

DOI: <https://doi.org/10.32626/2309-9763.2026-40-36-44>

УДК 37.091.26:54:373.5]:159.955

Цигвінцев Ігор,

Приватна установа «Інститут прикладної педагогіки»,

Київ, Україна

ORCID ID: <https://orcid.org/0009-0003-8554-4292>

ihor.com@ipp.org.ua

**ЕПІСТЕМІЧНА КОМПРЕСІЯ НАУКОВОГО МИСЛЕННЯ В УМОВАХ
ШВИДКІСНОГО СТАНДАРТИЗОВАНОГО ОЦІНЮВАННЯ:
ЛОНГІТЮДНИЙ АНАЛІЗ ТЕСТУВАННЯ З ХІМІЇ (2008–2025)**

Анотація. У статті здійснено теоретико-методологічний та лонгитюдний аналіз трансформації операціоналізації наукового мислення в системі високоставкового стандартизованого оцінювання з хімії в Україні упродовж 2008–2025 років. Дослідження охоплює період функціонування зовнішнього незалежного оцінювання (ЗНО) та національного мультипредметного тесту (НМТ) і спрямоване на реконструкцію імпліцитної моделі наукового мислення, яка інституціоналізується через структуру тестового інструмента.

У центрі уваги перебуває концепт епістемічної компресії, що визначається як структурне звуження спектра когнітивних та епістемічних операцій унаслідок скорочення тестового формату, часових ресурсів і переходу до цифрового середовища оцінювання. Теоретичне підґрунтя дослідження базується на сучасних підходах до аналізу наукового мислення, концепції валідності конструкту, теорії когнітивного навантаження та дослідженнях зворотного впливу стандартизованого оцінювання на освітню практику.

Методологія дослідження ґрунтується на лонгитюдному контент-аналізі офіційних психометричних звітів Українського центру оцінювання якості освіти за 2008–2025 роки. Розроблено трирівневу модель кодування, яка дає змогу аналізувати тип логічних операцій, характер когнітивної діяльності та ступінь епістемічної автономії, необхідної для виконання тестових завдань. Запропоновано аналітичну модель структурної епістемічної компресії, що пов'язує ємність тестового інструмента з часткою завдань, орієнтованих на багатокрокове пояснювальне міркування та інтерпретацію експериментальних ситуацій.

Результати дослідження свідчать про поступове зміщення епістемічного профілю тестування у напрямі алгоритмічної процедурності та домінування закритих тестових позицій. Встановлено, що скорочення кількості завдань і часу виконання створює ризик дрейфу конструкта хімічної компетентності, за якого інструмент оцінювання зберігає технічну надійність, але трансформує баланс когнітивних операцій, доступних для вимірювання. Отримані результати мають значення для теорії валідності стандартизованого оцінювання, освітньої політики та розвитку природничої освіти в умовах цифровізації.

Ключові слова: ЗНО; НМТ; наукове мислення; валідність конструкту; епістемічна компресія; хімічна освіта; стандартизоване оцінювання; високоставкове тестування; дрейф конструкта.

1. ВСТУП / INTRODUCTION

Постановка проблеми. У сучасних освітніх системах стандартизоване оцінювання виконує не лише функцію вимірювання навчальних досягнень. Воно виступає механізмом визначення того, що вважається знанням, компетентністю, і які типи інтелектуальної діяльності держава визнає легітимними. Особливо це стосується природничих дисциплін, де декларованою метою є формування наукового мислення, здатності конструювати гіпотези, аналізувати дані, пояснювати механізми та працювати з емпіричними доказами.

Проте в умовах високоставкового тестування (high-stakes testing) виникає структурна напруга. З одного боку, тест повинен бути надійним, швидким, стандартизованим і об'єктивним. З іншого – він має вимірювати складні когнітивні процеси, що за своєю природою є відкритими, контекстуальними і часто нечіткими. Чи здатен стандартизований інструмент одночасно забезпечувати технічну надійність і відображати складність наукового мислення – питання, яке виходить за межі локального контексту та має глобальне значення.

Національне тестування з хімії в Україні у 2008–2021 роках функціонувало у форматі класичного розгорнутого зовнішнього незалежного оцінювання: до 60 завдань, кілька типів відповідей, значна частка розрахункових і відкритих задач, тривалість виконання до 150 хвилин. Після 2022 року тест був трансформований у формат національного мультипредметного тесту (НМТ): 30 завдань, суттєве скорочення часу виконання, комп'ютеризоване середовище.

Формально декларується збереження перевірки компетентностей. Проте структурна компресія тесту, зменшення кількості завдань та часових ресурсів, ставить під питання збереження епістемічного спектра, доступного для вимірювання.

Ця стаття ставить принципове питання: яку модель наукового мислення фактично операціоналізує національний інструмент оцінювання з хімії, і чи відбувається структурне звуження епістемічного простору тесту внаслідок його компресії? Для відповіді вводиться концепт епістемічної компресії – процесу, внаслідок якого інструмент оцінювання поступово зміщується від репрезентації наукового конструювання до процедурного розпізнавання правильної відповіді. Таким чином, йдеться не про зниження рівня знань, а про можливу трансформацію того, що в системі освіти починає вважатися «науковим мисленням».

Аналіз останніх досліджень і публікацій. У сучасній педагогічній та когнітивній науці наукове мислення трактується як складна система інтелектуальних дій, що виходить за межі відтворення фактів або виконання алгоритмічних процедур. У працях Д. Кун [1], [2] наукове мислення описується як розвиток аргументативної координації між доказом і твердженням, що передбачає усвідомлення статусу знання як умовного та відкритого для перегляду. Р. Лоусон [3] підкреслює роль гіпотетико-

дедуктивного міркування. Дж. Сандовал [4] розглядає наукове міркування як епістемічну практику. Д. Чінн і Р. Малхотра [5] акцентують на регулятивних процесах, через які індивід визначає, що вважається прийнятним науковим поясненням.

Проблема набуває особливої ваги в умовах високоставкового стандартизованого оцінювання (high-stakes testing). У теоретичній моделі валідності конструкту С. Мессіка [6] валідність розуміється як інтегральна оцінка того, наскільки емпіричні дані та теоретичні аргументи підтримують інтерпретацію тестових результатів. У дослідженнях зворотного впливу оцінювання (washback effect) показано, що стандартизовані тести визначають зміст освітньої практики та структуру когнітивних дій [7]. Згідно з теорією когнітивного навантаження Дж. Свеллера [8], [9], багатокрокове міркування потребує значного ресурсу робочої пам'яті. Скорочення часу виконання завдань неминуче впливає на характер мисленнєвих операцій, які можуть бути активізовані.

Усі ці підходи сходяться в тому, що наукове мислення не може бути зведене до впізнавання правильної відповіді. Проте питання про те, як саме структурна компресія тестового інструмента впливає на операціоналізацію наукового мислення, залишається недостатньо дослідженим. Саме цій невирішеній проблемі присвячено дослідження.

2. МЕТА ТА ЗАВДАННЯ / AIM AND TASKS

Метою дослідження є теоретичне обґрунтування та емпірична реконструкція моделі наукового мислення, яка операціоналізується інструментом високоставкового стандартизованого оцінювання з хімії у 2008–2025 роках, а також перевірка припущення про наявність структурного звуження епістемічного простору тесту в умовах скорочення його формату.

Для досягнення мети поставлено такі **завдання**: 1) уточнити теоретичне розуміння наукового мислення як сукупності когнітивних і епістемічних актів; 2) запропонувати операціоналізацію конструкта через трирівневу схему кодування; 3) сформулювати аналітичну модель епістемічної компресії як структурної тенденції до звуження спектра мисленнєвих операцій за умов скорочення інструмента оцінювання.

3. МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ / RESEARCH METHODOLOGY

Дослідження ґрунтується на корпусі офіційних психометричних звітів Українського центру оцінювання якості освіти щодо результатів ЗНО та НМТ з хімії за 2008–2025 роки [10]. Об'єктом аналізу виступає структура тестового інструментарію в його інституційній репрезентації. Обрано лонгітюдний дизайн із елементами якісно-кількісного контент-аналізу.

Операціоналізація конструкта наукового мислення здійснюється на основі синтезу підходів до аналізу гіпотетико-дедуктивного міркування, інтерпретації доказів та пояснювального моделювання [1], [3], [5]. У межах дослідження розроблено трирівневу модель кодування.

Перший рівень фіксує тип логічної операції, яку завдання передбачає: репродукція знання, алгоритмічне застосування, дедуктивне прогнозування та індуктивне узагальнення.

Другий рівень операціоналізує тип когнітивної діяльності: формальні розрахункові операції, інтерпретація графічних або символічних репрезентацій, пояснення причинно-наслідкових зв'язків та аналіз експериментальної ситуації.

Третій рівень стосується ступеня епістемічної автономії – міри самостійності, з якою учасник конструює відповідь.

Кодування здійснюється на основі прикладів тестових завдань і описів їх когнітивного змісту з психометричних звітів. Для забезпечення об'єктивності передбачено незалежне кодування двома дослідниками з оцінюванням узгодженості за допомогою коефіцієнта каппа Коена.

Лонгітюдний вимір передбачає порівняння трьох періодів: розгорнутого формату тестування, переходу до компетентнісної парадигми та скорочення формату в умовах цифровізації. Для кожного року визначається частка завдань різних типів та розраховується інтегральний показник епістемічної компресії.

Обмеження методології: дослідження базується на агрегованих звітних даних і не передбачає доступу до індивідуальних відповідей учасників.

4. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ / RESEARCH RESULTS

4.1. Побудова аналітичної моделі.

Для інтерпретації лонгітюдних змін запропоновано модель структурної епістемічної компресії. В основі моделі лежить припущення, що тест є обмеженою вибіркою можливих когнітивних операцій. Залежною змінною виступає частка завдань із високим рівнем епістемічної автономії. Незалежними змінними є загальна кількість завдань, час виконання та ступінь закритості формату відповіді.

Базова аналітична гіпотеза: за умов зменшення структурної ємності тесту частка завдань, що вимагають високого рівня епістемічної автономії, має тенденцію до зменшення незалежно від декларативної специфікації компетентностей.

4.2. Механізм структурного обмеження.

Скорочення обсягу тесту створює ситуацію конкуренції між типами завдань за включення до специфікації. Перевагу отримують завдання, які забезпечують технічну надійність оцінювання, високу диференційну здатність і однозначність перевірки. Завдання, що передбачають складне пояснювальне міркування, мають більшу варіативність виконання і потенційно вищу похибку вимірювання.

4.3. Операційний критерій епістемічної компресії.

Епістемічна компресія фіксується за наявності трьох взаємопов'язаних ознак: 1) зменшення частки завдань із багатокроковим поясненням або інтерпретацією експерименту; 2) відносне зростання завдань алгоритмічного застосування; 3) зміна формату відповідей у бік домінування закритих тестових позицій.

4.4. Лонгітюдна реконструкція тенденцій.

Аналіз звітів за розгорнутий період тестування дозволяє реконструювати відносну стабільність представлення різних типів завдань у першому та другому часових інтервалах. Перехід до скороченого цифрового формату створює нову конфігурацію: зменшення кількості завдань об'єктивно обмежує можливість одночасного представлення широкого спектра мисленнєвих операцій.

4.5. Теоретична інтерпретація в межах моделі валідності.

З погляду валідності конструкту [6] це означає потенційний ризик дрейфу конструкта, коли операціоналізація хімічної компетентності поступово відхиляється від її теоретичної моделі. Тест зберігає надійність, але може змінювати спектр когнітивних операцій, які він здатний фіксувати.

4.6. Дискусія.

У теорії валідності С. Мессіка [6] підкреслюється, що валідність конструкту є динамічною характеристикою. Запропонована модель епістемічної компресії дає підстави припускати, що за умов скорочення тестового формату відбувається трансформація операціоналізації наукового мислення. У роботах Д. Кун [2] показано, що розвиток наукового мислення пов'язаний із поступовим оволодінням здатністю співвідносити гіпотези з доказами. Дослідження впливу тестування на навчання [7] показали, що учасники освітнього процесу орієнтуються на структуру і формат тесту.

Із точки зору освітньої політики це означає, що інструмент оцінювання виконує роль не лише вимірювача, а й регулятора епістемічної норми. Виникає ризик інституціоналізації алгоритмічного мислення як доміантної форми когнітивної діяльності.

Епістемічна компресія може розглядатися як побічний наслідок оптимізації надійності в умовах скороченого тестового інструмента. Вона не є свідомою редукцією змісту, а виникає внаслідок структурних обмежень, пов'язаних із часом, форматом і логістикою проведення оцінювання.

Результати не дають підстав для категоричних висновків про втрату валідності. Скоріше йдеться про зміну балансу епістемічних компонентів конструкта. Запропонована модель виконує евристичну функцію, дозволяючи перевести дискусію на рівень структурного аналізу операціоналізації знання.

5. ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ / CONCLUSIONS AND PROSPECTS FOR FURTHER RESEARCH

Результати лонгітюдного аналізу дозволяють зробити кілька принципових узагальнень. По-перше, структура тестування з хімії за досліджуваний період демонструє відносну стабільність алгоритмічного компоненту. По-друге, перехід до скороченого цифрового формату створює умови структурного обмеження епістемічного простору. По-третє, виявлена тенденція інтерпретується не як спрощення змісту, а як трансформація операціоналізації конструкта хімічної компетентності.

У теоретичному вимірі це означає ризик дрейфу конструкта, коли інструмент оцінювання зберігає технічну надійність, але змінює баланс епістемічних компонентів знання, що підлягають вимірюванню.

Обмеження: аналіз базувався на звітній документації без повного банку тестових завдань; відсутність доступу до індивідуальних відповідей обмежує можливість застосування моделей теорії відгуку на завдання; зміни у складі популяції учасників можуть впливати на психометричні показники.

Перспективи подальших досліджень включають: розширення корпусу аналізу за рахунок повних текстів тестових зошитів; поєднання контент-аналізу з когнітивними інтерв'ю учнів; розроблення формалізованого індексу епістемічної компресії для порівняльних міжнаціональних досліджень.

Аналіз трансформації національного інструмента оцінювання з хімії дозволяє побачити процес інституціоналізації певної моделі наукової раціональності. Питання про те, який тип мислення винагороджується системою оцінювання, є водночас питанням про те, який тип мислення відтворюється в освітній практиці.

6. СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ТА ТРАНСЛІТЕРАЦІЯ / REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Kuhn, D., 1991. *The Skills of Argument*. Cambridge: Cambridge University Press.
2. Kuhn, D., 2005. *Education for Thinking*. Cambridge, MA: Harvard University Press.
3. Lawson, A.E., 2005. What is the role of induction and deduction in reasoning and scientific inquiry? *Journal of Research in Science Teaching*, 42(6), 716–740.
4. Sandoval, W.A., 2005. Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education*, 89(4), 634–656.
5. Chinn, C.A., & Malhotra, B.A., 2002. Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218.
6. Messick, S., 1989. Validity. In: R.L. Linn, ed. *Educational Measurement*. 3rd ed. New York: Macmillan, 13–103.
7. Alderson, J.C., & Wall, D., 1993. Does washback exist? *Applied Linguistics*, 14(2), 115–129.
8. Sweller, J., 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
9. Sweller, J., Ayres, P., & Kalyuga, S., 2011. *Cognitive Load Theory*. New York: Springer.
10. Ukrainian Center for Educational Quality Assessment, 2008–2025. Reports on the results of external independent testing and National multi-subject test in chemistry. [online] Available at: <https://testportal.gov.ua/ofzvit/>.
11. DeMars, C., 2018. *Item Response Theory*. Oxford: Oxford University Press.
12. Hambleton, R.K., & Zenisky, A.L., 2016. *Translating and adapting tests for cross-cultural assessments*. In: S. Lane, M.R. Raymond and T.M. Haladyna, eds. *Handbook of Test Development*. 2nd ed. New York: Routledge, 166–182.
13. Joo, S.H., & Engelhard, G., 2018. Construct comparability across time. *Educational and Psychological Measurement*, 78(5), 789–809.

14. Lee, W.C., 2021. Examining construct shift in longitudinal assessments. *Educational Measurement: Issues and Practice*, 40(3), 48–57.
15. Park, S., & Liu, O.L., 2022. Examining construct drift over time in educational assessments. *Journal of Educational Measurement*, 59(2), 215–238.
16. Putnick, D.L., & Bornstein, M.H., 2016. Measurement invariance conventions and reporting. *Developmental Review*, 41, 71–90.
17. Van der Linden, W.J., 2018. *Handbook of Item Response Theory*. Boca Raton: CRC Press.
18. Zwick, R., 2017. *Who Gets In? Strategies for Fair and Effective College Admissions*. Cambridge, MA: Harvard University Press.
19. Zumbo, B.D., 2017. A handbook on the theory and methods of differential item functioning. *Language Testing*, 34(1), 7–25.

EPISTEMIC COMPRESSION OF SCIENTIFIC REASONING UNDER RAPID STANDARDIZED ASSESSMENT: A LONGITUDINAL ANALYSIS OF CHEMISTRY TESTING (2008–2025)

Ihor Tsyhventsev,

Public Institution “Institute of Applied Pedagogy”,

Kyiv, Ukraine

ORCID ID: <https://orcid.org/0009-0003-8554-4292>

ihor.com@ipp.org.ua

Abstract. The article presents a theoretical, methodological, and longitudinal analysis of the transformation of the operationalization of scientific reasoning within the system of high-stakes standardized chemistry assessment in Ukraine during 2008–2025. The study covers the period of the External Independent Testing (EIT/ZNO) and the National Multi-Subject Test (NMT) and aims to reconstruct the implicit model of scientific reasoning institutionalized through the structure of the testing instrument. The study focuses on the concept of epistemic compression, defined as the structural narrowing of the spectrum of cognitive and epistemic operations resulting from the reduction of test format, time resources, and the transition to a digital assessment environment. The theoretical framework of the research is grounded in contemporary approaches to scientific reasoning analysis, construct validity theory, cognitive load theory, and studies of the washback effect of standardized assessment on educational practice. The research methodology is based on a longitudinal content analysis of official psychometric reports issued by the Ukrainian Center for Educational Quality Assessment between 2008 and 2025. A three-level coding model was developed to analyze the type of logical operations, the nature of cognitive activity, and the degree of epistemic autonomy required for completing test tasks. An analytical model of structural epistemic compression is proposed, linking the capacity of the testing

instrument with the proportion of tasks oriented toward multi-step explanatory reasoning and the interpretation of experimental situations.

The findings demonstrate a gradual shift in the epistemic profile of assessment toward algorithmic procedurality and the dominance of closed-format test items. It was established that the reduction in the number of tasks and the time allotted for completion creates a risk of construct drift in chemistry competence assessment, whereby the testing instrument preserves technical reliability while transforming the balance of cognitive operations available for measurement. The obtained results are significant for the theory of construct validity in standardized assessment, educational policy, and the development of science education in the context of digitalization.

Keywords: External Independent Testing; ZNO; National Multi-Subject Test; NMT; scientific reasoning; construct validity; epistemic compression; high-stakes assessment; chemistry education; construct drift.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Kuhn, D., 1991. *The Skills of Argument*. Cambridge: Cambridge University Press.
2. Kuhn, D., 2005. *Education for Thinking*. Cambridge, MA: Harvard University Press.
3. Lawson, A.E., 2005. What is the role of induction and deduction in reasoning and scientific inquiry? *Journal of Research in Science Teaching*, 42(6), 716–740.
4. Sandoval, W.A., 2005. Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education*, 89(4), 634–656.
5. Chinn, C.A., & Malhotra, B.A., 2002. Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218.
6. Messick, S., 1989. Validity. In: R.L. Linn, ed. *Educational Measurement*. 3rd ed. New York: Macmillan, 13–103.
7. Alderson, J.C., & Wall, D., 1993. Does washback exist? *Applied Linguistics*, 14(2), 115–129.
8. Sweller, J., 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
9. Sweller, J., Ayres, P., & Kalyuga, S., 2011. *Cognitive Load Theory*. New York: Springer.
10. Ukrainian Center for Educational Quality Assessment, 2008–2025. Reports on the results of external independent testing and National multi-subject test in chemistry. [online] Available at: <<https://testportal.gov.ua/ofzvit/>>.
11. DeMars, C., 2018. *Item Response Theory*. Oxford: Oxford University Press.
12. Hambleton, R.K., & Zenisky, A.L., 2016. *Translating and adapting tests for cross-cultural assessments*. In: S. Lane, M.R. Raymond and T.M. Haladyna, eds. *Handbook of Test Development*. 2nd ed. New York: Routledge, 166–182.
13. Joo, S.H., & Engelhard, G., 2018. Construct comparability across time. *Educational and Psychological Measurement*, 78(5), 789–809.
14. Lee, W.C., 2021. Examining construct shift in longitudinal assessments. *Educational Measurement: Issues and Practice*, 40(3), 48–57.

15. Park, S., & Liu, O.L., 2022. Examining construct drift over time in educational assessments. *Journal of Educational Measurement*, 59(2), 215–238.
16. Putnick, D.L., & Bornstein, M.H., 2016. Measurement invariance conventions and reporting. *Developmental Review*, 41, 71–90.
17. Van der Linden, W.J., 2018. *Handbook of Item Response Theory*. Boca Raton: CRC Press.
18. Zwick, R., 2017. *Who Gets In? Strategies for Fair and Effective College Admissions*. Cambridge, MA: Harvard University Press.
19. Zumbo, B.D., 2017. A handbook on the theory and methods of differential item functioning. *Language Testing*, 34(1), 7–25.

Дата першого подання статті до публікації: 23.02.2026

Дата прийняття статті до публікації після рецензування: 16.03.2026

Дата публікації: 29.05.2026