметри вибірки та оцінюють можливу похибку обчислень. Якщо похибка перевищує допустимі межі, то необхідно буде збільшити обсяг досліджуваної вибірки, причому для зменшення похибки вдвічі, доведеться обсяг вибірки збільшити вчетверо. Цікавим результатом даної лабораторної роботи є те, що найменша похибка може виявитися у

вибірках, які не забезпечують задовільної оцінки генеральних параметрів. І, навпаки, відповідно до значення похибки у вибірках, параметри яких найменше відрізняються від генеральних, їхній обсяг доведеться значно збільшити (у наведеному прикладі принаймні до 100).

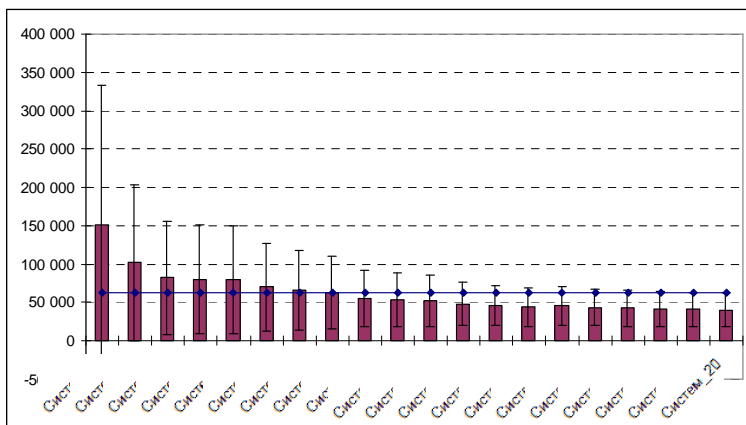


Рис. 2.

Наступна лабораторна робота присвячена дослідженню систематичних вибірок. На рис. 2, як і на попередньому, горизонтальною лінією зображено істинне середнє генеральної сукупності, а стовпцями з «вусиками» – вибіркові середні з довірчими інтервалами. У даному випадку вибірки порядку $k=20$ будувалися за сукупністю, впорядкованою за спаданням. Тенденція до спадання зберігається також і для обчислених середніх, однак у даному прикладі усі довірчі інтервали містять істинне значення.

При порівнянні ж середніх значень, отриманих за усіма побудованими вибірками можна отримати відносні похибки обчислення досліджуваних параметрів. Наприклад, для середніх отримали наступне:

Метод відбору	Спосіб оцінки	Значення	Відносна похибка
Випадкова	середнє	56618,14	9,65%
Систематична1	середнє	62399,70	0,42%

Звичайно, отримані результати не є абсолютними та універсальними, однак проведені лабораторні дослідження допоможуть майбутнім фахівцям здобути досвід формування репрезентативних вибірок та критично оцінювати результати, отримані у вибіркових обстеженнях. За сукупністю, наведеною у прикладі, усі міста можна також поділити на страти за кількістю населення та кластери за віднесеністю до певної області.

Для закріплення отриманих навичок корисним буде порівняння результатів самостійно проведеного суцільного обстеження та добір і обґрунтування найкращого плану вибіркового обстеження. Ми пропонуємо провести анкетування студентів факультету за анкетною, розробленою самостійно таким чином, щоб відповіді містили як числові дані, так і порядкові та номінативні, щоб можна було також оцінювати частки.

Ще одним з напрямків лабораторних робіт може бути дослідження ефективності та зміщення оцінок, отриманих у тих чи інших вибіркових дизайнах. Наприклад, можна дослідити, при якому обсязі вибірка буде більш репрезентативною, тобто забезпечить більш ефективні оцінки досліджуваних параметрів [8, с.29-31].

Порядок виконання подібного дослідження може бути таким:

1. Отримати приклад сукупності U для $N=8$ виду

i	1	2	3	...
y_i	y_1	y_2	y_3	...

Наприклад,

i	1	2	3	4	5	6	7	8
y_i	1	2	4	4	7	7	7	8

2. Для сукупності U обчислити сумарне (T) та середнє значення (μ), дисперсію (S^2).
3. З'ясувати, скільки різних випадкових вибірок обсягу n ($n=1\dots N$) можна отримати із заданої сукупності.
4. Побудувати усі можливі вибірки обсягу n .
5. Для кожної з побудованих вибірок обчислити сумарне значення ($t_{\text{виб}}$), середнє значення (y) та вибірккову дисперсію (s^2).
6. Обчислити для кожної вибірки оцінку \hat{t}_i сумарного значення сукупності U . Визначити імовірності отримання різних значень \hat{t} та побудувати гістограму розподілу оцінок сумарного:

k				
$P(\hat{t}_i = k)$				

7. Обчислити математичне сподівання, зміщення, дисперсію, коваріацію та стандартну похибку оцінки сумарного значення сукупності U .
8. Виконати пункти 4-7 для іншого значення n .
9. Порівняти дисперсії оцінок, отриманих для вибірок різного розміру.
10. Зробити висновки про ефективність, зміщення та точність оцінок сумарного значення сукупності U .

В результаті проведеного лабораторного дослідження буде з'ясовано, що сумарне значення, отримане за генеральною сукупністю, дорівнює 40.

При обсязі $n=4$ можна побудувати 70 різних випадкових вибірок. За їхніми результатами математичне сподівання сумарного значення становитиме також 40 (тобто такий відбір забезпечуватиме незміщену оцінку), а дисперсія оцінки сумарного становитиме $\approx 54,9$. Розподіл оцінок сумарного значення, отриманого за усіма 70-ма вибірками наведено на Рис. 3.

При обсязі $n=7$ можна побудувати 8 різних випадкових вибірок. За їхніми результатами математичне сподівання сумарного значення становитиме $\approx 39,99$ (тобто такий відбір забезпечуватиме також незміщену оцінку, хоча на Рис. 4 видно, що

у даному випадку ні для жодної вибірки оцінка не співпадає з точним значенням), але дисперсія оцінки сумарного становитиме лише $\approx 9,5$, тобто майже вшестеро менше, ніж для вибірок обсягу $n=4$. Тобто вибірки більшого розміру, принаймні при простому випадковому відборі, є більш ефективними, що безумовно слід враховувати при проведенні досліджень.

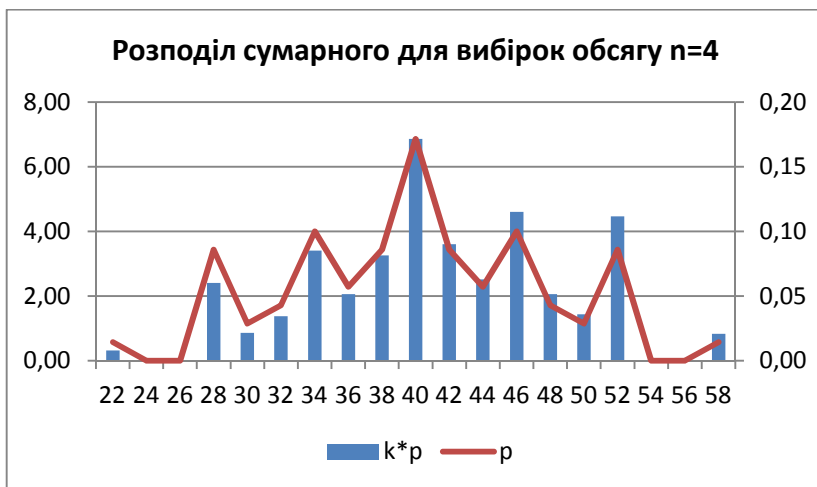


Рис. 3.

Для освітніх вимірювань особливо цікавим є кластерний відбір, оскільки в цій галузі лівова частка матеріальних ресурсів витрачається на методичне забезпечення, як то розробка та друк тестів, підготовку персоналу, оплату праці експертів та тих, хто буде проводити та оцінювати результати тестування. Крім того, специфіка освітнього процесу така, що зазвичай перевіряти якість та здійснювати освітній процес на суб'єктах, відібраних простим випадковим відбором, просто неможливо, або надзвичайно дорого.

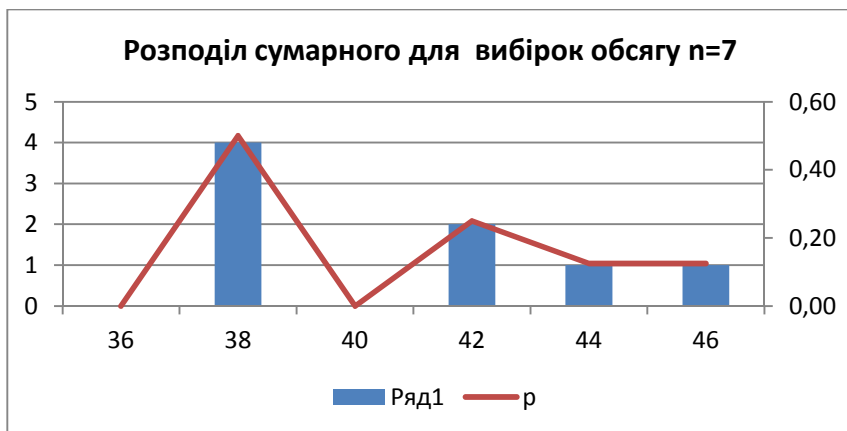


Рис. 4.

Література

1. Черняк О.І. Техніка вибірових досліджень. – К.: МІВВЦ, 2001. – 248 с.
2. Пархоменко В.М. Методи вибірових обстежень: Навчальний посібник. – К., 2001. – 148 с.
3. Василик О.І. Лекції з теорії і методів вибірових обстежень: навчальний посібник / О.І. Василик, Т.О.Яковенко. – К.: ВПЦ «Київський університет», 2010. – 208 с.
4. Кокрен У. Методы выборочного исследования. – М.: Финансы и статистика, 1976.
5. Джессен Р.Д. Методы статистических обследований. – М.: Финансы и статистика, 1985.
6. Всі міста України (перепис 5 грудня 2001 р.) [Електронний ресурс] – Режим доступу: http://ukrmap.org.ua/Naselenie_ukr.htm
7. Горицкий Ю.А., Перцов Е.Е. Практикум по статистике с пакетами StatGraphics, Statistica, SPSS [Ел. рес.]. – Реж. доступу: http://www.nsu.ru/matlab/Exponenta_RU/educat/systemat/goritskii/lr.a.sp.htm
8. Lohr Sh.L. Sampling: Design and Analysis –USA, 2010. – 609 p.

ОСОБЛИВОСТІ ВИКЛАДАННЯ ДИСЦИПЛІНИ «КОМП'ЮТЕРНІ СТАТИСТИЧНІ ПАКЕТИ»

Лупан І.В.

Розвиток сучасної теорії освітніх вимірювань неможливо уявити без відповідних математико-статистичних методів. Саме необхідність на практиці проводити об'єктивне оперативне оцінювання великої кількості людей, наприклад, випускників шкіл, спонукає науковців шукати відповідні математичні моделі. Виявляється, що ці моделі здебільшого імовірнісні, статистичні: для аналізу та оцінки існуючих та нових засобів, що використовувалися або будуть використовуватися в освітніх вимірюваннях, наприклад, тестів, необхідно збирати дані про результати пробних вимірювань, аналізувати їх статистичними методами. Це передбачає досить складну обробку великих обсягів даних. Опанування спеціального статистичного програмного забезпечення і становить основне завдання курсу «Комп'ютерні статистичні пакети», який включено до програми підготовки фахівців з освітніх вимірювань. Однак у викладанні даного курсу є певні особливості. Спробуємо їх прокоментувати.

По-перше, оскільки основним завданням курсу є ознайомлення слухачів з інструментарієм прикладних статистичних пакетів, існує небезпека притаманна усім інформатичним дисциплінам: це небезпека «кнопочної технології», тобто зведення навчального змісту до запам'ятовування комбінації кнопок, або послідовності вибору пунктів меню без свідомого розуміння виконуваних дій. Подолати цю проблему легко, піднявши теоретичний рівень курсу. Для цього слід розглядати не стільки інтерфейс прикладного пакета, скільки особливості застосування відповідних статистичних методів, вимоги, які накладаються на дані, алгоритми обчислення критеріїв. Можливо, не обов'язково робити це з математичними доведеннями, однак, перш ніж застосувати відповідний інструмент, студент має ознайомитися з його призначенням та специфікою. Саме на цих засадах нами розроблено посібник до даного курсу [1].

Другою проблемою є те, що статистичний інструментарій мають і універсальні математичні пакети, такі як Maple, Matematica та інші, і пакети загального призначення, такі як MS Excel, і, зрозуміло, вузькоспеціалізовані статистичні пакети, такі як SPSS, Statistica, – завданням же курсу є навчити слухачів використовувати необхідний інструментарій, користуючись наявним програмним забезпеченням. Вирішенням даної проблеми є також підвищення теоретичного рівня курсу в цілому: адже знаючи особливості застосування відповідного статистичного методу, легко зорієнтуватись в тому, як його реалізовано у конкретному пакеті. Тим не менше доцільно все ж при вивченні курсу знайомити студентів не з одним пакетом, а принаймні з пакетами із різних груп. Так у нашому курсі ми починаємо з пакета загального призначення MS Excel, оскільки його інтерфейс добре знайомий студентам ще зі школи і можна приділити більше уваги саме статистичним методам, а пізніше розглядаємо один із статистичних пакетів, поки що це SPSS. Вибір саме цих пакетів не принциповий: можна починати, наприклад, з OpenOffice Calc і продовжувати з пакетом Statistica. Однак практика показує, що у пакетах різних груп дійсно використовуються різні підходи до організації даних та інтерфейсів, тому знайомство бодай з двома різними пакетами надзвичайно доречно [1, 2]. До того ж складні статистичні процедури, необхідні кваліфікованому фахівцю з освітніх вимірювань, такі як дискримінантний, кластерний, канонічний, факторний аналіз тощо, реалізовані лише у спеціалізованих статистичних пакетах.

Також доречно знайомити студентів, по можливості, як з пропріетарним (платним), так і з безкоштовним програмним забезпеченням.

Курс складається з 10 тем, при вивченні яких передбачено ознайомлення з відповідними статистичними методами та виконання лабораторних робіт з використанням відповідного інструментарію статистичних пакетів.

До першого модуля включено процедури описової статистики та відповідні засоби комп'ютерних статистичних інструментів. Тут розглядаються теми:

1. Статистичні дані та їх обробка.

Класифікація типів змінних. Кількісні, якісні, ординальні та номінальні дані: кодування даних різних типів. Групування та ранжування даних. Візуалізація емпіричних даних.

2. Статистичні гіпотези та критерії.

Оцінка достовірності результатів статистичного дослідження. Рівень статистичної значущості. Помилки статистичного висновку. Потужність критеріїв. Класифікація методів статистичної обробки даних.

3. Інструменти статистичної обробки пакетів загального призначення (на прикладі табличного процесора MS Excel або OpenOffice.org Calc).

Класифікація статистичних пакетів. Інструменти статистичної обробки табличного процесора: статистичні функції, Пакет аналізу, Майстер діаграм. Визначення основних характеристик скалярних змінних: центрів групування та розсіювання значень, скошеності та гостроверхості розподілу і т.д., побудова гістограми та кумуляти розподілу, обчислення дискримінативної здатності тесту.

4. Призначення та особливості застосування спеціалізованих статистичних пакетів (на прикладі пакетів SPSS або Statistica).

Структура робочої області. Засоби введення та редагування даних і змінних. Перетворення даних. Редагування та збереження звітів. Побудова графіків. Виконання процедур описової статистики.

У другому модулі розглядаються базові статистичні методи, які застосовуються до порівняння двох вибірок або двох змінних. Засоби виконання перелічених нижче статистичних процедур є як у спеціалізованих пакетах, так і у табличних процесорах. Саме ці процедури найчастіше використовуються і у статистичному аналізі результатів тестування. Тут розглядаються теми:

5. Критерії порівняння двох вибірок.

Параметричні та непараметричні критерії. Критерії Фішера та Стюдента, критерій Крамера-Уелча. Непараметричні критерії

порівняння двох вибірок. Засоби порівняння двох вибірок у статистичних пакетах.

6. Критерії порівняння двох розподілів.

Задачі порівняння варіаційних рядів розподілу даних: порівняння емпіричного розподілу з теоретичним та з іншим емпіричним. Критерій χ^2 -Пірсона: умови застосування та обмеження критерію. Виконання обчислень та оцінка результату засобами статистичних пакетів.

7. Кореляційний та регресійний аналіз.

Характеристики статистичного зв'язку кількісних даних: індекс кореляції, коефіцієнт детермінації, кореляційне відношення. Парні рангові коефіцієнти кореляції. Кореляційний аналіз номінальних даних. Оцінка та перевірка на значимість характеристик статистичного зв'язку. Особливості виконання кореляційного аналізу у статистичних пакетах.

Постановка задачі регресійного аналізу. Вибір оптимальної регресійної моделі. Особливості виконання регресійного аналізу у статистичних пакетах.

8. Дисперсійний аналіз.

Моделі дисперсійного аналізу. Однофакторний дисперсійний аналіз. Визначення впливів головних ефектів та взаємодій. Візуалізація даних для дисперсійного аналізу. Особливості виконання дисперсійного аналізу у статистичних пакетах.

До третього модуля включено методи багатовимірною статистичного аналізу, які реалізовано лише у спеціалізованих пакетах. Однак саме ці методи є потужним інструментом нових актуальних досліджень у галузі освітніх вимірювань. У модулі розглядаються теми:

9. Задачі класифікації.

Простори об'єктів, спостережень, ознак. Основні постановки задач класифікації. Дискримінантний аналіз. Навчаюча вибірка. Дискримінантні функції. Задача кластерного аналізу. Види кластерного аналізу. Виконання кластерного та дискримінантного аналізу засобами статистичних пакетів.

10. Факторний та канонічний аналіз.

Аналіз головних компонент та факторний аналіз. Проблема кількості факторів. Методи факторного аналізу. Послідовність виконання факторного аналізу. Призначення канонічного аналізу. Особливості застосування та інтерпретування результатів канонічного аналізу. Виконання факторного та канонічного аналізу засобами комп'ютерних статистичних пакетів.

При удосконаленні курсу слід приділити увагу іншим інструментам статистичних пакетів, корисним у роботі фахівця з освітніх вимірювань, зокрема засобам аналізу надійності (Reliability Analysis), та спеціалізованим засобам для статистичної обробки результатів тестування, наприклад, ІТАР [3], а також на розробку завдань практичного спрямування. Деякі кроки у цьому напрямку вже здійснено, підтвердженням чого є наші публікації, присвячені застосуванням дискримінантного та канонічного кореляційного аналізу у дослідженнях в галузі освітніх вимірювань [4, 5].

Наведемо завдання лабораторних робіт з курсу «Комп'ютерні статистичні пакети».

Тема 1: «Описові статистики»

У теоретичних відомостях до лабораторної роботи розглядають поняття інтервального статистичного ряду та порядок його побудови (визначення кількості класових інтервалів та величини класового інтервалу), порядок ранжування значень вибірки та перевірку правильності ранжування, характеристики шкал вимірювання, статистики та перетворення, допустимі для кожної шкали.

Завданням лабораторної роботи передбачено виконання процедур описової статистики засобами пакетів MS Excel (або OpenOffice.org Calc) та SPSS або Statistica.

Завдання 1 (MS Excel)

1. Створити робочу книгу «Розподіл». На аркуші «Вибірки» ввести у стовпчик експериментальні дані (вибіри А та В). Скопіювати введені дані на аркуш «Параметри».
2. Обчислити кількість спостережень, мінімальне та максимальне, використовуючи стандартні функції MS Excel.

Обчислити розмах вибірки за формулою:
Розмах=максимальне – мінімальне.

3. Обчислити середні значення, моду, медіану, дисперсію, середнє квадратичне відхилення, асиметрію та ексцес для кожної вибірки.
4. Розрахувати похибки обчислення відповідно середнього, асиметрії та ексцеса за формулами:

$$\text{похибка} = \frac{S}{\sqrt{n}} \quad (1), \quad m_A = \sqrt{\frac{6}{n}} \quad \text{та} \quad m_E = 2 \cdot \sqrt{\frac{6}{n}}.$$

5. Скопіювати введені дані на аркуш «Обчислення», використавши команду Правка->Спеціальная вставка – >Связь (таким чином між даними аркушів «Вибірки» та «Обчислення» збережеться односторонній зв'язок: зміни даних на аркуші «Вибірки» відбиватимуться на аркуші «Обчислення»).
6. На аркуші «Обчислення» з даних вибірки В сформувати систематичну вибірку В_с з періодом 5 (скориставшись послугою Выборка з надбудови Анализ данных). Обчислити для неї середнє, дисперсію і стандартне квадратичне відхилення за формулами відповідно

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}, \quad S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad \text{та за}$$

допомогою функцій MS Excel.

7. Відсортувати дані вибірки В_с за зростанням.
8. Проранжувати вибірку В_с за допомогою функції РАНГ.СР (порядок ранжування = 1, тобто ранг 1 надати найменшому).

¹ Тут S – стандартне квадратичне відхилення, n – кількість спостережень у вибірці. Якщо обсяг генеральної сукупності (N) відомий,

то похибку обчислюють за формулою: $\text{похибка} = \frac{S}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$.

9. Перевірити правильність ранжування за формулою:

$$\sum R_i = \frac{N \cdot (N + 1)}{2}.$$

10. На аркуші «Розподіл» побудувати для кожної вибірки ряд розподілу частот за допомогою функції ЧАСТОТА, розрахувавши перед тим ширину класового інтервалу та визначивши границі класових інтервалів.
11. Для кожної частотної таблиці розрахувати і записати у сусідні стовпці відносну частоту, накопичену частоту та відносну накопичену частоту. Застосувати до відносних частот ПРОЦЕНТНИЙ формат.
12. До кожної частотної таблиці побудувати гістограму розподілу та кумуляту (до кожної таблиці окремий аркуш з графіками).
13. Зробити висновки про вид розподілів вибірок А та В за графіком. Проінтерпретувати отримані статистичні показники. Записати висновки у примітку.
14. До даних вибірки В застосувати процедуру *Описательная статистика* з надбудови *Анализ Данных*.

Завдання 2 (SPSS)

1. Завантажити пакет SPSS через меню ПРОГРАМИ, або запустивши файл C:\Program Files\SPSS\spsswin.exe.
2. Завантажити довільний файл *.sav, наприклад, файл Employee data.sav, з папки SPSS (File → Open → Data). Ознайомитися зі структурою файлу, параметрами змінних (тип, вид шкали, мітка, значення номінативної шкали), параметрами форматowanego виведення даних (ширина стовпця, вирівнювання).
3. Застосувати до змінних educ (рівень освіти у роках) та jobcat (категорія служби) процедури описової статистики пакету SPSS: Analyze → Descriptives та Analyze → Frequencies.
4. Зберегти звіт у форматі звіту SPSS, HTML та RTF.
5. Завантажити в SPSS дані з книги Розподіл.xls. Для цього слід, по-перше, виконати команду File → Open → Data, по-друге, вказати для завантажуваного файлу тип Excel, по-третє, вказати назву файла та аркуша з даними. Застосувати до них

процедури описової статистики (Descriptives) пакету SPSS. Порівняти результати, отримані в SPSS та MS Excel. Побудувати за досліджуваними даними гістограму з нормальною кривою (Histogram) та стовпцеві діаграми за частотою та відносною частотою. Зберегти звіт у форматі пакету SPSS та HTML.

6. Відсортувати дані за зростанням, застосувавши процедуру Data → Sort Cases. Вибрати змінну, за якою сортувати (Sort by:) та вказати порядок сортування (Ascending – за зростанням, або Descending – за спаданням).
7. Проранжувати значення змінної, застосувавши процедуру Transform → Rank cases. У діалоговому вікні вибрати змінну для ранжування, вказати порядок ранжування, надавши найменшому значенню ранг 1 (Assign rank 1 to smallest value) та порядок нарахування рангів однаковим значенням змінної (Ties → Rank Assigned to Ties вибрати Mean – обчислювати середнє арифметичне можливих рангів). Результатом буде нова змінна.

Завдання 3 (Statistica):

1. Завантажити пакет Statistica запуском файла C:\Program Files\StatSoft\STATISTICA6\statist.exe або через меню ПРОГРАМИ.
2. Завантажити довільний приклад з папки C:\...\STATISTICA6\Examples\Datasets, наприклад файл job_prof.sta, скориставшись командою File → Open головного меню програми.
3. Ознайомитися із структурою робочого вікна програми Statistica. Переглянути для кожної змінної її властивості, скориставшись командою Variable Specs з контекстного меню до заголовку змінної.
4. Ознайомитися з набором інструментів дескриптивної статистики (Statistica → Basic Statistics → Descriptive Statistics): для однієї із змінних прикладу застосувати послідовно усі процедури закладки Quick (Summary: Descriptive statistics, Frequency tables, Histograms, Bow&whisker plot). Ознайомитися із структурою, створеною в

- результаті застосування обчислювальних процедур робочої книги (Workbook). Зберегти робочу книгу.
5. На закладці Advanced вибрати необхідні параметри описової статистики та запустити процедуру обчислення (Summary: Descriptive statistics). Транспонувати отриману таблицю командою Data → Transpose → File головного меню. Зберегти таблицю у файл звіту. Для цього виконати команду контекстного меню виділеного об'єкту Extract as stand-alone window (виділити в окреме вікно) та Add to report (додати до звіту) головного меню.
 6. Застосувати процедури Frequency tables та Histograms закладки Normality, встановивши необхідну кількість класових інтервалів (Number of intervals). Додати результати їх застосування до створеного у попередньому пункті звіту. Зберегти звіт у форматах пакету Statistica, RTF та HTML.
 7. Зберегти файл з даними у форматі SPSS.
 8. Завантажити дані з робочої книги Розподіл.xls. Застосувати необхідні процедури описової статистики, побудувати необхідні графіки до однієї з вибірок. Створити та зберегти звіт.
 9. Впорядкувати значення змінної за зростанням: виділити заголовок змінної та застосувати процедуру Data → Sort.
 10. Проранжувати змінну. Оскільки Statistica не створює нової змінної, як SPSS, а замінює значення рангами, бажано спершу скопіювати ранжовану змінну (команда Copy Variables контекстного меню), та застосувати до копії команду Data → Rank.
 11. Порівняти інструментарій декількох статистичних пакетів. Зробити висновки.

Приклад виконання

На РРис. 2 подано результати застосування функції ЧАСТОТА (стовбець Частота у вибірці А), а також графічне представлення інтервального статистичного ряду розподілу у

вигляді гістограми частот та кумуляти (графіка відносних накопичених частот).



Рис. 2

	A	B	C	D	E	F
1		Вибірка А	Вибірка В	Вс	Ранг Excel	Ранг
2		98	96	109	1	1,5
3		110	109	109	1	1,5
4		110	101	123	3	3
5		111	109	125	4	4
6		120	112	148	5	5
7		125	123			
8		142	140			
9		142	125			
10		145	140			
11		150	148			
12	Кількість	10	10	5		
13	Сума	1253	1203	614	14	15
14	Мінімальне	98	96			
15	Максимальне	150	148			
16	Розмах	52	52			
17	середнє	125,300	120,300	122,800		
18	сер. Геом.	124,093	119,124			
19	сер. Гарм.	122,886	117,970			
20	Урізане	125,300	120,300			
21	Мода	110,000	109,000			
22	Медіана	122,500	117,500			
23	Дисперсія	333,567	317,789	255,200		
24	Відхилення	18,264	17,827	15,975		
25	Похибка сер.	5,776	5,637	7,144		
26	Асиметрія	0,033	0,286			
27	Похибка Ас.	0,775	0,775			
28	Ексцес	-1,621	-1,281			
29	Похибка Екс.	1,549	1,549			

Рис. 1

На Рис. 1 у стовпчиках В та С наведено вибірки А та В з результатами обчислення параметрів розподілів за допомогою функцій MS Excel.

Тема 2: «Порівняння параметрів двох вибірок»

У теоретичних відомостях до лабораторної роботи розглядають статистичні критерії Фішера-Снедекора (для порівняння дисперсій двох вибірок) та t-критерій Стьюдента (для порівняння середніх).

Завданням лабораторної роботи передбачено виконання порівняння експериментальних даних засобами пакетів MS Excel (або OpenOffice.org Calc) та SPSS або Statistica. Для аналізу можна використати дані, отримані в результаті тестування.

Завдання 1: порівняння параметрів двох незалежних вибірок.

1. В MS Excel у робочій книзі «СТЮДЕНТ» на аркуші «t-тест» сформувані дві вибірки значень обсягом 30-40 чисел (бажано щоб сформовані вибірки мали неоднаковий обсяг). Придумати статистичний сюжет до отриманих даних. Записати його у примітку до комірки А1.
2. Переконалися у підключенні Пакета Аналізу.
3. Обчислити дисперсії двох вибірок, з'ясувати, яка з них більша при простому порівнянні. Сформувані статистичні гіпотези стосовно порівнюваних дисперсій.
4. Виконати перевірку на рівність дисперсій двох вибірок, застосувавши Двухвыборочный F-тест для дисперсій (критерій Фішера-Снедекора) з Пакету Аналізу (Анализ данных). Зробити висновки про дисперсії двох вибірок.
5. Виконати порівняння дисперсій за допомогою стандартних функцій MS Excel:
визначити за функцією ФТЕСТ(масив1; масив2) двосторонню імовірність схожості двох сукупностей (2P);
визначити степені вільності: $k_1=n_1-1$ та $k_2=n_2-1$ (тут n_1 – кількість спостережень у вибірці з більшою дисперсією, n_2 – кількість спостережень у вибірці з меншою дисперсією);

- отримати значення односторонньої імовірності $P=2P/2$;
 визначити експериментальне значення критерію F , що відповідає отриманій імовірності за функцією
 $F_{РАСПОБР}(P; \text{степені_вільності}_1; \text{степені_вільності}_2)$;
 визначити критичні значення F для рівнів значущості $p=0,05$ та $p=0,01$.
6. Порівняти результати з результатами застосування процедури з Пакета Аналізу.
 7. Сформувати статистичні гіпотези стосовно середніх двох порівнюваних вибірок. Виконати для вибірок порівняння середніх: вибрати процедуру Двухвыборочный t-тест с одинаковыми дисперсиями або Двухвыборочный t-тест с различными дисперсиями, залежно від результатів F-тесту (див. п. 4). Зробити статистичні висновки.
 8. Отримати критичні значення t , застосувавши функцію СТЬЮДРАСПОБР(p , k), де p – це обраний рівень значущості, а k – степені вільності ($k = n_1 + n_2 - 2$).
 9. Отримати імовірність похибки першого роду для обчисленого значення t (за пунктом б), застосувавши функцію СТЬЮДРАСП(X , k , m), де X – обчислене у t-тесті значення критерію Стьюдента, k – степені вільності, m – кількість хвостів розподілу ($m=1$ для одностороннього критерію, $m=2$ для двостороннього критерію).
 10. Скопіювати дані з книги MS Excel таким чином: створити дві змінні, до першої змінної скопіювати досліджувані значення, а до другої записати номер вибірки (1 або 2). Поділ на вибірки у пакеті SPSS буде здійснено автоматично за значеннями другої змінної.
 11. Застосувати до введених даних процедуру SPSS Analyze → Compare Means → Independent Samples t-test. Вказати першу змінну як досліджувану (Test Var), а другу як групувальну (Grouped Var) та визначити її значення для різних груп (1 – для першої, 2 – для другої).
 12. Проаналізувати отримані результати, зробити висновки.
 13. Виконати порівняння середніх засобами пакету Statistica. Для даних, поданих у двох стовпцях (як в MS Excel), слід застосувати процедуру Basic statistics→t-test, independent,

by variables. Для даних, поданих як в SPSS (в один стовпчик + групувальна змінна), застосовують процедуру Basic statistics → t-test, independent, by groups.

14. Порівняти результати, отримані у різних пакетах.

Завдання 2: порівняння параметрів двох залежних вибірок.

1. На аркуші «**Парний t-тест**» ввести значення для парного тесту Стьюдента (спряжені дані, можливо результати двох тестів для однієї вибірки учнів).
2. Застосувати процедуру Парний двухвыборочный t-тест для середніх з Пакету Аналізу. Зробити статистичні висновки.
3. До тих самих даних застосувати Двухвыборочный t-тест. Порівняти результати з результатами парного тесту. Зробити висновки.
4. Для тих самих даних застосувати процедуру Analyze → Compare Means → Pared-Samples T Test пакета SPSS.
5. Ознайомитися з порядком застосування та налаштуваннями процедури Statistics → Basic statistics → t-test, dependent samples пакету Statistica.
6. Сформувати та зберегти звіт про виконання тесту, у якому зберегти вхідні дані та результати t-теста.

Тема 3: «Дисперсійний аналіз»

У теоретичних відомостях до лабораторної роботи розглядають одновимірний та двовимірний дисперсійний аналіз, умови їх застосування та критерії оцінювання якості дисперсійної моделі.

Завданням лабораторної роботи передбачено виконання дисперсійного аналізу для експериментальних даних засобами пакетів MS Excel (або OpenOffice.org Calc) та SPSS або Statistica та побудову відповідних графіків. Для аналізу можна використати дані, отримані в результаті тестування.

Завдання 1: однофакторний дисперсійний аналіз

1. На аркуші «**Однофакторний**» ввести значення для однофакторного дисперсійного аналізу.

2. Сформулювати статистичні гіпотези.
3. Виконати процедуру Однофакторний дисперсійний аналіз з Пакета аналізу.
4. Побудувати графік групових середніх.
5. Проінтерпретувати отримані результати, зробити статистичні висновки.

Завдання 2: однофакторний дисперсійний аналіз в пакетах SPSS та Statistica

1. Перенести дані лабораторної роботи в пакет SPSS. Додати змінну, що містить значення градацій контрольованого фактора.
2. В пакеті SPSS виконати однофакторний дисперсійний аналіз: Analyze → Compare Means → One-Way ANOVA, вказавши залежною змінною (Dependent List) основну досліджувану змінну, а змінну, що містить градації досліджуваного фактора як Factor.
3. В настройках (Options) встановити прапорці Descriptive, Homogeneity-of-variance та Means plot. Відповідно у звіті про дисперсійний аналіз отримаємо описову статистику для комірок дисперсійного комплексу, тест на однорідність дисперсій та графік середніх.
4. У пакеті Statistica оформити вхідні дані так само, як і в пакеті SPSS. Виконати процедуру Statistics → ANOVA, вибрати тип аналізу – One-Way ANOVA. У наступному діалоговому вікні вказати залежну змінну (Dependent variable) та факторну змінну і її рівні (в пакеті Statistica для аналізу можна вибрати не всі градації фактора).
5. На закладці Summary у наступному діалоговому вікні переглянути результати однофакторного аналізу: Cell statistics – результати описової статистики для кожної комірки дисперсійного комплексу, Univariate results – власне результати однофакторного аналізу, All effects/Graphs – графіки середніх.
6. Виконати однофакторний дисперсійний аналіз за допомогою послуги Statistics → Basic Statistics/Tables → Breakdown &

one-way ANOVA. На закладці Quick у діалоговому вікні Statistics by Groups Results отримати Summary Table of Statistics – результати описової статистики, Analysis of Variance – власне результати дисперсійного аналізу та Interaction plots – графіки середніх.

7. Порівняти результати застосування однофакторного дисперсійного аналізу в різних статистичних пакетах.

Завдання 3: двофакторний дисперсійний аналіз

1. Ввести значення для двофакторного аналізу.
2. Виконати процедуру Двухфакторный дисперсионный анализ с повторениями (аналіз без повторів здійснюється у разі, коли кожна комірка дисперсійного комплексу містить лише одне значення, наприклад, середнє по групі).
3. Побудувати графіки зміни середнього арифметичного для різних градацій впливаючих факторів. За графіком висунути припущення про наявність чи відсутність взаємодії факторів.
4. Проінтерпретувати отримані результати, зробити висновки.

Завдання 4: двофакторний дисперсійний аналіз в пакетах SPSS та Statistica

1. Перенести дані лабораторної роботи в пакет SPSS (добавити змінні, що містять значення градацій контрольованих факторів).
2. Виконати процедуру Analyze -> General Linear Model -> Univariate.
3. Перенести залежну змінну у поле Dependent Variable, а змінні-фактори у поле Fixed Factor(s).
4. У діалоговому вікні Options встановити прапорці Descriptive statistics та Estimates of effect size (оцінка величини ефекту).
5. Для побудови графіків вибрати кнопку Plots та додати два графіки, вказавши один із факторів як горизонтальну вісь (Horizontal Axis), а інший – як окрему лінію (Separate Lines).
6. Порівняти отримані результати з результатами попереднього завдання, зробити висновки.

Тема 4: «Кореляційний та регресійний аналіз»

У теоретичних відомостях до лабораторної роботи розглядають критерій лінійної кореляції Пірсона та критерій рангової кореляції Спірмена, критерії оцінювання достовірності отриманих коефіцієнтів, а також порядок побудови та оцінювання якості лінійної регресійної моделі.

Завданням лабораторної роботи передбачено здійснення кореляційного аналізу, побудову діаграми розсіювання та регресійної прямої засобами пакетів MS Excel (або OpenOffice.org Calc) та SPSS або Statistica.


Завдання 1: Кореляційний аналіз

Завдання 1а: кореляційний аналіз у пакеті MS Excel


1. Ввести експериментальні дані – стовпці X, Y, Z.
2. Для кожного стовпця обчислити кількість (СЧЕТ), середнє значення (СРЗНАЧ), дисперсію (ДИСП), середнє квадратичне відхилення (СТАНДОТКЛОН).
3. Обчислити коефіцієнти кореляції r_{xy} , r_{xz} , r_{yz} за формулою Пірсона (ПИРСОН).
4. Визначити достовірність коефіцієнта кореляції за критерієм Стьюдента:
 - а) обчислити значення критерію Стьюдента;
 - б) визначити р-значення (імовірність помилки I роду): $p = \text{СТЮДРАСП}(t, n-2, 2)$. Тут першим параметром є обчислене значення критерію, другим – кількість степенів вільності, а третім – кількість хвостів розподілу (2 – для двостороннього критерію та 1 – для одностороннього). Порівняти р-значення з прийнятим рівнем значущості ($\alpha=0,05$ або $\alpha=0,01$). При отриманому $p < \alpha$, гіпотеза H_0 відкидається і приймається H_1 , тобто кореляція достовірна. При $p > \alpha$ – кореляцію вважають недостовірною, тобто зв'язок між змінними, незалежно від абсолютної величини коефіцієнта кореляції, випадковий.
5. Побудувати точкові діаграми (діаграми розсіювання) для кожної пари змінних.

6. Виконати оцінку коефіцієнта кореляції змінних X та Y , обчисливши коефіцієнт кореляції Спірмена для частини основної вибірки (15-20 значень). Сформувати малочисельну вибірку випадковим чином, скориставшись процедурою **ВИБОРКА** з пакета Аналізу даних або згенерувавши 15-20 випадкових чисел за допомогою функції **СЛУЧМЕЖДУ(1; N)** (тут N – обсяг вибірки C) та скопіювавши на окремий аркуш записи з вибірки C з відповідними номерами. Обчислити для них коефіцієнт кореляції Спірмена за алгоритмом. Зробити висновки про його достовірність. Порівняти результат дослідження випадкової малочисельної вибірки з результатом для генеральної сукупності (тут вибірка C).
7. Обчислити для тієї ж малочисельної вибірки коефіцієнт кореляції Пірсона. Порівняти результати застосування різних алгоритмів. Зробити висновки.

Завдання 1б: кореляційний аналіз у пакеті SPSS

1. Для цього застосувати процедуру Analyze → Correlate → Bivariate, вибрати для порівняння усі змінні та вказати коефіцієнти кореляції, які слід обчислити: Пірсона (Pearson), τ -Кендала (Kendall's tau-b) та Спірмена (Spearman). У пункті Options можна вказати спосіб обробки пропущених даних та необхідність обчислення деяких параметрів дескриптивної статистики.
2. Побудувати діаграми розсіювання, застосувавши процедуру Graps → Scatterplot → Simple.
3. Побудувати тривимірну діаграму розсіювання Graps → Scatterplot → 3D. Вибрати найкращий ракурс тривимірної діаграми (відкрити редактор діаграм командою контекстного меню SPSS Chart Object → Open та скористатися інструментом 3D-Rotation ()).
4. Порівняти отримані результати з обчисленими в MS Excel. Зробити висновки.

Завдання 1в: кореляційний аналіз у пакеті Statistica

1. Ознайомитися з інструментарієм процедури Statistics → Correlation Matrices. Для даних вибірки С виконати на закладці Advanced/Plot операції Summary: Correlation matrix (обчислення коефіцієнта кореляції Пірсона для кожної пари вибратних змінних), Partitial correlation (часткова кореляція), Scatterplot matrix (матриця точкових діаграм), 3D scatterplots (тривимірна точкова діаграма). На тривимірній діаграмі підібрати найкращий ракурс зображення (кнопка  панелі інструментів графічних об'єктів). Результати застосування операцій внести до звіту та зберегти.
2. Застосувати критерії рангової кореляції (процедура Statistics → Nonparametrics → Correlation (Spearman, Kendall tau, gamma)).
3. Порівняти результати застосування трьох пакетів. Зробити висновки.

Завдання 2: Регресійний аналіз

1. За даними вибірки С побудувати рівняння регресії Y по X та X по Y, скориставшись функцією ЛИНЕЙН() MS Excel.
2. На відповідних точкових діаграмах за допомогою команди **Добавить линию тренда** побудувати графіки лінійної та деякої нелінійної регресії (підібрати таку нелінійну функцію, яка найкраще пояснює емпіричні дані, тобто має $R^2 \approx 1$). Висунути гіпотезу про те, яка функція краще апроксимує експериментальні дані.
3. За допомогою процедури Регрессия з Пакета Аналіза виконати лінійний регресійний аналіз Y по X. Зробити висновки.
4. За допомогою процедури Регрессия з Пакета Аналіза побудувати рівняння множинної лінійної регресії змінної Y від X та Z. Зробити аналіз отриманих результатів.
5. Ознайомитися з інструментами регресійного аналізу пакета SPSS (Analyze → Regression). У пакті SPSS виконати

регресійний аналіз Y по X та множинний регресійний аналіз Y по X та Z .

6. За допомогою процедури Analyze \rightarrow Curve Estimation побудувати до даних графіки лінійної та квадратичної регресії. Порівняти результати, з'ясувати, яка функція краще апроксимує дані.
7. Ознайомитися з інструментами регресійного аналізу пакета Statistica (Statistics \rightarrow Multiple Regression). Виконати регресійний аналіз Y по X у пакті Statistica.
8. Порівняти результати, отримані у трьох пакетах. Зробити висновки.

Приклади виконання

Приклад 1: Учні проходили тестування з трьох дисциплін. Результати тестування наведено у таблиці.

	A	B	C	D
1		Тест1 (X)	Тест 2 (Y)	Тест 3 (Z)
2	учень 1	30	20	18
3	учень 2	30	25	18
4	учень 3	30	20	18
5	учень 4	37	23	19
6	учень 5	49	30	21
7	учень 6	41	21	20
8	учень 7	25	16	15
9	учень 8	45	27	20
10	учень 9	40	25	17
11	учень 10	39	24	19

З'ясувати, чи існує зв'язок між змінними. Побудувати діаграми розсіювання.

Виконання:

У пакеті MS Excel Зручно подати результати обчислень під таблицею вхідних даних.

Нижче наведено формули, які слід ввести до комірок у даному випадку B14, C14, D14 та E14 відповідно для обчислення кореляції між змінними X та Y (R_{xy}).

Комірка	Призначення комірки	Формула
B14	Коефіцієнт кореляції за Пірсоном для першого та другого стовпців даних	=ПИРСОН(B2:B11;C2:C11)
C14	Емпіричне значення t-критерія для оцінки достовірності отриманого коефіцієнту кореляції: =B14*КОРЕНЬ(10-2)/КОРЕНЬ(1-B14^2)	
D14	Імовірність помилки I роду (α)	=СТЬЮДРАСП(C14;10-2;2)
E14	Результати порівняння α з обраним рівнем значущості та висновки	=ЕСЛИ(D14<0,05;»Н1»;»Н0«)

Для наведеного прикладу отримуємо:

	Кореляція (r)	t-критерій	alpha	Висновок
R_{xy}	0,826997062	4,160574433	0,003162722	Н1
R_{xz}	0,821279263	4,071553121	0,00357572	Н1
R_{yz}	0,718129688	2,918738829	0,019328577	Н1

Тобто в усіх випадках кореляція достовірна. Для Γ_{xy} та Γ_{xz} навіть на рівні значущості $\alpha=0,01$. А для Γ_{yz} лише на рівні значущості $\alpha=0,05$.

Діаграму розсіювання для змінних X та Y зображено на *Рис. 3*. На ній видно, що більшість точок майже лежить на прямій з додатним кутовим коефіцієнтом. Тобто за діаграмою можна припустити наявність сильного прямого зв'язку між змінними, що і підтверджується обчисленнями.

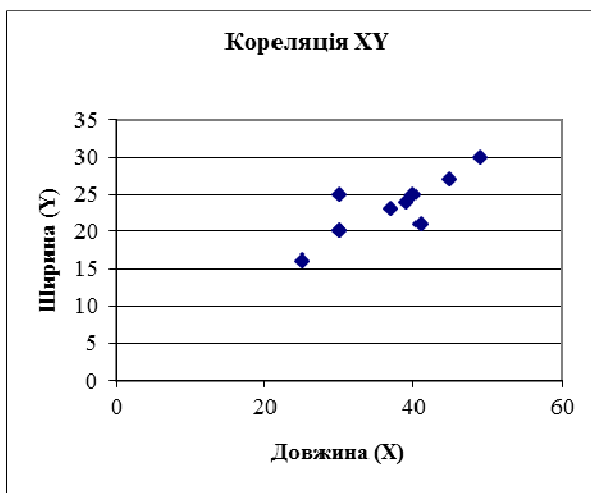


Рис. 3

Порядок обчислення коефіцієнта кореляції за формулою Спірмена для змінних X та Y представлено у наступній таблиці:

	A	B	C	D	E	F	G	H
1		Тест 1 (X)	Тест 2 (Y)	Тест 3 (Z)	Ранг X	Ранг Y	d	d ²
2	учень 1	30	20	18	3	2,5	0,5	0,25
3	учень 2	30	25	18	3	7,5	-4,5	20,25
4	учень 3	30	20	18	3	2,5	0,5	0,25
5	учень 4	37	23	19	5	5	0	0
6	учень 5	49	30	21	10	10	0	0
7	учень 6	41	21	20	8	4	4	16
8	учень 7	25	16	15	1	1	0	0
9	учень 8	45	27	20	9	9	0	0
10	учень 9	40	25	17	7	7,5	-0,5	0,25
11	учень 10	39	24	19	6	6	0	0

Сума значень останнього стовпця $\Sigma d^2=37$. Відповідно $r_s = 0,776$. Це значення менше за коефіцієнт кореляції за Пірсоном, однак воно теж достовірне (за критерієм Стьюдента) і свідчить про значний позитивний зв'язок між змінними, у даному разі – між результатами двох тестів.

Аналогічні результати будуть отримані і у пакетах SPSS або Statistica.

Приклад 2: Виконати регресійний аналіз за даними прикладу 1.

Виконання:

У пакеті MS Excel рівняння регресії можна побудувати за допомогою функції ЛИНЕЙН() та відповідної процедури Пакета Аналіза. Добір відповідної регресійної моделі також можна здійснити на діаграмі за допомогою послуги **Добавить линию тренда**.

Якщо на закладці Параметри вибрати настройки «показывать уравнение на диаграмме» та «поместить на диаграмму величину достоверности аппроксимации (R^2)», то вони будуть виведені на графіку (Рис. 4).

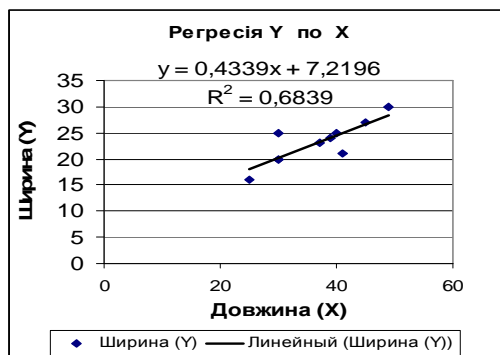


Рис. 4

За допомогою лінії тренда можна апроксимувати експериментальні значення, крім лінійної, також степеневою, експоненційною, поліноміальною або логарифмічною функцією.

За допомогою функції ЛИНЕЙН() можна отримати коефіцієнти регресійної прямої або повну статистику регресійного аналізу. У будь-якому разі дана функція є функцією-масивом (див. примітки).

У першому випадку результат буде розміщено у масиві з двох комірок. Наприклад, C14:D14. Вставимо туди формулу =ЛИНЕЙН(B2:B11; C2:C11; 1; 0). Отримаємо: C14=0,4339 – значення b ; D14=7,2196 – значення a . Тобто формула рівняння регресії буде така: $y=0,4339b+7,2196$.

У функції ЛИНЕЙН (Y; X; Конст.; Стат.) першим параметром є експериментальні значення залежної змінної (один стовбець); другим – експериментальні значення незалежних змінних (може бути декілька змінних, тобто стовпців); третій параметр вказує, чи потрібно, щоб коефіцієнт a дорівнював нулю (для $a=0$ вказують Конст.=0); параметр Стат включає або виключає обчислення повної статистики критерію.

При застосуванні формули =ЛИНЕЙН(B2:B11; C2:C11; 1; 1) отримаємо результат у два стовпці та 5 рядків. У клітинках відповідно розміщені:

Коефіцієнт b	0,4339	Коефіцієнт a	7,2196
Стандартна похибка обчислення b	0,1043	Стандартна похибка обчислення a	3,8911
Коефіцієнт детермінації R^2	0,6839	Стандартна похибка регресійних залишків	2,3927
Обчислене значення F-критерія	17,31	df (степені вільності)	8
Сума квадратів регресії	99,101	Сума квадратів залишків	45,799

При обчисленні множинної регресії результат подається у масив з п'яти рядків та $k+1$ стовпця (де k – кількість незалежних змінних).

Більш детально регресійний аналіз здійснюється за допомогою процедури Регрессия з Пакета Аналіза.

Тема 5: «Порівняння розподілів»

Статистичні методи, в основу яких покладено оперування з параметрами статистичного розподілу, такими, як, наприклад, середнє, дисперсія, та іншими, називають параметричними. Більшість параметричних методів розраховані на дані, розподіл яких близький до нормального. У випадках відхилення від нормальності не всі параметричні методи будуть коректно працювати, тому в таких випадках перевагу слід надавати менш вибагливим до виду розподілу даних непараметричним методам.

Для перевірки нормальності використовують як специфічні методи (окомірний, критерії асиметрії та ексцеса), так і більш універсальні методи, які застосовують також для вирішення інших статистичних задач (критерії згоди, наприклад, χ^2 -Пірсона та λ -Колмогорова-Смирнова).

У теоретичних відомостях до лабораторної роботи розглядають критерії асиметрії та ексцеса (за Н.О.Плохінським та Є.І.Пустильником), та критерій χ^2 -Пірсона.

Завданням лабораторної роботи передбачено засобами статистичних пакетів виконати порівняння емпіричного ряду з деяким теоретичним (рівномірним або нормальним), розрахувавши наперед частоти теоретичного ряду, тобто перевірити гіпотезу про вид розподілу емпіричного ряду, а також порівняти між собою два емпіричні ряди.

Ще одним із завдань є визначення кореляції між якісними ознаками, оскільки в цих задачах використовується критерій χ^2 -Пірсона.

Завдання 1: Оцінка нормальності емпіричного розподілу.

1. За емпіричними вибірками А та В з теми 1 обчислити емпіричні квантілі (0.05, 0.10, ..., 0.95) $\rightarrow P_k = \text{ПЕРСЕНТИЛЬ}$ (Емпіричний масив, К), де $K=0.05, 0.10 \dots 0.95$.
2. Для кожного квантіля обчислити нормалізоване значення $\rightarrow Z_k = \text{НОРМСТОБР}(K)$.
3. Побудувати точкову діаграму (P_k, Z_k).

4. Провести для отриманої сукупності точок лінійну лінію тренда. Візуально оцінити, наскільки графік (P_k, Z_k) відхиляється від прямої.
5. Зробити висновки про нормальність експериментальних вибірок та достовірність окомірного методу оцінки нормальності.
6. Обчислити асиметрію та ексцес для кожної вибірки. Порівняти емпіричний розподіл з нормальним за критичними значеннями асиметрії та ексцесу за критеріями Н.О.Плохінського та Є.І.Пустильника.
7. Порівняти отримані висновки з висновками, зробленими на основі застосування окомірного методу.
8. У пакеті SPSS побудувати квантильні графіки (Graphs → Q-Q Plots) та графіки накопичених частот (Graphs → P-P Plots). Порівняти значення асиметрії та ексцесу, обчислені процедурами описативної статистики пакету SPSS, з їхніми похибками обчислення. Порівняти отримані результати з результатами обчислень та побудов у пакеті MS Excel.
9. Виконати перевірку нормальності емпіричного розподілу за допомогою процедури пакета SPSS Analyze → Descriptive Statistics → Explore: при включенні прапорця Normality plots with tests буде проведено порівняння емпіричного розподілу з нормальним за тестом Колмогорова-Смирнова. Якщо значення похибки (Sig.) більше 0,05, то розподіл можна вважати нормальним.
10. Перевірити, як впливають на результати «теста нормальності» викиди, тобто екстремальні значення. Для цього «профільтрувати» вибірку за допомогою процедури Data → Select Cases та виконати для відфільтрованих даних попередній пункт завдання.
11. Побудувати такі ж графіки у пакеті Statistica:
 - Graphs → 2d Graphs → Normal Probability Plot
 - Quantile-Quantile Plot
 - Probability-Probability Plot.
12. Виконати у пакеті Statistica тест на нормальність емпіричного розподілу, за допомогою пунктів меню Statistica → Basic Statistics/Tables → Frequency Tables. На закладці Normality

діалогового вікна Frequency Tables слід встановити прапорець Kolmogorov-Smirnov test та натиснути кнопку Test for normality.

13. Зробити висновки про нормальність досліджуваних даних. Порівняти результати, отримані засобами різних статистичних пакетів.

Завдання 2: Порівняння емпіричного розподілу з деяким теоретичним.

Завдання 2а: порівняння емпіричного розподілу з теоретичним (нормальним):

1. За даними вибірки розрахувати частоти нормального розподілу.
 - 1.а) використовуючи функцію НОРМРАСП:

	A	B	C	D
1	X	Частота	Нормалізація	Частота нормального розподілу
2	111	3	=НОРМРАСП(A2, середнє, S, 0)	=C2*N*k
3	112	9	=НОРМРАСП(A3, середнє, S, 0)	=C3*N*k

Тут у стовпці А вказано значення X – верхню границю відповідного класового інтервалу.

У стовпці В – обчислені за формулою ЧАСТОТА частоти емпіричного розподілу.

У стовпці С – виконується нормалізація емпіричних значень за параметрами емпіричного розподілу – середнім та стандартним квадратичним відхиленням S (їх слід попередньо обчислити). При значенні четвертого параметра 0 (ХИБНЕ) формується значення вагової функції розподілу (функції щільності розподілу), а при значенні 1 (ІСТИНА) – інтегральної (накопичені відносні частоти). Для обрахунку частот слід обирати вагову функцію, тобто 0. У формулі можна посилатися на адреси комірок, які містять раніше обчислені параметри.

У стовпці D нормалізовані значення частот узгоджують з обсягом емпіричної вибірки, множачи на N – кількість значень в емпіричній вибірці, та k – розмір класового інтервалу. У

наведеному прикладі $k=1$ (різниця між сусідніми значеннями у стовпці змінної X).

1.б) використовуючи функції НОРМАЛІЗАЦІЯ та НОРМСТРАСП:

	A	B	C	D	E	F	G
1	X_i	X_{i+1}	Z_i	Z_{i+1}	$\Phi(Z_i)$	$\Phi(Z_{i+1})$	$(\Phi(Z_{i+1})-\Phi(Z_i))*N$
2	110	111	$-\infty$				
3	111	112					

Тут

X_i – нижня границя класового інтервалу;

X_{i+1} – верхня границя класового інтервалу;

Z_i та Z_{i+1} – нормалізовані значення границь класового інтервалу;

$\Phi(Z_i)$ та $\Phi(Z_{i+1})$ – значення функції Лапласа.

У таблиці $Z_i = \text{НОРМАЛІЗАЦІЯ}(X_i; \text{середнє}; \text{відхилення})$. Найменше Z_i та найбільше Z_{i+1} замінюють відповідно на $-\infty$ та $+\infty$ (тут на деяке велике за абсолютною величиною значення).

У таблиці можна замість функції Лапласа застосувати інтегральну функцію стандартного нормального розподілу

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz \quad (F(z) = \text{НОРМСТРАСП}(z)), \text{ яка пов'язана з}$$

функцією Лапласа рівнянням $F(z) = \frac{1}{2} + \Phi(z)$, оскільки константа не впливатиме на кінцевий результат.

Якщо ж використати інтегральну форму функції – НОРМСТРАСП (X_i ; середнє, S , 1), то можна буде обійтись без нормалізації (не обчислювати Z_i та Z_{i+1}).

Отримані на останньому кроці (множення на N) частоти не рекомендується округлювати, щоб не знижувати достовірність подальших обчислень.

2. На одному графіку побудувати гістограми емпіричного та теоретичного розподілів.

3. Порівняти ряд розподілу частот за емпіричною вибіркою з отриманим теоретичним (нормальним) рядом за допомогою критерія χ^2 Пірсона (використовуйте функцію ХИ2ТЕСТ для порівняння та функцію ХИ2ОБР для отримання критичних значень критерія для визначеного рівня значущості та степенів вільності);
4. Зробити висновки про те, чи можна вважати емпіричний ряд розподілу нормальним.

Завдання 2б: порівняння емпіричного розподілу з теоретичним (рівномірним)

1. Для експериментальних вибірок з теми 1 обчислити теоретичні частоти *рівномірного* розподілу.
2. На одній діаграмі побудувати гістограми емпіричного та теоретичного розподілів частот.
3. За допомогою функцій ХИ2ТЕСТ та ХИ2ОБР перевірити, чи можна емпіричний ряд вважати рівномірним.
4. Зробити висновки про рівномірність емпіричного розподілу.

Завдання 2в: порівняння емпіричного розподілу з теоретичним (рівномірним) в пакеті SPSS

1. Скопіювати значення вибірки В з лабораторної роботи 1.
2. Створити нову змінну.

2.а) За допомогою процедури Transform → Count, значеннями якої будуть номери інтервалів інтервального статистичного ряду частот вибірки В. Для цього у вікні процедури Count вказати назву нової змінної (Target Variable), перенести до списку Numeric Variables досліджувану змінну та визначити границі інтервалів (Define Value), вказавши їх у пункті Range *k* through *L* (тут *k* – розмір інтервалу, а *L* – найбільше включене значення, тобто, наприклад, «10, включаючи 140» означатиме діапазон від 131 до 140) або у пункті Range Lowest through *L* (діапазон включатиме значення, менші за *L*). Всі умови слід додати (Add) до списку Values to Count.

2.б) Або за допомогою процедури Transform → Recode → Into Different Variables. Для цього у вікні процедури Recode вказати

вхідну змінну (Input Variables), назву нової змінної (Output Variable), відкрити вікно Old and New Values, в якому вказати нові значення або назви діапазонів аналогічно до процедури Count. Всі умови слід додати (Add) до списку Old → New.

3. Застосувати до отриманої змінної непараметричний тест χ^2 : Analyze – Nonparametric Tests – Chi-Square (All categories equal).
4. Зробити висновки про рівномірність емпіричного розподілу.
5. За розрахованою змінною побудувати гістограму розподілу: Graphs → Bar Charts → Simple.
6. Порівняти емпіричний розподіл з теоретичними (нормальним та рівномірним) за допомогою процедури Analyze → Nonparametric → 1-Sample K-S (критерій Колмогорова-Смирнова). Порівняти отриманий у даній процедурі результат з тестом нормальності процедури Analyze → Descriptive Statistics → Explore.
7. Зробити висновки про рівномірність та нормальність експериментального розподілу. Порівняти результати з отриманими в MS Excel.

Завдання 2g: порівняння емпіричного розподілу з теоретичним (рівномірним) засобами пакету Statistica:

1. У програмі Statistica вибрати пункт меню Statistica → Distribution Fitting. У діалоговому вікні Distribution Fitting обрати характер розподілу – неперервний.

Вибір неперервного (Continuous Distribution) розподілу дозволить виконати порівняння емпіричного з нормальним (Normal), рівномірним на проміжку (Rectangular), а також іншими (Exponential, Gamma, Log-normal, Chi-square, Weibull, Gompertz). Вибір дискретного розподілу (Discrete Distribution) дозволить виконати порівняння з біноміальним, геометричним, та розподілами Бернуллі і Пуассона.

2. У діалоговому вікні Fitting Continuous Distributions вказати досліджувану змінну (Variable) та тип теоретичного розподілу (Distribution). На закладці Parameters (параметри) вказати

- кількість класових інтервалів для емпіричного розподілу (Numbers of categories), нижню границю (Lower limit) першого та верхню границю (Upper limit) останнього класового інтервалу.
3. На закладці Options необхідно виконати необхідні установки: слід встановити прапорець Combine Categories (об'єднувати категорії) у розділі Chi-Square test, встановити перемикач Yes (continuous) у розділі Kolmogorov-Smirnov test, обрати частотний графік, тобто графік-гістограму (Frequency distribution) та «сирі» частоти (Raw frequencies) у розділі Graph.
 4. Після налаштування слід натиснути кнопки Summary та Plot of observed and expected distribution на закладці Quick.
 5. Проаналізувати отримані результати. Порівняти з отриманими в інших пакетах.

Завдання 3: Порівняння двох та більше емпіричних розподілів.

1. За допомогою генератора випадкових чисел (функції СЛЧИС() або СЛУЧМЕЖДУ(A; B)) сформувати три ряди чисел однакового обсягу (50-80 значень): перший ряд – значення змінної A1 (у діапазоні 50-200 за інтервальною шкалою), другий ряд – значення змінної A2 (0-5 за порядковою шкалою), третій – значення змінної A3 (1-3 за номінативною шкалою). Скласти сюжет експериментального дослідження: придумати назви змінним та категоріям).
2. Скопіювати згенеровані **значення** на Аркуш2: за допомогою послуги ПРАВКА → СПЕЦИАЛЬНАЯ ВСТАВКА вставити лише значення². Сформувати статистичні гіпотези для порівняння емпіричних рядів розподілу.
3. Побудувати розподіли частот за ознакою A2 для рядів, визначених значеннями змінної A3 за допомогою послуги Данные → Сводная таблица.

² Генеровані випадкові числа мають властивість автоматично змінюватися.

У зведеній таблиці значення змінної A2 використати для назв стовпців (розподіли), а значення змінної A3 – для назв рядків (градації ознаки). До даних використати підрахунок кількості (у таблиці не повинно бути порожніх комірок! Якщо таке сталося, відкоректуйте вхідні значення та оновіть зведену таблицю).

4. Порівняти між собою отримані емпіричні розподіли. Зробити висновки про відмінності між ними. Проінтерпретувати отримані результати в термінах сюжету.

У пакеті SPSS

5. Скопіювати значення змінних A1, A2, A3 з Аркуша2.

6. Побудувати таблицю спряженості для змінних A2 (Row) та A3 (Column). Застосувати для цього процедуру Analyze → Descriptive Statistics → Crosstabs. Обчислити критерій χ^2 -Пірсона та коефіцієнт V Крамера (див. Завдання 4), зробивши відповідні настройки у вікні Crosstabs→Statistics.

7. Побудувати діаграми порівнюваних розподілів. Graphs → Bar Charts → Clustered. Визначити як вісь категорій (Category Axis) – змінну A2, а групувати за змінною A3 (Define Clusters by)³. Зробити висновки про відмінності між розподілами. Проінтерпретувати результати у термінах експериментального сюжету.

8. Порівняти з результатом обчислення критерія χ^2 -Пірсона в MS Excel.

Завдання 4: Визначення зв'язку (кореляції) між якісними ознаками.

1. Засобами MS Excel порівняти експериментальні ряди даних, обчислити кореляцію якісних ознак за формулою Крамера та зробити висновки про наявність або відсутність зв'язку між досліджуваними змінними. Проінтерпретувати отримані результати у термінах дослідження.

2. Побудувати діаграми порівнюваних розподілів.

3. У пакеті SPSS провести дослідження за даними файлу GSS93 subset.sav, який інсталується разом з пакетом. Файл

³ Або навпаки.

містить результати опитування громадської думки з низки питань. Зокрема «Чи потрібен закон про необхідність отримання в поліції дозволу на носіння вогнепальної зброї (Так/Ні)» (відповіді представлено у стовпці gunlaw). Дослідити зв'язок між відповідями на це запитання та статтю (sex), сімейним станом (marital) або віросповіданням (relig) опитуваних. Для цього вибрати процедуру Analyze → Descriptive... → Crosstabs та вказати в рядку (Row) змінну gunlaw (тобто залежну), а впливаючі змінні (тобто sex, relig, marital тощо) – у стовпці (Column). У вікні Statistics вибрати Chi-Square та Phi and Kramer's V. У вікні Cells вибрати Counts Observed та Expected (рахувати спостережувані та очікувані значення).

4. Зробити висновки про наявність або відсутність зв'язку між досліджуваними змінними.

Тема 7: «Дискримінантний аналіз»

Багатовимірні методи, одним з яких є дискримінантний аналіз (далі ДкА), представляють собою засновані на аналізі багатьох змінних емпіричні математичні моделі, прийнятні для інтерпретації у термінах досліджуваної предметної галузі. Однак теорію багатовимірних статистичних даних ще до кінця не розроблено, тому деякі з використовуваних багатовимірних методів не мають чіткої статистичної трактовки у плані концепції перевірки гіпотез, побудови довірчих інтервалів тощо.

Багатовимірні методи можна класифікувати за трьома критеріями: призначенням, способом співставлення даних та виглядом початкових емпіричних даних.

Класифікація за призначенням:

1. Методи прогнозування (ДкА та множинний регресійний аналіз) – дозволяють за множиною «відомих» змінних визначити значення «невідомої» змінної. У випадку ДкА «невідомі» змінні номінативні, а у випадку множинного регресійного аналізу – метричні.
2. Методи класифікації (варіанти кластерного аналізу та ДкА).
3. Структурні методи (факторний аналіз та багатовимірне шкалювання).

Класифікація за методом співставлення (початковим припущенням про структуру) даних:

1. Методи засновані на кореляційній моделі даних (припущенні, про узгоджену мінливість ознак у множини об'єктів) – факторний, множинний регресійний аналіз, частково ДкА.
2. Методи, побудовані на дистантній моделі (відмінності між об'єктами розглядаються як відстані між ними) – кластерний аналіз, багатовимірне шкалювання, частково ДкА.

Класифікація за виглядом вхідних даних:

1. Методи, що використовують тільки виміряні ознаки (множинний регресійний аналіз, факторний, ДкА).
2. Методи, вхідними даними для яких можуть бути попарні відмінності (схожість) об'єктів, оцінені групою експертів (кластерний аналіз, багатовимірне шкалювання).

У теоретичних відомостях до лабораторної роботи розглядають призначення та основні припущення дискримінантного аналізу, геометричну інтерпретацію методу, критерії оцінювання якості класифікації та інше.

Завданням лабораторної роботи передбачено виконання дискримінантного аналізу засобами пакетів SPSS або Statistica. У пакетах загального призначення, якими є табличні процесори MS Excel та OpenOffice.org Calc процедури дискримінантного аналізу не реалізовано.

Завдання 1: виконання дискримінантного аналізу засобами пакета SPSS

1. Серед прикладів до програми SPSS вибрати файл, який містить декілька числових та принаймні одну номінативну функцію (наприклад, cars.sav, world95.sav).
2. У головному меню SPSS вибрати пункт Analyze→Classify→Discriminant.
3. Вказати номінативну змінну як змінну, за якою здійснюється групування (Grouping Variable). Зазначити її мінімальне та максимальне можливі значення.
4. Визначити усі числові змінні, для яких перевіряється їхній внесок у прогнозування, як незалежні (Independent Variables).

5. Вибрати покроковий метод аналізу незалежних змінних (Use stepwise method).
6. У діалоговому вікні, яке відкривається кнопкою Method, вибрати критерії оцінки прогнозу Wilk's lambda, та встановити значення для статистик F-включення (Entry) та F-вилучення (Removal).
7. У діалоговому вікні, яке відкривається кнопкою Statistics вибрати методи статистичного аналізу вхідних даних, наприклад, аналіз середніх (Means) та однофакторний дисперсійний аналіз (Univariate ANOVAs), а також включити прапорець Unstandardized Function Coefficients (для виводу коефіцієнтів канонічних функцій).
8. У діалоговому вікні, яке відкривається кнопкою Classify, вказати,
 - що апіорні імовірності слід розраховувати на основі розмірів груп (Prior probabilities: Compute of group size);
 - що показувати слід результати для всіх об'єктів (Display: Casewise results) та вивести підсумкову таблицю (Display: Summary table);
 - що в системі координат канонічних функцій слід виводити графічні зображення усіх об'єктів (Plots: Combined groups). Якщо канонічних функцій дві, то цікавим буде графік Territorial map.
9. За отриманими результатами класифікації та апіорними імовірностями оцінити точність прогнозу.
10. Проаналізувати результати дискримінантного аналізу. Зробити висновки про якість прогнозування значень групуючої змінної від незалежних: яка із незалежних змінних має найбільший вплив на класифікацію? Чи всі обрані змінні мають вплив на групуючу змінну? Наскільки точний отриманий прогноз? Чи можна в подальшому класифікувати об'єкти лише за допомогою обраних змінних? Спробувати проінтерпретувати отримані канонічні функції.

Завдання 2: виконання дискримінантного аналізу засобами пакета Statistica

1. У пакеті Statistica виконати дискримінантний аналіз для тих самих даних: вибрати пункт меню Statistics → Multivariate

Exploratory Techniques → Discriminant Function Analysis. Вказати групуючу (Grouping) та незалежні (Independent) змінні. Вибрати покроковий аналіз (Advanced options (stepwise analysis)).

2. У наступному діалоговому вікні (Model Definition) на закладці Advanced налаштувати параметри дискримінантного аналізу:
 - а) вибрати метод Forward stepwise – при якому на кожному кроці до моделі залучатиметься змінна з найбільшим значенням статистики F-включення (процес буде припинено, коли не знайдеться жодної змінної зі значенням F-включення більшим за порогове).
 - При виборі методу Standart – до моделі буде включено усі незалежні змінні. При методі Backward stepwise змінні зі значенням статистики F-виключення меншим за порогове будуть виключатися з моделі.
 - б) встановити порогове значення толерантності – якщо хоча б одна незалежна змінна матиме значення толерантності менше за порогове, дискримінантний аналіз не проводитиметься.
 - в) встановити порогові значення F-включення та F-виключення (завжди обов'язково $F_{\text{to enter}} > F_{\text{to remove}}$).
 - г) встановити кількість кроків для аналізу (якщо кількість кроків буде мала, то аналіз припиниться раніше, ніж змінні досягнуть визначених у попередньому пункті порогових значень F включення/виключення).
 - д) виводити результати аналізу покрокові (At each step) або тільки підсумки (Summary only).
3. У наступному діалоговому вікні (Discriminant Function Analysis Results) переглянути результати дискримінантного аналізу. Порівняти їх з результатами, отриманими у пакеті SPSS. Зробити висновки.

Тема 8: «Кластерний аналіз»

У теоретичних відомостях до лабораторної роботи розглядають призначення та методи кластерного аналізу, методи визначення міри схожості об'єктів та методи об'єднання кластерів. Завданням лабораторної роботи є виконання ієрархічного

кластерного аналізу та аналізу k-середніх (ітеративного) для перевірки результатів ієрархічної кластеризації.

Завдання 1: Ієрархічний кластерний аналіз

1. Отримати дані з файлу⁴.
2. Виконати ієрархічний кластерний аналіз у пакеті SPSS: Analyze → Classify → Hierarchical Cluster ... У діалоговому вікні кластерного аналізу вказати змінні (Variable(s)), для яких виконуватиметься кластеризація.
3. Встановити параметри кластеризації: у діалоговому вікні Plots... включити прапорець Dendrogram; у діалоговому вікні Method вибрати метод об'єднання у кластери (Cluster Method) – за замовченням встановлено метод Between-group linkage, – та міру відстані між об'єктами (Measure) – за замовченням встановлено Squared Euclidian distance. Зберегти результати кластеризації у вигляді окремої змінної, вказавши у діалоговому вікні Save необхідну кількість кластерів (Single solution).
4. Виконати кластерний аналіз з іншими параметрами: мірою відстані та методом об'єднання кластерів.
5. Зробити висновки, проінтерпретувати отримані результати, оцінити (візуально) достовірність отриманої кластеризації.
6. Виконати ієрархічний кластерний аналіз у пакеті Statistica: Statistics → Multivariate Exploratory Technics → Cluster Analysis → Joining (Tree Clustering). У діалоговому вікні встановити параметри кластеризації: змінні (Variables), метод об'єднання (Amalgamation rule), міру відстані (Distance measure), вид подання вхідних даних⁵ (Input file) та для яких об'єктів здійснювати кластеризацію: для досліджуваних

⁴ Для виконання завдання рекомендується попередньо зібрати відповідний статистичний матеріал. Якщо таких матеріалів немає, можна використати данні з файлів, що додаються до статистичного пакета SPSS, або вибірку C, вважаючи кожен стовпець окремою змінною.

⁵ Дані можуть бути представлені у вигляді рядків (Raw data) або матриці відстаней (Distance Matrix).

об'єктів (Cases (rows)) чи для змінних⁶ (Variables (columns)).

7. У наступному діалоговому вікні на закладці Advanced переглянути результати кластеризації.
8. Порівняти процедури проведення та результати кластерного аналізу, виконаного засобами двох пакетів.

Завдання 2: Ітеративний кластерний аналіз

1. Для даних з попереднього завдання визначити кількість кластерів (наприклад, 2-4).
2. Виконати k-кластерний аналіз засобами пакета SPSS: Analyze – > Classify –> K-Means Cluster Analysis. У діалоговому вікні встановити параметри кластеризації: змінні (Variables) та кількість кластерів (Number of clusters. У допоміжному діалоговому вікні Iterate встановити кількість ітерацій (Maximum Iterations) та включити прапорець Use running means – коригувати середні. Включити всі прапорці у допоміжних діалогових вікнах Save та Options.
3. За даними таблиці Final Cluster Centers побудувати лінійний графік середніх для кожного кластера.
4. Виконати k-кластерний аналіз засобами пакета Statistica (K-means clustering), встановивши у відповідному діалоговому вікні параметри кластеризації: змінні (Variables), об'єкти кластеризації (Cluster), кількість кластерів (Number of clusters), кількість ітерацій (Number of iterations) та спосіб визначення початкових центрів кластерів⁷ (Initial cluster centers).
5. Порівняти процедури проведення та результати кластерного аналізу, виконаного засобами двох пакетів. Зробити висновки.

⁶ Кластеризація змінних можлива лише для даних, поданих у вигляді рядків (Raw data).

⁷ За замовченням встановлено “Sort distances and take observations at constant intervals”.

Література

1. Лупан І.В., Авраменко О.В. Комп'ютерні статистичні пакети: навчально-методичний посібник. – Кіровоград: «КОД», 2010. – 216 с.
2. Лупан І.В. Програмне забезпечення курсу «Комп'ютерні статистичні пакети». // Матеріали міжнародної науково-практичної конференції «Освітні вимірювання в інформаційному суспільстві». – К.: НПУ, 2010. – С. 56.
3. Лісова Т.В. Програмні засоби для аналізу результатів тестування.// Наукові записки НДУ ім. М.Гоголя. Психолого-педагогічні науки. – 2011. – №7. – С.46-52.
4. Лупан І.В. Дослідження дискримінуючої здатності результатів зовнішнього незалежного оцінювання. //Науковий часопис національного педагогічного університету імені М.П.Драгоманова. Серія №5. Педагогічні науки: реалії та перспективи. – Випуск 20: збірник наукових праць. – К.: Вид-во НПУ ім. М.П.Драгоманова, 2009. – С.87-94.
5. Лупан І.В., Халецька З.П., Чеча В.О. Використання канонічного кореляційного аналізу у педагогічних дослідженнях. //Наукові записки НДУ ім. М.Гоголя. Психолого-педагогічні науки. – 2011. – №10. – С.63-68.

РОЗРОБКА КОМП'ЮТЕРНИХ МОДЕЛЕЙ ПЕДАГОГІЧНОГО ТЕСТУВАННЯ – СКЛАДОВА ПІДГОТОВКИ МАГІСТРІВ ОСВІТНІХ ВИМІРЮВАНЬ

Паращук С.Д.

Сучасний стан розвитку інформаційно-комп'ютерних технологій сприяє впровадженню в навчальних закладах систем автоматизованої діагностики, зокрема комп'ютерного педагогічного тестування. Такі системи дозволяють значно підвищити ефективність функцій контролю навчальних досягнень. Для організації педагогічного тестування вищі навчальні заклади можуть використовуватися власні автоматизовані системи контролю або програмні засоби, які пропонуються на ринку програмних продуктів. Системи тестування, що є на ринку, зазвичай різної якості та призначення. Серед вільно розповсюджуваних є системи, що дозволяють організувати педагогічне тестування на певному рівні, але їх зазвичай потрібно доробляти, щоб врахувати специфіку організації контролю в конкретному навчальному закладі.

Фахова підготовка магістрів освітніх вимірювань передбачає формування у них умінь використовувати комп'ютерні технології в педагогічному тестуванні. Зокрема, магістри вивчають можливості окремих існуючих комп'ютерних систем тестування та отримують практичні навички їх використання для організації педагогічного контролю. Зауважимо, що магістри навчаються працювати з такими системами як користувачі, тобто на користувачькому рівні, а не рівні програміста. Це означає, що вони не повинні мати фахову підготовку з інформатики для здобуття освітньо-кваліфікаційного рівня магістра.

Існуючі на ринку комп'ютерні системи тестування мають різні можливості, але всі вони підтримують можливість створення та редагування тестів і організувати сам процес тестування. Деякі з них також можуть автоматично виконувати статистичний аналіз результатів тестування. Зазвичай, в комп'ютерних системах

передбачені певні сценарії проведення тестування, які певним чином пов'язані з моделями педагогічного тестування. Вже на етапі планування тесту розробник тесту повинен передбачати модель тестування. Це означає, що магістри освітніх вимірювань повинні володіти знаннями про існуючі моделі та використовувати їх на практиці. Адже від вибору моделі суттєво залежить процес проведення комп'ютерного тестування та й вибір самої системи тестування.

На сьогодні існують декілька класифікацій моделей педагогічного тестування. Метою даної статті визначення змісту навчального матеріалу про моделі комп'ютерного педагогічного тестування як складової підготовки магістрів освітніх вимірювань.

У статті Е.Г. Матвієвскої [1] дається наступне визначення «*модель педагогічного тестування* – це схема пред'явлення тестових завдань і оцінювання результатів тестування». З даного визначення випливає, що модель визначається двома складовими – алгоритмом подання тестових завдань випробовуваному та методами оцінювання цих завдань. У цій же статті запропонована така критеріально-рівнева типологія моделей: класична, класична з урахуванням складності, за зростанням складності, з розділенням знань за рівнями засвоювання, з урахуванням часу відповіді, з обмеженням часу, адаптивна, на нечіткій математиці.

Л.В. Зайцева, Н.О. Прокоф'єва виділяють такі «методи організації контролю знань» [2]:

- 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 відповідей студента ... за заздалегідь розробленим сценарієм ... указують наступне запитання залежно від правильності відповіді...», «контроль на основі моделі 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»).

Також у статті О.Г. 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) [2].

В основі наведених вище класифікацій лежать алгоритми подання тестових завдань випробовуваному. Такі алгоритми в сучасних умовах можна реалізувати практично лише при комп'ютерному тестуванні. Окрім того, доцільно ввести поняття комп'ютерної моделі педагогічного тестування, яке пов'язане з можливостями сучасних систем 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 може бути 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.

Як відмічалось, 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. Таким чином першоосновою для класифікації моделей комп'ютерного педагогічного тестування повинна виступати структурна організація бази тестових завдань.

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. Окрім того, ці моделі не охоплюють всі можливі варіанти організації структури бази завдань, але практично дозволяють реалізувати різні методи діагностики, які відмічалися вище.

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, отримують 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.

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 отримують 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 p , і знову випадковим чином вибираємо s завдань із $(p+1)$ -ої групи. Якщо студент завжди дає правильні відповіді на вибрані завдання, продовжуємо обхід з кроком p 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 p 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 $(h-p)$ -ої групи. Тоді

повертаємося до $(h-p+1)$ -ої групи завдань i , починаючи з неї, працюємо за алгоритмом 1. Зараховуємо студенту правильні відповіді на $s \cdot (h-p)$ завдань із груп $1, \dots, h-p$, а відповіді на решту завдань отримаємо 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 nom 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, внісши 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.

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 і, 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 g і f залежні, тобто з g проведена дуга до f . Припустимо, що під час тестування студент відповів невірно на завдання g і вірно на f . Оскільки f залежить від g , то оцінка завдання f повинна залежати не тільки від його ваги, а також від відповіді на g . 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 $1, 2, \dots, k$. 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 реалізувати елементи адаптивного контролю.

Висновок. Запропоновані три види моделей комп'ютерного педагогічного тестування не охоплюють всіх можливих варіантів їх реалізації в системах тестування. Однак їх знання та розуміння дозволять магістру з освітніх вимірювань розробляти різні варіанти тестів та по-різному організувати тестування. Зазвичай вибір моделі повинен відбуватися вже на етапі планування тесту. Знання моделі дозволить також правильно вибрати саму комп'ютерну систему тестування. Деякі з магістрів можуть мати фахову освіту з інформатики, а значить в майбутньому також можуть стати розробниками власної системи тестування.

Література

1. Матвиевская Е.Г. Педагогические измерения. Опыт и проблемы исследования [Електронний ресурс] // CREDO NEW теоретический журнал. – <http://credonew.ru/content/view/769/33/>.
2. Колгатін О.Г. Автоматизована педагогічна діагностика у сучасному університеті [Електронний ресурс]. – <http://www.ime.edu-ua.net/em8/content/08kogumi.htm>.

ЯК ВИБРАТИ КОМП'ЮТЕРНУ СИСТЕМУ ТЕСТУВАННЯ ЗНАНЬ

Фетісов В.С.

Поява комп'ютерів спричинила появу нових технологій у різних галузях наукової і практичної діяльності людини. Однією з них стала освіта, яка сама по собі є потужною інформаційною сферою, що володіє величезним досвідом використання різноманітних інформаційних систем. Тому, звичайно, що освіта швидко відгукнулася на можливості, які з'явилися завдяки появи комп'ютерів. Зараз неможливо уявити собі будь-яку сферу освіти, де не використовуються комп'ютерні технології, у складі яких найважливіше місце займають інформаційні навчальні системи. У свою чергу, одним з найбільш поширених напрямків використання таких систем є такі, що виконують одночасно функцію перевірки засвоєння і практичного застосування студентами і учнями одержаних під час навчального процесу знань, які називають **комп'ютерними системами тестування знань (КСТЗ)**.

На ринку програмного забезпечення пропонується велика кількість КСТЗ, які розрізняються за самими різними критеріями. І це дуже добре, оскільки тим, хто буде працювати з ними, є з чого вибирати. Проте при цьому виникає проблема вибору: яка ж система, власне кажучи, нам потрібна і чим саме слід керуватися під час її вибору? Для цього, на думку автора, слід керуватися певними вимогами, які і пропонуються у статті.

Ці вимоги відбивають точку зору автора, який у минулому був розробником КСТЗ.

Загальні вимоги до КСТЗ

КСТЗ повинна виконувати свою найважливішу функцію – *забезпечення ефективного тестування знань*. Але ефективне тестування знань складається з багатьох складових. Одні з них стосуються КСТЗ у цілому, інші – її конкретних модулів. При

цьому – звичайно – до КСТЗ у конкретних умовах будуть висуватися конкретні вимоги. Але серед них можна виокремити кілька основних вимог, яким практично завжди має відповідати КСТЗ:

1. *Універсальність*. КСТЗ має використовуватися як для внутрішнього, так і для зовнішнього тестування. Можливість використання КСТЗ для внутрішнього і зовнішнього тестування досягається гнучкою побудовою системи, що передбачає розподіл її користувачів за ролями, підтримкою різних систем і шкал оцінювання тощо. Наприклад, для користувача з роллю «свій студент» слід вести систему оцінок, яка відповідає навчальному плану навчального закладу, а для зовнішнього користувача можна застосувати іншу, наприклад, накопичувальну систему оцінювання успішності тестування.

2. *Підтримка безпечного і універсального механізму керування користувачами з розподілом прав доступу*. Це питання тісно пов'язане з функцією універсальності. КСТЗ повинна дозволяти мати користувачів з різними правами, що забезпечують доступ до різних функцій і складових системи. Наприклад, можна виділити такі типи користувачів: «свій» студент, студент, який не є студентом навчального закладу, де використовується система, викладач, розробник тестів, адміністратор.

3. *Захищеність інформаційної бази тестів і результатів тестування*, доступ до яких повинні мати тільки користувачі з відповідними правами.

4. *Шифрування* інформаційного наповнення тестів, принаймні це стосується відповідей. Останні три вимоги складають те, що можна назвати *інформаційною безпекою* тестів.

5. *Підтримка основних форм тестових завдань* (завдання із простим і складним множинним вибором, відкритої форми, завдання на встановлення відповідності та завдання на відновлення правильної послідовності).

6. *Підтримка двох типів проведення тестування*. За своїм функціональним призначенням тести можуть бути двох типів: для проведення соціологічних або психологічних тестів і для проведення тестів знань (контрольних тестів). За першим

варіантом особа, яка тестується, відповідає на *всі* завдання тесту. За результатами тесту за заздалегідь визначеними правилами підраховується сума балів, який відповідає певний висновок. За другим варіантом особа відповідає на *заздалегідь визначену кількість випадково відібраних контрольних завдань у тестовій формі* для визначених при налаштуванні розділів, тем.

7. Підтримка різних режимів тестування:

8. Екзаменаційний режим. У цьому режимі за невірні відповіді віднімаються бали, а окремі контрольні завдання дозволяється пропускати, за що бали не додаються і не віднімаються.

9. Режим самотестування або тренажера. Під час тестування виводяться повідомлення про помилки або рецензії на відповідь.

10. Режим навчання. Може містити вступ і пояснення до завдання, різноманітні навчальні матеріали у текстовому або мультимедійному вигляді.

11. Можливість імпорту тестів у текстовий формат і формат HTML, а також експорту тестів з електронних текстових версій паперових тестів або з інших КСТЗ.

12. Наявність української (російської) локалізації.

13. Наявність докладної документації українською (російською) мовою.

14. Технічна *підтримка* КСТЗ.

Крім цих основних вимог, до конкретних КСТЗ можуть бути висунуті також такі.

1. Вартість самої системи, якщо вона є пропрієтарною. Одночасно, слід враховувати, що потужні КСТЗ можуть використовувати додаткові програмні засоби, наприклад, платні СУБД (Microsoft SQL Server), тому – за їх наявності – слід врахувати вартість придбання відповідного програмного забезпечення.

2. Якщо КСТЗ планується використовувати на національному або міждержавному рівнях, то вона має відповідати державним і міжнародним стандартам, наприклад, для другого варіанту відповідати стандарту IMS QTI (Information Model Specification Question & Test Interoperability). У цьому випадку

виникає питання підтримки КСТЗ різних операційних систем (багатолатформеність), оскільки зовсім не є фактом, що всі користувачі будуть використовувати ту саму операційну систему. Більше того, на Україні, як і у всьому світі, спостерігається стійка тенденція до розширення використання UNIX-подібних операційних систем, тобто вільного програмного забезпечення.

3. КСТЗ, які мають за мету отримати широке розповсюдження у навчальному процесі, мають підтримувати стандарт збереження освітнього контенту SCORM. SCORM (Sharable Object Reference Model) – це набір стандартів і специфікацій, що був розроблений для дистанційного навчання. Він, зокрема, містить вимоги до організації навчального матеріалу. Згідно концепції SCORM навчальний матеріал має подаватися відносно невеликими блоками, які можуть бути використані у різних навчальних курсах і використовуватися у системах дистанційного навчання незалежно від того, у межах якої системи вони були створені. Отже SCORM дозволяє забезпечити сумісність компонентів курсів і можливість їх багаторазового використання.

4. Не є обов'язковою вимога відсутності прив'язування до конкретної предметної галузі знань. Але інколи як раз потрібна саме спеціалізована оболонка для складання тестових завдань за допомогою якої можна, наприклад, здійснювати молекулярне моделювання у хімії, будувати електричні схеми у фізиці і та ін.

5. Можливість створення за допомогою КСТЗ *паралельних завдань*. Для цього застосовуються так звані *фасети*. Фасет – це форма запису кількох варіантів того самого завдання.

6. Окремим питанням є здатність КСТЗ відповідати вимогам *адаптивного тестування*. Такого роду тестування враховує індивідуальні здібності особи, яка тестується, шляхом аналізу відповідей на тестові завдання. Якщо відповідь правильна, то вважається, що рівень підготовки особи, яка тестується, вищий за складність тестового завдання і вона здатна вирішувати завдання не тільки цієї складності, а ще й вище, в іншому випадку – ні. Адаптивне тестування враховує індивідуальні особливості і відповідно підвищує або зменшує рівень складності подальших

тестових завдань. В результаті особа, яка проходить тестування, самостійно встановлює для себе певний рівень складності завдань. Звичайно, що й оцінювання завдання більш високої складності відповідно має оцінюватися більш високими балами.

У свою чергу конкретні вимоги можна застосувати і до кожного модуля КСТЗ.

Вимоги до модуля створення, підготовки і редагування тестових завдань

1. Створення необмеженої кількості тестів, тем, завдань і варіантів відповідей на них.

2. Максимально просте і зручне формування та модифікація наповнення тестів, у тому числі керування тестовими завданнями: зміна їх порядку, вставка, переміщення, вилучення і копіювання, а також пошук і відбір (фільтрування) їх за розділом, темою, завданням.

3. Перевірка цілісності створюваного тесту, тобто наявність засобів контролю відповідності тексту завдань і варіантів відповідей, коректність посилань на ресурси і т. ін.

4. Здійснення орфографічного контролю.

5. Наявність базових функцій з форматування тексту завдань і варіантів відповідей, у тому числі вибір накреслення, розмірів, шрифтів і кольору; використання таких видозмін як верхні і нижні індекси, закреслення і застосовування їх на рівні слів, речень, окремих завдань, відповідей; створення списків.

6. Визначення ваги тестового завдання.

7. Формування індивідуальної інструкції з виконання для кожного тестового завдання.

8. Можливість додавання мультимедійних об'єктів, у першу чергу графічних зображень, а також відео та аудіо ресурсів. Зрозуміло, наприклад, що графічні зображення є необхідною складовою тестів в хімії, біології, фізики тощо. Крім цього, у деяких випадках виникає потреба додавання OLE-об'єктів (діаграми, креслення, таблиці, формули).

9. Можливість формування додаткових форм тестових

завдань.

10. Якщо планується використовувати КСТЗ у режимі навчання, то доцільно мати у складі КСТЗ можливість додавання текстових файлів, зміст яких є навчальною складовою.

11. Друк паперового варіанту тесту.

Зручною можливістю модуля є наявність його *portable* версії, тобто такої програми, що не вимагає попередньої інсталяції і може бути завантажена з будь-якого носія, в тому числі змінного. Це дозволяє розробнику тестів використовувати будь-який комп'ютер для формування тесту, а особі, яка проводить тестування – для проведення тестування без необхідності попередньої інсталяції програми.

Доцільно надати розробнику тесту мати можливість формувати:

1. загальний заголовок, стислу анотацію, стислу інформацію щодо автора (розробника тесту);
2. загальну інструкцію з виконання;
3. нотатки (будь-які зауваження до тесту розробника, що бачить тільки він).

Настроювання модуля має містити таке:

1. Можливість розташування завдань і варіантів відповідей на них як випадковим чином, так і строго послідовно.
2. Можливість застосування різних *шкाल оцінювання*. Оптимальним варіантом є можливість застосування будь-якої шкали, але для конкретного навчального закладу може бути достатнім наявність тих шкал, які застосовують в ньому застосовують, наприклад, 5-бальну систему оцінювання.
3. Обмеження кількості завдань за темами.

Вимоги до модуля проведення тестових завдань

1. Головною вимогою до модуля тестування є *ергономічність*. Інтерфейс оболонки для проходження тестування має бути максимально зручним. Особа, яка проходить тестування, має витратити мінімум часу на вивчення інтерфейсу і пошук потрібних для відповіді елементів (наприклад, полів введення).

Саме вікно модуля повинне містити тільки потрібні для тестування елементи. Оптимальним варіантом є можливість настроювання інтерфейсу «під себе» шляхом вилучення службових елементів вікна (панелі інструментів, допоміжної текстової інформації, годинника, що відображає час проходження тестування).

2. Захист тесту на рівні завдання пароля для обмеження доступу до загальних настроювань модуля.

3. Вибір організації тестування: у локальному режимі – на кожному комп'ютері окремо або в Інтернеті.

4. За умови організації тестування у локальному режимі зручно мати portable-версію модуля, що дозволяє провести тестування на будь-яких комп'ютерах, не витрачаючи часу на інсталяцію КСТЗ.

5. Експорт результатів тестування у вигляді текстового файлу для подальшого аналізу, імпорт в електронні таблиці тощо.

6. Якщо за закінченням тестування передбачається публікація його результатів в Інтернеті, то модуль має містити можливість експорту результатів тестування у форматі HTML.

Найважливіше значення мають питання, пов'язані з аналізом результатів тестування:

1. Збирання, збереження і подання у зручній формі різноманітної статистичної інформації результатів тестування. Найкращим варіантом є наявність засобів (так званих «конструкторів форм»), за допомогою яких можна самостійно будувати потрібні звіти результатів тестування.

2. Інтелектуальний аналіз тестів з метою оцінки якості, придатності окремих завдань у тестовій формі, виявлення некоректно сформульованих або неправильних завдань.

Модуль повинен також надавати користувачеві здійснювати низку настроювань, серед яких найбільш важливими можна назвати такі:

1. Вибір режиму тестування.

2. Завдання кількості тестових завдань.

3. Встановлення обмеження за часом.

4. Можливість вибору варіанту формування завдань у тестовій формі та (або) відповідей: у випадковому порядку або

послідовно відповідно до порядку їх розташування у тесті.

5. Можливість завдання способу відбору завдань з тестової вибірки, що формується з кількох тем. Наприклад, повинна бути можливість відбору до вибірки завдань у тестовій формі пропорційно їх кількості за темами.

6. Можливість врахування рівня складності під час формування тестової вибірки.

7. Великий відсоток осіб з погіршеним зором потребує наявності оперативної зміни масштабу відображення інформації.

Можна навести також низку другорядних настроювань. Наприклад, дехто з осіб, які проводять тестування, мають бажання чути звуковий сигнал за закінченням тестування, деякі особи, які проходять тестування, бажають бачити час тестування. А під час аналізу результатів тестування доцільно надати можливість користувачеві самостійно визначати кольорову інтерпретацію правильних, частково правильних і помилкових відповідей.

Вимоги до модуля адміністрування

1. Можливість контролювати процес тестування у реальному режимі часу.

2. Захист результатів тестування.

3. Можливість призупинення тестування, запам'ятавши його результати, і продовжити через певний час з використанням збережених даних.

4. За умовою організації тестування в Інтернеті слід визначитися, чи є необхідність у виділеному сервері.

5. Простота надбудови, масштабування системи.

6. Якщо передбачається розташувати тести на серверах в Інтернеті, то подальшого поглиблення набувають питання, пов'язані з інформаційною безпекою тестів, зокрема, можливість шифрування тестів безпечними алгоритмами шифрування.

Зрозуміло, що перелік цих вимог можна розширити, оскільки вони відбивають особисту точку зору автора, але знайомство з ними беззаперечно буде корисним для користувачів комп'ютерних систем тестування знань.

Література

1. Фетісов В.С. Основні вимоги до комп'ютерних систем тестування знань (КСТЗ) // Наукові записки. Серія «Психолого-педагогічні науки» / Фетісов В.С.; за заг. ред. проф. Є. І. Коваленко – Ніжин: Видавництво НДУ ім. Гоголя, 2011, №10, с. 101-105.
2. Фетісов В.С. Основные требования к компьютерным системам тестирования знаний (КСТЗ). – Педагогические измерения, 2011, № 3, с. 39-48.

Додатки

Додаток І. Програма вступного фахового випробування для вступників за освітньо-кваліфікаційним рівнем «Магістр» спеціальності 8.18010022 «Освітні вимірювання» у НПУ імені М.П. Драгоманова

Пояснювальна записка

Програма призначена для підготовки абітурієнтів до складання вступного фахового випробування в Інститут інформатики Національного педагогічного університету імені М.П. Драгоманова на спеціальність «Освітні вимірювання» освітньо-кваліфікаційного рівня «магістр».

Підготовка магістра з освітніх вимірювань у вищому педагогічному навчальному закладі базується на прогностичній характеристиці педагога як суб'єкта культури, який персоніфікує цілісний соціокультурний досвід людства. Фундаментальними засадами навчання з освітніх вимірювань студентів виступають: інтегральність, аксіологічність, феноменологічно-герменевтичний аспект освіти, полікультурність, толерантність, *педагогічне спрямування*. Це дозволяє забезпечити адекватність сучасної підготовки магістра меті, завданням і принципам професійної діяльності фахівців з освітніх вимірювань.

Характерна особливість педагогічної системи підготовки фахівців зі спеціальності «Освітні вимірювання» в НПУ імені М.П. Драгоманова полягає в забезпеченні професійної підготовки фахівця, чия професійна діяльність розгортається в сфері освіти та науки, яка є актуальною на сучасному ринку праці.

Реалізація зазначених завдань й особливостей навчання студентів за спеціальністю «Освітні вимірювання» у Національному педагогічному університеті імені М.П. Драгоманова забезпечить їх готовність до фахового виконання професійних функцій. Магістри з освітніх вимірювань отримують

Додаток I

поглиблену теоретичну підготовку в галузі педагогічного оцінювання, тестування та моніторингу якості освіти. Це необхідно для забезпечення реальних потреб розвитку сучасної освіти України. Реформаційним процесам ще бракує кваліфікованих фахівців центрів тестування та моніторингу якості освіти, фахівців в галузі освітніх вимірювань, здатних до самостійної творчої роботи в освітній сфері. Відсутність в Україні підготовки магістрів за спеціальністю «Освітні вимірювання» з врахуванням поширення поля їхньої діяльності на навчальний процес у вищій та середній школі, адміністративні функції у закладах середнього рівня, забезпечення діяльності освітніх організацій та установ визначають гостру потребу у підготовці керівників центрів моніторингу якості освіти, методистів.

Магістри цього профілю готуватимуться до науково-викладацької, просвітницької та методичної діяльності у освітній та науковій сферах. Вони можуть працювати викладачами дисциплін циклу «Освітні вимірювання» у вищих навчальних закладах 3 і 4 рівня акредитації, а також керівниками та фахівцями центрів моніторингу якості освіти як окремих навчальних закладів, так і регіональних та всеукраїнських центрів.

Національний педагогічний університет імені М.П.Драгоманова, виходячи з потреб державного та регіонального ринків праці, планує комплексну підготовку керівників центрів моніторингу якості освіти та фахівців-методистів не лише для педагогічної роботи, але й для професійної діяльності в управлінській та моніторингових сферах за двома циклами.

Підготовка магістра здійснюється відповідно з навчальним планом, який сприяє наданню достатнього обсягу знань за спеціальністю «Освітні вимірювання».

ЗМІСТ ПРОГРАМИ

Модуль 1. Педагогіка

Змістовий модуль 1. Основи педагогіки

Педагогіка в системі наук про людину. Методологія і методи педагогічних досліджень. Розвиток, соціалізація і виховання особистості. Педагогічний процес як система і цілісне явище. Закономірності і принципи педагогічного процесу.

Змістовий модуль 2. Теорія освіти і навчання (дидактика)

Предмет і основні категорії дидактики. Процес навчання, його структура. Різні підходи до навчання, їх характеристика. Закономірності та принципи навчання. Зміст освіти як фундамент базової культури особистості. Методи і засоби навчання та активізації пізнавальної діяльності учнів. Організаційні форми навчання. Урок в сучасній школі. Контроль, оцінювання та облік успішності учнів. Засоби навчання.

Змістовий модуль 3. Управління педагогічними системами.

Сутність і загальні принципи управління педагогічними системами. Основні функції управління навчальним процесом. Взаємодія соціальних інститутів в управлінні педагогічною системою. Підвищення кваліфікації і атестація педагогічних працівників.

Модуль 2. Психологія

Змістовий модуль 1. Психологія навчання

Психологічне поняття про навчання. Навчання і розвиток. Психологічні механізми навчання. Психологічна характеристика форм навчання (традиційної, проблемної, програмованої, розвиваючої та ін.). Психологічні та психофізіологічні вимоги до організації навчального процесу.

Поняття про навчання. Учіння як процес і як діяльність. Структура навчальної діяльності. Мотиваційний компонент учіння. Психологічні особливості формування навчальних умінь та навичок.

Навчальність як важлива характеристика суб'єктів учбової діяльності. Віковий та індивідуальний підхід у навчанні. Вміння вчитися самостійно як інтегральна якість особистості. Психологічні чинники успішності в учінні.

Змістовий модуль 2. Психологія педагогічної діяльності

Загальна характеристика педагогічної діяльності. Індивідуальний стиль діяльності педагога. Психологія педагогічного оцінювання. Психологічний аналіз уроку. Психологія педагогічних здібностей та умінь.

Виховна функція педагогічного спілкування. Психологія педагогічного спілкування. Психологічна характеристика педагогічного такту. Вплив Я-концепції педагога на його професійне спілкування.

Модуль 3. Педагогічні комунікації

Основні риси, функції сучасного вчителя. Педагогічна майстерність, її елементи. Сутність, функції педагогічної комунікації.

Складові невербальної комунікації. Основні засоби невербальної комунікації.

Мовне спілкування як факт комунікації. Функції мовлення вчителя.

Особливості, функції педагогічного спілкування. Педагогічне спілкування як діалог. Бар'єри спілкування. Ускладнення, що виникають у процесі спілкування. Структура педагогічного спілкування. Стили педагогічного спілкування. Установка вчителя. Уміння слухати. Правила педагогічного спілкування.

Взаємозв'язок між членами колективу в спільній діяльності. Соціально-психологічний клімат у педагогічному колективі.

Конфлікти як засіб регулювання міжособистісних стосунків. Конфлікти у школі. Педагогічний конфлікт: структура, сфера, динаміка. Конфлікт у взаємодії «вчитель-учні». Міжособистісні конфлікти у педагогічному колективі.

Конфлікт у педагогічній взаємодії. Поведінка вчителя у конфліктній ситуації. Методи вирішення конфліктів.

Модуль 4. Основи інформатики

Змістовий модуль 1. Системи опрацювання текстів

Системи опрацювання текстів, їх класифікація та функції. Завантаження текстового редактора. Призначення та система вказівок текстового редактора. Введення тексту з клавіатури. Редагування тексту. Перевірка орфографії. Робота з текстовими файлами. Використання буферу обміну. Пошук інформації в середовищі текстового редактора. Робота з контекстами – пошук, заміна, виділення, перенесення, копіювання, форматування. Робота з об'єктами в середовищі текстового редактора. Робота із шрифтами. Форматування документу. Друкування тексту. Шаблони текстів та робота з ними. Використання таблиць в текстах. Структура документа.

Змістовий модуль 2. Електронні таблиці

Електронні таблиці (ЕТ) та їх призначення. Введення й редагування числової, формульної та текстової інформації. Робота

з файлами ЕТ. Координати комірок. Діапазон комірок. Опрацювання табличної інформації: копіювання, редагування, вилучення, переміщення, форматування. Пошук інформації в середовищі ЕТ. Виведення табличної інформації на друк. Використання функцій та операцій для опрацювання інформації, поданої в таблиці. Ділова графіка. Побудова діаграм і графіків на основі табличної інформації. Використання логічних функцій для опрацювання табличної інформації.

Створення в електронній таблиці бази даних, впорядкування та пошук потрібної інформації в середовищі ЕТ. Фільтрування даних. Об'єднання даних. Аналіз даних в середовищі ЕТ.

Змістовий модуль 3. Мережа Інтернет.

Глобальна мережа Інтернет. Інформаційний зв'язок в Інтернет. Сервер та робоча станція. Апаратні, програмні та інформаційні складові сучасної мережі Інтернет. Поняття про ресурси Інтернет. Ідентифікація комп'ютерів в мережі. Адресація в Інтернет. Протоколи передавання інформації. Принципи функціонування глобальної мережі. Провайдери та інформація, необхідна для під'єднання до мережі Інтернет. Різні способи під'єднання комп'ютерів до глобальної мережі. Основні послуги глобальної мережі Інтернет: гіпертекстові сторінки, електронна пошта, телеконференції, файлові архіви, інтерактивне спілкування. Види та основні можливості програмного забезпечення для роботи в глобальній мережі Інтернет.

Електронна пошта. Принципи функціонування. Поштові стандарти. Електронна адреса. Основні можливості поштових програм для роботи з електронними повідомленнями: створення електронного повідомлення, відправлення, відправлення копій, приєднання файлів до повідомлень, одержання повідомлення, створення відповіді на електронне повідомлення тощо. Адресна книга.

Поняття про телеконференції. Програмне забезпечення та конфігурація телеконференцій. Різні групи телеконференцій. Правила організації і проведення телеконференцій. Пошук потрібних груп новин та підписка на них. Відправлення повідомлень в телеконференції.

Поняття про гіпертекстовий документ. Служба перегляду гіпертекстових сторінок – World Wide Web (WWW – всесвітня

Додаток I

павутина). Організація інформації, принципи навігації. WWW-сервери. Адреса Web-сторінки. Перегляд та створення Web-сторінки. Пошукові машини. Пошук інформації в Інтернет.

Програми-браузери. Правила роботи та налагодження. Використання різноманітних сторінок кодування. Файлові ресурси в мережі Інтернет. Інтерактивне спілкування в Інтернет.

Модуль 5. Математика

Змістовий модуль 1. Функції

Види функцій: лінійна, квадратична, показникові, степенева. Способи задання функції. Побудова графіків функцій та їх аналіз.

Змістовий модуль 2. Елементи теорії ймовірностей та математичної статистики

Основні поняття теорії ймовірностей. Подія і ймовірність. Відносна частота події. Ймовірність події. Операції над подіями. Ймовірності суми та добутку подій. Умовна ймовірність та незалежність подій.

Елементи комбінаторики. Перестановки, розміщення, комбінації (без повторень). Біном Ньютона.

Статистика та її методи. Статистичні таблиці. Ряди розподілу. Наочне подання статистичного матеріалу.

Середні значення. Мода, медіана, математичне сподівання, дисперсія.

КРИТЕРІЇ

оцінювання відповідей абітурієнтів Інституту інформатики на вступному фаховому випробуванні з «Освітніх вимірювань»

<i>Рівень</i>	<i>Кількіс на харак терис тика рівня</i>	<i>Характеристика відповідей абітурієнта</i>	
		<i>на питання теоретичного змісту</i>	<i>на питання практичного змісту</i>
Низь-кий	100-123 бали	Абітурієнт не усвідомлює змісту питання білету, тому його відповідь не має безпосереднього відношення до поставленого питання. Наявна повна відсутність уміння міркувати.	У абітурієнта недостатньо розвинуті логічні та аналітичні уміння, уміння узагальнення.

Задовільний	124 - 149 балів	Відповіді на питання білету носять фрагментарний характер, характеризуються відтворенням знань на рівні запам'ятовування. Абітурієнт поверхово володіє умінням міркувати, його відповіді супроводжуються другорядними міркуваннями, які інколи не мають безпосереднього відношення до змісту запитання	Абітурієнт проаналізував причини виникнення ситуації, однак не зміг чітко сформулювати методи її вирішення.
Достатній	150-169 балів	У відповідях на питання білету допускаються деякі неточності або помилки непринципового характеру. Абітурієнт демонструє розуміння навчального матеріалу на рівні аналізу властивостей. Помітне прагнення абітурієнта логічно розмірковувати при відповіді на питання білета.	Питання розгорнуто в цілому правильно, але наявні окремі неточності у визначенні методів розв'язання проблемної ситуації.
Високий	170-200 балів	Абітурієнт дає повну і розгорнуту відповідь на питання білету. Його відповіді свідчать про розуміння навчального матеріалу на рівні аналізу закономірностей, характеризуються логічністю і послідовністю суджень, без включення випадкових і випадання істотних з них.	Завдання вирішено правильно, відповідь розгорнута. Абітурієнт показав розвинуті уміння аналізувати, обґрунтовувати, узагальнювати

**Додаток II. Програма виробничої практики
магістрів спеціальності 8.18010022
«Освітні вимірювання» у НПУ імені
М.П. Драгоманова**

Укладачі програми:

доктор педагогічних наук, професор Сергієнко В.П.,
кандидат педагогічних наук, доцент Тополя Л.В.

Вступ

Виробнича практика студентів передбачена державним освітнім стандартом і є невід'ємною складовою професійної підготовки у вищій школі. Вона забезпечує встановлення безпосереднього зв'язку між теорією та практикою, оскільки саме в умовах реальної професійної діяльності відбувається інтеграція теоретичних знань і практичних умінь студентів, що зумовлює оволодіння ними складовими професійної компетентності.

Програма складена у відповідності до Законів України «Про освіту», «Про вищу освіту», «Положення про проведення практики у вищих навчальних закладах України» (затверджено наказом Міністерства освіти України від 08.04.1993 р. № 93) і «Методичних рекомендацій по складанню програм практики студентів вищих навчальних закладів України» (лист Міністерства освіти України від 14.02.1996 р. № 315/97) та адаптована до вимог та принципів Болонського процесу.

У відповідності з головними напрямами Болонського процесу, що полягають у лібералізації вищої школи, практика студентів-магістрів організовується з дотриманням принципів певної свободи у виборі місця та способу її проходження, але й одночасної відповідальності за свій вибір. Адже тільки внутрішньо вільна людина, котра опанувала мистецтво прийняття рішень і відповідальності за них, може бути по-справжньому свідомим громадянином, а також конкурентоспроможним, мобільним професіоналом. Адже студент, який особисто вибрав стратегічний освітній напрям на основі особистого інтересу та вподобань, і навчається зовсім інакше. Дотримання такої вимоги Болонського процесу створює можливість для магістрів бути мобільнішими,

брати участь у спільних освітніх моніторингових проектах, зокрема і європейських, підвищує їх конкурентоспроможність на європейському та світовому ринку праці.

Ефективність виробничої практики обумовлюється теоретичною обґрунтованістю системи підготовки студентів до практики, її навчальним і виховним характером, комплексним підходом до визначення змісту, завдань, форм і методів організації та проведення практики, забезпеченням послідовності, системності та наступності на різних етапах її проведення.

Основними завданнями виробничої практики магістрів є оволодіння студентами сучасними методами і формами професійної діяльності; формування професійних умінь і навичок, необхідних для самостійного вирішення педагогічних завдань в умовах реальної діяльності; виховання потреби систематичного професійного самовдосконалення.

Для студентів 5-го курсу професійного спрямування «Освітні вимірювання» рівня підготовки – магістр (8.18010022) згідно з навчальним планом передбачено проведення виробничої практики у 9-му семестрі тривалістю 6 робочих тижнів.

Виробнича практика проходить в Українському центрі оцінювання якості освіти.

Організацію практичної підготовки майбутніх фахівців регламентують положення та накази про проведення практики студентів вищих навчальних закладів України, затверджені наказом Міністерства освіти і науки, молоді та спорту України.

1. Цілі і завдання практики

Виробнича практика студентів спеціальності «Освітні вимірювання» - це невіддільна складова освітньо-професійної підготовки магістрів, основним завданням якої є закріплення теоретичних знань, отриманих у процесі навчання, формування у майбутнього магістра професійного вміння приймати самостійні рішення в певних соціальних умовах, оволодіння студентами сучасними методами, формами організації своєї діяльності.

У період практики закладаються основи досвіду професійної діяльності, практичних умінь і навичок, професійних якостей особистості майбутнього фахівця з освітніх вимірювань за освітньо-кваліфікаційним рівнем «Магістр».

Додаток II

Виробнича практика — це пошукова, практична і дослідницька робота студентів, що передбачає формування та удосконалення професійних умінь і навичок на основі ознайомлення з функціями фахівця з освітніх вимірювань та керівника закладу з оцінювання якості освіти під час роботи у відповідних установах і організаціях; аналіз та узагальнення фактичного матеріалу для підготовки магістерської роботи.

Завдання практики:

- ✓ поглиблення, закріплення та практичне використання теоретичних знань з фахових дисциплін;
- ✓ формування професійної ідентифікації майбутніх фахівців з освітніх вимірювань, їх професійної самосвідомості;
- ✓ формування вмінь і навичок для реалізації діагностичної, контролюючої та коригувальної функцій керівника;
- ✓ формування та закріплення вмінь складати програму спостереження на основі самостійно сформульованих гіпотез;
- ✓ вивчення технології проведення опитування, інтерв'ю, анкетування, контролю, коригування, перевірки;
- ✓ ознайомлення з технікою застосування методу аналізу документів.

Відповідно до специфіки бази практики практикантам ставляться різні завдання професійної діяльності, що забезпечать формування у магістрів умінь і навичок, передбачених Галузевими стандартами вищої освіти для спеціальності «Освітні вимірювання».

Після проходження науково-дослідної та педагогічної практики студент ***повинен знати:***

- ✓ етапи організації та реалізації професійної діяльності;
- ✓ принципи побудови, планування, організації керівництва закладом з оцінювання якості освіти, контролю та коригування результатів його діяльності;
- ✓ етапи реалізації та проведення заходів з освітніх вимірювань у закладах освіти;
- ✓ особливості проведення заходів з контролю якості освіти;
- ✓ принципи формування висновків і рекомендацій.

Специфіка діяльності конкретної бази практики вимагає від студентів-практикантів сумлінного виконання відповідних завдань для якісного формування відповідних професійних навичок. Після закінчення практики магістр **повинен вміти:**

- ✓ застосовувати загальнотеоретичні та прикладні знання під час розроблення
- ✓ конкретних проектів з діагностичної проблематики;
- ✓ використовувати різні підходи для організації конкретних моніторингових заходів;
- ✓ здійснювати якісну експертизу та оцінювання діяльності закладів освіти;
- ✓ представляти результати практичної діяльності у наукових звітах;
- ✓ планувати проведення заходів з моніторингу;
- ✓ застосовувати математико-статистичні методи для моніторингу якості освіти, для презентації результатів діяльності; використовувати відповідні кількісно-якісні методи емпіричного аналізу отриманих даних;
- ✓ інтерпретувати отримані результати відповідно до поставлених завдань;
- ✓ використовувати комп'ютерні технології у професійній діяльності.
- ✓ здійснювати методичну роботу з навчання вчителів і викладачів засобом моніторингу якості освіти;
- ✓ забезпечувати та здійснювати моніторингову діяльність у середній і вищій школах, спеціальних закладах та установах;
- ✓ створювати засоби для методичного забезпечення моніторингу якості освіти;
- ✓ розробляти рекомендації до створення тестових завдань з оцінювання знань, умінь і навичок учнів і студентів;
- ✓ аналізувати результати педагогічного контролю з метою виявлення резервів покращення якості освіти та оптимізації професійної діяльності навчально-виховних закладів;
- ✓ проводити експертне оцінювання діяльності учнів, студентів, учителів, викладачів зокрема та навчально-виховного закладу в цілому.

2. Зміст та завдання практики

2.1. Порядок проходження практики

Практика є складовою процесу підготовки магістрів в Інституті інформатики і проводиться на базі сучасних навчально-виховних закладів усіх рівнів, у центрах моніторингу якості освіти, науково-дослідних інститутах, у діагностичних центрах тощо. Вона є важливим етапом підготовки фахівців у межах майбутньої спеціальності та набуття досвіду самостійної діяльності. Практика студентів передбачає безперервність і послідовність її проведення для забезпечення одержання достатнього обсягу практичних знань і умінь відповідно до характеристик освітньо-кваліфікаційного рівня «Магістр» і є обов'язковою для студентів усіх форм навчання.

Магістерська практика студентів регламентується відповідними документами НПУ імені М. П. Драгоманова та Інституту інформатики.

Практика складається з **трьох етапів**.

1. Підготовчий етап.

На цьому етапі викладачі інституту розробляють програму магістерської виробничої практики. На початку семестру, в якому відбувається практика, студенти ознайомлюються з програмою, на її основі разом з керівником-методистом розробляють індивідуальні плани роботи з урахуванням специфіки установи, де відбуватиметься практика. Керівник практики перевіряє і коригує плани згідно з функціональними обов'язками і темою магістерської роботи.

За тиждень до початку практики кафедра комп'ютерної інженерії організовує настановну конференцію, де визначаються мета і завдання практики. У цей період студенти розподіляються за викладачами (методистами), які керуватимуть практикою.

2. Основний етап.

Робочий день практиканта визначається правилами трудового розпорядку та режимом роботи установи або організації, де він проходить практику. Кожен практикант працює за індивідуальним планом. План магістерської виробничої практики складається так, щоб максимально відтворити всі напрямки практичної діяльності фахівця з освітніх вимірювань. З огляду на зазначене, індивідуальний план студента-практиканта

повинен бути спрямований на формування певних умінь і навичок у таких сферах діяльності:

1) професійно-освітня робота (навчання здійснювати моніторинг якості освіти у навчальних закладах, підвищення професійної культури учасників навчально-виховного або трудового процесу, розкриття змісту роботи і виконання завдань служб моніторингу освіти тощо);

2) науково-дослідна робота (організація та здійснення магістерського дослідження, виконання якісного і кількісного аналізу, обґрунтування висновків тощо);

3) практична самостійна професійна діяльність (вивчення умов діяльності закладів моніторингу якості освіти, встановлення та діагностика якості надання освітніх послуг навчально-виховними закладами, з'ясування проблем, що утруднюють діяльність закладів освіти, створення та аналіз засобів з перевірки, контролю та оцінювання якості навчання для навчальних закладів різних рівнів акредитації тощо);

4) прогностична діяльність (розроблення, апробація і застосування тестових моделей перевірки якості навчання для різних навчально-виховних закладів з різних навчальних дисциплін; проектування змісту і напрямів індивідуального моніторингу особистості суб'єкта навчально-виховного процесу та складання на цій основі рекомендацій щодо стратегії його подальшої діяльності тощо);

5) консультативна діяльність (консультативна допомога в розв'язуванні проблем моніторингу освіти).

3. Підсумковий етап.

Після завершення практики проводиться підсумкова конференція, де студенти захищають звіти про виконання програми магістерської виробничої практики, підтверджуючи їх необхідними документами.

2.2. Щоденник практиканта

Щоденник розроблено за формою, визначеною "Положенням про практику студентів НПУ імені М. П. Драгоманова", і містить такі складові:

- ✓ перша сторінка, де визначено вид практики студента, навчальний підрозділ, курс, групу, спеціальність;
- ✓ направлення на практику;

Додаток II

- ✓ календарний план проходження практики;
- ✓ звіт про виконання програми практики;
- ✓ відгук керівника практики від бази практики з оцінкою роботи студента;
- ✓ відгук керівника практики від інституту про роботу студента;
- ✓ основні положення практики;
- ✓ правила ведення й оформлення щоденника.

2.3. Форми контролю

Поточний контроль здійснюється методистом упродовж проходження студентами виробничої практики шляхом аналізу та оцінювання їх систематичної роботи з виконання завдань практики. Підсумковий контроль здійснюється наприкінці практики шляхом оцінювання цілісної систематичної виробничої діяльності студентів. Під час оцінювання враховується рівень теоретичної підготовленості магістра, якість виконання завдань ним практики, рівень оволодіння вміннями, що стосуються моніторингу освітніх процесів, ставлення студента до виконання обов'язків, якість оформлення звітної документації.

Оцінювання відбувається на основі Рейтингового регламенту Інституту інформатики за Шкалою відповідності (додаток А).

2.4. Індивідуальні завдання

Ознайомитися із: загальними принципами діяльності Українського центру оцінювання якості освіти (УЦОЯО), його керівництвом і підрозділами; розробленими та апробованими моделями зовнішнього незалежного оцінювання (ЗНО) навчальних досягнень учнів загальноосвітніх навчальних закладів для створення умов рівного доступу до вищої освіти; банком тестових завдань; способами поширення знань про зовнішнє незалежне оцінювання серед учнівського загалу і педагогічної громадськості тощо.

Вивчити питання, що стосуються:

- підготовки до здійснення зовнішнього незалежного оцінювання;
- проведення моніторингу якості освіти;
- проведення соціально-психологічних досліджень та атестації педагогічних працівників;

- ведення реєстру осіб, які пройшли зовнішнє незалежне оцінювання;
- оформлення і видачі особам, які пройшли зовнішнє незалежне оцінювання, сертифікатів;
- надання інформації про результати зовнішнього незалежного оцінювання за запитом загальноосвітніх навчальних закладів та вищих навчальних закладів.

Підготувати та взаємооцінити тестові завдання з однієї з дисциплін (за вибором студента) шкільного курсу.

3. Звіт про виконання завдань практики

Звіт про проходження практики студенти оформлюють самостійно і подають керівникам практики від базової установи та від Інституту для підпису. Керівник практики від інституту складає рецензію, де зазначає ступінь виконання студентом магістратури програми практики, рівень оволодіння методами дослідження і рівень набутих навичок.

Інші документи про практику (щоденник проходження практики, пропозиції та матеріали, підготовлені студентами і прийняті для використання у практичній роботі бази практики, характеристика на кожного студента) також мають бути підписані керівником від бази практики і завірені печаткою. Документи передаються викладачеві, який здійснював керівництво магістерською практикою від інституту.

Захист звіту про практику відбувається на підсумковій конференції.

4. Підведення підсумків практики

Зшитий або скріплений скобами звіт та щоденник про практику (з відгуком керівника практики від організації) протягом 3-х днів після завершення практики здають керівникові практики від кафедри, який перевіряє правильність їхнього оформлення, підписує і передає на кафедру.

Для проведення захисту практики на засіданні кафедри затверджують комісію у складі 3-х викладачів кафедри. Захист практики проводять у 10-денний термін після її завершення.

Підведення підсумків практики здійснюється шляхом індивідуального захисту практики кожним студентом. Під час

Додаток II

захисту студент розповідає комісії про проходження практики, формулює завдання практики і пояснює їхнє виконання; інтерпретує отримані результати, відстоює обрану ним методику виконання завдань.

За результатами захисту виставляють диференційовану оцінку, яку заносять у залікову книжку. Критерії оцінювання такі: вичерпність та правильність виконання завдань практики, якість оформлення звіту, переконливість відповідей студента під час захисту, відгук керівника практики від організації.

Студента, який не здав звіту про практику або отримав на захисті незадовільну оцінку, направляють на практику повторно, він здійснює перездачу практики за встановленою процедурою або ж його відраховують з Університету.

Оцінювання відбувається за шкалою, наведеною у таблиці.

За шкалою ECTS	За шкалою університету	Визначення	Оцінка за національною шкалою	
			Екзамен	Залік
A	90 – 100	Відмінно	5 (відмінно)	Зараховано
B	80 – 89	Дуже добре	4 (добре)	
C	70 – 79	Добре		
D	65 – 69	Задовільно	3 (задовільно)	
E	60 – 64	Достатньо		
FX	35 – 59	Незадовільно з можливістю повторного складання	2 (незадовільно)	Незараховано
F	1 – 34	Незадовільно з обов'язковим повторним курсом		

Зразок оформлення листа про проходження практики

Директору _____
(повна назва інституту)

_____ (прізвище та ініціали)

_____ (прізвище та ініціали студента)

Група _____

Спеціальність _____

Форма навчання _____

Прошу дозволити проходження _____ в термін
(вказати вид практики)

з “_” _____ по “_” _____ на базі _____
(повна назва закладу)

Керівник практики від _____
(повна назва бази практики)

_____ (прізвище та ініціали, посада)

Погоджено: _____
(підпис керівника практики (прізвище та ініціали) від бази практики)

“_” _____
(дата)

_____ (підпис керівника практики (прізвище та ініціали) від навчального закладу)

“_” _____
(дата)

МП

_____ (підпис студента)

“_” _____
(дата)

Лист надсилається в інститут на ім'я директора інституту за 1,5 – 2 місяці до початку практики.

Зразок оформлення титульної сторінки звіту про виконання завдань практики

_____ (вид практики)
студентом спеціальності _____ Інституту інформатики

ЗВІТ

про виконання програми _____ практики

_____ (вид практики)
студента _____ (прізвище, ім'я, по батькові)

Групи _____

Спеціальність _____

Кваліфікаційний рівень _____

База практики _____ (повна назва)

Керівник практики від бази практики
_____ (посада, прізвище, ініціали)

Керівник практики від інституту
_____ (посада, прізвище, ініціали)

Звіт захищений
“ ___ ” _____ 200_ р.
_____ (підпис)

Рейтинговий регламент інституту інформатики

Кредитно-модульна система – це модель організації навчального процесу, яка ґрунтується на поєднанні двох складових: модульної технології навчання та кредитів (залікових одиниць) і охоплює зміст, форми та засоби навчального процесу, форми контролю якості знань та вмінь і навчальної діяльності студента в процесі аудиторної та самостійної роботи. Кредитно-модульна система має за мету поставити студента перед необхідністю регулярної навчальної роботи протягом усього семестру з розрахунком на майбутній професійний успіх.

Навчальний модуль – це логічно завершена, відносно самостійна, цілісна частина навчального курсу, сукупність теоретичних та практичних завдань відповідного змісту та структури з розробленою системою навчально-методичного та індивідуально-технологічного забезпечення, необхідним компонентом якого є відповідні форми рейтингового контролю.

Кредит (залікова одиниця) – це уніфікована одиниця виміру виконаної студентом аудиторної та самостійної навчальної роботи (навчального навантаження), що відповідає 36 годинам робочого часу.

Рейтинг (рейтингова оцінка) – це кількісна оцінка досягнень студента за багатобальною шкалою в процесі виконання ним заздалегідь визначеної сукупності навчальних завдань.

Рейтингова система оцінювання (PCO) – це система визначення якості виконаної студентом усіх видів аудиторної та самостійної навчальної роботи та рівня набутих ним знань та вмінь шляхом оцінювання в балах результатів цієї роботи під час поточного, модульного (проміжного) та семестрового (підсумкового) контролю, з наступним переведенням оцінки в балах у оцінку за традиційною національною шкалою та шкалою ECTS.

PCO передбачає використання поточної, контрольної, підсумкової, підсумкової семестрової модульних рейтингових оцінок, а також екзаменаційної та підсумкової семестрових рейтингових оцінок.

Додаток II

Поточна модульна рейтингова оцінка складається з балів, які студент отримує за певну навчальну діяльність протягом засвоєння даного модуля – виконання та захист індивідуальних завдань (розрахунково-графічних робіт, рефератів тощо), лабораторних робіт, виступи на семінарських та практичних заняттях тощо.

Контрольна модульна рейтингова оцінка визначається (в балах та за національною шкалою) за результатами виконання модульної контрольної роботи з даного модуля.

Підсумкова модульна рейтингова оцінка визначається (в балах та за національною шкалою) як сума поточної та контрольної модульних рейтингових оцінок з даного модуля.

Підсумкова семестрова модульна рейтингова оцінка визначається (в балах та за національною шкалою) як сума підсумкових модульних рейтингових оцінок, отриманих за засвоєння всіх модулів.

Екзаменаційна рейтингова оцінка визначається (в балах та за національною шкалою) за результатами виконання екзаменаційних завдань.

Залікова рейтингова оцінка визначається (в балах та за національною шкалою) за результатами виконання всіх видів навчальної роботи протягом семестру.

Підсумкова семестрова рейтингова оцінка визначається як сума підсумкової семестрової модульної та екзаменаційної (залікової – у випадку диференційованого заліку) рейтингових оцінок (в балах, за національною шкалою та за шкалою ECTS).

Підсумкова рейтингова оцінка з дисципліни, яка викладається протягом декількох семестрів, визначається як середньозважена оцінка з підсумкових семестрових рейтингових оцінок у балах з наступним її переведенням у оцінки за національною шкалою та шкалою ECTS. Зазначена підсумкова рейтингова оцінка з дисципліни заноситься до додатку до диплому фахівця.

Якщо студент з поважних причин, що підтверджено документально, був відсутній на заняттях, він має право на одне перескладання з можливістю отримання максимальної кількості балів. Термін перескладання визначається викладачем.

Якщо впродовж семестру студент пропустив значну кількість занять, не має оцінок за виконання модулів, у

відповідних графах «Відомості обліку успішності КМСОНП» виставляються «0», а у графі екзамену – відмітка про не допуск до нього.

Залік та екзамен є формами підсумкового контролю результатів навчання студентів і мають на меті перевірку системності засвоєння програмового матеріалу, цілісності бачення навчального курсу, рівня осмислення знань та набуття умінь, їх комплексного застосування у практичній діяльності, діагностування ефективності самостійної навчальної роботи студентів.

Для запобігання репродуктивного характеру перевірки знань та умінь при проведенні заліку та екзамену передбачається використання теоретичних і практичних запитань, задач різного ступеня складності, а також творчих завдань з метою виявлення можливих рівнів засвоєння студентами змісту навчального курсу.

Відмітка «зараховано» виставляється студенту при умові виконання вимог, передбачених у змісті навчальної програми, а саме:

- регулярному відвідуванні лекційних, практичних, лабораторних занять або їх негайному відпрацюванні, своєчасному складанні колоквіумів, лабораторних робіт та інших форм поточного контролю, виконанні контрольних і розрахункових робіт з позитивними результатами;

- поглибленні набутих знань у процесі самостійної роботи, опрацюванні наукової інформації;

- засвоєнні змісту навчального курсу в обсязі, передбаченому галузевим стандартом вищої освіти.

Відмітка «не зараховано» виставляється студенту в тому випадку, якщо його навчальна діяльність не відповідає вище означеним вимогам.

Додаток III. Програма комплексного державного екзамену для магістрів спеціальності 8.18010022 «Освітні вимірювання» у НПУ імені М.П. Драгоманова

Укладачі програми:

доктор педагогічних наук, професор Сергієнко В.П.,
доктор фізико-математичних наук, професор Малежик М.П.,
кандидат педагогічних наук, доцент Тополя Л.В.,
аспірантка Кухар Л.О.,
методист Макаренко О.Л.

Пояснювальна записка

Кваліфікаційний іспит студентів спеціальності «Освітні вимірювання» здійснюється державною екзаменаційною комісією після завершення навчання на освітньо-кваліфікаційному рівні «магістр» для встановлення фактичної відповідності рівня підготовки вимогам кваліфікаційної характеристики. Програма кваліфікаційного іспиту включає предмети науково-предметної підготовки (класичні тестові моделі та їх застосування, основи конструювання тестів, методологія та методи досліджень в освіті, організація управління навчальним процесом у ВНЗ).

Кваліфікаційний іспит має виявити фахові знання студентів.

Державний екзамен проводиться за білетами, затвердженими кафедрою комп'ютерної інженерії. Кожен білет містить три завдання: завдання 1 – з дисциплін класичні тестові моделі та їх застосування, основи конструювання тестів, завдання 2 – з дисциплін методологія та методи досліджень в освіті, організація управління навчальним процесом у ВНЗ, завдання 3 – практичне завдання.

ОСНОВИ КОНСТРУЮВАННЯ ТЕСТІВ

Метою є перевірка сформованості професійної компетентності майбутніх фахівців з освітніх вимірювань з питань тестології та конструювання тестів, перевірка навичок складати та використовувати тестовий інструментарій для оцінювання якості освіти за сучасними програмами та результатами національних і

міжнародних порівняльних досліджень якості освіти, володіти методами та засобами опрацювання результатів тестування та оцінювання якості тестових завдань і тесту в цілому.

Відповідь магістра повинна засвідчити ґрунтовні наукові усвідомлені знання відповідно до таких вимог:

знати:

- типи тестових завдань;
- форми і види тестових завдань, їх переваги та недоліки;
- класифікацію тестів за різними ознаками;
- основи технології розроблення і коригування тестів;
- форми і можливості досягнення поставлених цілей педагогічного контролю;
- особливості процедури проведення тестування, в т.ч. психологічні;
- основи опрацювання та інтерпретації результатів тестування;

вміти:

- формулювати цілі та принципи складання тесту і тестових завдань (при використанні тесту для поточного, проміжного та підсумкового контролю);
- використовувати таксономію Б. Блума при формулюванні цілей і складати матрицю тесту;
- проектувати тест, який відповідає поставленим цілям, на основі логічного структурування навчального матеріалу;
- складати специфікацію тесту, до якої входять матриця тесту, опис типів завдань, які використовуються для оцінювання досягнень студентів, а також особливості процедури тестування;
- проводити експертизу завдань, тесту в цілому і робити висновки про змістову валідність тесту;
- робити висновки щодо валідності тесту на основі завдань і вносити корективи;
- робити осмислений вибір тестів;

мати уявлення про:

- сучасні тенденції розвитку освіти у світі і проблеми оцінювання її якості;
- становлення і розвиток теорії та практики психологічних і

Додаток III

- педагогічних тестів;
- відмінності нормативно-орієнтованих і критеріально-орієнтованих тестів;
- характер тестів;
- способи розв'язання проблеми постановки мети навчання у вітчизняній та зарубіжній (таксономія Б. Блума) педагогіці;
- можливості побудови педагогічних тестів за аналогією з відомими психологічними;
- проблеми, пов'язані з використанням тестів;
- способи використання педагогічних тестів у зарубіжній вищій школі.

Основи теорії тестових завдань та тестів

Болонська декларація, система академічних кредитів ECTS, якість освіти, якість підготовки, педагогічне вимірювання, педагогічні тести, педагогічна діагностика, критерії якості, критерій інформативності, інформативність тесту, рівні розуміння тесту, тест як інструмент вимірювання рівня знань, переваги та недоліки тестування, традиційні та сучасні засоби контролю знань, форми перевірки знань, види контролю, види контролю за змістом навчального матеріалу, моніторинг якості освіти, політика з оцінювання та тестування, характеристики якості національної освіти, стандарти якості, учасники процесу оцінювання, категорії учасників тестування.

Тестологія, понятійний апарат тестології, різні підходи до визначення поняття тест, тестування, класична теорія конструювання тестів (СТТ), переваги та недоліки, сучасна теорія конструювання тестів (IRT), переваги та недоліки, педагогічний тест, класифікація педагогічних тестів.

Зміст та етапи створення тесту

Схема педагогічних цілей, таксономія Блума, рівні освітніх цілей, класифікація когнітивної сфери за Блумом, Безпалько, Скаткіним, дистрактор, претестове завдання, педагогічне вимірювання, нормативно орієнтовані та критеріально-орієнтовані тести, особливості їх конструювання та застосування, види тестування, комп'ютерне тестування, адаптивне комп'ютерне тестування, види тестування за способом проведення, якість тесту.

Проблеми складання тестових завдань, гомогенні тести, гетерогенні тести, інтегративні тести, класифікація тестів за процедурою формування черги подання, класифікація тестів за кількістю претендентів, види тестового контролю, види тестування, класифікація навчальних цілей, таксономія, постановка і конкретизація цілей навчання, категорії навчальних цілей в когнітивній сфері, рівні засвоєння знань за І. Лернером, класифікація рівнів засвоєння знань за В. Безпалько, компетентність, компетенція, домени компетентності, ключові компетентності.

Проектування стандартизованого тесту

Етапи створення тесту, визначення мети тестування, етапи робіт по створенню критеріально- та нормативно-орієнтованих тестів, етапи створення тестів навчальних досягнень для тестів різного рівня застосування, стандартизація тесту, варіативність змісту тестів, паралельні варіанти тестів, фасет.

Вимоги до складання тестових завдань, правила для складання завдань на доповнення, навчальні тести, контрольні тести, формати тестових завдань, принципи добору відповідей, створення тестових завдань різних когнітивних рівнів, залежність видів і форм тестів від специфіки навчальної дисципліни, види відкритих тестових завдань, види закритих тестових завдань, тестові завдання на встановлення відповідності, тестові завдання на встановлення правильної послідовності, тестові завдання на доповнення, тестові завдання з розгорнутою відповіддю, вимоги до завдань в тестовій формі, валідність, складність, надійність, стійкість тесту, шкалювання, репрезентативність тесту, значущість, дискримінантність, достовірність, науковість, несуперечність тесту.

Аналіз та представлення результатів тестування

Статистичні характеристики тесту, етапи системи комплексної експертизи якості тестових матеріалів, експертне оцінювання якості тестових завдань, матриця результатів тестування, розрахунок основних математико-статистичних характеристик, коефіцієнт кореляції, підходи визначення вагових коефіцієнтів, змістова валідність, валідність за віковою диференціацією, конструктивна, критеріальна, поточна, порівняльна, прогностична валідність, чинники, що знижують

Додаток III

валідність результатів оцінювання, теорія однопараметричного оцінювання Раша.

Шкали, стандартизація шкали, шкалювання, номінальні шкали, порядкові шкали, інтервальна шкала, шкалювання результатів тестових вимірювань, шкали процентильних рангів, шкала первинних рангів, етапи стандартизації тесту, критерії точності тесту, репрезентативність.

Інструкція до тесту, розроблення інструкцій, стандартизація процедури та умов тестування, нормативна база Зовнішнього незалежного оцінювання.

КЛАСИЧНІ ТЕСТОВІ МОДЕЛІ ТА ЇХ ЗАСТОСУВАННЯ

Метою навчального курсу “Класичні тестові моделі та їх застосування” є ознайомлення майбутніх спеціалістів в галузі освітніх вимірювань із класичними математичними моделями оцінювання якості освіти, методологічними та теоретичними основами тестового контролю, сферами застосування, перевагами та недоліками класичної теорії тестування, аналізом результатів педагогічних вимірювань.

Відповідь магістра повинна засвідчити ґрунтовні наукові усвідомлені знання відповідно до таких вимог:

знати:

- основні математичні моделі класичної теорії тестування;
- поняття матриці відповідей, первинні бали;
- процедури та способи оцінювання надійності тесту;
- алгоритм усереднювання значень функції вимірювання.
- точність початкових вимірювань;
- розподільна здатність тесту, надійність та валідність тесту;
- матриця відповідей;
- алгоритми і методи оцінки рівня підготовленості тестованих та характеристик тестових завдань у рамках класичної теорії тестування;
- розподіл дистракторів;
- види, критерії та джерела підвищення валідності тесту;
- порядкова шкала та метрична шкала;
- остаточний бал учасників тестування;
- диференціація учасників тестування;
- основні методи шкалювання результатів освітнього

- вимірювання;
- інформаційна функція тесту;
- характеристична функція тесту.

уміти:

- проводити статистичну перевірку гіпотез тестування;
- проводити порівняння емпіричної і теоретичної імовірності успіху;
- перевіряти рівномірність розподілу дистракторів;
- перевіряти значущість розбіжності різних результатів тестування на метричній та на порядковій шкалі;
- здійснювати перевірку паралельності варіантів тесту;
- здійснювати шкалювання результатів тестування;
- перетворювати єдину метричну шкалу у нормовану;
- виконувати порівняння тестів при конструюванні;
- вирівнювати різні варіанти тесту на єдиній метричній шкалі у процесі математичної обробки результатів.

володіти:

- методами розроблення занять по підготовці суб'єктів навчання до іспитів;
- навиками роботи з комп'ютерними пакетами програм по обробці результатів тестування.

Введення в теорію вимірювань

Тестові оцінки як дискретні змінні. Нормальний розподіл. Нескінченні популяції та неперервні змінні (дійсні числа та шкали вимірювання, рівні вимірювальних шкал) Підходи до шкалювання в розробленні тестів. Цілі застосування тестової оцінки. Побудова вибірки змісту, конструювання та рецензування завдань. Схеми оцінювання завдань. Описова статистика для недихотомічних змінних. Дескриптивна статистика для дихотомічних змінних. Дисперсія складової оцінки.

Надійність

Класична модель істинної оцінки. Показник надійності та коефіцієнт надійності. Надійність елементів тестової оцінки. Стандартна похибка вимірювання. Фактори, що впливають на коефіцієнт надійності. Наближене обчислення істинних оцінок. Надійність різниці оцінок.. Введення коефіцієнта генералізації для однофасетних досліджень. Генеральні сукупності з фіксованими

Додаток III

фасетами. Використання критеріально-орієнтованих вимірювань. Теорія надійності для класифікацій за категоріями майстерності.

Валідність

Змістовна валідність. Критеріально-орієнтована валідизація. Конструктна валідність. Коефіцієнти валідності для істинних оцінок. Часткова кореляція. Множинна регресія. Дискримінантний аналіз. Корельовані фактори. Модель факторного аналізу.

МЕТОДОЛОГІЯ І МЕТОДИ ДОСЛІДЖЕНЬ В ОСВІТІ

Наукові основи педагогічних досліджень.

Методологія експерименту та методи опрацювання й аналізу результатів експерименту. Методи аналізу змісту та результатів діяльності. методи дослідження на емпіричному та методологічному рівнях. Поняття про педагогічні дослідження, їх роль і місце у навчально-виховному процесі та методи педагогічних досліджень. Організаційні основи педагогічних досліджень. Методологія експерименту. Опрацювання й оформлення результатів вимірювань та похибки вимірювань та їх класифікація. Принципи педагогічних досліджень. Психолого-педагогічні аспекти досліджень. Сучасні тенденції розвитку методів досліджень в освіті. Педагогічні дослідження в Україні у світлі інтеграції в єдиний європейський та світовий освітній простір.

Методика організації та проведення педагогічних досліджень.

Принципи планування та проведення педагогічного дослідження. Методологічні основи наукових досліджень. Середнє арифметичне значення результатів вимірювань. Середня квадратична похибка середнього арифметичного значення вимірюваної величини. Особливості опрацювання результатів малої кількості вимірювань. Розрахунки похибок посередніх вимірювань. Подання результатів вимірювань.

Проблема вимірювань у соціальних науках

Теоретичні конструкти і операційні визначення емпіричного дослідження. Види шкал: номінальна, порядкова, інтервальна, шкала відношень. Графік розподілу. Допустимі статистики в кожній шкалі: опис середньо-групових даних, можливості порівняння індивіда з групою, вимір залежностей в кожній шкалі. Описова статистика: міри центральної тенденції (мода, медіана,

середнє арифметичне) та міри мінливості (розкид, відхилення, середнє відхилення, стандартне відхилення). Загальне уявлення про кореляцію та кореляційні процедури. Кореляційне дослідження. Оцінка кореляції: графічна та математична, коефіцієнт кореляції. Тлумачення кореляції. Коефіцієнт кореляції та причинно-наслідковий зв'язок. Коефіцієнт детермінації. Регресивний аналіз. Лінія регресії. Тлумачення регресії. Нормальна крива розподілу. Нормальна крива та ймовірність. Вибірка, помилки вибірки. Стандартна похибка виміру. Проктильні норми. Нормалізація та нелінійні перетворення.

Практика обстежень в педагогіці, психології та соціології.

Етика досліджень.

Етика досліджень, проблеми добору, процедури опрацювання, інтерпретації та збереження інформації. Вплив умов тестування на результат. Спеціальні заходи тестування в спеціальних групах респондентів. Забезпечення мотивації тестування. Комп'ютерне тестування. Фальшування при тестуванні. Забезпечення безпечності тестування та його результатів. Звіт про тестування. Об'єктивний підхід в дослідженнях. Вимір здібностей, історія виміру інтелекту, теорії інтелекту. Монофакторні теорії структури інтелекту. Мультифакторні теорії структури інтелекту. Кубічні моделі інтелекту. Теорія інтелекту Р.Кеттелла: плинний та кристалізований інтелект. Теорія інтелекту Стернберга. Тестування когнітивної сфери: шкали Стенфорд-Біне, шкали Векслера.

**ОРГАНІЗАЦІЯ УПРАВЛІННЯ НАВЧАЛЬНИМ ПРОЦЕСОМ У
ВИЩОМУ НАВЧАЛЬНОМУ ЗАКЛАДІ**

Метою навчального курсу «Організація управління навчальним процесом у вищому навчальному закладі» є опанування нормативно-правовою базою організації навчального процесу у ВНЗ та вивчення його особливостей.

Завдання курсу – аналіз державної політики в галузі вищої освіти України, вивчення особливостей, завдань, структури, змісту, форм і методів організації навчального процесу у ВНЗ.

У результаті вивчення дисципліни студент повинен **знати:**

Додаток III

- зміст та структуру вищої освіти України;
- принципи і систему управління та нормативно-правову базу організації навчального процесу у ВНЗ;
- суть процесу навчання та його особливості у ВНЗ;
- принципи, закономірності та умови планування і організації навчального процесу у ВНЗ;
- форми організації навчального процесу і вимоги до них у ВНЗ;
- основні види навчальних та індивідуальних занять, самостійної роботи та практичної підготовки студентів;
- контрольні заходи та умови їх організації у ВНЗ;
- вимоги до науково-методичного забезпечення навчального процесу у ВНЗ.

вміти:

- впроваджувати в процесі своєї педагогічної діяльності принципи освітньої політики України;
- раціонально планувати та організовувати навчальний процес у ВНЗ;
- розробляти науково-методичне забезпечення навчального процесу у ВНЗ;
- складати навчальний план підготовки фахівців за напрямом базової освіти, індивідуальний план НПП, індивідуальний навчальний план студента та іншу документацію;
- організовувати педагогічну співпрацю з громадськістю, державними організаціями, підприємствами тощо;
- приймати оптимальні педагогічні рішення та впроваджувати ефективні шляхи педагогічного управління навчальним процесом у ВНЗ.

Теоретичні та методологічні основи управління навчальним процесом у ВНЗ

Структура та зміст вищої освіти України

Структура, стан, тенденції та перспективи розвитку вищої освіти України.

Багатоступенева система вищої освіти. Освітні та освітньо-кваліфікаційні рівні. Освітньо-професійна програма, структурно-

логічна схема та інші нормативні документи підготовки фахівців у ВНЗ. Нормативна та вибіркова частини змісту освіти у ВНЗ.

Зміст вищої освіти як науково-обґрунтована система дидактично та методично оформленого навчального матеріалу для різних освітніх і кваліфікаційних рівнів.

Управління навчальним процесом у ВНЗ

Державна політика у галузі вищої освіти. Система та основні функції державного управління вищою освітою України.

Державні органи управління вищою освітою (Міністерство освіти і науки, молоді та спорту України; Центральні органи виконавчої влади України, яким підпорядковуються ВНЗ; Вища атестаційна комісія; Державна акредитаційна комісія; органи громадського самоврядування) та їх функції.

Управління і внутрішній контроль ВНЗ, особливості їх організації. Види, форми і методи внутрішнього контролю та управління. Структурні підрозділи ВНЗ та принципи управління ними. Вчена рада, наглядова рада, робочі та дорадчі органи, органи громадського та студентського самоврядування у ВНЗ.

Нормативно-правова база організації навчального процесу у ВНЗ

Завдання ВНЗ у світлі Державної національної програми “Освіта. (Україна XXI ст.)” та Національної доктрини розвитку освіти в Україні.

Основні документи, які регламентують нормативно-правову базу діяльності ВНЗ та підготовки сучасного фахівця з вищою освітою: Закони України “Про освіту” та “Про вищу освіту”, укази Президента України, постанови Кабінету Міністрів України, накази Міністерства освіти і науки України та інші акти законодавства з питань вищої освіти.

Суть процесу навчання та його особливості у ВНЗ

Рівень організації навчального процесу у ВНЗ та його взаємозв’язок і залежність від умов розвитку суспільства.

Навчальний процес у ВНЗ як система організаційних і дидактичних заходів, спрямованих на реалізацію змісту освіти на певному освітньому або кваліфікаційному рівні відповідно до державних стандартів освіти.

Закономірності та основні принципи організації навчального процесу у ВНЗ.

Форми організації навчального процесу та вимоги до них у ВНЗ

Форми навчання та організації навчального процесу у ВНЗ

Форми навчання у ВНЗ (денна (стаціонарна), вечірня, заочна, дистанційна, екстернат), їх характеристика та умови навчання студентів.

Форми організації навчального процесу у ВНЗ: навчальні заняття, виконання індивідуальних завдань, самостійна робота студентів, практична підготовка, контрольні заходи.

Основні види навчальних занять у ВНЗ

Основні види навчальних занять у ВНЗ: лекція, лабораторне, практичне, семінарське, індивідуальне заняття, консультація.

Лекція як основна форма проведення навчальних занять у ВНЗ, призначених для засвоєння теоретичного матеріалу. Дидактична мета, вимоги до тематики та змісту лекцій. Лектор, його рівень, права і обов'язки.

Тематика, основні етапи підготовки та умови проведення лабораторних занять. Дидактична мета, основні етапи і вимоги до студентів та науково-педагогічних працівників під час проведення практичного заняття.

Семінарські заняття, їх види та особливості проведення на різних навчальних курсах.

Індивідуальні заняття та завдання, контроль їх виконання

Зміна методології процесу навчання відповідно до кредитно-модульної системи організації навчального процесу та його переорієнтація на індивідуально-диференційовану та особистісно-орієнтовану форму.

Індивідуальна робота студента у сучасному ВНЗ та, її роль та місце в оволодінні студентами навчальним матеріалом, особливості та вимоги до її організації.

Види, визначення обсягу часу, змісту, форм та методів проведення і забезпечення навчально-методичними засобами індивідуальних навчальних занять.

Консультації та особливості і нормативи для їх розрахунку і проведення для студентів різних форм навчання у ВНЗ. Індивідуальні завдання (реферати, розрахункові та графічні роботи, курсові та дипломні проекти тощо), їх тематика, керівництво підготовкою, захист та оцінювання.

Науково-дослідна робота студентів в системі їх професійної підготовки.

Курсові та дипломні проекти, їх мета, тематика, порядок виконання та захисту.

Організація самостійної роботи студентів у ВНЗ

Самостійна робота студента у сучасному ВНЗ, її роль та місце в оволодінні студентами навчальним матеріалом, особливості та вимоги до її організації.

Регламентування навчального часу та графік самостійної роботи студентів відповідно до кредитно-модульної системи організації навчального процесу у ВНЗ.

Вимоги до сучасного навчально-методичного забезпечення самостійної роботи студентів у ВНЗ.

Роль науково-педагогічного працівника та обов'язки ВНЗ щодо організації самостійної роботи студентів на сучасному етапі розвитку вищої освіти.

Організація практичної підготовки студентів у ВНЗ

Практична підготовка студентів, її дидактична мета і зміст. Структуру, види, бази практики студентів ВНЗ.

Керівництво, організація, проведення і контроль практики студентів у ВНЗ.

Науково-методичне, нормативне і матеріальне забезпечення практичної підготовки студентів у ВНЗ.

Контрольні заходи та їх організація у ВНЗ

Роль і місце контрольних заходів в організації навчального процесу у ВНЗ. Форми поточного та підсумкового контролю засвоєння студентами навчального матеріалу.

Мета, особливості та умови організації поточного контролю під час проведення практичних, лабораторних та семінарських занять у ВНЗ.

Семестровий контроль та державна атестація студента як основні види підсумкового контролю у ВНЗ. Форми семестрового контролю: семестровий екзамен, диференційований залік та залік з конкретної навчальної дисципліни. Екзаменаційна сесія, терміни її проведення та тривалість. Шкала та форми оцінювання результатів складання екзаменів та диференційованих заліків. Умови повторного складання екзаменів та заліків у ВНЗ.

Додаток III

Державні атестація студента, її дидактична мета, основні форми, порядок проведення та документального оформлення. Основні завдання державної екзаменаційної (кваліфікаційної) комісії, умови формування її персонального складу, підбір і призначення голови.

Планування та організація навчального процесу

Умови планування та графік навчального процесу у ВНЗ

Загальна характеристика кредитно-модульної системи організації навчального процесу у ВНЗ. Принципи організації навчального процесу в умовах КМСОНП.

Розподіл навчального навантаження за типами курсів, рівнями і модулями. Модульна структура навчального плану підготовки за кваліфікацією.

Тривалість навчального року, канікул, весняного та осіннього семестрів. Планування та терміни теоретичного навчання, екзаменаційних сесій, канікул, навчальної практики, виробничої (переддипломної практики), дипломного проектування та Державної атестації студентів відповідно до освітньо-кваліфікаційного рівня.

Навчальне навантаження студента

Обсяг навчального навантаження студента. Академічний кредит як залікова одиниця трудомісткості навчальної роботи.

Облікові одиниці навчального часу студента: академічна година, навчальний день, тиждень, семестр, курс, рік та нормативні вимоги до них.

Планування і облік навчального часу студента. Розклад та порядок відвідування студентами навчальних занять.

Індивідуальний навчальний план студента, його структура і порядок укладання.

Робочий час науково-педагогічного працівника

Основні види та обсяг роботи науково-педагогічного працівника. Індивідуальний план роботи науково-педагогічного працівника.

Норми часу для планування, розрахунку та обліку навчальної роботи науково-педагогічних працівників.

Перелік основних видів та облік, методичної, наукової та організаційної роботи науково-педагогічного працівника.

Науково-методичне забезпечення навчального процесу у ВНЗ
Державні стандарти освіти

Система стандартів вищої освіти: державний стандарт вищої освіти, галузеві стандарти вищої освіти та стандарти вищої освіти ВНЗ.

Державний стандарт вищої освіти як сукупність норм, які визначають вимоги до освітнього та освітньо-кваліфікаційного рівня. Складові державного стандарту вищої освіти, їх функції та роль.

Галузеві стандарти вищої освіти: освітньо-кваліфікаційна характеристика, освітньо-професійна програмам, засоби діагностики якості вищої освіти. Стандарти вищої освіти ВНЗ: перелік спеціалізацій за спеціальностями, варіативні частини ОКХ випускників ВНЗ, варіативні частини ОПП, варіативні частини засобів діагностики якості вищої освіти, навчальні плани та програми.

Навчальні плани та програми навчальних дисциплін

Навчальний план як документ, що визначає графік навчального процесу, перелік, послідовність та час вивчення навчальних дисциплін, форми навчальних занять та терміни їх проведення, а також форми проведення підсумкового контролю.

Структура навчального плану, основні вимоги до його складання. Методика розробки навчального плану на основі державного стандарту освіти. Базовий та робочий навчальний план, їх структура та призначення. Навчальна програма дисципліни, основні вимоги до її структури та змісту.

Науково-методичний комплекс дисципліни та його структурі компоненти.

Підручники та навчальні посібники

Підручник та навчальний посібник як основні види навчальної літератури у ВНЗ, їх функції та структурні компоненти.

Критерії відбору навчального матеріалу, вимоги до структури та обсягу підручників і навчальних посібників.

Порядок затвердження та рецензування підручників і навчальних посібників, надання грифу Міністерством освіти та науки України, а також іншими центральними органами виконавчої влади, що мають у своєму підпорядкуванні ВНЗ.

Додаток III

Інструктивно-методичні матеріали до семінарських та практичних занять

Врахування галузевих особливостей підготовки фахівців у ВНЗ та відповідних вимог до підготовки навчальної літератури.

Вимоги до структури та обсягу інструктивно-методичних матеріалів до семінарських і практичних занять.

Науково-методичні вимоги до складання інструктивно-методичних матеріалів до семінарських і практичних занять.

Основна література

1. Аванесов В. С. Математические модели педагогического измерения / В. С. Аванесов. – М.: Исслед. центр, 1994. – 26 с.

2. Аванесов В.С. Тесты в социологическом исследовании. – М., 1982. – 199 с.

3. Анастаси А. Психологическое тестирование. / Анастаси А., Урбина С. – С–Пб.: Питер, 2002. – 688с.

4. Грабар М.Й. Применение математической статистики в педагогических исследованиях. Непараметрические методы. / Грабар М.Й., Краснянская К.А. – М.: Педагогика, 1977.– 136 с.

5. Ефремова Н.Ф. Тестовый контроль в образовании: учебное пособие. – М. : Логос, Университетская книга, 2007. – 386 с.

6. Зовнішнє незалежне оцінювання в освіті України. Курс лекцій: Навч. посібник. / Кашина Г.С., Сергієнко В.П. – Луцьк, 2010. – 115 с.

7. Ким В.С., Тестирование учебных достижений. - Уссурийск, 2007. Режим доступа http://uss.dvfu.ru/static/kim_testing_monograph/src/contents.php

8. Клайн П. Введение в психометрическое проектирование. Справочное руководство по конструированию тестов. –Киев: ПАН Лтд, 1994. -184 с.

9. Конструювання тестів. Курс лекцій: Навч. посібник. / Кухар Л.О., Сергієнко В.П. – Луцьк, 2010. – 182 с.

10. Крокер Л., Алгина Дж.. Введение в классическую и современную теорию тестов. Москва, Логос, 2010.

11. 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.

12.Смирнова–Трибульська Є.М. Дистанційне навчання з використанням системи *Moodle*. Посібник для вчителів. Видавництво «Айланта», – 2007 – 505 с.

13.Франчук В.М. MOODLE (Тести). Посібник для студентів інформатичних спеціальностей педагогічних університетів. – К.: НПУ імені М.П. Драгоманова, 2009. – 55 с.

14.Цехмістрова Г.С., Форменко Н.А. Управління в освіті та педагогічна діагностика: Навчальний посібник для студентів вищих навчальних закладів. - К.: Видавничий Дім, 2005. - 280 с.

Додаткова література

1. Булах І.Є., Мруга М.Р. Створюємо якісний тест. Навчальний посібник. / К.: Майстер-клас, – 2006 –160 с.

2. Глас Дж., Стэнли Дж. Статистические методы в педагогике и психологии. -М.: Прогресс, 1976. -495 с.

3. Жалдак М.І., Кузьміна Н.М., Михалін Г.О. Теорія ймовірностей і математична статистика: Підручник для студентів фізико-математичних спеціальностей педагогічних університетів. – Вид. 2, перероб. і доп. / Жалдак М.І., Кузьміна Н.М., Михалін Г.О. – Полтава: «Довкілля-К», 2009 – 500 с.

4. Кадневский В.М. История тестов: Монография. М. : Народное образование, 2004. 464 с.

5. Ковальчук В.В., Мойсеев Л.М. Основи наукових досліджень: Навч.посібник.- 3-е вид., перероб. і допов. - К.: ВД „Професіонал“,2005. – 240 с.

6. Лукіна Т.О. Моніторинг якості освіти: теорія і практика / Лукіна Т.О.– К.: Вид. дім “Шкільний світ”: Вид. Л.Галіцина, 2006. –128 с. – (Б-ка “Шкіл. світу”).

7. Методичні рекомендації зі складання тестових завдань / В.П. Сергієнко, Л.О. Кухар. – К., НПУ, 2011. – 41 с.

8. Мойсеюк Н.Є. Педагогіка. Навчальний посібник. 3-є видання, доповнене, 2001 р. - 608 с

9. Моніторинг стандартів освіти / За ред. Альберта Тайджмана і Т.Невілла Послтвейна. - Львів: Літопис, 2003. - 328 с.

10.Моніторинг якості освіти: світові досягнення та українські перспективи. За заг. ред. О.І.Локшиної. – К.: К.І.С., 2004. – 128 с.

11.Моніторинг якості освіти: світові досягнення та українські перспективи / За заг. ред. О.І. Локшиної – К.: К.І.С, 2004. –128 с.

Додаток III

12. Навчально-методичний комплекс програм та документів щодо підготовки магістрів зі спеціальності 8. 18010022 «Освітні вимірювання» / За ред. професора В.П. Сергієнка – К.: Видавництво НПУ імені М.П. Драгоманова, 2011. – 164 с.

13. Навчально-методичний комплекс фахової підготовки бакалаврів спеціальності 6.040302 «Інформатика*» Спеціалізація «Освітні вимірювання» / За ред. В.П. Сергієнка. – Луцьк: 2010. – 86 с.

14. Новиков Д.А. Статистические методы в педагогических исследованиях (типовые случаи). - М.: МЗ-Пресс, 2004. – 67 с.

15. Педагогічне оцінювання і тестування. Правила, стандарти, відповідність. Наукове видання / Я.Я.Болюбаш, І.Є.Булах, М.Р.Мруга, І.В.Філончук.– К.: Майстер-клас, 2007.– 272 с.

16. Підласий І.П. Діагностика та експертиза педагогічних проєктів. Навчальний посібник для студентів вищих навчальних закладів. / Підласий І.П. – К.: Україна, 1998. – 344 с.

17. Руденко В.М., Руденко Н.М. Математичні методи в психології: підручник / В.М., Руденко Н.М. Руденко. – К.: Академвидав, 2009. – 384 с. (Серія «Альма-матер»). – ISBN078-966-8226-81-6.

18. Сергієнко В.П., Франчук В.М.. Методичні рекомендації зі створення тестових завдань та тестів у системі управління навчальними матеріалами MOODLE / Сергієнко В.П., Франчук В.М.. – К.: НПУ імені М.П. Драгоманова, 2012. – 58с.

19. Слєпкань З.І. Наукові засади педагогічного процесу у вищій школі: навч.посібник. -К.: Вища шк., 2005. - 239 с.: іл.

20. Франчук В.М., Тимченко А.А., Триус Ю.В., Стеценко І.В., Оксамитна Л.П., Заспа Г.О., Тупицький Д.П., Тьорло Ю.В., Герасименко І.В. Інформаційно-аналітична система контролю та оцінювання навчальної діяльності студентів ВНЗ. Монографія - Черкаси: МакЛаут, 2010. - 300 с.

21. Франчук В.М., Франчук Н.П. Використання системи управління навчальними ресурсами MOODLE у навчальному процесі. Матеріали міжнародної науково-практичної конференції "Освітні вимірювання в інформаційному суспільстві". - К.: НПУ, 2010. - 110 с.

22. Франчук В.М., Франчук Н.П. Система управління навчальними ресурсами MOODLE. Конструювання та

адміністрування тестів. ABSTRACTS. Second International Summer School "Education Measurements: Teaching, Research and Practice". - Nizhyn, 2010. .– 56–57 с.

23.Цехмістрова Г.С. Основи наукових досліджень. Навчальний посібник. – Київ: Видавничий Дім «Слово», 2003. – 240 с.

24.Чельшкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. /Чельшкова М.Б.- М.:Логос, 2002. - 432с

Дистанційні курси

1. <http://www.moodle.ii.npu.edu.ua>
 - Основи педагогічного оцінювання
 - Зовнішнє незалежне оцінювання
 - Конструювання тестів
2. <http://moodle.ndu.edu.ua>
 - Основи педагогічних вимірювань і моніторингу якості освіти
 - Комп'ютерні технології у тестуванні
 - Педагогічні вимірювання та моніторинг якості освіти
 - Математико-статистичні методи в педагогічних вимірюваннях

Додаток IV. Програма вступного фахового випробування для вступників за освітньо-кваліфікаційним рівнем «Магістр» спеціальності 8.18010022 «Освітні вимірювання» у КДПУ імені Володимира Винниченка

Пояснювальна записка

Фаховий вступний іспит з основ технологій освітньої діяльності до магістратури за спеціальністю 8.18010022 Освітні вимірювання передбачає визначення рівня засвоєння навчальних програм (що розкривають структуру цього іспиту) для різних природничо-математичних, технічних напрямів підготовки та ряду гуманітарних спеціальностей, розроблених згідно національних стандартів вивчення технологій навчальної діяльності у вищій школі. Вступний іспит проводиться з метою: а) організувати якісний та фаховий процес відбору осіб на здобуття освітньо-кваліфікаційного рівня магістр за цією спеціальністю; б) врахувати специфіку підготовки абітурієнтів (ОКР бакалавр) різних природничо-математичних, технічних напрямів підготовки та ряду гуманітарних спеціальностей. Тому випробування має базуватися на відповідних навчальних програмах бакалаврату, призначених для вищих навчальних закладів, які здійснюють підготовку фахівців з різних спеціальностей. Таим чином, вступне випробування з основ технологій освітньої діяльності до магістратури за спеціальністю 8.18010022 Освітні вимірювання визначає рівень володіння претендентами на навчання освітніми технологіями у контекстах:

- *Володіння інформаційно-комунікаційними технологіями* як складовою частиною освітніх технологій взагалі – поняття про апаратну частину ПК та про їх програмне забезпечення, уміння працювати з операційною системою, володіння текстовим редактором, електронними таблицями, системами управління базами даних, презентаціями, системами автоматичної обробки документів, системами автоматичної конвертації документів в електронний вигляд, зі скануванням зображень та їх сегментацією, з процесами розпізнавання тексту, з системами автоматичного

перекладу та перевірки правопису, з інформаційними технологіями Internet;

- *Володіння діловою українською мовою* – широта лексичного запасу, рівень володіння базовою українською граматику, навички української розмовної мови та культури усного спілкування; володіння сталими виразами та кліше, характерними для ділової української мови; рівень формування вміння працювати з особовими та інформаційними документами;

- *Володіння основами організації наукових досліджень* – знання і вміння щодо наукових досліджень, що є системоутворюючим фактором організації та проведення майбутнім магістром наукових досліджень у сфері предметної діяльності, а саме знання загальнотеоретичних положень наукових досліджень: методологічні принципи наукових досліджень і доцільність їх застосування в дослідницькій роботі; методи наукових досліджень та їх застосування в наукових дослідженнях; загальнотеоретична структура та підходи до проведення науково-дослідної роботи;

- *Володіння термінологією та закономірностями філософії* – розуміння світогляду як духовно-практичного способу освоєння світу, соціальної спрямованості філософського знання, духовно-творчої природи філософії, поняття про структуру процесу пізнання, про форми та методи пізнання, про проблему істини в теорії пізнання, тлумачення логіки як наука про засоби та форми мислення, поняття як форми мислення, судження та його логічної структури, використання законів логічного мислення, засвоєння співвідношення законів формальної та діалектичної логіки, поняття про умовиводи, їх роль та структуру, про типологію умовиводів, про доведення, їх роль та структуру;

- *Володіння фактами та причинно-наслідковими зв'язками з історії України* – знання та розуміння складного процесу формування та розвитку багатомільйонного українського народу, його діяльності в соціально-економічній, духовній, політичній і державній сферах з давніх-давен до сьогодення, розгляд історії України в тісному взаємозв'язку з глобальними історичними процесами, з історією її найближчих сусідів, з якими у різні часи українці перебували у складі різних держав;

Додаток IV

• *Володіння основами економічних знань* – трактування економіки, як сфери суспільного життя; розуміння змісту господарських відносин на певній території; поняття про виробничі ресурси, їх класифікацію, про виробництво як процес суспільної праці, про еволюцію виробництва, розуміння суті та структури економічної системи, знання основних типів економічних систем: ринкова економіка «вільної конкуренції», адміністративно-командна, змішана, традиційна системи; уявлення про сутність, передумови формування та функції ринку, про принципи ринкової економіки, про становлення і еволюцію грошових відносин, про підприємництво, його суть, принципи і форми функціонування, про циклічні коливання економічного розвитку;

• *Володіння елементами теорії ймовірностей та математичної статистики* – поняття про елементи комбінаторики, про основні поняття теорії ймовірностей: стохастичний експеримент (випадковий дослід), випадкова подія, ймовірність, про статистичне та класичне означення ймовірності, основні терміни та задачі математичної статистики, про можливі варіанти дослідження вибірки.

Питання з курсу ОНД

1. Охарактеризувати наукову діяльність, як інтелектуальну творчу діяльність.
2. Теоретичні та методологічні принципи науки.
3. Фундаментальні наукові дослідження.
4. Визначення мети, завдань, об'єкта і предмета дослідження.
5. Наукова інформація та способи її пошуку.
6. Комп'ютер як інструмент науково-дослідної роботи.
7. Пошук бібліографічної інформації в каталогах і картотеках.
8. Реферат як форма навчальної та науково-дослідної роботи.
9. Структура кваліфікаційної роботи.
10. Види науково-дослідної роботи студентів.

Питання з курсу Ділова українська мова

1. Офіційно-діловий стиль. Найважливіші риси, які визначають діловий стиль.
2. Основні вимоги до мовлення.

3. Класифікація документів.
4. Основні правила оформлення документів.
5. Документація щодо особового складу. Заява.
6. Документація щодо особового складу. Характеристика.
7. Документація щодо особового складу. Автобіографія.
8. Інформаційні документи. Доповідні і пояснювальні записки.
9. Обліково-фінансові документи. Доручення.
10. Обліково-фінансові документи. Розписка.

Питання з курсу Філософія

1. Філософія, її роль у житті людини і суспільства
2. Філософське розуміння буття
3. Матерія: генезис, сутність, основні властивості
4. Духовність, дух, душа, ідеальне. Поняття свідомості та її структура
5. Чуттєве та раціональне пізнання, їх форми та взаємозв'язок
6. Наукове пізнання, його рівні, форми та методи
7. Поняття як форми мислення
8. Судження та його логічна структура
9. Закони логічного мислення.
10. Умовиводи, їх роль та структура. Типологія умовиводів

Питання з курсу Історія України

1. Географічне положення, природні умови, населення сучасної України. національна символіка.
2. Київська Русь, її історичне значення.
3. Виникнення українського козацтва. Запорізька Січ: устрій, побут.
4. Початок українського національного відродження в Україні (перша половина XIX ст.) Т.Г.Шевченко.
5. Українська культура другої половини XIX ст.. Становлення театру корифеїв.
6. Встановлення Радянської влади в Україні. Політика воєнного комунізму.
7. Українське національно-культурне відродження 1920-х років.
8. Німецький окупаційний режим і рух опору в Україні.

Додаток IV

9. Україна на міжнародній арені. Основні напрямки зовнішньої політики в сучасних умовах.
10. Українська діаспора та Україна.

Питання з курсу Економіка

1. Економіка як сфера суспільного життя. Основна проблема економіки. Економічний зміст виробництва, розподілу, обміну і споживання.
2. Обмеженість виробничих ресурсів. Виробництво та продукт економічної діяльності.
3. Натуральне і товарне виробництво як історичні форми суспільного виробництва.
4. Основні типи економічних систем: ринкова економіка «вільної конкуренції», адміністративно-командна, змішана, традиційна системи.
5. Структура та інфраструктура ринку. Економічний механізм ринку.
6. Гроші у функціонуванні ринкової системи. Види грошей.
7. Доходи у ринковій економіці: заробітна плата, рента, позичковий відсоток і підприємницький прибуток.
8. Підприємництво, його суть, принципи і форми. Об'єктивні основи виникнення і розвитку підприємницької діяльності.
9. Макроекономічна нестабільність: циклічність економічного розвитку, безробіття, інфляція.
10. Макроекономічне регулювання: сутність, цілі, методи.

Питання з курсу Теорія ймовірностей та математична статистика

1. Елементи комбінаторики. Комбінаторні правила суми та добутку.
2. Перестановки, розміщення, комбінації без повторень.
3. Випадковий дослід. Випадкові події. Дії над подіями.
4. Частота події та її властивості. Статистичне означення ймовірності.
5. Класичне означення ймовірності. Властивості ймовірності.
6. Геометричні ймовірності.
7. Основні поняття та задачі математичної статистики.

8. Генеральна сукупність і вибірка. Варіаційний та статистичний ряд. Полігон частот.
9. Мода і медіана статистичного ряду. Середнє значення вибірки.
10. Інтервальний статистичний ряд. Гістограма. Середнє значення інтервального статистичного ряду.

Питання з курсу Інформатика

1. Редагування тексту засобами текстового процесору. Навести приклади.
2. Форматування тексту засобами текстового процесору. Навести приклади.
3. Правила пошуку та заміни фрагментів тексту та спеціальних символів у документі засобами текстового процесору.
4. Робота з таблицями засобами текстового процесору.
5. Робота з графічними об'єктами текстового процесору.
6. Створення та використання шаблонів текстового процесору.
7. Створення та використання стилів текстового процесору.
8. Створення автоматичного змісту документу засобами текстового процесору.
9. Вставка колонтитулів, нумерація сторінок засобами текстового процесору. Попередній перегляд та друк документу засобами текстового процесору.
10. Створення предметного покажчика засобами текстового процесору.
11. Уведення інформації в комірки електронної таблиці. Формати даних електронної таблиці.
12. Використання арифметичних операцій засобами табличного процесору. Уведення формул та функцій засобами табличного процесору.
13. Редагування та форматування даних в комірках табличного процесору.
14. Відносні, абсолютні та мішані адреси комірок табличного процесору.
15. Копіювання, перенесення та об'єднання комірок табличного процесору. Копіювання і перенесення формул та функцій табличного процесору.
16. Побудова діаграм засобами табличного процесору.

Додаток IV

17. Використання автофільтру електронної таблиці. Використання розширеного фільтру електронної таблиці.
18. Сортування даних засобами табличного процесору.
19. Призначення та створення комп'ютерної презентації.
20. Копіювання (переміщення) слайдів комп'ютерної презентації. Вставка графічних об'єктів комп'ютерної презентації.
21. Оформлення слайдів комп'ютерної презентації за допомогою кольорів, фону, шаблонів.
22. Налаштування анімації, звука та порядку слідування об'єктів комп'ютерної презентації. Демонстрація комп'ютерної презентації.
23. Вставка й використання керуючих кнопок комп'ютерної презентації.
24. Вставка гіперпосилань комп'ютерної презентації.
25. Основні служби (сервіси) *Internet*. Навести приклади. Використання Web-браузера для завантаження та пошуку Web-документів.
26. TCP/IP протокол, IP-адреси. Навести приклади.
27. HTTP протокол, URL-адреси, доменні адреси комп'ютерів в *Internet*. Навести приклади.
28. Пошукові системи *Internet*. Прийоми ефективного пошуку даних.
29. Команди простого пошуку пошукових систем. Навести приклади.
30. Команди (оператори) мови запитів пошукових систем. Навести приклади.

**Додаток V. Програма асистентської практики
магістрів спеціальності 8.18010022
«Освітні вимірювання» у КДПУ імені
Володимира Винниченка**

Передмова

Соціально-економічні та науково-технічні зміни, які відбуваються у суспільстві, вимагають перебудови як освіти в цілому, так і професійної освіти зокрема. Час ставить завдання підготовки спеціаліста, здатного ініціативно, творчо й по-науковому мислити, вміти самостійно поповнювати свої знання та успішно використовувати їх у своїй практичній діяльності, що дасть змогу йому досить ефективно працювати в нових умовах, забезпечуючи розвиток суспільства. На підготовку високоосвічених фахівців з освітніх вимірювань і спрямована програма асистентської практики магістрантів.

Практична підготовка студентів є обов'язковим компонентом освітньо-професійної програми для здобуття кваліфікаційного рівня «магістр», є логічним продовженням теоретичного навчання магістра і завершальним етапом його професійно-педагогічної підготовки. Вона проводиться у навчальних закладах II-IV рівнів акредитації в умовах максимально наближених до реальної майбутньої професійної діяльності.

Мета практики – підготовка до цілісного виконання функцій фахівця з освітніх вимірювань, викладача дисциплін з освітніх вимірювань, до проведення системи навчально-виховної роботи з студентами, формування у студентів-практикантів досвіду педагогічної та оціночної діяльності, виховання морально-етичних якостей викладача вищого навчального закладу, індивідуального творчого стилю педагогічної діяльності, потреби в самоосвіті.

Особливістю асистентської практики є те, що крім педагогічної складової (магістранти вникають у коло реальних проблем професійної праці викладача, вивчають зміст і обсяг його роботи), практиканти повинні виконати ряд завдань освітньої складової програми практики, що стосуються моніторингу якості

Додаток V

навчальних досягнень та рівня підготовки студентів, створення банку тестових завдань, їх калібрування.

Перед початком асистентської практики проводиться настановна конференція, на якій магістрам повідомляють мету практики, пояснюють її завдання та зміст, а також порядок проходження. По закінченню практики проводиться підсумкова конференція за результатами всіх видів робіт, виконаних магістрантами.

До керівництва педагогічною практикою від університету залучаються досвідчені викладачі, які мають стаж роботи у вищих навчальних закладах не менше 3-х років.

Відповідно до мети ставляться завдання практики:

- а) поглибити й закріпити теоретичні знання набуті студентами в університеті й навчити застосовувати ці знання на практиці в навчально-виховній роботі з студентами (учнями, слухачами курсів);
- б) озброїти практикантів умінням спостерігати й аналізувати навчально-виховну роботу, яку проводять з студентами;
- в) навчити магістрантів, спираючись на знання фахових методик, психології, педагогіки, проводити навчально-виховну роботу з студентами з урахуванням їхніх вікових та індивідуальних особливостей, працювати з студентським колективом, формувати комунікативні уміння, толерантність та ін..;
- г) підготувати практикантів до самостійного проведення практичних (лабораторних) занять із застосуванням різноманітних методів, що активізують пізнавальну діяльність студентів;
- д) навчити магістрантів використовувати на практиці тестові технології оцінювання, створювати критерії оцінювання;
- е) удосконалити уміння практикантів конструювати тематичні завдання з дисциплін фаху у тестовій формі з закритими відповідями, проводити тестування слухачів курсів або студентів (учнів), здійснювати математико-статистичну обробку результатів тестування та їх аналіз, навчити правильного оформлення відповідної документації, включаючи статистичну довідку за отриманими результатами проведеного тестування;

- є) розвивати й закріплювати у студентів повагу до педагогічної професії викладача, стимулювати прагнення до вивчення спеціальних і педагогічних дисциплін, удосконалення своїх педагогічних здібностей для підготовки до творчого розв'язання завдань національної програми “Освіта” (“Україна XXI століття”).

Загалом асистентську практику можна орієнтовно поділити на *три етапи*:

- 1) **підготовчий** – ознайомлення студентів з програмою практики (мета, завдання, зміст практики), формою ведення звітної документації; знайомство з студентами, куратором групи, за якою закріплений студент; адаптація у студентському колективі;
- 2) **основний** – робота магістрантів на посаді викладача дисциплін з освітніх вимірювань, виконання функцій фахівця з освітніх вимірювань, а також науково-дослідна робота, пов'язана з педагогічною та оціночною діяльністю;
- 3) **підсумковий** – підбиття підсумків асистентської практики, оформлення та здача звітної документації, оцінка виконаної магістрантами роботи загалом та кожним практикантом зокрема.

1. ЗАГАЛЬНІ ЗАВДАННЯ ТА ВИМОГИ

На *VI курсі* студенти проходять *асистентську практику* в навчальних закладах II-IV рівнів акредитації як викладачі дисциплін з освітніх вимірювань та фахівці з освітніх вимірювань. Студенти розподіляються на практику по одному (у окремих випадках по 2-3) в групу.

Тривалість практики на *VI курсі* – 4 тижні.

Асистентська практика передбачає (на I тиждні):

1. Обов'язкову присутність студента на настановній конференції з питань асистентської практики.
2. Зустріч з адміністрацією факультету, закріплення практиканта за групою, знайомство його з студентами, куратором групи.
3. Знайомство із специфікою навчально-виховного процесу даного навчального закладу.
4. Вивчення плану роботи викладача з предмету (навчальний і робочий плани, а також плани роботи гуртків, факультативів, проблемних і методичних груп, об'єднань, семінарів тощо).
5. Визначення та опрацювання літератури, знання якої необхідні для викладання відповідної дисципліни.
6. Відвідування й аналіз усіх занять з дисциплін освітніх вимірювань, окремих занять з інших дисциплін, годин куратора у закріпленій групі для вивчення студентського колективу.
7. Ознайомлення з досвідом педагогічної та оціночної діяльності викладачів, вивчення різних методів та прийомів активізації навчально-пізнавальної діяльності студентів.
8. Вивчення документації по організації та проведенню тестування та специфіки математико-статистичної обробки його результатів, плану написання статистичної довідки.
9. Складання індивідуального плану навчально-методичної та виховної роботи *на весь період практики* та затвердження його у групового методиста.

Магістранти під час практики повинні виконати такі завдання:

- протягом перших 2 днів практики скласти індивідуальний план роботи на весь період практики, ознайомитися з групою студентів (учнів, слухачів курсів), з якими вони будуть

працювати, з документацією, що веде й оформляє викладач кафедри;

- протягом практики вести щоденник психолого-педагогічних спостережень;
- розробити конспект лекції з дисципліни, передбаченої навчальним планом спеціалізації чи спеціальності «Освітні вимірювання», та конспекти 2-3 практичних (семінарських, лабораторних) занять, узгодити їх з фаховим методистом практики (методист затверджує конспекти і дає дозвіл на проведення заняття, або його частини);
- провести 2-3 практичних (семінарських, лабораторних) заняття та зробити їх самоаналіз;
- відвідати 3-4 заняття студентів-практикантів, взяти участь у їх обговоренні;
- провести позакласне заняття, пов'язане з підготовкою студентів (учнів, слухачів курсів) до тестування;
- сконструювати тематичні завдання з дисциплін фаху у тестовій формі з закритими відповідями, провести тестування слухачів курсів або студентів (учнів), здійснити математико-статистичну обробку результатів тестування та їх аналіз;
- написати статистичну довідку за отриманими результатами проведеного тестування;
- написати звіт про проходження практики.

Всі заняття, проведені студентами оцінюються з позицій:

- вміння самостійно підготуватися та провести заняття;
- досягнення мети заняття;
- вміння творчо підходити до організації та проведення занять, використання інноваційних технологій навчання;
- вміння володіти аудиторією;
- техніка мови;
- вміння зробити підсумок заняття.

Присутність керівника практики на кожному занятті, що проводить студент-практикант є обов'язковою, як і подальший їх аналіз.

Тестування, проведене студентами-практикантами, оцінюється з урахуванням: організаційно-технологічного забезпечення, проведення тестування, структурування тестових

Додаток V

завдань, якості математико-статистичної обробки результатів тестування, їх аналізу, якості написання статистичної довідки.

Студенти також працюють над дипломною роботою, постійно консультуються з науковим керівником, періодично звітують про виконану роботу з теми дослідження.

Звітна документація. Після закінчення асистентської практики, магістранти повинні здати на випускову кафедру таку документацію:

1. Індивідуальний план роботи студента-практиканта.
2. Щоденник психолого-педагогічних спостережень.
3. Текст розробленої тематичної лекції та розгорнуті плани-конспекти практичних (семінарських, лабораторних) занять, проведених у вузі з підписом та оцінкою методиста з фаху.
4. Самоаналіз проведених занять.
5. Конспект позакласного заходу.
6. Сконструйовані завдання у тестовій формі, їх математико-статистичну обробку, аналіз та зроблені висновки.
7. Статистичну довідку, написану за результатами виконаного тестування.
8. Звіт про проходження практики.
9. Витяг із засідання кафедри (фахової циклової комісії) вузу II-IV рівня акредитації з аналізом результатів проходження науково-педагогічної практики магістрантом та указанням оцінки у вигляді цифрового балу.

Відгук про роботу магістранта складає фаховий методист, який перевіряє здану практикантом документацію та виставляє загальну оцінку з асистентської практики, враховуючи:

- 1) оцінку якості проведених занять,
- 2) оцінку організації і проведення тестування та аналізу його результатів,
- 3) оцінку проведеної позааудиторної навчально-виховної роботи з студентами (учнями),
- 4) оцінку за оформлення документації,
- 5) оцінку за ставлення до практики.

Підсумки практики. Підведення підсумків проводиться на засіданні випускової кафедри та на підсумковій науково-практичній конференції з асистентської практики, в якій обов'язково беруть участь всі магістранти, які пройшли практику у

відповідних закладах освіти. На підсумкову конференцію з асистентської практики студенти готують тематичні доповіді, в яких розкривають вивчений ними та власне здобутий досвід педагогічної та оцінної діяльності, своїх фахових знахідок й пропозицій щодо поліпшення проведення практики

За результатами всіх видів робіт, які виконує студент під час практики, виставляється *диференційований залік*. Загальна оцінка за наслідками асистентської практики не виставляється у випадку невиконання хоча б одного із завдань.

При виставленні загальної оцінки враховується характеристика (оцінка), яку дає фаховий методист, та результати науково-дослідної роботи студента, які можуть бути використані ним для виконання дипломної роботи, для підготовки доповіді, виступу на підсумковій науково-практичній конференції з практики, науково-методичної статті тощо.

Обов'язки магістранта в період асистентської практики

1. Дотримуватися всіх правил внутрішнього розпорядку навчального закладу, в якому проходить практика, розпоряджень керівників закладу й асистентської практики, методистів факультету. У разі невиконання цих вимог студент може бути не допущеним до подальшого проходження асистентської практики.
2. Відповідно до програми практики скласти індивідуальний план навчально-виховної роботи з студентами (учнями, слухачами курсів). План затверджується фаховим та груповим методистом.
3. Працювати в навчальному закладі щоденно не менше шести годин згідно з індивідуальним планом. Виконувати всі види робіт, передбачених програмою практики. У разі відсутності (перебування на конференції, хвороба, тощо) обов'язково попередити керівників практики.
4. Щоденно вести відповідну документацію, оформити та вчасно здати звітну документацію.

Обов'язки фахового методиста асистентської практики студентів-магістрантів

Методист з фаху асистентської практики повинен:

Додаток V

- брати участь у настановному інструктажі студентів на факультеті (перед початком практики) і в підсумковій конференції (при завершенні практики);
- у випадку неявки студентів на практику, порушення ними трудової дисципліни, нехтування своїми обов'язками, а також виникненні конфліктних ситуацій повідомити про це письмово факультетського керівника практикою, завідувача випускової кафедри і деканат;
- планувати навчально-виховну роботу студентів-практикантів, надавати методичну допомогу у підготовці і проведенні занять за фахом, організації роботи із студентським колективом, позанавчальною роботою тощо;
- проаналізувати конспекти занять, розроблені практикантом, прийняти рішення про допуск практиканта до проведення заняття або його частини;
- відвідувати заняття студентів-практикантів (практичні, семінарські, лабораторні заняття та ін.), брати участь у їх обговоренні, оцінювати і виставляти відповідний бал у заліковій лист проходження практики;
- надавати практикантам допомогу у виборі теми виховного заходу, розробці плану-конспекту, завізувати його (до початку проведення), бути присутнім на заліковому виховному заході, брати участь у його обговоренні й виставити диференційовану оцінку за його проведення і оформлення у заліковий лист;
- надавати консультації щодо конструювання тематичних завдань з дисциплін фаху у тестовій формі з закритими відповідями, організації та проведення тестування, математико-статистичної обробки результатів тестування, їх аналізу, написання статистичної довідки; оцінити роботу практиканта, як фахівця з освітніх вимірювань;
- *(якщо він є науковим керівником магістерської роботи практиканта)* під час асистентської практики контролювати роботу студентів та надавати методичну допомогу в організації та зборі експериментального матеріалу для магістерської роботи, консультації щодо виконання індивідуального науково-дослідного (творчого) завдання, оцінити його;

- надавати консультації щодо оформлення звітної документації, враховуючи терміни її здачі, оцінити якість її оформлення практикантом;
- спільно з іншими керівниками узгодити підсумкову оцінку;
- на засіданні кафедри доповідати про стан проходження, а також про підсумки практики.

ДОДАТКИ

Додаток 1

Форма індивідуального плану

Індивідуальний план роботи студент складає протягом першого тижня проходження асистентської практики. Добре продуманий індивідуальний план свідчить про самостійність та активність практиканта. Чітке виконання індивідуального плану дисциплінує студента, організовує його діяльність, привчає до порядку.

Титульна сторінка індивідуального плану роботи має такий вигляд:

Фаховий методист
(прізвище та ініціали)
“Затверджую” (підпис)
“ ___ ” _____ 20__р.

Індивідуальний план роботи
студента-практиканта VI курсу
спеціальності 8.18010022 «Освітні вимірювання»
фізико-математичного факультету
КДПУ імені Володимира Винниченка
(прізвище, ім'я, по батькові)

На першій сторінці плану магістрант вказує:

Заклад освіти _____
Група _____
Викладач з фаху _____
Куратор групи _____
Термін практики з _____ до _____

Додаток V

Далі фіксує необхідні відомості про групу (студентський колектив), розклад дзвінків, розклад занять.

Потім розділяє кожну сторінку на 6 днів, вказує дні та числа, зміст роботи, відмітки про виконання.

У плані треба подати такі напрямки:

1. Навчальна робота (відвідування, підготовка й проведення занять, виготовлення наочності, дидактичних матеріалів, перевірка письмових робіт студентів і т.ін.).
2. Позааудиторна робота (підготовка й проведення додаткових занять, консультацій, засідань гуртків тощо).
3. Виховна робота (бесіди з студентами, відвідування, підготовка й проведення годин куратора, диспутів та ін.).
4. Робота фахівця з освітніх вимірювань (конструювання тематичних завдань з дисциплін фаху у тестовій формі з закритими відповідями, організація й проведення тестування слухачів курсів або студентів (учнів), здійснення математико-статистичної обробки результатів тестування та їх аналіз, написання статистичної довідки за отриманими результатами проведеного тестування).
5. Наукова робота (вивчення передового досвіду викладачів, проведення психолого-педагогічних спостережень, педагогічних експериментів, аналіз результатів анкетування й тестування учнів і т.ін.).
6. Настановна та підсумкова конференції (участь, виступ, оформлення документації тощо).

Додаток 2

Вимоги до оформлення плану-конспекту заняття

Титульна сторінка плану-конспекту має такий вигляд:

Фаховий методист

(прізвище та ініціали)

“Затверджую”

(підпис)

“___” _____ 20__р.

КОНСПЕКТ
 ПРАКТИЧНОГО (ЛАБОРАТОРНОГО) ЗАНЯТТЯ
 з _____ у _____ групі
 навчального закладу освіти _____
 студента-практиканта VI курсу
 спеціальності 8.18010022 «Освітні вимірювання»
 фізико-математичного факультету
 КДПУ імені Володимира Винниченка
 (прізвище, ім'я, по батькові)

Дата проведення

“ ___ ” _____ 20_р.

Примітка: без перевірки конспекту викладачем або методистом студент до заняття не допускається.

План-конспект може мати таку схему:

1. Тема заняття.
2. Цілі й завдання заняття (освітні, виховні, розвиваючі).
3. Обладнання (наочність, ТЗН тощо).
4. Тип заняття.
5. Структура заняття (план).
6. Зміст заняття (послідовний опис роботи на уроці за його етапами з нотуванням відведеного часу в хвиликах).

I. Організація аудиторії (1–2 хв.).

II. Перевірка домашнього завдання (... хв.).

Форма роботи, формулювання питань і очікуваних відповідей на них з указанням прізвищ студентів, яких намічено запитати.

III. Актуалізація опорних знань студентів (... хв.).

IV. Формування умінь та навичок, досвіду діяльності.

Вироблення умінь розв'язувати задачі, формування навичок самоконтролю. Опис вправ з усіма можливими варіантами їх розв'язання. Вказати форму організації діяльності студентів під час виконання вправ, розв'язування задач тощо.

VI. Завдання додому (... хв.).

Вказується, **що** і **як** буде задаватися. Усі запропоновані студентам задачі й вправи повинні бути розв'язані. Розв'язання чітко й правильно оформлене.

Додаток V

VII. Підбиття підсумків заняття (... хв.).

(З'ясувати, чи досягнута мета заняття, чи розв'язані поставлені завдання.)

7. Використана література.

Підпис практиканта _____

Підпис викладача _____

Додаток 3

Орієнтовна схема аналізу заняття

I. Загальні відомості про заняття.

Тема заняття. Цілі й завдання заняття. Обладнання. Тип заняття.

II. Початок заняття. Організаційна частина.

1. Якими прийомами викладач-практикант забезпечив добрий початок заняття (впевненість, доброзичливість, чіткість, ввічливість, вимогливість і т.ін.).
2. У чому проявився невдалий початок заняття, які причини цього (непідготовленість викладача, неорганізованість його, метушливість, в'ялість тощо).
3. Увага студентів на початку заняття (привітання, робоча поза, організація робочого місця викладача й студентів тощо).

III. Перевірка домашнього завдання.

1. Методика проведення перевірки домашнього завдання: форми організації діяльності студентів (фронтальна, групова, індивідуальна), види опитування (перевірка наявності домашнього завдання, математичний диктант тощо).

Постановка питань – чіткість, правильність, повнота, однозначність, лаконічність.

Відповідь студентів – правильність, повнота, самостійність, чіткість, ясність, культура мови, розуміння.

Увага студентів під час опитування, їх діяльність – доповнення до відповіді, виправлення її. Прийоми активізації діяльності студентів, що використовує викладач.

Увага викладача до відповідей студентів.

2. Ступінь виявлення знань, навичок й умінь студентів – всебічність, глибина.
3. Правильність, об'єктивність оцінки викладачем відповідей студентів. Мотивування оцінки.
4. Узагальнення викладачем відповідей студентів.

IV. Закріплення нових знань, формування навичок та умінь, досліду діяльності.

1. Форма закріплення вивченого матеріалу (просте повторення, питання, розв'язування задач, самостійна робота, виконання вправ, робота з літературою тощо).
2. Активізація викладачем самостійності та пізнавальної активності студентів у момент закріплення (порівняння, зв'язок з раніше вивченим, висновки, узагальнення та ін.).
3. Оцінка одержаної по каналу “зворотного зв'язку” інформації про ступінь розуміння й засвоєння студентами групи навчального матеріалу. Організація роботи із слабкими, середніми й здібними студентами по закріпленню вивченого матеріалу.

V. Завдання додому.

1. Характер і зміст теоретичного й практичного завдання; посиленість та відповідність змісту заняття, інші завдання (поточне повторення), інші види завдань.
2. Диференціація завдань за ступенем складності.
3. Місце подачі домашнього завдання на занятті, інструктивні вказівки викладача.

VI. Підсумок заняття.

Підсумки проведеної роботи на занятті, вчасність закінчення заняття, організованість студентів у кінці заняття.

VII. Аналіз діяльності викладача-практиканта на уроці.

Знання фактичного матеріалу; вміння володіти увагою аудиторії, організовувати роботу, підвищувати активність, інтерес, увагу, тримати дисципліну, враховувати індивідуальні особливості кожного студента; стиль і тон у роботі (жвавий, рішучий, повільний, байдужий); педагогічний такт; спостережливність, вдумливість, витримка, винахідливість, емоційність; зовнішній вигляд, охайність; культура мовлення, поза, міміка, жестикуляція.

VIII. Аналіз діяльності студентів на уроці.

Додаток V

Дії студентів щодо досягнення мети; ставлення до викладача; пізнавальний інтерес, рівень працездатності; вміння виділяти провідні ідеї, аналізувати, оцінювати підсумки своєї діяльності та одногрупників, розвиток мислення; культура розумової праці; поведінка, дисципліна, чим пояснюється.

IX. Загальні висновки.

1. Чи досягнуті цілі заняття. Раціональність використання часу.
2. Відповідність заняття сучасним вимогам.

X. Висновки і пропозиції.

Досягнення мети, чи оптимальним шляхом йшов процес її досягнення; виконання плану заняття; виховне, освітнє і розвиваюче значення заняття; що цінного в цьому. Поради викладачеві, як закріпити й удосконалити позитивні моменти заняття чи подолати недоліки.

Критерії оцінки заняття

Заняття, проведені студентами-практикантами оцінюються за вузівською (4-бальною) системою.

Оцінка **“відмінно”** ставиться за заняття, проведене на високому науковому та методичному рівні, тобто за заняття, на якому:

- 1) повністю розв’язані усі поставлені перед ним навчально-виховні завдання;
- 2) використані оптимальні методи й прийоми роботи, сучасні технології навчання;
- 3) усі студенти засвоїли теорію й навчилися її застосовувати, у них сформовані навички й уміння з теми;
- 4) максимально активізувалася навчально-пізнавальна діяльність студентів, їх самостійна робота;
- 5) витримана розмірність частин уроку, своєчасно задана домашня робота;
- 6) використані необхідні й можливі наочні посібники й технічні засоби навчання;
- 7) об’єктивно виставлені й аргументовані оцінки;

- 8) практикант не допустив жодної фактичної, методичної та психолого-педагогічної помилки і показав високу загальну, педагогічну й мовну культуру.

Відмінно проведене заняття – це заняття яскраве, живе, оригінальне, емоційне, заняття, яке не залишає студентів байдужими.

Оцінка **“добре”** ставиться за заняття, проведене на високому науковому та належному методичному рівні, тобто за ефективне, змістовне заняття, на якому:

- 1) повністю розв’язані навчально-виховні завдання;
- 2) використані традиційні методи й прийоми роботи;
- 3) усі студенти засвоїли теорію й в деякій мірі навчилися її застосовувати, тобто у них сформовані певні навички й уміння з теми;
- 4) активізувалася самостійна робота студентів;
- 5) використані необхідні наочні посібники й технічні засоби навчання;
- 6) студент допустив незначні фактичні й методичні помилки;
- 7) недостатньо витримана розмірність частин уроку (затянуте опитування, несвоєчасно задане домашнє завдання; нечітко підведений підсумок заняття, не завжди переконливо аргументовані оцінки).

Добре проведене заняття – це заняття, що задовольняє усім основним вимогам, які пред’являються до відмінного заняття, але відрізняється недостатньою чіткістю у роботі практиканта, незначними упущеннями організаційного характеру.

Оцінка **“задовільно”** ставиться за заняття, проведений на належному науково-методичному рівні, на якому:

- 1) основні навчально-виховні завдання розв’язані, мета заняття загалом досягнута;
- 2) студенти зрозуміли новий матеріал і набули необхідних навичок й умінь;
- 3) витримана розмірність основних частин заняття;
- 4) використовувалися наочні посібники й технічні засоби навчання;
- 5) в основному правильно виставлені й прокоментовані оцінки студентам;

Додаток V

- б) несвоєчасно й не досить чітко задане домашнє завдання;
- 7) не підведений підсумок уроку;
- 8) спостерігалися незначні порушення вимог педагогіки й психології та методичні помилки, що знижували ефективність заняття.

Задовільне заняття – це заняття в принципі ефективне, яке дало позитивний результат; це заняття, під час проведення якого практикант показав добрі знання предмету й уміння правильно організувати навчально-виховну роботу. У той же час на занятті допущені деякі незначні порушення вимог педагогіки, психології й методики, робота на занятті носить не досить чіткий характер.

Оцінка **“незадовільно”** ставиться за неефективне, слабе, незадовільне заняття, тобто заняття, на якому:

- 1) поставлені навчально-виховні завдання не розв’язані (студенти не зрозуміли матеріал, не набули необхідних навичок й умінь з теми);
- 2) практикант допускав грубі фактичні, методичні та інші помилки;
- 3) не активізувалася робота студентської аудиторії; порушена розмірність основних частин заняття, а також допускалися інші помилки, що негативно вплинули на його результати.

Проведення заняття акумулює у собі всі навчальні досягнення студента-практиканта. Тому аналіз та оцінка занять, проведених практикантом – це аналіз та оцінка його знань і навичок з предмету, методики його викладання, педагогіки й психології, аналіз та оцінка його умінь реалізовувати ці знання й навички на практиці. Тому під час оцінювання уроку враховується весь комплекс знань, навичок й умінь, ерудиція студента, рівень його професійно-педагогічної, мовної та загальної культури. Корисно, щоб заняття обговорювалися й аналізувалися при участі групи магістрів-практикантів, тому одним із завдань практики є відвідування практикантом занять, що проводять інші практиканти, та участь у їх обговоренні. Дуже важливою в процесі аналізу є правильна оцінка студентом своєї роботи, вміння його самостійно визначати позитивне та побачити недоліки в реалізації розробленого заняття

Орієнтовна схема конспекту години куратора

Титульна сторінка конспекту має такий вигляд:

Фаховий методист
(прізвище та ініціали)
“Затверджую” (підпис) .
“ ___ ” _____ 20__р.

КОНСПЕКТ ГОДИНИ КУРАТОРА
проведеної у _____ групі
навчального закладу освіти _____
студента-практиканта VI курсу
спеціальності 8.18010022 «Освітні вимірювання»
фізико-математичного факультету
КДПУ імені Володимира Винниченка
(прізвище, ім'я, по батькові)

Дата проведення
“ ___ ” _____ 20__р.

Конспект години куратора (навчально-виховного заходу)
може мати таку схему:

1. Тема, форма заходу (бесіда, диспут тощо).
2. Обґрунтування обраної теми й форми роботи (актуальність, відповідність інтересам, запитам та рівню інтелектуального розвитку студентів).
3. Цілі й завдання заходу (освітні, виховні, розвиваючі).
4. Обладнання (наочність, ТЗН тощо).
5. Організація підготовки й проведення заходу.
6. Зміст (хід проведення заходу).
7. Методика проведення заходу.
8. Рекомендована література для студентів.
9. Використана студентом-практикантом література.
10. Самоаналіз проведеного заходу.

**Звітна документація
щодо організації та проведення тестування
повинна налічувати:**

1. інструкцію для викладача-практиканта, що проводитиме тестування, яка повинна містити пояснення процедури, режиму, правил проведення тестування;
2. інструкцію для студентів (учнів, слухачів курсів), що тестуються, яка повинна містити пояснення процедури, режиму, правил виконання тестових завдань, умов оформлення робіт;
3. сконструйовані тематичні завдання з дисциплін фаху у тестовій формі з закритими відповідями (та в іншій формі) разом з бланками для відповідей на них;
4. змістовний опис етапів математико-статистичної обробки результатів тестування, їх аналіз;
5. список використаної студентом-практикантом літератури;
6. статистичну довідку, написану за результатами проведеного тестування.

У статистичній довідці слід вказати інформацію про організацію й проведення тестування:

1. Тема, форма тестування (комп'ютерне, письмове тощо).
2. Цілі й завдання тестування.
3. Обладнання (наочність, технічні засоби навчання, програмне забезпечення тощо).
4. Вид інструктажу (усний, письмовий, ін.).
5. Типи (види) тестових завдань.
6. Результати тестування у вигляді оціночних балів та їх процентного співвідношення (бажано також проілюструвати графічно).
7. Загальні висновки (чи вдало сконструйовані тестові завдання; рівень підготовки тестованих; вказати завдання, які виконали переважна більшість (меншість) учнів (студентів чи слухачів курсів); чи були конфліктні ситуації під час тестування чи при підготовці до нього; пропозиції, побажання практиканта та ін.).

Звіт студента-практиканта

Кожен студент на завершення асистентської практики пише звіт про свою роботу, який, разом з іншою звітною документацією, характеризує ступінь виконання завдань асистентської практики.

Звіт студента-практиканта складається з таких розділів: навчальна робота, виховна робота, робота фахівця з освітніх вимірювань, висновки. Обов'язково перевіряється методистами асистентської практики.

Титульна сторінка звіту має такий вигляд:

Фаховий методист
(прізвище та ініціали)
“Затверджую” (підпис) .
“ ___ ” _____ 20__р.

ЗВІТ

про виконану роботу
студента(ки)-практиканта(ки) VI курсу
спеціальності 8.18010022 «Освітні вимірювання»
фізико-математичного факультету
КДПУ імені Володимира Винниченка
(прізвище, ім'я, по батькові)
що проходив(ла) асистентську практику
у групі №__ навчального закладу _____
у період з _____ до _____ 20__року

Кіровоград 20__

Орієнтовна схема звіту практиканта

I. Навчальна робота

1. Загальні відомості про місце і час проходження асистентської практики, коротка характеристика навчального закладу (випускової кафедри): професорсько-викладацький склад, обладнання, особливості навчально-виховного процесу.
2. Характеристика групи, в якій студент проходив практику. Успішність з математичних та ін. дисциплін. Відповідність рівня знань, умінь і навичок програмовим вимогам. Причини слабкої успішності окремих студентів.
3. Відвідання занять: а) із спеціальності; б) з інших предметів.
4. Планування і проведення занять студентами:
 - а) кількість проведених занять, їх критична оцінка: чи вдалося досягти поставлених цілей, завдань заняття (якщо ні, вказати причину); труднощі, які були при плануванні, у підготовці та в процесі проведення заняття або частини його, як їх подолали;
 - б) засоби, що використовувалися для активізації розумової діяльності студентів на уроці;
 - в) які навчальні матеріали виготовив, оцінка їх та роль цих матеріалів на занятті.
5. Якість перевірки письмових робіт.
6. Робота з студентами, що не встигають з предметів (як ви оцінюєте результати додаткових занять, їх кількість, скільки студентів було охоплено).
7. Індивідуальна та диференційована робота з студентами, її результати.
8. Позааудиторна робота з фаху (тематика та форми проведених заходів, їх кількість; ознайомлення з роботою гуртка, проведення додаткових занять, консультацій та ін.). Якість проведеної роботи.
9. Труднощі у виконанні завдань навчальної роботи, шляхи їх подолання.
10. Оволодіння уміннями і навичками проведення навчальної роботи.
11. Загальні висновки про навчальну роботу.

II. Виховна робота

1. Група, до якої прикріплений студент, особливості планування роботи.
2. Робота, проведена на допомогу куратору, студентському колективу (її конкретні форми й види).
3. Тематика, якість проведених залікових виховних заходів.

4. Відвідання позакласних виховних заходів, проведених кураторами й практикантами.
5. Труднощі у підготовці й проведенні виховної роботи у групі.
6. Оволодіння уміннями й навичками організації і проведення виховної роботи.
7. Загальні висновки про виховну роботу.

III. Робота фахівця з освітніх вимірювань

1. Бесіди, проведені практикантом під час організації й проведення тестування у групі (вказати, які саме).
2. Розробка й підготовка відповідної документації (інструкції, бланки відповідей та ін. (вказати конкретно).
3. Використання технічних засобів навчання, пакетів прикладних програм тощо.
4. Якість сконструйованих тематичних завдань з дисциплін фаху у тестовій формі з закритими відповідями (оцінюється за результатами математико-статистичної обробки).
5. Методика проведення тестування слухачів курсів або студентів (учнів).
6. Математико-статистичні методи обробки результатів тестування (вказати конкретно).
7. Оформлення статистичної довідки за отриманими результатами проведеного тестування (які труднощі виникали при написанні, оцініть її якість)
8. Який матеріал дібрано для написання кваліфікаційної роботи з фаху.
9. Вивчення психолого-педагогічної, методичної, математико-статистичної літератури під час практики, результати опрацювання її.

IV. Висновки

1. Що дала асистентська практика магістранту.
2. Позитивне в організації і проведенні практики.
3. Недоліки (конкретно які), виявлені під час практики, шляхи їх усунення (Ваші пропозиції випусковій кафедрі).

Особистий підпис _____

Підпис викладача _____

Підпис куратора групи _____

Підпис методиста з фаху _____

**Додаток VI. Питання до державного екзамену з
освітніх вимірювань для
магістрантів спеціальності
8.18010022 «Освітні вимірювання» у
КДПУ імені Володимира Винниченка**

Вимірювання в освіті

1. Педагогічне вимірювання. Латентні змінні. Компоненти процесу педагогічних вимірювань. Інструмент вимірювання.
2. Типи вимірювання. Основні характеристики нормо-орієнтованого та критеріально-орієнтованого вимірювання. Відмінності між нормо-орієнтованими і критеріально-орієнтованими тестами.
3. Класифікація навчальних цілей. Таксономія Блума.
4. Поняття завдання у тестовій формі, тестового завдання, тесту. Вимоги до завдань у тестовій формі. Форми тестових завдань. Переваги та недоліки різних форм тестових завдань.
5. Завдання з вибором однієї та декількох правильних відповідей. Принципи формулювання та оцінювання.
6. Завдання на встановлення відповідності та правильної послідовності. Завдання відкритої форми з короткою та розгорнутою відповіддю. Принципи формулювання та оцінювання.
7. Поняття моніторингу. Комплекс моніторингових показників (індикаторів). Поняття якості освіти, її структура. Основні групи показників якості освіти.
8. Поняття моніторингу якості освіти. Моніторинг як кваліметричний інструмент. Етапи проведення моніторингового дослідження.

Моніторинг якості освіти

1. Поняття якості освіти, її структура. Основні групи показників якості освіти.
2. Поняття моніторингу якості освіти. Комплекс моніторингових показників (індикаторів). Міжнародні системи критеріїв і показників якості шкільної освіти.
3. Моніторинг як кваліметричний інструмент. Етапи проведення моніторингового дослідження.

4. Міжнародні моніторингові порівняльні дослідження якості освіти. Організатори міжнародних досліджень якості освіти (TIMSS, PISA, PIRLS, SITES).

Комп'ютерні технології в тестуванні

1. Види, типи, етапи, вимоги, характеристики комп'ютерного тестування.
2. Інформаційна безпека процедур масового комп'ютерного тестування.
3. Сервіси Web 2.0, та їх використання в освіті та освітніх вимірюваннях.
4. Програмні засоби для локального тестування, загальні характеристики, приклади використання.
5. Мережеве програмне забезпечення для проведення тестування. Система myTestX.
6. Мережеве програмне забезпечення для проведення тестування. Веб орієнтовані системи. Система x-TLS.
7. Загальна характеристика VLE Moodle, створення курсів.
8. Середовище Moodle, банк тестових завдань, створення тесту.
9. Аналіз тестових завдань та засоби оцінювання в середовищі Moodle.
10. Комп'ютерне адаптивне тестування. Види адаптації, алгоритми формування тесту

Методика викладання освітніх вимірювань

1. Розвиток освітніх вимірювань як окремої галузі. Складові освітніх вимірювань та зв'язок з іншими галузями. Проблеми освітніх вимірювань.
2. Методична система навчання освітніх вимірювань.
3. Методи досліджень у галузі освітніх вимірювань.
4. Вимоги до професійної відповідальності при оцінюванні. Стандарти освітнього та психологічного тестування. Кодекс справедливого тестування в освіті.
5. Стандарти компетентності вчителів з оцінювання учнів. Ключові компетенції фахівця з освітніх вимірювань. Освітньо-кваліфікаційна характеристика фахівця з освітніх вимірювань.
6. Методика навчання математико-статистичних основ освітніх вимірювань. Методика проведення лекційних та практичних занять з дисциплін математико-статистичного циклу.

Додаток VI

7. Методика навчання розробки та використання тестів. Використання засобів дистанційного навчання. Методика використання комп'ютерних технологій навчання і тестування.
8. Методика навчання основ освітніх вимірювань та моніторингу в галузі освіти. Методика проведення лекційних та семінарських занять організаційно-педагогічного циклу.

Конструювання тестів

1. Класифікації педагогічних тестів.
2. Основні етапи конструювання педагогічного тесту.
3. Планування змісту тесту та загальні принципи його відбору.
4. Методичне оснащення тесту та його складові. Склад інструкції для особи, яка проводить тест, та склад інструкції для випробовуваного.
5. Системи завдань в тестовій формі. Ланцюгові завдання. Приклади.

Основи педагогічного оцінювання

1. Загальнопедагогічні вимоги, що покладено в основу оцінювання навчальних досягнень учнів. Рівні навчальних досягнень учнів, якісна характеристика рівнів і загальні критерії їх оцінювання.
2. Історія, сучасний стан та тенденції розвитку систем оцінювання в освіті. Принципи і функції педагогічного оцінювання.
3. Зовнішнє незалежне оцінювання: зміст і процедура. Організаційно-технологічне забезпечення зовнішнього незалежного оцінювання.
4. Основні види і методи педагогічного оцінювання.

Класичні тестові моделі

1. Основні положення класичної теорії тестування. Паралельні та еквівалентні тести.
2. Кореляційна матриця. Коефіцієнт кореляції Пірсона, φ - коефіцієнт кореляції, коефіцієнт бісеріальної кореляції та їх інтерпретація.
3. Розрахунок основних математико-статистичних характеристик тестових завдань та тесту у рамках класичної теорії.
4. Поняття надійності тесту в класичній теорії. Коефіцієнти надійності Спірмана – Брауном та α -Кронбаха та їх порівняння.

5. Процедури оцінювання надійності тесту. Формули для обчислення коефіцієнта надійності.
6. Поняття валідності тесту. Змістова, критеріально-орієнтована, конструктна валідності та процедури для їх оцінювань.

Математично-статистичні методи в освітніх вимірюваннях

1. Знаходження законів розподілу випадкових величин на основі дослідних даних педагогічних вимірювань.
2. Вибірковий метод Оцінка результатів вибіркового спостереження у педагогіці та психології, розповсюдження їх на генеральну сукупність. Мала вибірка.
3. Загальні методи вивчення зв'язку соціально-педагогічних явищ і процесів. Непараметричні показники зв'язку. Рангові коефіцієнти зв'язку.

Моделі та методи IRT

1. Основні припущення IRT. Однопараметрична модель Раша. Латентні параметри. Логіти. Характеристичні криві.
2. Дво- та три параметричні моделі Бірнбаума: недоліки та переваги. Порівняння моделей Бірнбаума з моделлю Раша.
3. Матриця відповідей. Первинні бали: обчислення, статистичні характеристики, властивості.
4. Обчислення оцінок латентних параметрів на основі емпіричних даних. Стандартні помилки вимірювання. Обчислення стійких оцінок латентних параметрів методом найбільшої правдоподібності.
5. Побудова характеристичних кривих для досягнень учасників тестування (PSS) та для складності тестових завдань (ISS) на основі емпіричних даних. Два підходи теорії параметризації тестів. Перевірка гіпотези про відповідність емпіричних даних моделі Раша
6. Інформаційні функції тестового завдання та системи тестових завдань у різних моделях тестування
7. Роздільна здатність тесту (PST) Оцінка дискримінаційної здатності тесту. Аналіз дистракторів
8. Надійність. Середня квадратична помилка вимірів. Методи обчислення коефіцієнта надійності. Ретестова надійність
9. Валідність: класифікація, математичне тлумачення. Міра валідності завдань тесту. Метод підвищення контрастності груп

Додаток VI

10. Класифікація шкал: кількісні та якісні шкали. Порядкові, метричні, нормовані та номінальні школи у педагогічних вимірюваннях.
11. Порядкові шкали у педагогічному тестуванні. Відсоткова шкала. Процентильна шкала. Нормалізовані шкали типу $N(\mu; \sigma^2)$. Шкали нормалізованих процентилів. Нормалізовані шкали з постійним кроком. Ітераційна процедура для обчислення виставкових балів
12. Метричні шкали. Перенесення латентних параметрів на єдину метричну шкалу. Перетворення єдиної метричної шкали в нормовану. Остаточний бал учасників тестування