

*standpoint of cooperative communication (according to G. Grice). The approbation under the conditions of a true Russian language learning in secondary schools of several regions has proved the efficiency of the measuring didactic procedure.*

*It is an acute issue to determine the communicative minimum which is a set of knowledge and skills that a young individual needs for a successful communication in different social contexts. The minimum will serve as a ground for creating a complex of communicative mini-tasks (4-5 in each case) for independent high school graduates evaluation.*

**Keywords:** *communicative competence, independent high school graduates evaluation, communicative mini task, communicative minimum.*

**Ковальчук Ю. О.**  
**Ніжинський державний університет імені Миколи Гоголя**  
**(Ніжин, Україна)**

### **ВИКОРИСТАННЯ ОСВІТНІХ ВИМІРЮВАНЬ В СИСТЕМАХ МОНІТОРИНГУ ЯКОСТІ ОСВІТИ: ВИКЛИКИ ТА ЗАСТЕРЕЖЕННЯ**

*В Україні назріла нагальна необхідність створення цілісних та ефективних систем моніторингу якості освіти, заснованих на передових положеннях та методах теорії освітніх вимірювань. В статті розглядаються деякі труднощі та застереження, які можуть виникнути на шляху створення та застосуванні подібних моніторингових програм.*

**Ключові слова:** *моніторинг якості освіти, вибіркового метод, блочно-матричний дизайн тестів, теорія тестів IRT, маргінальний метод оцінювання параметрів тестових завдань та осіб.*

В Україні назріла нагальна необхідність створення цілісних та ефективних систем моніторингу якості освіти, заснованих на положеннях та методах сучасної теорії освітніх вимірювань. Вирішення цього завдання пов'язане з багатьма різноплановими проблемами – від загальних пізнавальних та методологічних до суто технічних. Розгляду деяких з них присвячена ця робота.

Використання процедур та методів об'єктивних вимірювань в моніторингових програмах розпочалося в першій половині ХХ століття спочатку в США, а потім в інших розвинених країнах. Починаючи з 80-х років, теоретичною базою цих програм стала сучасна теорія вимірювання латентних характеристик Item Response Theory (IRT), основи якої розробили G. Rasch [18], F. Lord, M. Novick [14], R. Hambleton [13] та інші. Пізніше їх ідеї були розвинені для специфічних умов групо-орієнтованих вимірювань в працях таких дослідників, як J. Campbell [11], S. Messick [17], A. Fitzpatrick [12], W. Yen [19], R. Adams, M. Wilson, W. Wang [9], R. Brennan [10] та ін. Отримані результати використовуються в сучасних відомих моніторингових системах національного та міжнародного рівнів: NAEP, NAAL, TIMSS, PIRLS, PISA, IALS. В [16] здійснено загальний огляд світової практики використання групо-орієнтованого оцінювання в моніторингових програмах, та пов'язаних з цим проблем. Питання популяризації та застосування об'єктивних вимірювань в освіті на пострадянському просторі розглядалися у роботах Н. Б. Челишкової [8], Ю. М. Неймана та В. А. Хлебнікова [4], В. І. Звоннікова [3], В. С. Кіма [5], та інших. Серед вітчизняних авторів, які досліджують даний напрямок, слід назвати О. І. Ляшенка, Я. Я. Болюбаша, І. Є. Булах, М. Р. Мругу [1], С. А. Ракова [7], Ю. О. Жука. Але, незважаючи на наявність значної кількості розрізнених публікацій, на даний час Україна не має власних моніторингових програм, заснованих на положеннях сучасної теорії освітніх вимірювань.

В термінах теорій систем та системного аналізу галузь освіти в будь-якій країні є складною відкритою системою. Її складність виражається у можливості виділення в ній

ієрархії підсистем, а відкритість означає взаємодію з зовнішнім соціально-економічним та культурним середовищем на вході та на виході. Зокрема, на виході важливим аспектом є успішність осіб, що навчалися, у їх подальшій професійній діяльності. Взаємодія елементів цієї системи та її підсистем між собою, а також з зовнішнім середовищем, може характеризуватися рядом кількісних та якісних індикаторів. Ця взаємодія може більш чи менш успішно фіксуватися математико-статистичними методами у термінах кореляційного зв'язку, який, проте, не дає інформації про причинно-наслідкову залежність.

Будь-яке моніторингове дослідження в освіті проводиться з метою прийняття управлінських та навчальних рішень. З точки зору класичної теорії прийняття рішень, ця задача є *неструктурованою*. Це означає, що рішення не можуть прийматися на основі одних лише математичних моделей системи та її підсистем, і їх ефективність не може бути однозначно передбачуваною. На передній план виходять досвід та інтуїція осіб, що приймають рішення, їх світоглядні та соціо-культурні особливості. Подібна ситуація, хоча й значно простіша, спостерігається для задачі валідизації освітнього вимірювання. Так, М. Кейн (М. Kane) в [15] вказує на те, що процес валідизації тесту є процесом практичної аргументації, і теоретичним підґрунтям цього процесу може слугувати теорія практичної аргументації С. Тулміна. Основну ідею цієї теорії можна сформулювати як необхідність пред'явлення свідчень на користь того чи іншого твердження, причому жодне із свідчень не може бути прямим доказом твердження. Чим більше свідчень, тим більш імовірним постає твердження, проте воно практично ніколи не може вважатися повністю вірогідним. З точки зору математика, подібна ситуація може бути описана як наявність лише необхідних, але не достатніх умов для доведення теореми.

Із сказаного випливає, що система моніторингу якості освіти повинна бути, по-перше, комплексною, по-друге – системою практичної аргументації, з притаманними подібній системі вадами. Роль освітнього вимірювання як умовно об'єктивного джерела інформації при цьому лише зростає. Але поєднання об'єктивних даних із суб'єктивними породжує додаткові виклики та застереження, з якими повинні бути добре обізнані особи, які приймають рішення.

Загальною проблемою можна вважати схильність людини видавати бажане за дійсне, плутаючи твердження-факти з твердженнями-гіпотезами. По суті, фактами можна вважати лише прямі дані об'єктивного вимірювання, валідність якого підтверджена (знову ж таки, у межах системи практичної аргументації), а також деякі контекстуальні дані, такі як відповіді осіб на запитання анкети чи дані про освіту та фах викладачів, наявність у них наукового ступеню тощо. Наприклад, твердження *“крива розподілу балів тесту рівня навчальних досягнень має вершину, зсунуту вправо”* може бути фактом. На основі цього факту можна сформулювати, як мінімум, дві різні гіпотези: (1) тест надто простий для цільової популяції тестованих; (2) група опитуваних в цілому володіє вищим рівнем навчальних досягнень у порівнянні з цільовою популяцією. Обидві гіпотези можуть бути підтвержені або відхилені лише на основі додаткового дослідження з залученням інших, зовнішніх щодо отриманих при тестуванні, даних. Ще приклад. Автор, ознайомившись із загальними результатами ЗНО з математики випускників шкіл міста Ніжина, звернув увагу на несподівано низькі результати однієї із шкіл. Невже рівень викладання математики у цій школі настільки понизився? Виявилось, що з певного часу на території цієї школи почав функціонувати фізико-математичний ліцей, куди перейшли вчитися кращі учні школи, і результати ЗНО його випускників тепер подаються у звітах окремо. Головну причину знайдено, але де гарантія, що присутність відповідної контекстуальної інформації буде закладена у моніторинговій програмі? Таким чином, факти, будучи, по суті, просто вхідними даними, мало придатні для прийняття рішень, усі інші твердження – це гіпотези, перевірка яких потребує додаткових даних, і, можливо, додаткових *локальних* досліджень.

Перераховані у цих прикладах гіпотези багатьма можуть сприйматися як доконані факти. У цій проблемі слід виділяти три аспекти. Перший аспект лежить у площині логіки. Другий аспект – притаманні людям в цілому відхилення від раціональної поведінки, які психологи називають когнітивними спотвореннями. Вони досліджувалися багатьма авторами, починаючи з Арістотеля, (див., наприклад, [6]). Теорія перспектив Д. Канемана та А. Тверського, розроблена для врахування типових відхилень від раціональної поведінки для прийняття рішень у сфері економіки, є мало придатною для використання в програмах моніторингу якості освіти. Проте особам, які приймають рішення в галузі освіти, слід бути обізнаними з суттю проблеми.

Третій аспект пов'язаний з особистими бажаннями та вподобаннями осіб, які приймають рішення. Цей аспект є найбільш складним і непередбачуваним.

Використання ЗНО, як і будь-яких інших вимірювань, для дослідження якості освіти може здійснюватися лише у поєднанні з іншими даними. Прикладом залучення додаткових даних, а саме вибіркового даних про успішність навчання студентів ВНЗ на першому курсі, є дослідження прогностичної валідності тестів ЗНО, яке проводилося протягом кількох останніх років. Однак ці дослідження ніяк не пов'язані з дослідженням якості освіти. Серед проблем, які стають на заваді використання ЗНО у програмах моніторингу якості освіти, слід виділити те, що кожен конкретний тест проходять не всі випускники школи, а лише вибірка випускників, яка не є репрезентативною. Гіпотеза про те, що даний тест обирають найбільш успішні з даної дисципліни випускники, яка дозволила б узагальнити результати тестів на всю цільову популяцію, потребує збору відповідних свідчень.

Від цих найбільш загальних проблем перейдемо тепер до більш “технічних”. J. Mazzeo, S. Lazer та M. J. Zieki в огляді [16] на прикладі відомих національних та міжнародних моніторингових програм (NAEP, TIMSS, PISA та ін.) описують основні особливості використання об'єктивних вимірювань як складових елементів моніторингових систем. Зокрема, вони вказують на те, що вимірювання у цьому випадку є орієнтованим на групи, а не на окремих індивідуумів. Основними ознаками, які відрізняють групо-орієнтоване оцінювання від індивідуального, та породжуваними ними проблемами є наступні.

1. Вимірюваннями охоплюється популяція із значними відмінностями у навчальних програмах, а також соціально-побутових, економічних, культурних умов навчання. Це призводить до необхідності покривати набором тестових завдань з кожної дисципліни більш широку предметну область, що, в свою чергу, робить практично неможливим тестування кожного учасника моніторингової програми одним і тим же “довгим” тестом. У більшості відомих в світі моніторингових програмах ця проблема вирішується введенням так званого блочно-матричного дизайну тестів, яким передбачається, що різні учасники тестування отримуватимуть різні частини загального “довгого” тесту. Унаслідок цього загострюється проблема порівнюваності результатів тестування.

2. Звіти про результати тестування повинні містити групові, а не індивідуальні характеристики, а саме такі, як середній бал, відсотки під або над певними пороговими балами шкали, загальні частотні розподіли результатів по групах. Самі групи утворюються, виходячи із контекстуальних даних дослідження, що виступають у ролі незалежних змінних, і проблемою є їх велика кількість.

Практично всі відомі у світі моніторингові програми для подолання проблем групо-орієнтованого тестування використовують методи та процедури Item Response Theory (IRT), причому при розрахунку параметрів тестових завдань застосовується *маргінальний* метод. Це відразу породжує ряд труднощів, серед яких виділимо наступні.

1. IRT є досить складною для розуміння та застосування теорією. Особливо це стосується варіанту використання в рамках цієї теорії маргінального методу оцінки параметрів тестових завдань та осіб. Отже, для правильного дизайну, опрацювання та

інтерпретації результатів тестування потрібні будуть фахівці найвищої кваліфікації. Також потрібне буде спеціальне комп'ютерне програмне забезпечення.

2. Виникнуть труднощі з обґрунтуванням висновків дослідження для осіб, які приймають рішення в галузі освіти, для широкого загалу освітян, для громадськості. Наприклад, більшості потенційних споживачів результатів дослідження нічого не говоритимуть такі поняття як стандартна похибка вимірювання, довірчий інтервал тощо. Швидше за все, доведеться готувати кілька варіантів звітів про результати моніторингового дослідження, адаптованих до різних груп зацікавлених осіб.

3. Застосування методів IRT потребують досить великих об'ємів вибірки опитуваних. Цей параметр залежить також від того, який саме варіант блочно-матричного дизайну тестування буде обрано. Якщо, наприклад, множина тестових завдань з деякої дисципліни буде розбита на 5 тестів, і буде вирішено, що для забезпечення прийнятної надійності вимірювання кожен з тестів має виконувати не менше 1000 осіб, то загалом тестуванням має бути охоплено 5000 осіб.

Існують також загальні для будь-якої моніторингової програми технічні труднощі: необхідність достатнього стимулювання учасників до демонстрації оптимальної поведінки, поступовий перехід від паперової форми тестування до комп'ютерної (ця обставина утруднює порівняння результатів різних років у довготривалих моніторингових програмах), вибір методики інтерпретації порогових балів шкали у термінах знань і умінь і донесення цієї інформації до громадськості тощо.

Певний самостійний інтерес становить питання можливості поєднання результатів моніторингового дослідження з даними уже існуючих процедур вимірювання, таких, як ЗНО чи TIMSS. Щодо ЗНО, потрібно враховувати дві невирішені на сьогодні проблеми, які не дозволяють ні вважати ЗНО самостійною моніторинговою програмою, ні використовувати його результати в інших моніторингових програмах. Перша проблема полягає у тому, що, внаслідок використання методу еквіпроцентильної нормалізації при підрахунку тестових балів, цей вид оцінювання є повністю нормо-орієнтованим, а значить, окремі тестові бали не несуть жодної інформації про знання та вміння осіб, які отримали ці бали. Для подолання цієї проблеми потрібно завершити процес шкалування, поєднавши бали шкали із змістом тестів. Це можна зробити відомими методами, такими, як метод узагальнення даних про знання та вміння, пов'язані з окремими балами шкали (Scale Anchoring). Друга проблема полягає у тому, що вибірка учнів, які склали тест ЗНО з певного предмету, не є репрезентативною, а отже, інформацію про рівень успішності учасників ЗНО не можна узагальнити на всю популяцію випускників школи. Цю проблему можна подолати, провівши додаткове дослідження щодо співвідношення розподілів рівнів успішності учасників ЗНО та всіх випускників, задіявши таку додаткову інформацію, як шкільні оцінки з даної дисципліни чи середній бал атестату. Якщо при цьому вдасться, наприклад, довести, що у ЗНО з даної дисципліни брали участь ті випускники, які мають вищий рівень з цієї дисципліни, ніж ті, хто не брав участі у ЗНО, можна цілком надійними математичними методами поширити результати ЗНО на всю популяцію випускників.

**Висновки.** Програми моніторингового дослідження якості освіти можуть і повинні будуватися навколо процедур об'єктивного вимірювання рівня навчальних досягнень. Разом з тим, слід добре розуміти ті проблеми, які при цьому виникають, щоб уникнути помилок при інтерпретації результатів та прийнятті рішень.

#### *Використана література:*

1. Булах І. Наукове обґрунтування підходів до шкалювання результатів зовнішнього незалежного оцінювання в Україні / І. Є. Булах, О. І. Ляшенко, М. Р. Мруга, Л. І. Серета // Вісник ТІМО. – 2009. – № 4. – С. 29–34.
2. Булах І. Створюємо якісний тест : навчальний посібник / І. Є. Булах, М. Р. Мруга. – К. : Майстер-клас, 2006. – 160 с.

3. Звонников В. 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.
4. Нейман Ю. Введение в теорию моделирования и параметризации тестов / Ю. М. Нейман, В. А. Хлебников. – М., 2000. – 172 с.
5. Ким В. Тестирование учебных достижений. [Монография] / В. С. Ким. – Уссурийск : Изд-во УГПИ, 2007. – 214 с.
6. Макрэйни Д. (2012). Психология глупостей. Заблуждения, которые мешают нам жить / Д. Макрэйни. – М. : Альпина Бизнес Букс. – 344 с.
7. Раков С. Системи оцінювання й шкалювання результатів національних стандартизованих вступних іспитів до ВНЗ (аналіз світового досвіду) / С. А. Раков // Вісник ТІМО. – 2009. – № 4. – С. 23–28.
8. Чельшкова М. Теория и практика конструирования педагогических тестов / М. Б. Чельшкова. – М. : Логос, 2002. – 432 с.
9. Adams R. The Multidimensional Random Coefficients Multinomial Logit Model / R. J. Adams, M. Wilson, W. Wang // Applied Psychological Measurement. – 1997. – 21. – P. 1–23.
10. Brennan R. Some Problems, Pitfalls, and Paradoxes in Educational Measurement / Robert L. Brennan // Educational Measurement: Issues and Practice. – 2001. – 20(4). – P. 6–18.
11. Campbell J. Students selecting stories: the effects of choice in reading assessment: results from the NAEP Reader special study of the 1994 National Assessment of Educational Progress / Jay R. Campbell, Patricia Donahue. – Educational Testing Service, National Center for education Statistics, U.S. Dept. of Education, Office of Educational Research and Improvement, 1997. – 98 p.
12. Fitzpatrick A. Social Influences in Standard-Setting: The Effects of Social Interaction on Group Judgments / A. R. Fitzpatrick // Review of Educational Research. – 1989. – 59. – P. 315–328.
13. Hambleton R. Fundamentals of Item Response Theory / R. K. Hambleton, H. Swaminathan, H. J. Rogers. – Newbury Park, CA : Sage, 1991. – 175 p.
14. Lord F. Statistical theories of mental test scores / F. M. Lord, M. R. Novick. – Reading MA : Addison-Wesley, 1968. – 560 p.
15. Kane M. T. Validation // Educational Measurement. Edited by Robert L. Brennan – 4th ed. – ACE Praeger, 2006. – P. 17-65.
16. Mazzeo J. Monitoring Educational Progress with Group-Score Assessments / J. Mazzeo, S. Lazer, M. J. Zieki // Educational Measurement. Edited by Robert L. Brennan – 4th ed. – ACE Praeger, 2006. – P. 17-65.
17. Messick S. National Assessment of Educational Progress Reconsidered: A New Design for a New Era. (NCES 15-TR-20) / S. J. Messick, A. E. Beaton, F. M. Lord. – Washington, DC : National Center for Education Statistics, 1983.
18. Rasch G. Probabilistic models for some intelligence and attainment tests (expanded edition) / G. Rasch. – Chicago : University of Chicago Press, 1980. – 199 p.
19. Yen W. Scaling Performance Assessments: Strategies for Managing Local Item Dependence / W. M. Yen // Journal of Educational Measurement. – 1993. – 30(3). – P. 187–213.

**Ковальчук Ю. А. Использование образовательных измерений в системах мониторинга качества образования: вызовы и предостережения.**

*В Украине назрела насущная необходимость создания целостных и эффективных систем мониторинга качества образования, основанных на передовых положениях и методах теории образовательных измерений. В статье рассматриваются некоторые трудности и предостережения, которые могут возникнуть на пути создания и применения подобных мониторинговых программ.*

**Ключевые слова:** мониторинг качества образования, выборочный метод, блочно-матричный дизайн тестов, теория тестов IRT, маргинальный метод оценки параметров тестовых заданий и тестируемых.

**Kovalchuk Y. O. Using the educational measurement in monitoring the quality of education: challenges and cautions.**

*In Ukraine, there is an urgent need for coherent and effective systems of monitoring the education quality, based on the best provisions and methods of the theory of educational measurement. This article discusses some of the difficulties and caveats to come in the way of development and application of such monitoring programs.*

**Keywords:** monitoring the quality of education, the sampling method, the block-matrix test design, Item Response Theory, marginal estimating of parameters of test items and individuals.