

УДК 303.2,37.091.26, 519.2

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

Авраменко О.В.

Наведено стислий огляд основних наукових та навчально-методичних видань з математичного моделювання в галузі освітніх вимірювань.

Ключові слова: освітні вимірювання, IRT, модель Раша, тестування.

Выполнен короткий обзор основных научных и учебно-методических изданий по математическому моделированию в области образовательных измерений.

Ключевые слова: образовательные измерения, IRT, модель Раша, тестирование.

The article presents a brief overview of the main scientific and methodological publications on mathematical modeling in the field of educational measurement.

Key words: educational measurement, IRT, Rasch model, testing.

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 р. Метою четвертого видання є оновлення новими поколіннями дослідників і 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, оцінці діяльності, встановлення

стандартів продуктивності і технології в тестуванні. Останній розділ включає в себе розділи, присвячені повторному мовному тестуванню, тестування на підзвітності в школах типу K-12, стандартизовані оцінки індивідуальних досягнень у школах типу K-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 усіх 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 проблему валідності тестів і викладено 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 шкалювання, 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 та 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, застосованості їх в умовах наявності комп'ютерної підтримки. В додатках наведений 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 у 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 [3]. Проаналізовано процес управління якістю освіти, його компоненти та наукове забезпечення. Показана роль у ньому педагогічних вимірювань. Охаракте-

ризовані сучасні вимірювачі якості освіти, розкриті вимоги до них у контексті проблем управління якості освіти.

В Інтернет-виданні В.С.Аванесова [1] подано початки теорії педагогічних вимірювань, наведено означення педагогічного тесту, класифіковано тести на традиційні та нетрадиційні. Автором висвітлені питання про зміст тесту, принципи розробки, ступінь складності, вимоги до тестів, докладно описані завдання різних типів.

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 вузькоспеціальними математичними або статистичними виданнями, у яких представлено суто математично-статистичні методи, на яких базується сучасна теорія освітніх вимірювань.

Метою статті є 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 у галузі статистики та педагогічних вимірювань.

Деякі приклади освітніх вимірювань та їх аналіз у межах моделі Раша наведено O.D.Duncan [32]. Основи теорії Раша, сучасні розробки у межах 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 у результаті експертного аналізу тестування окремих характеристик та деяких інших ситуацій. Вимірювання виконується у межах моделі Раша. Виконано узагальнення моделі вимірюванням із 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, Байєсовські оцінки тощо. Наведено покрокове розв'язання прикладів.

У книзі A.Boomsma, M.A.J. van Duijn & T.A.B.Snijders [26] розглянуто деякі аспекти

параметричної та непараметричної IRT, у межах 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] являє 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 ТПРТ (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 праці викладені тільки найпростіші ідеї і методи ТПРТ, для розуміння яких не потрібно глибоких знань математики. Книга складається з 6 розділів: 1) основні поняття і термінології; 2) оцінювання латентних параметрів; 3) оцінювання точності параметризації; 4) статистична перевірка гіпотез тестування; 5) елементи аналізу регресії та кореляції; 6) шкалювання результатів тестування та 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. У першій главі наведено основні відомості про педагогічне тестування, у другій главі представлено форми, структура та принципи формування тестових завдань. Основним методом статистичної обробки результатів тестування у межах класичної теорії тестів присвячена третя глава монографії. У четвертій главі висвітлено 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. У першій частині розглядаються теоретичні аспекти вимірювання латентних параметрів. Розглянуто класичну теорію тестування і теорію

вимірювання латентних параметрів, наведено їх порівняння. В другій частині монографії досліджується точність вимірювання латентних параметрів залежно від числа індикаторних змінних, числа їх градацій та ступеня кореляції. В третій частині наведені приклади вимірювання латентних змінних у галузі навчання: рівня знань із педагогіки, якості випускної кваліфікаційної роботи, толерантності. На базі методології вимірювання латентних параметрів аналізується якість текстових завдань із вибором правильної відповіді. Четверта частина присвячена застосуванню системи вимірювання латентних параметрів у 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 ін.

Монографія С.А.Позднякова [12] та його дисертаційне дослідження [13] присвячені дослідженню точності вимірювання латентних параметрів в освіті. Проаналізовано існуючі методи оцінювання інтегральних показників (латентних параметрів) у 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 залежно 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 із метою збільшення якості освіти в Росії присвячено працю Т.С.Анісімової [2].

Монографія F.B.Baker & S.-H.Kim [23] докладно описує більшість нещодавно розроблених моделей IRT, дає докладні пояснення алгоритмів, які можуть бути використані для оцінки параметрів за різних моделей IRT. Широко переглянуте і розширене 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 й алгоритмами комп'ютерних програм для оцінки параметрів, які сприяють створенню чіткого розуміння вимог процедур IRT оцінки.

Теорію освітніх вимірювань моделлю Раша та основи застосування цієї теорії викладено у фундаментальному підручнику E.V.Smith & M.S.Smith [69]. Зокрема модель Раша та нові моделі ймовірнісної 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.Челишкової [15].

У підручнику М.Б.Челишкової [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.Челишкової [14] присвячений теорії, методології та технології адаптивного тестування в освіті.

У навчальному посібнику В.І.Звонникова та М.Б.Челишкової [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. У посібнику міститься коротка інформація про деякі інноваційні засоби аутентичного оцінювання, до яких, у першу чергу, відносять портфоліо. Наводяться різні методи шкалювання даних тестування. Окремий розділ присвячено проблемам використання тестових технологій в освіті. Сфери їх застосування пов'язуються з експериментом з уведення ЄДЕ і з моніторингом якості шкільної освіти. Всього в посібнику 15 глав, кожна з яких завершується і вправою для самостійного виконання. У кінці книги наведений список літератури для більш детального вивчення порушених питань. У цілому в посібнику робиться акцент на сучасний стан практики використання тестів у Росії.

У посібнику В.І.Звонникова та М.Б.Челишкової [6] розглянуто проблеми оцінки якості результатів навчання під час атестації студентів у контексті компетентнісного підходу. Викладено ключові положення теорії педагогічних вимірювань. Розкрито основні підходи до розробки вимірювачів. Висвітлено класичний та сучасний підходи до аналізу якості тестових завдань і тестів. Подано словник термінів.

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

Навчальний посібник F.B.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. У якомусь сенсі показано тільки "те, що потрібно знати", а не всі деталі теорії. Книга дає змогу інтерпретувати результати випробувань, які були проаналізовані в межах IRT. Проте для того, щоб застосовувати теорію у практичних наборах потрібно вивчати більш просунуті книги із застосування 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 розраховано на студентів, які цікавляться процесом тестування та іншими методами вимірювання в освіті.

Основні правила роботи у середовищі RUMM, призначеного для обробки результатів тестування у межах моделі Раша, наведено 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, та 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 із 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 мова йде 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. Описано багатопараметричні моделі Раша та моделі Самеїма. Визначено, як здійснити вибір моделі (за якими критеріями та відмінності в моделювальних результатах).

У невеликій за обсягом праці M.H.Moulton [61] представлено теоретичні основи неодновимірної моделі шкалювання, за результатами аналізу практичних прикладів доведено переваги цієї моделі, висвітлено перспективи подальших досліджень.

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

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

Наближені методи обчислення та кластерний аналіз у 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 & X.Wang [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 у межах 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 байєсівської статистики, байєсівські 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 байєсівського підходу.

З основами застосування теорії ланцюгів Маркова та методу Монте Карло до стохастичного симулювання байєсівських висновків можна ознайомитись у підручнику D.Gamerman & H.F.Lopes [38]. Елементи байєсівського аналізу даних представлено у працях A.Gelman, J.B.Carlin, H.S.Stern & D.B.Rubin [39].

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

У посібнику L.Yao [81] описано користувацькі можливості програмного засобу BMIRT, 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 у соціальних науках.

У книзі M.J.Kolen & R.L.Brennan [51] наведено методи випробування для стандартизованих тестів у сфері освіти та психології з метою гарантування їх взаємозамінності у 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 праці E.B.Andersen [18]. Застосування моделі Раша, а також узагальнені принципи та підходи в соціальних та освітніх вимірюваннях представлені в працях D.Andrich [21] та [22]. Прикладні аспекти сучасної теорії тестування у психологічних дослідженнях наведено у праці 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 B.D.Wright & G.N.Masters [79] та B.D.Wright & M.H.Stone [80] представлені поширення моделі Раша освітніх вимірювань для оцінки службової діяльності та деяких інших категоріальних даних.

Праця G.Jr.Engelhard & M.Wilson [35] висвітлює теоретичні та практичні аспекти вимірювань, представлено цікаві приклади використання на практиці основних теоретичних закономірностей.

У підручнику G.S.Ingebo [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 у визначенні їх навичок. Результати представлено у доступній формі. Підручник розрахований на учнів, батьків, учителів, адміністраторів і наукових працівників, тому інформація надана у популярній формі.

У монографії T.G.Bond & C.M.Fox [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 N.Bezruczko [24].

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

Наведений у праці 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 і потребує розширення та доповнення.

Література

1. Аванесов В. С. Теория и методика педагогических измерений : материалы публикаций в открытых источниках и Интернет / В. С. Аванесов. – ЦТ и МКО УГТУ-УПИ, 2005. – 98 с.
2. Анисимова Т. С. Измерение латентных переменных в образовании / Т. С. Анисимова. – М. : Исследовательский центр проблем качества подготовки специалистов, 2004. – 148 с.
3. Звонников В. И. Измерения и качество образования / В. И. Звонников. – М. : Логос, 2006. – 312 с.
4. Звонников В. И. Измерения и шкалирование в образовании / В. И. Звонников. – 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. Челышкова. – М. : Издательский центр “Академия”, 2007. – 224 с.
6. Звонников В. И. Контроль качества обучения при аттестации / В. И. Звонников, М. Б. Челышкова. – М. : Логос, 2011. – 272 с.
7. Ким В. С. Тестирование учебных достижений : монография / В. С. Ким. – Уссурийск : Изд-во УГПИ, 2007. – 214 с.
8. Крокер Л. Введение в классическую и современную теорию тестов / Л. Крокер, Дж. Алгина. – М. : Логос, 2010. – 668 с.
9. Маслак А. А. Измерение латентных переменных в образовании и других социально-экономических системах: теория и практика / А. А. Маслак. – Славянск-на-Кубани : Изд. центр СГПИ, 2007. – 424 с.
10. Морев И. А. Образовательные информационные технологии : учебное пособие / И. А. Морев. – Владивосток : Изд-во Дальневост. ун-та, 2004.
11. Ч. 2. Педагогические измерения. – 2004. – 174 с.
12. Нейман Ю. М. Введение в теорию моделирования и параметризации педагогических тестов / Ю. М. Нейман, В. А. Хлебников. – М. : Прометей, 2000. – 168 с.
13. Поздняков С. А. Исследование точности измерения латентных переменных в образовании / С. А. Поздняков. – Славянск-на-Кубани : Издательский центр СГПИ, 2007. – 118 с.
14. Поздняков С. А. Метод и алгоритмы измерения латентных переменных при управлении в образовательных системах : автореф. дисс. на соиск. уч. степ. канд. технич. н. / Поздняков С. А. – Курск, 2009.
15. Челышкова М. Б. Адаптивное тестирование в образовании (теория, методология, технология) / М. Б. Челышкова. – М. : Исследовательский центр проблем качества подготовки специалистов, 2001. – 165 с.
16. Челышкова М. Б. Разработка педагогических тестов на основе современных математических моделей : учебное пособие / М. Б. Челышкова. – М. : Исследовательский центр проблем качества подготовки специалистов, 1995. – 32 с.
17. Челышкова М. Б. Теория и практика конструирования педагогических тестов / М. Б. Челышкова. – М. : Логос, 2002. – 432 с.
18. Allen M. J. Introduction to Measurement Theory / M. J. Allen, W. M. Yen. – Belmont, CA : Wadsworth, 1979.
19. Andersen E. B. The logistic model for m answer categories : mathematical models for social psychology / E. B. Andersen ; in W. E. Kempf & B. H. Repp (Eds.). – Vienna, Austria : Hans Huber, 1997.
20. Andersen E. B. Discrete statistical models with social science applications / E. B. Andersen. – Amsterdam : North-Holland, 1980.
21. Andrich D. Advanced Social and Educational Measurement / D. Andrich. – Perth : Murdoch University, 2001. – 128 p.
22. Andrich D. Rasch Models for Development / D. Andrich. – London : Sage Publications, Inc., 1988. – 94 p.
23. Baker F. B. The Basics of Item Response Theory / F. B. Baker. – Portsmouth NH : Heinemann Educational Books, 1985. – 131 p.
24. Baker F. B. Item response theory: parameter estimation techniques / F. B. Baker, S.-H. Kim. – 2nd ed. – New York : Marcel Dekker, 2004.
25. Bezruczko N. Rasch Measurement in Health Sciences / N. Bezruczko. – Maple Grove, Minnesota : JAM Press, 2005. – 483 p.
26. Bond T. G. Applying the Rasch model. Fundamental Measurement in the Human Sciences / T. G. Bond, C. M. Fox. – Mahwah, New Jersey : Lawrence Erlbaum Associates, Inc., Publishers, 2001. – 255 p.
27. Boomsma A., van Duijn M. A. J., Snijders T. A. B. Essays on item response theory. – New York : Springer Verlag, 2001.
28. Brennan R. Educational Measurement. – Westport, CT : Praeger, 2006. – 796 p.
29. Dorans N. J. Linking and aligning scores and scales / N. J. Dorans, M. Pommerich, P. W. Holland. – Springer, New York, 2007.
30. Drasgow F. Polytomous IRT [Special issue] / F. Drasgow // Applied Psychological Measurement. – 19(1). – 1995.
31. Duncan O. D. Rasch measurement: Further examples and discussion / O. D. Duncan ; in C. F. Turner & E. Martin (Eds.), Surveying subjective phenomena, Volume 2. – New York : Russell Sage Foundation, 1984.
32. Ebel R. L. Essentials of educational measurement / R. L. Ebel, D. A. Frisbie. – Prentice-Hall, 1986. – 360 p.
33. Embretson S. E. Item response theory for psychologists / S. E. Embretson, S. P. Reise. – Mahwah, NJ : Lawrence Erlbaum Associates, 2000.

35. Engelhard G. Jr., Wilson M. Objective measurement: Theory into practice, Vol. 3 / G. Jr. Engelhard, M. Wilson. – Norwood, New Jersey : Ablex, 1996.
36. Fischer G. H. Rasch models: Foundations, recent developments, and applications / G. H. Fischer, I. W. Molenaar. – New York : Springer-Verlag, 1995.
37. Fox J.-P. Bayesian Item Response Modeling: Theory and Applications / J.-P. Fox. – New York : Springer, 2010. – 313 p.
38. Gamerman D., Lopes H. F. Markov chain Monte Carlo: Stochastic simulation for Bayesian inference / D. Gamerman, H. F. Lopes. – 2nd edition. – Chapman & Hall, Boca Raton, FL, 2006.
39. Gelman A. Bayesian data analysis / A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin. – 2nd ed. – New York : Chapman & Hall, 2003.
40. Getting Started RUMM 2010. Rasch Unidimensional Measurement Models. – Perth : RUMM Laboratory Ltd, 2001. – 87 p.
41. Gronlund N. E. Measurement and evaluation in teaching / N. E. Gronlund. – Macmillan Pub. Co., 1965. – 420 p.
42. Gronlund N. E. Measurement and evaluation in teaching, Sixth edition / N. E. Gronlund, R. L. Linn. – New York : Macmillan, 1990.
43. Haladyna T. M. Developing and Validating Multiple-choice Test Items / T. M. Haladyna. – 3 edition. – Routledge, 2004. – 320 p.
44. Hambleton R. K. Application of Item Response Theory / R. K. Hambleton. – Vancouver, BC : Educ Res Inst B.C., 1983.
45. Hambleton R. K. Item Response Theory: Principles and Applications / R. K. Hambleton, H. Swaminathan. – Norwell, MA : Kluwer-Nijhoff Publishing, 1985.
46. Hambleton R. K. Fundamentals of Item Response Theory / R. K. Hambleton, H. Swaminathan, H. J. Rogers. – Newbury Park, CA: Sage, 1991.
47. Ingebo G. S. Probability in the Measure of Achievement / G. S. Ingebo. – Chicago : MESA Press, 1997. – 148 p.
48. Karabatsos G. The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory / Karabatsos G. ; 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 / H. Kelderman. – 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 / J.-P. Kim. – Michigan State University, East Lansing, MI, 2001.
51. Kolen M. J. Test equating, scaling, and linking : Methods and practices / M. J. Kolen, R. L. Brennan. – 2nd edition. – Springer, New York, 2004. – 574 pp.
52. Lazarsfeld P. F. Latent Structure Analyses. Houghton Mifflin Co / P. F. Lazarsfeld, N. W. Henry. – Boston, NY, 1969. – 292 p.
53. Linacre J. M. A user's guide and manual to WINSTEPS: Rasch-model computer programs / J. M. Linacre. – Chicago : Linacre, 2004.
54. Linacre J. M. Many-Facet Rasch Measurement / J. M. Linacre. – Chicago : MESA Press, 1994. – 149 p.
55. Linn R. L. Educational Measurement / ed. R. L. Linn. – 3 edition. – New York : Macmillan, 1988.
56. Linn R. L. Measurement and assessment in teaching / R. L. Linn, N. E. Gronlund. – Seventh edition. – New York : Macmillan, 1995.
57. Linn R. L. Measurement and assessment in teaching / R. L. Linn, N. E. Gronlund. – 8th edition. – Upper Saddle River, NJ: Prentice Hall, 2000.
58. Linn R. L. Measurement and assessment in teaching / R. L. Linn, M. D. Miller. – 9th edition. – Upper Saddle River, NJ: Prentice Hall, 2005.
59. Lord F. M. Application of Item Response Theory to Practical Testing Problems / F. Lord. – Hillsdale N.J. Lawrence Erlbaum Ass., 1980. – 266 p.
60. Miller M. D. Measurement and assessment in Teaching / M. D. Miller, R. L. Linn, N. E. Gronlund. – 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 / M. H. Moulton. – Educational Data Systems, Morgan Hill, CA, 2004. – 22 p.
62. Ntzoufras I. Bayesian Modeling Using WinBUGS / I. Ntzoufras. – 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) / S. J. Osterlind. – Springer; 2nd edition, 1997. – 352 p.
64. Ostini R. Polytomous item response theory models / R. Ostini, M. L. Nering. – Australia : Measured Progress, 2006. – 120 p.
65. Rasch G. Probabilistic models for some intelligence and attainment tests / G. Rasch ; expanded edition, with foreword and afterword by Benjamin D. Wright). – Chicago : University of Chicago Press, 1980. – 199 p.
66. Reckase M. D. Multidimensional Item Response Theory / M. D. Reckase. – New York : Springer, 2009. – 354 p.
67. Shavelson R. J. Generalizability Theory: A Primer / R. J. Shavelson, N. M. Webb. – Newbury Park, CA : Sage, 1991.

68. Sijtsma K. Introduction to Nonparametric Item Response Theory / K. Sijtsma, I. W. Molenaar. – London : SAGE Publications, 2002. – 168 p.
69. Smith E. V. Introduction to Rasch Measurement. Theory, Models and Applications. Maple Grove / E. V. Smith, M. S. Smith. – Minnesota : JAM Press, 2004. – 689 p.
70. Smith R. M. IPARM: Item and person analysis with the Rasch model / R. M. Smith. – Chicago, Illinois : MESA Press, 1991.
71. Smith R. M. Applications of Rasch measurement / R. M. Smith. – Chicago, Illinois : MESA Press, 1992.
72. Spray J. Identifying nonuniform DIF in polytomously scored test items. (RR 941) / J. Spray, T. Miller. – Iowa City, IA : American College Testing Program, 1994.
73. Thissen D. Test scoring. Lawrence Erlbaum Associates / D. Thissen, H. Wainer. – Mahwah NJ, 2001.
74. van der Linden W.J. Linear models of optimal test design. Springer / van der Linden W.J. – New York, 2005.
75. van der Linden W. J. Handbook of modern Item Response Theory (IRT) / van der Linden W. J., Hambleton R. K. (Eds.). – New York : Springer-Verlag, 1997. – 510 p.
76. Wainer H. Testlet response theory and its applications / H. Wainer, E. T. Bradlow, X. Wang. – New York : Cambridge University Press, 2007. – 267 p.
77. Wilson M. Measurement: A constructive approach / M. Wilson. – Mahwah, New Jersey : Lawrence Erlbaum Associates, 2003.
78. Wilson M. Constructing Measures: An Item Response Modeling Approach / M. Wilson. – Mahwah, New Jersey : Lawrence Erlbaum associates, 2005. – 228 p.
79. Wright B. D. Rating Scale Analysis / B. D. Wright, G. N. Masters. – Chicago : MESA PRESS, 1982. – 206 p.
80. Wright B. D. Best Test Design / B. D. Wright, M. H. Stone. – Chicago, MESA PRESS, 1979. – 222 p.
81. Yao L. BMIRT: Bayesian multivariate item response theory. – CTB/McGraw-Hill, Monterey, CA, 2003.