

МЕТОДИКА ВИЗНАЧЕННЯ АЛЬТЕРНАТИВНОЇ ШКАЛИ ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ ТЕСТОВОГО КОНТРОЛЮ ЗНАТЬ СТУДЕНТІВ ВІДПОВІДНО ДО МОДЕЛІ RASCH-IRT

У статті розглянуто можливості застосування альтернативної моделі оцінювання результатів тестового контролю знань; надано методичні рекомендації щодо впровадження і використання шкали оцінювання на базі моделі Раша (Rasch-IRT); проаналізовано практичне застосування елементів математико-статистичної теорії вимірювань (Item Response Theory) з метою коригування тестових завдань.

Ключові слова: тестовий контроль, шкала оцінювання, адаптивний тест, модель Раша (Rasch-IRT), статистична обробка, математико-статистична теорія вимірювань.

Проблема об'єктивного оцінювання рівня знань студентів за умов інформатизації навчальних закладів та впровадження автоматизованих комп'ютерних навчально-контролюючих систем є актуальною як з технічного, так і суто методичного погляду [1, с. 8, 16]. Оцінювання – метод визначення відповідності очікуваного результату факторам впливу, вимогам, стандартам, що належить до психометричних педагогічних вимірювань, базованих на математичній статистиці [2, с. 47, 58, 156].

Проблемність оцінювання зумовлена тим фактом, що результати проходження тестового контролю перебувають під впливом багатьох факторів, які призводять не лише до мінливості показників, а й до зсуву шкали статистичного оцінювання [3, с. 43].

Мета статті – проаналізувати можливості практичного застосування елементів математико-статистичної теорії вимірювань (Item Response Theory) і моделі Раша (Rasch) для оцінювання результатів тестового контролю знань на основі альтернативної шкали, у тому числі з метою коригування тестових завдань і збереження результатів оцінювання та уникнення процедури повторного тестування.

Об'єктом дослідження є результати тестового контролю закритого типу з основ вищої математики та математичної статистики. Обсяг вибірки: за кількістю відповідей – 7150; за кількістю респондентів (студентів) – 143.

Метод дослідження: загальна описова статистика, методи факторного та дискримінантного аналізу; методи, базовані на математико-статистичній теорії вимірювань (Item Response Theory) і моделі Раша (Rasch).

Незалежно від форми проведення тестового контролю, його результат є статистичною оцінкою, на яку впливають регульовані та нерегульовані фактори, серед яких фактор якості, надійності, валідності тестових завдань або пострегульований фактор, який за умови закритого тесту не впливає на вже отримані результати оцінювання; якості навчальної підготовки, включно із технічними (ретрансляційними) знаннями та пізнавально-аналітичними здібностями студента; випадковості щодо отримання правильної (неправильної) відповіді та інші непередбачувані фактори.

Таким чином, оцінка є випадковою та відносною величиною, що характеризується інтервальним значенням. І вона не може сприйматися як абсолютно достовірний остаточний результат.

Відповідно до класичної теорії психометричних педагогічних вимірювань, використовують статистичну шкалу оцінювання, яка конструюється на основі загального статистичного розподілу всієї сукупності результатів проходження тесту й розглядається як традиційна [3, с. 46]. У більшості випадків за основу беруть нормальну форму розподілу результатів, що є цілком обґрунтованою початковою гіпотезою в разі використання єдиного уніфікованого тесту, який базується на стандартизованих завданнях. Формальний підхід вимагає перевірки відповідності розподілу результатів щодо нормальності (перевірка початкової гіпотези). Процес оцінювання має конкурентно-конкурсні ознаки і є відносним – статистична оцінка є віднесеною до загальної (генеральної) сукупності результатів. Критичним елементом статистичної обробки при цьому є визначення шкали, за якою відбувається оцінювання і, власне, встановлюється кінцева оцінка [6, с. 109]. У випадку поточного контролю знань шкала оцінювання залежить, певною мірою, від характеристик тесту: обсягу, що визначається загальною кількістю тестових завдань; складності, змістовності та розподілу завдань щодо навчальних розділів; форми, що визначається типом тестових завдань, їх варіативністю. До основних характеристик належать якість, надійність, валідність [1]. За таких умов, визначення форми розподілу з метою використання нормальної шкали оцінювання є більш проблемним, відповідальним та критичним. Альтернативна до традиційної шкала оцінок може бути сконструйована на основі широко вживаної моделі Раша (Rasch model) [4, с. 9, 51; 7, с. 349]. Модель була розроблена для оцінювання знань, навичок, відношень та сприйняття. Модель Раша може бути використана в поєднанні з так званою теорією Item Response Theory (IRT). IRT розглядають як “розширену статистичну теорію тестів” або “математико-статистичну теорію вимірювань” [5, с. 165; 8, с. 230]. Вибір IRT методу аналізу, який фактично асоціюється з моделлю Раша, зумовлений універсальним характером методики визначення шкали статистичних оцінок. Модель Раша дає змогу позбутися або зменшити вплив таких недоліків, які, поперше, пов’язані з представленням даних на упорядкованій нормальній шкалі та неможливістю забезпечення пропорційності оцінювання. По-друге, отримані тестові результати є суб’єктивними, ситуаційними та залежними як від студента, так і від того чи іншого завдання тесту. Останнє може означати, що заміна тестових запитань, їх коригування, зміна обсягу вибірки завдань або обсягу респондентів призводять до появи нового результату.

Модель Раша ґрунтується на таких положеннях: імовірність позитивного вибору відповіді залежить від ставлення (attitudes) студента до певного питання та від ступеня відповідності (agreeability) цього запитання дійсності. Імовірність визначається якістю підготовки студента і якістю підготовлених тестових завдань. Таким чином, імовірність надання правильної відповіді залежить від “здібностей” студента (B), включно із його

знаннями та навичками, і залежить від складності тестового запитання (D) [9, с. 271, 287; 10]. Отримана функція ймовірності P_i (B, D) є складовою функцією зазначених аргументів і забезпечує умовну об'єктивність результатів тестування та певну незалежність результату одного респондента від статистичних відповідей усіх інших. На основі функції P_i (B, D) може бути побудована характеристична S-подібна крива, яка представляє результат, відображає підготовленість студента, якість тестових завдань і тесту взагалі. Функція Раша дає змогу визначити додатковий параметр (логіт) випадковості відгадування правильних відповідей.

Одиницею шкали за моделлю Раша є значення $d = \ln [P/(1-P)]$, де P – ймовірність правильної відповіді на питання; значення P може бути замінено на частоту правильних відповідей. Модель Раша оцінює складність кожного питання в одиницях d і забезпечує статистику відповідності. Звіт результатів кожного студента містить як кількість (відсоток) правильних відповідей та логіт результату, так і розрахунок коефіцієнта альфа Кронбаха для кожного завдання. До технічних переваг упровадження моделі Раша належить комп'ютерна реалізація, наприклад, програмне забезпечення MiniSteps, WinSteps, BiLOG, WinGen.

У результаті дослідження ми отримали такий результат: відбираючи найбільш прості або складні завдання, результат залишався в допустимих межах загального розподілу; результати оцінювання стають більш об'єктивними за рахунок позбавлення простих запитань (припущення) та симетричності розподілу логітів запитань. У разі використання традиційних методик оцінювання, коригування вимагає зміни тесту й проведення повторного тестування. Таким чином, вимірювання за моделлю Раша (Rasch-IRT) стає індивідуалізованим, стандартним, незалежним. Найважливішим є те, що вимірювання за моделлю умовно не залежить від вибірки тестових завдань і респондентів та від розподілу індивідуальних відповідей і складу тестових запитань.

Модель Раша дає змогу оцінювати складність кожного тестового завдання та здатність студента щодо надання правильної відповіді, виражену в одиницях інтервальної шкали; забезпечує розподіл завдань залежно від їх ваги. Все це дає змогу вважати модель Раша найбільш придатною для отримання об'єктивних результатів тестування та здійснювати коригування тестових завдань без повторного тестування.

Висновки. Запропоновано використовувати на практиці комплексну оцінку, яка включає: шкалу оцінювання традиційного типу нормального розподілу та альтернативну (додаткову) шкалу за моделлю Раша (Rasch-IRT); характеристичні ознаки тесту в цілому (якість, валідність, надійність) і визначену оцінку-результат. Представлена альтернативна шкала оцінок-логітів результатів тестового контролю є вірогідною.

Загальний статистичний аналіз на основі моделі Раша (Rasch-IRT) дав змогу ефективно та оптимально отримати детальний результат-звіт, включно із розподілами статистичних показників результатів тестування; зробити висновок щодо рівня складності тестових завдань і на основі інтервальних значень та характеристичних функцій здійснити коригування

окремих тестових завдань. Представлений спосіб обробки надає можливість коригувати результати тестування без повторного проведення тесту. Наведена методика оцінювання не надає остаточної оцінки за традиційною шкалою, оцінки тесту за шкалою логітів є відносними.

Список використаної літератури

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. Звонников В. И. Современные средства оценивания результатов обучения : учеб. пособ. для студ. высш. учеб. заведений / В. И. Звонников, М. Б. Челышкова. – Москва : Академия, 2007. – 224 с.
3. Ким В. С. Тестирование учебных достижений : монография / В. С. Ким. – Уссурийск : Изд-во УГПИ, 2007. – 214 с.
4. Andrich D. Rasch Models for Development / D. Andrich. – London : Sage Publications, Inc., 1988. – 94 p.
5. 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.
6. Brennan R. Educational Measurement / R. Brennan. – Westport : Praeger, 2006. – 796 p.
7. Fischer G. H. Rasch models: Foundations, recent developments, and applications / G. H. Fischer, I. W. Molenaar. – New York : Springer-Verlag, 1995. – 655 p.
8. Hambleton R. K. Item Response Theory: Principles and Applications / R. K. Hambleton, H. Swaminathan. – Norwell : Kluwer-Nijhoff Publishing, 1985. – 360 p.
9. Karabatsos G. The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory / G. Karabatsos // Introduction to Rasch measurement / E. Smith & R.M. Smith (Eds.). – Gainesville : JAM Press, 2004. – Ch.16. – 353 p.
10. 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.

Стаття надійшла до редакції 01.02.2015.

Прокопченко А. Е. Методика определения альтернативной шкалы оценивания результатов тестового контроля знаний студентов в соответствии с моделью Rasch-IRT

В статье рассмотрены возможности применения альтернативной модели оценивания результатов тестового контроля знаний; даны методические рекомендации по внедрению и использованию шкалы оценивания на базе модели Раша (Rasch-IRT); проанализировано практическое применение элементов математико-статистической теории измерений (Item Response Theory, IRT) с целью корректировки тестовых заданий.

Ключевые слова: *тестовый контроль, шкала оценивания, адаптивный тест, модель Раша (Rasch-IRT), статистическая обработка, математико-статистическая теория измерений.*

Prokopchenko O. Method of Determination of Alternative Scale of Evaluation of Results of Test Control of Knowledges of Students in Accordance With Model of Rasch-IRT

Adaptive testing or adaptive knowledge measuring is one way to test knowledge of students. The article reveals the problems of test knowledge control in modern educational establishments. The article presents comparative analysis and methodical recommendations of the using the alternative scale of evaluation of test results. The alternative method is considered actual among others due to certain competitive advantages. Methods and instruments of classic theory apart big possibilities do not solve many tasks that gives modern level of education development. In this paper is considered the conduction of computer

adaptive testing using mathematical-statistical tools of Item Response Theory, which aims to determine quickly the level of full-time students and distance education students learning knowledge, and to identify gaps in their knowledge.

Methodical recommendations are given in relation to introduction and use of scale of evaluation on the base of Rasch model; practical application of elements of statistical theory of measuring (Item Response Theory) is analyzed with the purpose of tests tasks correction. Rasch model allows estimating the complication of every test question and ability of student in relation to the grant of right answer, shown in units of interval scale; provides the division of tasks depending on their weight. In accordance with a Rasch model procedure of the mathematical data processing is considered; the example of rich in content analysis and computer processing and evaluation of results of passing of test is given from bases of higher mathematics and mathematical statistics.

The statistical analysis in combination with the modern models of presentation of testing results allows more objectively and optimum to determine evaluation descriptions of educational process. Descriptive statistical estimation must be complex and alternative. The represented method of evaluation does not give final estimation on a traditional scale, are estimations of test after a scale logits are relative.

Key words: *test control, scale of evaluation, adaptive test, Rasch-IRT model, statistical processing.*