

ЗАСТОСУВАННЯ МОДЕЛІ РАША ДЛЯ АНАЛІЗУ ШКАЛИ ОЦІНЮВАННЯ КОМПЕТЕНТНОСТІ ТЕХНІЧНИХ ЕКСПЕРТІВ

У статті подано результати дослідження відомих шкал для оцінювання компетентності технічних експертів із застосуванням моделі Раша. Представлено основні особливості побудови моделі Раша, розглянуті математичні основи моделі Раша для дихотомічних і багатомірних даних. Розглянуто основні результати аналізу компетентності експертів у сфері технічного регулювання із застосуванням моделі Раша.

Ключові слова: компетентність експертів, шкала оцінювання, модель Раша, характеристична крива, програмний засіб.

Вступ

Математична модель Раша знайшла широке застосування для створення нових або для перегляду існуючих шкал у різних сферах діяльності [1]. Особливістю моделі Раша є трансформація отриманих вхідних даних в інтервальну шкалу натуральних логарифмів. До процесу логарифмічного перетворення вхідні дані у цій моделі взагалі не розглядаються.

Модель Раша [2–4] забезпечує отримання придатних результатів застосуванням декількох статистик щодо адекватності отриманих даних, певної діагностичної інформації. Ця модель представляє параметри тестування на єдиній загальній лінійній шкалі, що дозволяє здійснювати спеціальну інтерпретацію вхідних даних.

1 Аналіз публікацій та досліджень

Наукові публікації для моделі Раша охоплюють доволі велику кількість сфер діяльності [2–9], однак практично відсутні публікації, що охоплюють сферу технічного регулювання.

У праці [10] модель Раша визначається як порівняння результатів досліджуваних на шкалі натуральних логарифмів. Математичну сторону і саму теорію Г. Раша успішно розвиває Д. Андріч [2, 3]. Якщо дані відповідають моделі Раша, то в результаті вони представляються на інтервальній шкалі, яка стійка до втрати деяких первинних даних. Тому модель є методом об'єктивного масштабування даних. Розроблено кілька програмних засобів (ПЗ), зокрема [11], що дозволяють проводити необхідні розрахунки за моделлю Раша, а також дати відповідну оцінку придатності даних для використання моделі.

2. Постановка завдання

Метою дослідження є здійснення порівняльного аналізу відомих шкал для оцінювання компетентності технічних експертів із застосуванням

моделі Раша. Результатом порівняння повинні стати висновки щодо ефективності шкал оцінювання з наведенням отриманих характеристик моделі Раша.

3. Методи дослідження

Модель Раша може бути застосована до оцінювання в широкому колі сфер діяльності. Багато оцінок у певних сферах пов'язані з чітко визначеною групою осіб, що реагують на сукупність предметів для оцінки. Як правило, відповіді на поставлені пункти пропонованого тесту оцінюються 0, 1 (для двох упорядкованих категорій), або 0, 1, 2 (для трьох упорядкованих категорій), або 0, 1, 2, 3 (для чотирьох упорядкованих категорій).

Відповіді на питання тесту додаються між елементами тесту, щоб дати кожній залученій до оцінювання особі єдиний загальний бал. Цей загальний бал підсумовує відповіді на всі пункти тесту, а особа з більш високим загальним балом, ніж інша вважається, що відображає більшу кількість оціночної змінної. Підсумовуючи десятки елементів тесту з метою подання єдиної оцінки для особи, мається на увазі, що ці елементи призначені для вимірювання однієї змінної, яка часто називається однозначною змінною.

Модель Раша впливає з більш фундаментальної вимоги: порівняння двох осіб не залежить від того, які елементи можуть використовуватися в рамках елементів, що оцінюють ту саму змінну. Тобто, модель Раша розглядається як критерій структури відповідей, а не просто статистичного опису відповідей.

Аналіз вхідних даних за моделлю Раша дає низку деталей для перевірки того, чи додані результати в цих даних виправдано. Це називається тестом відповідності між даними та моделлю. Звичайно, дані ніколи не відповідають ідеальній моделі, і важливо враховувати придатність даних

для моделі стосовно використовуваних для застосування загальних балів [12–14].

Якщо дані адекватно відповідають цілям моделі, то аналіз за моделлю Раша лінеаризує загальну оцінку, яка обмежена рівнем 0 і максимальною оцінкою для елементів тесту до вимірювань. Лінеаризоване значення є місцем розташування особи на одновірному континуумі – значення називається параметром моделі та в одновірному контексті може бути тільки одним числом.

Модель Раша належить до моделей теорії реагування на предмети (IRT). Співвідношення між ймовірністю успіху до елемента та прихованою рисою описується функцією, яка називається характеристичною кривою елемента (ICC) або функцією відповіді елемента (IRF), яка приймає S-форму (рис. 1). Вона показує зв'язок між загальною оцінкою на тест та оцінкою місця розташування суб'єкта [12–14].

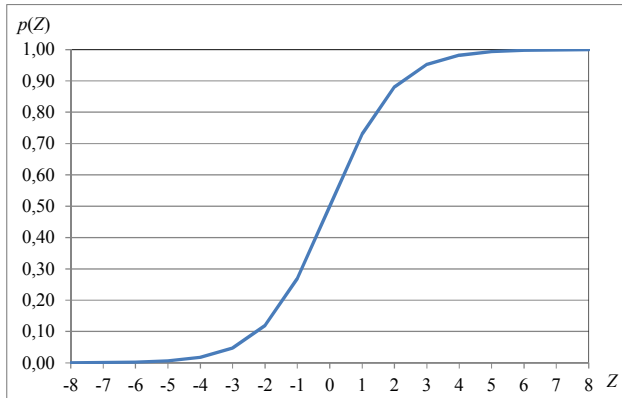


Рисунок 1 – Характеристична крива елемента

Чим вища здатність суб'єкта відносно складності елемента тесту, тим більша вірогідність правильної відповіді на цей елемент. Коли розташування суб'єкта по прихованій ознаці дорівнює складності елемента, за визначенням є 0,5 ймовірність правильної відповіді у моделі Раша.

Психометрична модель Раша концептуалізує шкалу вимірювання як лінійку (рис. 2) [12]. Елементи розташовані повздовж шкали вимірювань відповідно до їх складності: менш складні елементи розташовуються ліворуч, а найбільш важкі – справа. Стрілки вказують місце розташування предметів А, В і С на континуумі здатності.

Суб'єкти також можуть бути розташовані на тій же шкалі вимірювань і розташовуються відповідно до їх здатності: з низькою здатністю розташовані ліворуч, а з вищою здатністю – праворуч. Менш складні елементи можуть бути

успішно досягнуті більш здатними суб'єктами.

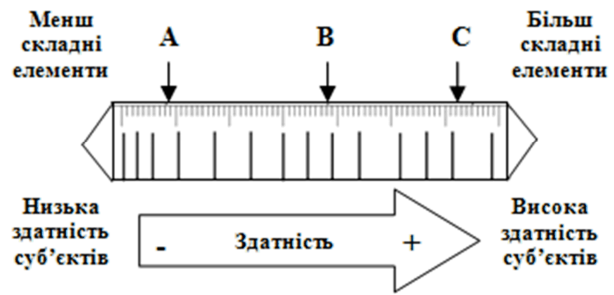


Рисунок 2 – Умовне представлення здатності континууму

Первинні бали мають невідомі проміжки між ними. Модель Раша будує оцінки реальних інтервалів складності елемента та особистої здатності шляхом створення лінійних шкал. У цьому процесі значення елемента калібруються, а особистісні якості вимірюються на загальному континуумі, який містить приховану рису.

Якщо рейтинг елемента відсутній, модель оцінює ймовірний рейтинг суб'єкту, не вказуючи відсутні дані. Одночасно, неправдоподібність відображення певного елемента оцінюється за рівнем статистики. Це порівняння того, що насправді отримано, і те, що передбачає модель, повинно отримуватися на основі оціночних заходів [12–14].

4. Математичний апарат моделі Раша

Дихотомічна випадкова величина $X_{ni} = x \in \{0, 1\}$, де $x = 1$ позначає позитивну відповідь і $x = 0$ – негативну відповідь на певну позицію оцінки. У моделі Раша для дихотомічних даних ймовірність результату $X_{ni} = 1$ дається

$$\Pr\{X_{ni} = 1\} = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)}, \quad (1)$$

де β_n – здатність суб'єкта n ; δ_i – складність елемента.

Таким чином, $\Pr\{X_{ni} = 1\}$ – це ймовірність успіху при взаємодії між відповідним суб'єктом і пунктом оцінки.

Багатомірна модель Раша, яка є узагальненням дихотомічної моделі, може бути застосована в контекстах, в яких послідовні цілі бали є категоріями збільшення рівня або величини прихованої риси. Шкала Лікерта, яка іноді використовується для багатомірної моделі Раша, містить п'ять типів відповідей: «категорично не згоден» (може бути призначено 0); «не згоден» (1); «не згоден і не погоджуюся» (2); «погоджуюся» (3);

«абсолютно згоден» (3).

Часткова кредитна модель Раша [3, 5] має однакову математичну структуру, але отримана з іншої вихідної точки пізніше і виражена дещо іншою формою. Випадкова величина $X_{ni} = x \in \{0, 1, \dots, m_i\}$ може приймати цілі значення від 0 до максимуму m_i – максимальної оцінки для пункту i .

Для часткової кредитної моделі Раша ймовірність результату $X_{ni} = x$ є

$$\Pr\{X_{ni} = x, x > 0\} = \frac{\exp \sum_{k=1}^x (\beta_n - \tau_{ki})}{1 + \sum_{j=1}^{m_i} \exp \sum_{k=1}^j (\beta_n - \tau_{ki})}, \quad (2)$$

$$\Pr\{X_{ni} = 0\} = \frac{1}{1 + \sum_{j=1}^{m_i} \exp \sum_{k=1}^j (\beta_n - \tau_{ki})}, \quad (3)$$

де τ_{ki} – k -й поріг розташування пункту i на латентному континуумі; β_n – розташування суб'єкта n в одному континуумі.

Ці рівняння однакові, як

$$\Pr\{X_{ni} = x\} = \frac{\exp \sum_{k=0}^x (\beta_n - \tau_{ki})}{\sum_{j=0}^{m_i} \exp \sum_{k=0}^j (\beta_n - \tau_{ki})}, \quad (4)$$

де значення τ_{0i} вибирається для зручності обчислень, тобто $\sum_{k=0}^{m_i} (\beta_n - \tau_{ki}) \equiv 0$.

Графік, який показує зв'язок між ймовірністю певного елемента як функцією місцезнаходження суб'єкта, називається кривою ймовірності елемента (СРС).

Логіт (logit) є ключовим елементом для ймовірнісної моделі Раша. Функція логіт є зворотною для сигмоїдної «логістичної» функції або логістичного перетворення, що використовується в математиці [10, 15]. Під вимірюванням у моделі Раша розуміється місце розташування (зазвичай у логітах) за прихованою змінною.

Логіт числа p – ймовірність, визначається за формулою

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = -\log\left(\frac{1}{p} - 1\right). \quad (5)$$

Величина $p/(1-p)$ – це відповідний коефіцієнт і логіт ймовірності є логарифмом шансів.

Статистики Infit і Outfit – найбільш широко використовувані діагностичні статистики моделі Раша. Графіки Раша і таблиці використовують нормалізовані невважені середні квадрати, так що графіки є симетричними центрованими до нуля [11].

5. Результати аналізу компетентності технічних експертів із застосуванням моделі Раша

У працях [16–17] описані методики оцінювання компетентності технічних експертів із застосуванням різних методів. Для всіх цих методик використані одні і ті ж критерії оцінювання компетентності експертів, які вставляють певну шкалу.

Для реалізації зазначених методів оцінювання компетентності технічних експертів застосовані наступні критерії: К1 – освіта і науковий рівень у сфері технічного регулювання; К2 – загальний стаж роботи; К3 – досвід роботи у сфері технічного регулювання; К4 – досвід роботи експертом у сфері технічного регулювання; К5 – робота на посаді. За кожним із критеріїв були визначені відповідні бальні оцінки.

Для оцінювання компетентності технічних експертів проведено опитування із застосуванням спеціально розробленої анкети на основі розглянутих критеріїв оцінювання компетентності. До оцінювання було залучено 32 фахівці зі стандартизації.

Отримані первинні дані про зазначених експертів оброблено із застосуванням моделі Раша і ПЗ MINISTER 4.0.1 [11]. Результати перетворення вхідних первинних даних за елементами (критеріями) та за суб'єктами (експертами) у вимірювання Раша наведені відповідно у табл. 1 і 2.

У табл. 1 і 2 результати вимірювання за елементами і за суб'єктами подані у логітах у порядку їх зменшення. Похибка вимірювання подається на основі моделі Раша, тобто це оціночне значення, яке при додаванні та відніманні з вимірювання у логітах дає мінімальну відстань до того, як різниця стане значущою.

У стовпчиках статистик Infit і Outfit знаходяться параметри, що характеризують відповідність даних моделі Раша: MNSQ (mean-square statistic) – значення, що характеризують рівень випадковості результатів або невідповідність даних моделі вимірювання; ZSTD – стандартизовані значення MNSQ, тобто ймовірність середньоквадратичної статистики, вираженої як z-статистика (середньоквадратичне відхилення).

Таблиця 1 – Результати перетворення даних за критеріями

Критерій	Загальний бал	Вимірювання	Похибка вимірювання	Статистика Infit		Статистика Outfit	
				MNSQ	ZSTD	MNSQ	ZSTD
K1	202	32	0,59	1,08	0,4	1,14	0,6
K5	203	32	0,57	1,11	0,5	1,12	0,6
K4	240	32	-0,33	0,78	-0,8	0,69	-0,9
K2	241	32	-0,36	1,30	1,1	1,24	0,8
K3	245	32	-0,47	0,65	-1,3	0,80	-0,5
Середнє значення	226,20	32	0,00	0,98	0,0	1,00	0,1
Стандартне відхилення	19,40	0,0	0,47	0,23	0,9	0,21	0,7

Таблиця 2 – Результати перетворення даних за експертами

Експерт	Загальний бал	Вимірювання	Похибка вимірювання	Статистика Infit		Статистика Outfit	
				MNSQ	ZSTD	MNSQ	ZSTD
29	44	2,76	0,89	0,86	0,4	1,64	0,9
30	44	2,76	0,89	0,83	0,3	1,39	0,7
01	43	2,22	0,63	0,51	-0,3	0,67	0,1
13	42	1,90	0,52	0,73	-0,1	0,65	-0,1
17	41	1,66	0,46	0,28	-1,2	0,29	-0,9
19	41	1,66	0,46	0,28	-1,2	0,29	-0,9
21	41	1,66	0,46	0,28	-1,2	0,29	-0,9
23	41	1,66	0,46	2,00	1,3	1,35	0,7
04	39	1,30	0,40	0,84	-0,1	0,73	-0,2
12	39	1,30	0,40	0,09	-2,5	0,13	-1,9
25	39	1,30	0,40	0,84	-0,1	0,73	-0,2
02	38	1,15	0,38	1,93	1,4	1,94	1,4
05	38	1,15	0,38	0,62	-0,5	0,57	-0,6
20	38	1,15	0,38	1,52	1,0	1,34	0,7
22	38	1,15	0,38	0,79	-0,2	0,74	-0,2
10	37	1,00	0,37	1,38	0,8	1,28	0,6
16	37	1,00	0,37	0,68	-0,4	0,68	-0,4
26	37	1,00	0,37	0,96	0,1	0,92	0,1
08	36	0,87	0,36	0,71	-0,4	0,70	-0,4
31	36	0,87	0,36	1,00	0,2	1,04	0,3
27	35	0,74	0,36	0,37	-1,3	0,36	-1,3
09	33	0,49	0,35	1,10	0,4	1,08	0,3
14	32	0,37	0,35	0,95	0,1	0,95	0,1
07	31	0,24	0,35	1,48	0,9	1,48	0,9
15	31	0,24	0,35	0,74	-0,3	0,64	-0,5
06	29	-0,01	0,36	0,93	0,1	1,07	0,3
11	28	-0,14	0,37	1,80	1,3	1,78	1,2
03	27	-0,29	0,38	1,36	0,7	1,55	1,0
28	26	-0,44	0,39	0,50	-0,7	0,48	-0,8
32	26	-0,44	0,39	1,36	0,7	1,35	0,7
18	25	-0,60	0,41	2,25	1,6	2,40	1,8
24	19	-1,84	0,51	1,02	0,2	1,44	0,8
Середнє значення	35,30	0,87	0,43	0,97	0,0	1,00	0,1
Стандартне відхилення	6,20	0,98	0,13	0,53	0,9	0,53	0,8

MNSQ також називається відносним χ^2 -квадратом чи нормованим χ^2 -квадратом. Середньозважена статистика відповідності – це χ^2 -квадратна статистика, поділена на її ступені сво-

боди. Для ймовірності $p \leq 0,05$ (двосторонній розподіл), $ZSTD > |1,96|$.

Найбільш очікувані значення MNSQ знаходяться поблизу 1,0. Найбільш якісними і

відповідними вважаються значення MNSQ у межах від 0,5 до 1,5. Значення нижче 1,0 показують, що дані чи занадто передбачувані, чи надмірно передбачувані, чи переоцінені дані моделі. Значення вище 1,0 вказують на занадто непередбачувані дані чи недооцінені дані моделі.

Значення більші за 1,5 вказують на невизначеність та «шум» (надмірна непередбачуваність даних) у вхідних даних, значення менші за 0,5 теж небажані, бо свідчать про «інформаційну перевантаженість» елементу.

Прийнятними є значення MNSQ від -2,0 до +2,0. Значення MNSQ за модулем більші 2,0 розглядаються як такі, що не відповідають моделі вимірювання і не можуть бути використані при аналізі результатів. Аналіз починають із питань з високим значенням MNSQ.

Отримані значення MNSQ для критеріїв для статистики Infit знаходяться у межах від 0,65 до 1,3, а для статистики Outfit від 0,69 до 1,24. Це свідчить про те, що всі ці значення є прийнятними для аналізу за моделлю Раша і немає наявності «шуму» у вхідних даних.

Отримані значення MNSQ для експертів для статистики Infit знаходяться у межах від 0,28 до 2,00, окрім значення 2,25 для експерта 18, а для статистики Outfit від 0,29 до 1,94, окрім значення 2,40 для того ж експерта 18. Це свідчить про те, що всі ці значення є прийнятними для аналізу за моделлю Раша, окрім даних для експерта 18. Зважаючи на це вважається доцільним видалення даних експерта 18 з подальшого аналізу.

Аналіз праць [16–18] показує, що експерт 18 також показав негативні результати оцінювання його компетентності при застосуванні інших методів оцінювання.

На рис. 3 побудовані характеристичні криві для всіх експертів, що оцінювалися, за всіма елементами із застосуванням ПЗ MINISTEP 4.0.1. Аналіз взаємного розміщення цих кривих допомагає вдосконалити оцінювання як систему критеріїв. У нашому випадку більшість кривих зосереджені в області середньої та меншої за середню складності. Характеристичні криві практично рівномірно заповнюють весь інтервал від -3 до +3 логітів при максимально допустимому діапазоні від -5 до +5 логітів, що свідчить про гарну погодженість встановлених для оцінювання експертів критеріїв.

По кожному критерію і оцінювання в цілому можна отримати графічне представлення відповідності даних обраній моделі, яка представлена на рис. 4. Отримані дані свідчать про присутність кореляції з даними по обраній моделі.

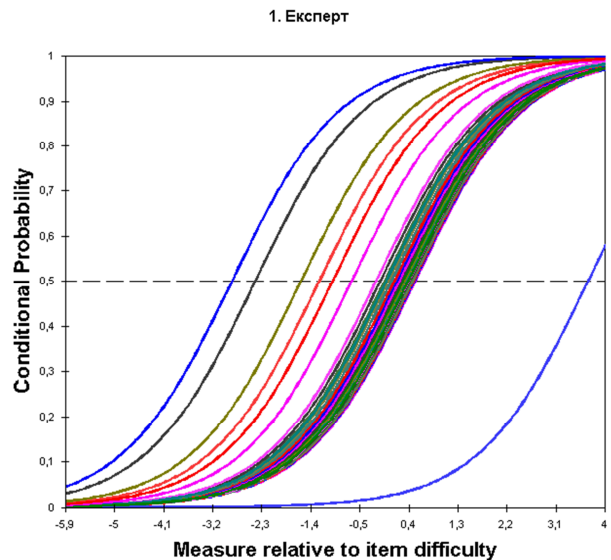


Рисунок 3 – Характеристичні криві для експертів

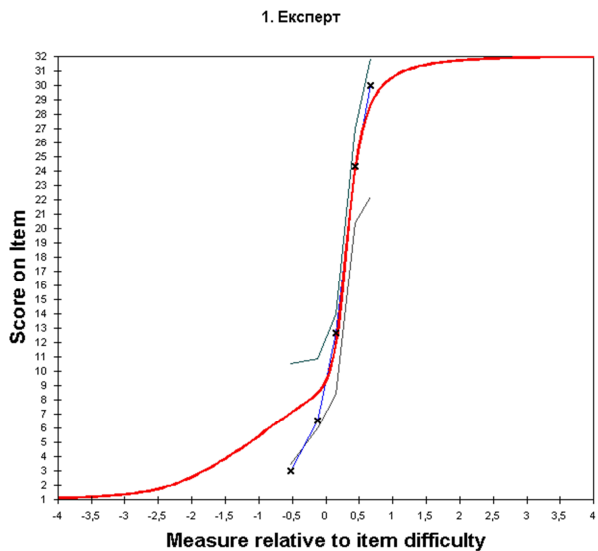


Рисунок 4 – Відповідність отриманих даних характеристичній криві для експертів

На рис. 5 представлені результати оцінювання за моделлю Раша стосовно кожного з п'яти критеріїв для експертів.

Висновки

Отримані результати показали можливість застосування моделі Раша для аналізу шкали оцінювання компетентності технічних експертів.

Аналіз отриманих за багатомірною моделлю Раша результатів з оцінювання компетентності технічних експертів зі стандартизації показав, що обрана шкала критеріїв для експертів відповідає встановленим до моделі Раша вимогам, отримані дані вимірювань за цією моделлю дозволяють розрахувати встановлені статистики як для критеріїв, так і для експертів, що оцінені.

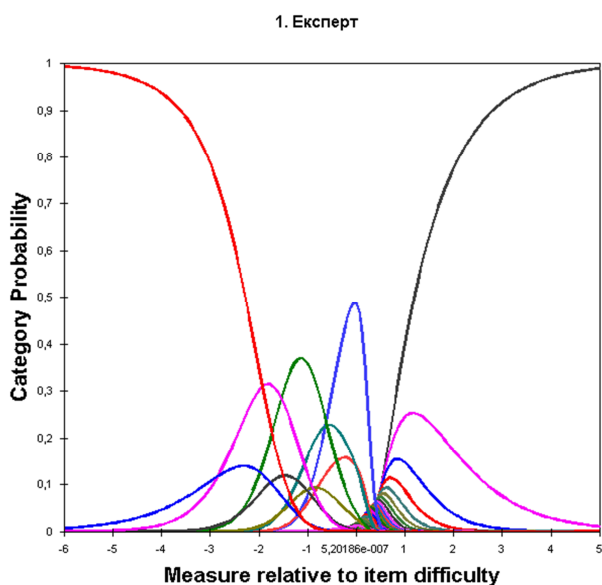


Рисунок 5 – Результати оцінювання стосовно кожного з п'яти критеріїв

Лише один з тридцяти двох оцінених експертів має дані, які непридатні для аналізу за моделлю Раша, що свідчить про низьку його компетентність. Ці результати корелюються з результатами щодо цього ж експерта, отриманими при застосуванні інших методів оцінювання компетентності експертів.

Список використаних джерел

1. Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institution for Educational Research, 1960.
2. Andrich D. Rasch models for measurement. Beverly Hills: Sage Publications, 1988.
3. Andrich D. Sufficiency and conditional estimation of person parameters in the polytomous Rasch model. *Psychometrika*. – 2010. – 75(2). – P. 292–308.
4. Linacre J. M. Rasch dichotomous model vs. One-parameter Logistic Model. *Rasch Measurement Transactions*. – 2005, 19:3, 1032.
5. Masters G. N., Keeves J. P. (Eds.). *Advances in measurement in educational research and assessment*. – New York: Pergamon, 1999.
6. Bond T. G., Fox C. M. *Applying the Rasch model: fundamental measurement in the human sciences*, 2nd edn. Mahwah: Lawrence Erlbaum Associates, 2007.
7. Hagquist C., Bruce M., Gustavsson J. P. Using the Rasch model in nursing research: an introduction and illustrative example. *Int. J. Nurs. Stud.* – 2009, 3. – P. 380–393.
8. Ehlan A. H., Kucukdeveci A. A., Tennant A. The Rasch Measurement Model. In: Franco Franchignoni (Ed). *Research Issues in Physical & Rehabilitation Medicine*. Pavia: Maugeri Foundation. – 2010. – P. 89–102.
9. Wright B. D., Linacre J. M. A measurement is the quantification of a specifically defined comparison. *Rasch Measurement Transactions*. – 1987, 1:1. – P. 5–4.
10. Деменченко О. Г. Математические основы Rasch Measurement // *Педагогические измерения* – 2010. – № 1. – С. 3–21.
11. A User's Guide to WINSTEPS@MINISTEP Rasch-Model Computer Programs. Program Manual 4.0.0 by John M. Linacre, 2017.
12. Evaluation scales in rehabilitation – <http://www.rehab-scales.org>.
13. Rasch Model/Rasch Analysis: Definition, Examples. Statistics How To – <http://www.statisticshowto.com/rasch-model>.
14. Institute for Objective Measurement, Inc. – <http://www.rasch.org/rmt/rmt101k.htm>.
15. Cramer, J. S. *Logit Models from Economics and Other Fields*, Cambridge University Press. – 2003. – P. 13.
16. Величко О. М. Порівняльний аналіз результатів оцінки компетентності експертів різними програмними засобами / О. М. Величко, Т. Б. Гордієнко, Л. В. Коломієць // *Металургическая и горнорудная промышленность*. – 2015. – № 1 (292). – С. 136–140.
17. Величко О. М., Коломієць Л. В., Гордієнко Т. Б., Шевцов А. Г., Карпенко С. Р., Габер А. А. Групове експертне оцінювання та компетентність експертів / За загал. ред. д-ра техн. наук О. М. Величка – Одеса: ВМВ, 2015. – 285 с.
18. Величко О. М., Гордієнко Т. Б., Коломієць Л. В., Карпенко С. Р. Удосконалення методів і засобів оцінювання компетентності експертів у сфері технічного регулювання / За загал. ред. д-ра техн. наук О. М. Величка – Одеса: «Апрель», 2016. – 152 с.

Надійшла до редакції 24.11.2017

Рецензент: д.т.н., проф. Коломієць Л.В., Одеська державна академія технічного регулювання та якості, Одеса.

О. Н. Величко, д.т.н., Т. Б. Гордиенко, д.т.н.

ПРИМЕНЕНИЕ МОДЕЛИ РАША ДЛЯ АНАЛИЗА ШКАЛЫ ОЦЕНКИ КОМПЕТЕНТНОСТИ ТЕХНИЧЕСКИХ ЭКСПЕРТОВ

В статье представлены результаты исследования известных шкал для оценки компетентности технических экспертов с применением модели Раша. Представлены основные особенности построения модели Раша, рассмотрены математические основы модели Раша для дихотомических и многомерных данных. Рассмотрены основные результаты анализа компетентности экспертов в сфере технического регулирования с применением модели Раша.

Ключевые слова: компетентность экспертов, шкала оценивания, модель Раша, характеристическая кривая, программное средство.

О. М. Velychko, DSc, Т. В. Gordiyenko, DSc

APPLICATION OF RASCH MODEL FOR ANALYSIS OF THE SCALE OF COMPETENCE ASSESSMENT OF THE TECHNICAL EXPERTS

The article presents the results of research of known scales for assessing the competence of technical experts using the model of Rasch. The main features of constructing the Rasch model are presented, the mathematical foundations of the Rasch model for the dichotomous and multidimensional data are considered. The main results of the analysis of the competence of experts in the field of technical regulation using the model of Rasch are considered.

Keywords: expert competence, assessment scale, Rasch model, characteristic curve, software tool.