



Юрій Жук – доктор педагогічних наук, доцент, завідувач відділу моніторингу та оцінювання якості загальної середньої освіти Інституту педагогіки НАПН України, м. Київ, Україна.

Коло наукових інтересів: педагогічна кваліметрія; середовища навчання.

✉ zhukyrij@gmail.com

ORCID <https://orcid.org/0000-0002-6932-2484>

Лідія Ващенко –

кандидат педагогічних наук, старший науковий співробітник, старший науковий співробітник відділу моніторингу та оцінювання якості загальної середньої освіти Інституту педагогіки НАПН України, м. Київ, Україна.

Коло наукових інтересів: особливості внутрішнього контролю якості загальної середньої освіти; технології оцінювання навчальних досягнень з біології здобувачів закладів загальної середньої освіти; тестові технології оцінювання.

✉ vaschenko_ls@ukr.net

ORCID <https://orcid.org/0000-0002-0637-2142>



Світлана Науменко – кандидат педагогічних наук, старший науковий співробітник, старший науковий співробітник відділу моніторингу та оцінювання якості загальної середньої освіти Інституту педагогіки НАПН України, м. Київ, Україна.

Коло наукових інтересів: якість загальної середньої освіти; моніторинг якості освіти; міжнародні порівняльні дослідження якості загальної середньої освіти; тестові технології.

✉ sveta_naum@ukr.net

ORCID <https://orcid.org/0000-0002-8279-4427>

УДК 373.5.091.26.012:159.955.072

<https://doi.org/10.32405/2411-1317-2023-3-63-78>

ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ СФОРМОВАНОСТІ КАУЗАЛЬНОГО МИСЛЕННЯ ЗДОБУВАЧІВ ЗАГАЛЬНОЇ СЕРЕДНЬОЇ ОСВІТИ

Анотація. Авторами запропоновано й апробовано методику оцінювання рівня сформованості каузального мислення, засновану на тестових технологіях. Методика полягає у використанні

критеріально-орієнтованого тесту, що містить дві групи запитань, формулювання яких ініціює різну розумову діяльність учнів. В одній групі запитань запропоновані для вибору варіанти можливих причинно-наслідкових зв'язків між процесами, явищами чи подіями, що вивчаються в даній предметній галузі (каузально-орієнтована група запитань). Інша група запитань спрямована на знання раніше вивчених понять і явно не містить прохання вказати причинно-наслідкові зв'язки між ними (не каузально-орієнтована група запитань). Обидві групи запитань представлені тестовими завданнями із вибором відповіді із низки запропонованих варіантів.

З метою дослідження властивостей тесту в його структурі використовувалися опції «складність» та «вагаюся». Отримані в результаті експерименту дані дозволяють зробити висновок, що запропонована структура тесту показала достатню здатність диференціювати результати педагогічних вимірювань щодо каузально-орієнтованих та не каузально-орієнтованих запитань, що, своєю чергою, дає можливість порівняти в учнів рівні засвоєння навчального матеріалу та сформованості каузального мислення під час використання одного тесту. Короткий формат тесту уможливило його використання безпосередньо під час реального освітнього процесу. Складання тестових завдань доволі просто реалізувати вчителю самостійно відповідно до цілей контролю контекстного навчання.

Ключові слова: здобувачі освіти; каузальне мислення; тестовий інструментарій; критеріально-орієнтований тест.

Постановка проблеми. Важливою умовою повоєнного відновлення України є створення закладами загальної середньої освіти сприятливих умов для формування особистості, спроможної самостійно приймати відповідальні рішення та прогнозувати їх наслідки.

Уміння «ухвалювати рішення, що передбачає здатність обирати способи розв'язання проблем на основі розуміння причин та обставин, які призводять до їх виникнення, досягнення поставлених цілей із прогнозуванням та урахуванням можливих ризиків і наслідків» Державний стандарт базової середньої освіти (2020) визначає як наскрізне для всіх ключових компетентностей (Постанова Кабінету Міністрів України, 2020).

Ключову роль у формуванні цього вміння відіграє рівень розвитку каузального мислення, яке проявляється в уміннях знайти і пояснити (каузальне міркування) причинно-наслідкові зв'язки між подіями, явищами або процесами, характерними для даної предметної області. Встановлення причинно-наслідкових зв'язків із спостережень є класичною проблемою в наукових висновках та центральною частиною людського навчання. Люди можуть вивчати причинно-наслідкову структуру в різних умовах, на основі інформації, представленої у різних форматах: спостереження за частотами спільного виникнення між причинами і наслідками, взаємодії між фізичними об'єктами або шаблонами просторового чи часового збігу (Griffiths & Tenenbaum, 2009).

Головне припущення каузального підходу полягає в тому, що люди здатні робити висновки, найбільш інваріантні причинно-наслідковим зв'язкам для досягнення своїх цілей на основі спостережуваних подій (Holyoak & Cheng, 2011). Наша концепція причинно-наслідкового судження узгоджується як з науковим використанням поняття причинно-наслідкового зв'язку, так і з спостережуваними причинними судженнями навчених і ненавчених міркувань.

Особливої актуальності набуває проблема експериментальних досліджень взаємозв'язків між причинно-наслідковими способами мислення (каузальним мисленням) та академічною успішністю здобувачів закладів загальної середньої освіти, оскільки однією з причин низьких академічних досягнень може бути низький рівень сформованості в них уміння встановлювати причинно-наслідкові зв'язки у навчальних предметах.

Для проведення таких досліджень, очевидно, необхідний педагогічний інструментарій, який надасть можливість оцінювати рівень сформованості каузального мислення (хоч би на рівні точкової оцінки) та порівнювати його з результатами академічних досягнень учнів.

Аналіз останніх досліджень і публікацій. Причинно-наслідковому мисленню, його психічним механізмам, співвідношенню із загальним інтелектом людини, моделюванню процесів

причинно-наслідкових міркувань та багатьом іншим аспектам цієї проблеми присвячена значна кількість публікацій. У більшості праць зауважується, що причинно-наслідкові міркування дозволяють людині структурувати хаотичний потік подій в осмислені форми. Тобто каузальне мислення дає можливість враховувати досвід минулого для досягнення особистісних цілей (Buchsbaum et al., 2015; Buehner & Cheng, 2005; Griffiths & Tenenbaum, 2009; Le Pelley et al., 2017; Pazzani et al., 1987).

Результати експериментального вивчення каузального мислення висвітлені в праці зарубіжних учених М. Мустакіма (Mustakim M.), Дж. Манасюра (Mansyur J.), А. Хатібе (Hatibe A.), М. Різала (Rizal M.) та С. Н. Кахару (Kaharu S.N.) (Mustakim et al., 2020). Аналізуючи особливості навчально-пізнавальної діяльності здобувачів освіти у процесі розв'язування фізичних задач, дослідники зробили висновок, що саме каузальне міркування відіграє важливу роль під час вирішення проблеми (завдання).

У дослідженні Д. Рассела (Russell D.) (Russell, 1982) акцентовано увагу на розробленні шкали причинно-наслідкових зв'язків, яка оцінює каузальне сприйняття з точки зору локусу причинності, стабільності та керованості. Отримані з використанням цієї шкали результати вказують на відмінності у сприйнятті здобувачами причин успіхів та невдач. При цьому атрибути успіху, на противагу атрибутам невдач, сприймаються як більш внутрішні, стабільні та контрольовані.

Одним із прикладів вимірювального інструменту, орієнтованого на конкретне дослідження, є оцінювальний тест осмисленого причинно-наслідкового мислення, описаний у роботі Х. Берканта (Berkant H.) (Berkant, 2009). Метою цього дослідження було з'ясування залежності здатності учнів до осмисленого каузального мислення від рівня їхньої академічної успішності, від здатності розуміти прочитане та від статі. Автором зроблено висновки, що рівень успішності учнів та рівень розуміння прочитаного є значущими предикторами здатності здобувачів до осмисленого каузального мислення незалежно від статі. З огляду на це, навчальна діяльність може бути спроектована на основі взаємозв'язку між осмисленим каузальним мисленням, успішністю та розумінням прочитаного.

У праці Ч. Чен (Chen C.), Л. Бао (Bao L.), Д. Фрітчмана (Fritchman J.) і Х. Ма (Ma H.) (Chen et al., 2021) описано дослідження особливостей причинно-наслідкових міркувань здобувачів закладів загальної середньої освіти та коледжів Китаю щодо розуміння третього закону Ньютона. Автори виходили з того, що під час вивчення фізики причинно-наслідкові зв'язки відіграють важливу роль у розвитку концептуального розуміння та навичок вирішення проблем. Результати дослідження виявили наявність чотирьох рівнів розвитку причинно-наслідкових міркувань, кожен із яких розкриває набір унікальних шляхів мислення щодо використання асиметрії в міркуваннях для визначення причинно-наслідкового зв'язку.

У роботі І. Шоу (Shou Y.) та М. Смітсона (Smithson M.) (Shou & Smithson, 2015) обґрунтовано, що на середнє значення і на однорідність відповідей може суттєво впливати тип запитання. При цьому відповіді на запитання, яке вимагає від респондентів передбачити наслідки можливої причини, можуть бути значно нижчими й різноманітнішими, ніж їхні відповіді на запитання, в якому їм пропонується діагностувати причину результату. Цей висновок підтверджує нашу гіпотезу, що запитання в каузальній орієнтації мотивує учня до каузального роздуму. Адаже через недостатні знання роздуми набувають напряму передбачуваного характеру щодо причинно-наслідкового зв'язку, зазначених у запитанні причини і результату. Проте методика дослідження І. Шоу та М. Смітсона заснована на онлайн-опитуванні респондентів і не пристосована для використання в реальному освітньому процесі.

У працях вітчизняних авторів підтверджено залежність результатів тестування від особливостей побудови тестів. Експериментально встановлено, що форма тестових завдань і відповідей як зовнішньоорганізовані конструкції тесту є сенсорно-аферентним стимулом операційного образу діяльності, який виявляється у формуванні інтересу та мотивації до виконання завдання (Гривко та Жук, 2019), фактор «форма запитання» впливає на результати тестування (Гривко & Zhuk, 2020), порядок розміщення завдань у тесті має менший вплив на результативні показники виконання тесту, ніж профільна спеціалізація здобувачів (Жук та Ващенко, 2020), результати

тестування залежать від формату тестових завдань, освітнього рівня учнів та їхньої профільної спеціалізації (Гривко, 2020; Жук та Ващенко, 2019).

Аналіз праць, присвячених проблемі причинно-наслідкового (каузального) мислення, дозволяє зробити висновок, що причинність як відношення між причиною і наслідком, є складною проблемою, яка потребує дослідження в різних аспектах.

Формулювання цілей статті. Метою статті є узагальнення результатів експериментального дослідження щодо використання критеріально-орієнтованого тестового інструментарію для вивчення рівня сформованості каузального мислення у випускників закладів загальної середньої освіти.

Виклад основного матеріалу дослідження.

Методика дослідження.

Дослідження рівня сформованості каузального мислення у здобувачів закладів загальної середньої освіти здійснювалося в освітньому процесі з біології, оскільки саме в цьому курсі приділяється особлива увага формуванню вміння встановлювати та аналізувати причинно-наслідкові зв'язки, що відповідає функціональному підходу до навчання цього предмету (Постанова Кабінету Міністрів України, 2020).

Каузальне пояснення у біології зосереджується на механізмі, за допомогою якого здійснюється біологічний процес. У шкільному курсі біології, як і в дидактично-орієнтованій моделі біологічної науки, формування в учнів причинно-наслідкових (каузальних) зв'язків між біологічними процесами є важливим завданням навчання. Під час навчання біології зокрема, та природничих наук у цілому, мають бути збалансовані всі аспекти освітнього процесу: учні мають не лише здобувати знання, а й розвивати метакогнітивні навички та навички критичного мислення (розуміння процесів) (Naimnule & Corebima, 2018).

Найоптимальнішим інструментарієм для оцінювання сформованості каузального мислення з точки зору об'єктивності, мобільності та досвіду використання може бути критеріально-орієнтований тест. Його застосування дозволяє в одному точковому вимірі виявляти та порівнювати рівні засвоєння учнями навчального матеріалу й сформованості в них каузального мислення. Для використання такого тесту в реальному освітньому процесі загальна кількість тестових завдань не повинна перевищувати тридцять.

Педагогічні спостереження показують, що найбільш поширеною формою умовиводу, якою користуються учні на етапі формування гіпотези щодо прийняття рішення відносно відповіді на каузально-орієнтоване запитання, є простий категоричний силогізм (Жук, 1998).

Характерною рисою умовиводу такого типу є те, що в ньому деяке загальне знання застосовується до одиничного або часткового випадку на основі встановлення зв'язку даного часткового або одиничного випадку із загальним. Тобто, можна говорити про те, що використання простого категоричного силогізму показує вміння учня побачити в даному конкретному випадку (конкретному запитанні) ті закономірності (зв'язки, особливості, властивості), які описані в теорії, та використати цю теорію (загальне) для розв'язування конкретного запитання (часткове).

Отже, вміння учня використовувати умовиводи, посилення та висновки, які також представлені категоричними судженнями, є необхідною (але, вочевидь, не достатньою) умовою для правильного формування гіпотези щодо відповіді на каузально-орієнтоване запитання. Знання учнем властивостей закономірностей, що описують ці властивості, дозволяє йому робити категоричні умовиводи. Якщо умовиводи мають ймовірнісний характер, то це свідчить про незнання учнем теоретичного матеріалу, тобто про невідповідність його до відповіді на дане запитання. Це виявляється в обиранні учнем відповіді «вагаюся» у форматі бінарної змінної (binary variable).

Обираючи опцію «вагаюся», учень висловлює сумнів щодо своєї можливості вибрати (сформулювати) правильну відповідь. Сумніви відносяться до метакогнітивного моніторингу, який проявляється у метакогнітивному контролі та визначає здатність людей планувати, відстежувати та оцінювати свою когнітивну діяльність (Flavell, 1979). Метакогнітивний контроль розглядається у багатьох дослідженнях як фактор регуляції освітньої діяльності (Rivers et al., 2020).

У низці досліджень стверджується, що навички наукового процесу (science process skills) під час навчання будуть більш значущими, коли в учнів разом із навичками критичного мислення будуть сформовані й метакогнітивні навички. Це пов'язано з тим, що навички наукового процесу використовуються для набуття і використання знань та для вирішення проблем й аналізу результатів рішень (Padilla, 1990).

Сумніви у правильності вибору відповіді в учнів проявляються у післятестовий період. Дослідженню цього явища присвячена велика кількість праць (McMorris et al., 1987; Prinsell et al., 1994; Tiffin-Richards et al., 2022), результати яких неоднозначні й найчастіше трактуються по-різному. Проте більшість дослідників роблять висновок, що всі учасники дослідження, які сумнівалися у відповіді на запитання та мали можливість її виправити, покращують свої академічні показники (Bauer et al., 2007; Casteel, 2014; Hanna, 2010).

Після однозначної класифікації елементів умови запитання на основі конкретних знань з використанням умовиводу типу простого категоричного силлогізму, подальші логічні дії використовують умовні умовиводи (або низку умовиводів) типу «якщо..., тоді...», тобто таке складне судження, яке помилкове тоді й лише тоді, коли попереднє судження (антецедент) істинне, а інше (концеквент) – помилкове (Moss, 2018). У логіці таке судження має назву імплікативне.

Імплікація відіграє важливу роль у міркуваннях та висновках, оскільки передбачає причинно-наслідковий зв'язок між даними та висновком (каузальна імплікація) й істинність каузальної імплікації залежить від змісту висловлювання (Laleye, 2020).

Використання опції «складність» у форматі бінарної змінної дозволяє відстежувати суб'єктивне розуміння учнями рівня складності міркувань, які необхідні для ухвалення рішення щодо вибору правильної відповіді про причинно-наслідкові зв'язки між процесами, подіями чи явищами, які містяться в умові (запитанні) тестового завдання.

Головною метою нашого дослідження було з'ясування можливості використання тестових технологій для вивчення особливостей формування каузального мислення здобувачів профільної освіти в освітньому процесі з предметів природничо-наукового циклу. Вивчалася питання можливості створення такого вимірювального інструменту (тесту), який дасть учителям інформацію для порівняння рівня сформованості в учнів каузального мислення (каузального міркування) та рівня структури знань (рівня засвоєння програмного матеріалу), що сформувалися у здобувачів у процесі вивчення фрагмента предметної галузі, та який можна використовувати в освітньому процесі. Однією з важливих рис цього тесту є його компактність (мала кількість завдань), відносна простота способів обробки та аналізу результатів тестування.

У дослідженні взяли участь учні чотирьох 11-х класів ($N = 90$; 53,3% дівчата, 46,7% юнаки) різної профілізації (хіміко-біологічна, фізико-математична, економічна та гуманітарна) закладів загальної середньої освіти України. Хіміко-біологічна і фізико-математична спеціалізації представляють природничо-науковий напрямок, економічна і гуманітарна спеціалізації – гуманітарний напрямок. Середній вік учнів був 16,93 років ($SD = 0,64$).

Учасників було відібрано за принципом доступності та типовості. Така не випадкова (неймовірна) вибірка була єдиною можливою і визначалася умовами експерименту, які повністю залежали від дослідників. Як відомо, для такої вибірки складно встановити її репрезентативність, проте вона дозволяє більш глибоко, якісно та всебічно розкрити предмет дослідження.

Дослідження проводилося в умовах реального освітнього процесу і як вимірювальний інструментарій було обрано критеріально-орієнтований тест. Завдання тесту не виходили за межі програмного матеріалу шкільного курсу біології. Тест був розроблений авторами дослідження та пройшов експертизу низки фахівців у галузі педагогіки.

Відповідно до цілей дослідження в тесті використовувалися два типи формулювань запитань у тестових завданнях. У першому типі запитань респонденту пропонувалося встановити (вказати) причинно-наслідкові зв'язки між процесами чи явищами (подіями). Цей тип запитань можна умовно назвати каузально-орієнтованими (КО). У другому типі запитань учням пропонувалося обрати правильну відповідь без вимоги вказати причинно-наслідкові зв'язки. Відповідно, цей тип

запитань можна умовно назвати не каузально-орієнтованими (НКО). Тест містив дві системи (групи) тестових завдань, кожна з яких мала різну цільову функцію. Каузально-орієнтована система тестових завдань складалася із 16 завдань, не каузально-орієнтована – з 15. Окрім цього передбачалося, що учень самостійно буде вказувати складність (1 – складне, 0 – просте) кожного тестового завдання. Згідно із обраною методикою зазначена респондентом заявлена оцінка складності тестових завдань дозволила порівняти результати виконання груп тестових завдань (КО і НКО).

Істотною відмінністю від формату критеріально-орієнтованих тестів, які традиційно використовуються для оперативного оцінювання ступеня засвоєння учнями певного фрагмента предметної галузі, що вивчається, стало введення опції «вагаюся», яка, по суті, означала відмову респондента від прийняття рішення щодо вибору правильної відповіді на запитання тестового завдання.

Тест, який використовувався у дослідженні, уможливив у процесі одного виміру з'ясувати взаємозалежність трьох параметрів: (1) результат виконання тесту («оцінка»); (2) заявлений рівень складності («складність»); (3) відмова від прийняття рішення щодо вибору правильної відповіді («вагаюся»). Оцінка параметрів здійснювалася методом визначення групових середніх значень за кожним параметром, групою спеціалізації (природничо-наукова, гуманітарна) та групою завдань (каузально-орієнтовані, не каузально-орієнтовані).

Щоб оцінити показник надійності (внутрішньої узгодженості) авторського тесту, використовувався коефіцієнт Альфа Кронбаха (Cronbach's alpha, α), величина якого, визначена за результатами експерименту (0,67), була визнана адекватною для даного дослідження.

Значимість відмінності середніх значень досліджуваних вибірок визначалася із застосуванням однофакторного дисперсійного аналізу (перевірка на нормальність розподілу вибірок проводилася за методом Колмогорова-Смірнова). Практична значимість розміру ефекту (ES) впливу фактора «тип запитання» на результати виконання учнями тесту визначалася за методикою Д. Коена (Cohen's d). Такий підхід, у цілому, відповідає вимогам Американської психологічної асоціації (American Psychological Association) до публікацій результатів емпіричних досліджень у галузі поведінкових наук.

Відповідно до методики Д. Коена, розмір ефекту для педагогічних досліджень визначається як: $ES < 0,20$ – несуттєвий; $0,20 \leq ES < 0,50$ – малий; $0,50 \leq ES < 0,80$ – середній; $0,80 \leq ES$ – великий.

Розрахунки проводилися за масивами даних результатів індивідуальних відповідей учасників дослідження, на основі яких формувалися матриці по кожному параметру. На основі матриць параметрів формувалися матриці за ознакою спеціалізації (природничо-наукова, гуманітарна) та матриці за групами завдань (каузально-орієнтовані, не каузально-орієнтовані).

У процесі обробки результатів дослідження використовувався табличний процесор MS Excel 10.0 та статистичні пакети Statistica 6.0 і SPSS 15.0.

Ця стаття є спробою представити метод вимірювання впливу фактора «каузальність», який реалізовано у відмінності формулювань запитань у групах (системах) тестових завдань, і, таким чином, визначити рівень впливу даного фактора на результати тестування.

Описаний у статті метод вимірювання заснований на поєднанні визначення кореляції між змінними у кожній системі (внутрішньосистемні) та між однойменними змінними у різних системах. Кожна система характеризується взаємодією трьох параметрів, величини яких експериментально визначаються методом групових середніх, а можливість обчислення кореляції між однойменними параметрами систем визначено належністю однієї групи респондентів до цих систем.

Результати дослідження.

Одним із показників якості освіти є когнітивні досягнення учнів, які вимірюються балами за тести та іспити. Результат виконання тесту традиційно ототожнюється з оцінкою за фрагмент учебного матеріалу, який перевіряється з використанням даного тесту.

Результати тестування після перевірки було подано у бінарній формі ($a_{i,j} = \{1 - \text{правильна відповідь}, 0 - \text{не правильна відповідь}\}$) та зведено в матрицю $|a_{i,j}|$ ($i = \overline{1,n}; j = \overline{1,m}$), де n – кількість тестових завдань (запитань) у тесті, m – кількість учасників тестування. Результат виконання тесту j -м учнем обчислюється за формулою (1):

$$\bar{A}_j = \frac{1}{n} \sum_{i=1}^n a_{ij} \quad (1)$$

Такий метод усереднення дозволяє обчислити частку правильних відповідей незалежно від кількості тестових завдань (запитань) у тесті і може бути представлений у відсотках максимального можливого результату оцінювання.

Середня групова оцінка за відповіді на запитання тесту обчислюється за формулою (2):

$$\bar{A}_{\text{гр}} = \frac{1}{m} \sum_{j=1}^m \bar{A}_j \quad (2)$$

Середня оцінка відповіді на i -е запитання тесту обраховується за формулою (3):

$$\bar{A}_i = \frac{1}{m} \sum_{j=1}^m a_{ij} \quad (3)$$

Результати розрахунків наведено у таблиці 1.

Таблиця 1

Середні групові результати тестування

Профіль	Група тестових завдань	M	SD	Критерій Фішера	d	r
Природничо-науковий	КО	0,3177	0,1115	$F(3,95) = 16,73;$ $p < 0,05$	0,836	0,614
	НКО	0,3981	0,0779			
Гуманітарний	КО	0,3308	0,1010	$F(3,95) = 4,19;$ $p < 0,05$	0,447	0,559
	НКО	0,3759	0,1008			

M – середнє значення результату тестування, SD – стандартне відхилення, d – розмір ефекту (за Д. Коеном), r – коефіцієнт кореляції.

За даними табл. 1 можна зробити висновок, що між групами каузально-орієнтованих (КО) та не каузально-орієнтованих (НКО) завдань існують статистично значущі відмінності щодо результатів виконання відповідних фрагментів тесту для обох спеціалізацій старшокласників, які брали участь у дослідженні. Результати виконання групи НКО завдань тесту вищі, ніж групи КО завдань для обох спеціалізацій здобувачів. Так, у групі респондентів природничо-наукового профілю середнє значення результату тестування за КО завдання становить 0,3177, за НКО завдання – 0,3981, у групі респондентів гуманітарного профілю – відповідно 0,3308 та 0,3759.

Статистично значущої різниці між результатами виконання КО і НКО завдань тесту для обох профілів не було виявлено: критерій Фішера для групи КО завдань становив: $F(3,95) = 0,33$; $p = 0,536$; для групи НКО завдань – $F(3,96) = 1,38$; $p = 0,242$.

Поняття складності тестового завдання, групи тестових завдань або тесту загалом визначається обраною моделлю вимірювання (Bratfisch et al., 1972; Chang, 2015) та складності, яку було обраховано за результатами виконання тесту або окремих тестових завдань (Lalor et al., 2018). Методика даного дослідження дозволяє порівняти заявлену (суб'єктивну) складність та складність, обраховану за результатами виконання тесту або окремих тестових завдань і яку ми називатимемо «реальна складність».

Заявлена учнями складність тестового завдання визначалася на основі суб'єктивних відчуттів здобувачів: під час виконання тесту кожен учень самостійно вказував складність, на його погляд, кожного тестового завдання (1 – складне, 0 – просте).

Матрицю відповідей учнів на опцію «складність» ($c_{i,j}$ | $i = \overline{1,n}$; $j = \overline{1,m}$), яка є заявленою учнями складністю щодо тестового завдання на основі суб'єктивного відчуття, було сформовано

на основі бінарності відповідей відповідно до вимог тестування ($c_{i,j} = \{1 - \text{складне завдання}, 0 - \text{просте завдання}\}$).

Реальна складність тестового завдання обчислювалася на основі даних щодо результату виконання учнями цього завдання (Taib & Yusoff, 2014). Індекс складності i -го запитання k -го тестового завдання показує, наскільки простим є це завдання, визначаючи частку учнів, які правильно відповіли на нього (формула (4)):

$$\bar{I}_{d_i} = \frac{1}{m} \sum_{j=1}^m a_{ij} \quad (4)$$

Згідно із цієї формулою, складніші завдання призведуть до значень, близьких до 0,0. Отже, складність завдання (question difficulty, \bar{D}_i) визначається за формулою: $\bar{D}_i = 1 - \bar{I}_{d_i}$ згідно з якою складніші завдання призведуть до значень, близьких до 1,0. Такий метод, на нашу думку, дає можливість порівняти реальну складність запитання (\bar{D}_i) із заявленою складністю (claimed difficulty, C_j).

Аналогічно до такого підходу ми обраховували заявлену складність тесту для j -го учня: $\bar{C}_j = 1 - \bar{A}_j$, де \bar{A}_j – оцінка j -го учня за тест.

Обчисливши середнє значення заявленої складності тесту за формулою (5), ми порівняли реальну складність із заявленою.

$$\bar{C}_d = \frac{1}{m} \sum_{j=1}^m \bar{C}_j \quad (5)$$

Результати розрахунків подано у таблиці 2.

Таблиця 2

Середня групова складність груп тестових завдань

Профіль	Група тестових завдань	Заявлена складність (\bar{C}_{d_k})	Реальна складність (\bar{D}_k)
Природничо-науковий	КО	0,5208	0,6698
	НКО	0,3438	0,6317
Гуманітарний	КО	0,4524	0,6702
	НКО	0,3373	0,6409

Між групами каузально-орієнтованих (КО) і не каузально-орієнтованих (НКО) тестових завдань існують статистично значущі відмінності щодо заявленої складності запитань у тестових завданнях відповідних фрагментів тесту для обох спеціалізації старшокласників: природничо-наукова спеціалізація: $F(3,9) = 11,05$; $p < 0,05$; $d = 0,713$; гуманітарна спеціалізація: $F(3,95) = 3,97$; $p < 0,05$; $d = 0,435$. Статистично значущої різниці між результатами заявленої складності груп запитань у тестових завданнях фрагментів тесту однойменних орієнтацій для обох спеціалізації старшокласників не було виявлено: група КО запитань: $F(3,95) = 1,31$; $p = 0,245$; група НКО запитань: $F(3,95) = 0,01$; $p = 0,899$.

Дані табл. 2 дають можливість зробити висновки: 1) реальна складність для КО та НКО груп тестових завдань практично однакова (середнє значення відношення $D_{ко}/D_{нко} = 1,05$); 2) заявлена складність для КО та НКО груп тестових завдань значно відрізняються (середнє значення відношення $S_{ко}/S_{нко} = 1,43$); 3) у групах НКО тестових завдань (у респондентів природничо-наукового та гуманітарного профілів) реальна складність удвічі більша за заявлену складність: середнє значення відношення реальної складності до заявленої складності (D/C) для груп НКО тестових завдань становить 1,87, для груп КО тестових завдань – 1,38. Це підтверджує хорошу здатність диференціювати результати вимірювань із використанням запропонованої структури тесту.

Матриця відповідей на опцію «вагаюся» $|b_{i,j}|$ ($i = \overline{1,n}$; $j = \overline{1,m}$), яку, в цьому випадку, можна трактувати як відмову від відповіді на запитання тесту, формується на основі бінарності відповідей відповідно до інструкції до тесту.

Усереднений результат відмови j -м учнем від відповіді на запитання тесту, усереднений груповий результат відмови від відповіді на запитання тесту та усереднений груповий результат відмови від відповіді на i -те запитання тесту обчислюються відповідно за формулами (6, 7, 8):

$$\bar{B}_j = \frac{1}{n} \sum_{i=1}^n b_{ij} \quad (6)$$

$$\bar{B}_{\text{гр}} = \frac{1}{m} \sum_{j=1}^m \bar{B}_j \quad (7)$$

$$\bar{B}_i = \frac{1}{m} \sum_{j=1}^m b_{ij} \quad (8)$$

Аналогічно були обраховані параметри груп запитань каузально-орієнтованого і некаузально-орієнтованого типів. Результати розрахунків наведено у таблиці 3.

Таблиця 3

Середні групові значення вибору опції «вагаюся»

Профіль	Група тестових завдань	M	SD	Критерій Фішера	d	R
Природничо-науковий	КО	0,3646	0,2331	$F(3,94) = 14,93;$ $p < 0,05$	0,789	0,631
	НКО	0,2099	0,1647			
Гуманітарний	КО	0,3435	0,1959	$F(3,95) = 4,82;$ $p < 0,05$	0,422	0,539
	НКО	0,2582	0,2079			

M – середнє значення вибору опції «вагаюся», SD – стандартне відхилення, d – розмір ефекту (за Д. Коеном), r – коефіцієнт кореляції.

Між групами каузально-орієнтованих і не каузально-орієнтованих запитань існують статистично значущі відмінності щодо відмови від прийняття рішення для обох профілів: природничо-науковий: $F(3,94) = 14,93$; $p < 0,05$; $d = 0,789$; гуманітарний профіль: $F(3,95) = 4,82$; $p < 0,05$; $d = 0,422$ (див. табл. 3). При цьому статистично значущої різниці між результатами вибору опції «вагаюся» між учнями різних профілів не спостерігається: група КО запитань: $F(3,95) = 0,22$; $p = 0,637$; група НКО запитань: $F(3,96) = 1,51$; $p = 0,225$.

У ході дослідження були обраховані моделі множинної регресії для визначення, якою мірою прямі вимірювання заявленої складності запитань передбачають відмову учнів від прийняття рішення щодо вибору правильної (на їхню думку) відповіді на запитання тестового завдання. Дані рівняння регресії подано у таблиці 4.

Таблиця 4

Таблиця рівняння регресії

Профіль	Група тестових завдань	$p < 0,05$
Природничо-науковий	КО	$y = 0,3494 + 0,4701x, F(4)=7,22; R^2 = 0,1354$
	НКО	$y = 0,1547 + 0,9036x, F(4)=31,38; R^2 = 0,4055$
Гуманітарний	КО	$y = 0,2494 + 0,5909x, F(4)=8,77; R^2 = 0,1715$
	НКО	$y = 0,1438 + 0,7496x, F(4)=26,01; R^2 = 0,394$

y – «вагаюся», x – «складність».

Дані табл. 4 показують, що в усіх випадках спостерігається позитивна залежність. Проте у групах НКО запитань ця залежність дещо більша, ніж у групах КО запитань.

Малі коефіцієнти детермінації, особливо для груп КО запитань (див. табл. 4), вказують на той факт, що відмова від відповіді на запитання тестового завдання залежить не лише від його складності, а й від інших факторів. Виявлення цих факторів потребує подальших досліджень.

На рисунках 1 та 2 показано структурні зв'язки у групах респондентів різної спеціалізації та зазначено величини внутрішньосистемних кореляцій і кореляцій між однойменними змінними у різних системах.

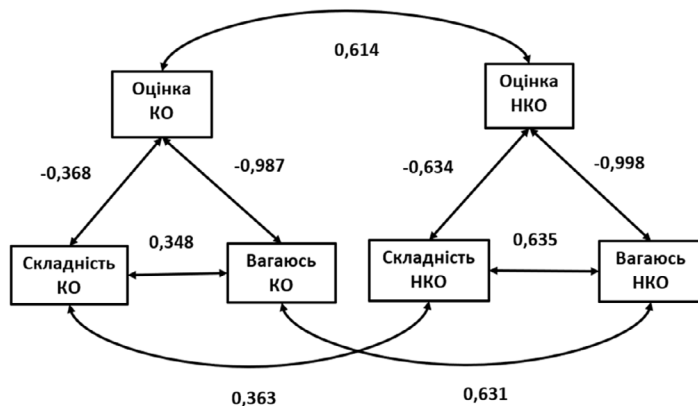


Рис. 1. Структура кореляційних зв'язків у групі респондентів природничо-наукового профілю.

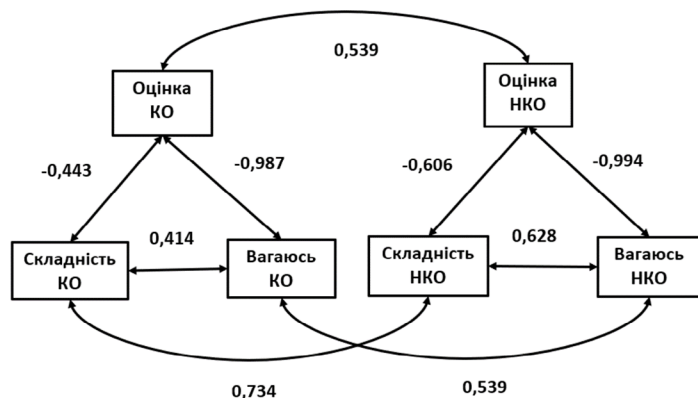


Рис. 2. Структура кореляційних зв'язків у групі респондентів гуманітарного профілю.

З рис. 1 та рис. 2 видно, що високі від'ємні кореляції (антикореляції) (понад $-0,9$) існують між параметрами «оцінка» та «вагаюся» у каузально-орієнтованих і не каузально-орієнтованих групах завдань та у групах респондентів природничо-наукового та гуманітарного профілів. Тобто: чим нижчі результати учня за тестове завдання (і за КО, і за НКО завдання), тим більше він «вагається» щодо відповіді на нього, а чим вищі результати за тестове завдання – тим менше він вагається.

Високі кореляції (понад $0,5$) також існують між групами КО і НКО завдань у параметрах «оцінка» та «вагаюся» у групах респондентів природничо-наукового та гуманітарного профілів, між параметрами «складність» і «вагаються» у групі НКО завдань для здобувачів обох профілів та

у параметрі «складність» між групами КО і НКО завдань лише у групі респондентів гуманітарного профілю, а висока антикореляція ($-0,6$) – між параметрами «оцінка» і «складність» у групі НКО завдань в обох групах спеціалізації респондентів (див. рис. 1, рис. 2). Між усіма іншим показниками існує або низька кореляція ($0,3-0,4$), або низька антикореляція. При цьому між параметрами «оцінка» і «складність» у групах КО та НКО завдань спостерігається антикореляція.

Висновки дослідження та перспективи подальших розвідок. Експериментальне дослідження результативності використання критеріально-орієнтованих тестів для вивчення рівня сформованості каузального мислення в учнів профільної школи на прикладі навчального предмету «біологія» проведено в Україні вперше.

Його результати показують, що каузальне мислення, сформоване на основі досвіду повсякденного життя, відрізняється від каузального мислення, яке формується на основі вивчення причинно-наслідкових структур, які можуть бути досить абстрактними у цій предметній галузі.

Причинно-наслідкові зв'язки між процесами, подіями та явищами та вивчаються в курсі біології зокрема, та в предметах природничих наук, у цілому, мають засвоєні учнями. Недостатність їх засвоєння ініціює механізми каузального мислення, які виявляються в каузальному роздумі про можливі причини зв'язків, зазначені у запитанні. Можна припустити, що знання, що є в учня, і каузальне мислення роблять свій внесок у прийняття рішення про вибір опцій «вагаюся» і «складність».

Регресійний аналіз показує зв'язок між параметрами «складність» і «вагаюся» (див. табл. 2, 3, 4). При цьому у групах НКО завдань заявлена складність має відносно більший вплив на відмову здобувача від відповіді на тестове завдання (див. табл. 4, рис. 1, 2). У групах КО завдань спостерігається менша кореляція між заявленою складністю та відмовою від відповіді на запитання. Одним із пояснень цього може бути саме використання учнем навичок каузального мислення під час аналізу запитання. Це підтверджує гіпотезу, що запропонований формат тесту може бути представлений як вимірвальний інструмент для оцінювання рівня сформованості каузального мислення учня.

Судження типу еквівалентності (кон'юнкції імплікацій), що відповідають вислову «тоді й лише тоді, коли...», та широко використовуються в математиці, у біології присутні переважно у «прихованому» вигляді. Під час використання таких логічних операцій спостерігається хибна імплікація, коли учень із істинного висловлювання (умови, посилення) робить неправильний висновок.

Встановлено статистично значущі відмінності між групами каузально-орієнтованих та не каузально-орієнтованих завдань тесту у всіх параметрах, що досліджувалися: результати виконання фрагментів тесту, складність запитань тестових завдань, відмова від прийняття рішення на запитання тестового завдання в даних фрагментах (групах) тесту.

У цьому дослідженні увагу зосереджено на каузальних структурах, які уявляються респонденту як відносно прості, оскільки вже вивчалися раніше. Під час відповіді на каузально-орієнтоване запитання учень очевидно спирається на наявні у нього знання у цій предметній області. Іншими словами, пояснення окремої події має здійснюватися на основі наявних у респондента знань про причину події. Природньо припустити, що правильна відповідь на каузально-орієнтоване запитання обумовлена не лише рівнем сформованості в учня каузального мислення, а й наявністю у нього конкретних знань. Проте експериментальні дані, отримані на таких малих вибірках, показують суттєві відмінності у результатах відповідей на різні типи запитань (КО і НКО).

Деякі відмінності в результатах тестування за всіма параметрами для груп респондентів різної спеціалізації можуть пояснюватися відмінністю у програмах навчання. В обох групах спеціалізації здобувачів кореляційні зв'язки між параметрами, що вивчалися, для груп не каузально-орієнтованих запитань більші, ніж для каузально-орієнтованих. Привертає увагу той факт, що кореляція для груп КО і НКО запитань щодо змінної (параметру) «складність» у групі респондентів гуманітарного профілю вдвічі більша, ніж у групі респондентів природничо-наукового профілю (див. рис. 1, 2).

Неправильні відповіді на каузально-орієнтовані запитання тестових завдань можуть свідчити про те, що в учнів сформувалися неправильні уявлення щодо причинно-наслідкових зв'язків між процесами, подіями та явищами, які вивчаються у конкретному фрагменті предметної галузі. Причинами цього можуть бути: 1) пропуск послідовних логічних операцій – помилка діяльності; 2) неправильне розуміння (або трактування) явища, яке описано (або спостерігається), – помилка розпізнавання; 3) неправильне віднесення явища, яке описано (або спостерігається), до відомих теорій – помилка класифікації. З'ясування ступеня впливу перерахованих причин на особливості формування каузального мислення вимагає додаткових досліджень.

Подальшого вивчення потребують питання впливу інших факторів (окрім параметру «складність») на відмову здобувачів від ухвалення рішення щодо вибору правильної (на їхню думку) відповіді на каузально-орієнтоване запитання.

Використані джерела

- Гривко, А. В. (2020). Вплив послідовності виконання тестових завдань на результати тестування: огляд і узагальнення актуальних досліджень. У І. Б. Вашеняк (Гол. ред.), *Оцінювання досягнень здобувачів загальної середньої освіти: сучасні виклики та перспективи: збірник наукових праць за матеріалами Всеукраїнської науково-практичної конференції* (с. 32–38). Видавництво ХОІППО. <https://lib.iitta.gov.ua/726253/>.
- Гривко, А. В., & Жук, Ю. О. (2019). Використання засобів ІКТ у процесі експериментального дослідження емотивно-оцінного ставлення учнів до різних форм тестових завдань з української мови. *Інформаційні технології і засоби навчання*, 70(2), 285–297. <https://doi.org/10.33407/itlt.v70i2.2621>. <https://journal.iitta.gov.ua/index.php/itlt/article/view/2621>.
- Жук, Ю. О. (1998). Використання елементів формальної логіки у процесі розв'язування навчальної фізичної задачі. У *Матеріали науково-практичної конференції «Дидактичні проблеми фізичної освіти в Україні»* (с. 63–65). Чернігівський ДДП.
- Жук, Ю. О., & Ващенко, Л. С. (2019). Вплив форми тестових завдань та профільної спеціалізації класів на результати виконання здобувачами середньої освіти тесту з біології. *Український педагогічний журнал*, (2), 90–99. <https://doi.org/10.32405/2411-1317-2019-2-90-99>. <https://uej.undip.org.ua/index.php/journal/article/view/428>.
- Жук, Ю. О., & Ващенко, Л. С. (2020). Особливості впливу композиції тесту з біології на результативні показники тестування старшокласників. *Український педагогічний журнал*, (1), 20–31. <https://doi.org/10.32405/2411-1317-2020-1-20-31>. <https://uej.undip.org.ua/index.php/journal/article/view/369>.
- Постанова Кабінету Міністрів України № 898 «Про деякі питання державних стандартів повної загальної середньої освіти» (2020, Вересень 30). *Верховна Рада України. Законодавство України*. <https://zakon.rada.gov.ua/laws/show/898-2020-%D0%BF#Text>.
- Bauer, D., Kopp, V., & Fischer, M. R. (2007). Answer changing in multiple choice assessment change that answer when in doubt – and spread the word!. *BMC Medical Education*, 7, 28 (2007), <https://doi.org/10.1186/1472-6920-7-28>.
- Berkant, H. G. (2009). An Investigation of Students' Meaningful Causal Thinking Abilities in Terms of Academic Achievement, Reading Comprehension and Gender. *Educational Sciences: Theory and Practice*, 9(3), 1149–1165. <https://eric.ed.gov/?id=EJ858922>.
- Bratfisch, O., Borg, G., & Dornič, S. (1972). Perceived Item-Difficulty in Three Tests of Intellectual Performance Capacity. *Reports from the Institute of Applied Psychology, the University of Stockholm*, 29. <https://eric.ed.gov/?id=ED080552>.
- Buchsbaum, D., Griffiths, T. L., Plunkett, D., Gopnik, A., & Baldwin, D. (2015). Inferring action structure and causal relationships in continuous sequences of human action. *Cognitive Psychology*, 76, 30–77. <https://doi.org/10.1016/j.cogpsych.2014.10.001>.
- Buehner, M. J., & Cheng, P. W. (2005). Causal Learning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 143–168). Cambridge University Press. <https://psycnet.apa.org/record/2005-09680-007>.
- Casteel, C. A. (2014). Answer Changing on Multiple-Choice Test Items Among Eighth-Grade Reader. *The Journal of Experiential Education*, 59(4), 300–309. <https://doi.org/10.1080/00220973.1991.10806568>.

- Chang, S. L. (2015). Students' Perceived Test Difficulty, Perceived Performance and Actual Performance of Oral Tests. 23. 1225–1242. https://www.researchgate.net/publication/285206775_Students'_perceived_test_difficulty_perceived_performance_and_actual_performance_of_oral_tests.
- Chen, Ch., Bao, L., Fritchman, J. C., & Ma, H. (2021). Causal reasoning in understanding Newton's third law. *Physical Review Physics Education Research*, 17, 010128. <https://doi.org/10.1103/PhysRevPhysEducRes.17.010128>.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>.
- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological review*, 116(4), 661–716. <https://doi.org/10.1037/a0017201>. <https://psycnet.apa.org/record/2009-18254-002>.
- Hanna, G. S. (2010). To Change Answers or Not to Change Answers: That is the Question. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 62(9), 414–416. <https://doi.org/10.1080/00098655.1989.10114108>.
- Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, 62, 135–163. <https://doi.org/10.1146/annurev.psych.121208.131634>.
- Hryvko, A., & Zhuk, Yu. (2020). The research of the effect of different question forms and their combination variability on the results of testing. *SHS Web Conf.*, 75, 04004. <https://doi.org/10.1051/shsconf/20207504004>.
- Lalaye, S. A. (2020). The logic of language and the counterintuitiveness of the logical tool of material implication. *Sophia: An African journal of Philosophy*, 19(2), 220–228. <https://www.ajol.info/index.php/sophia/article/view/213691>.
- Lalor, J. P., Wu, H., Munkhdalai, T., & Yu, H. (2018). Understanding Deep Learning Performance through an Examination of Test Set Difficulty: A Psychometric Case Study. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 4711–4716). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D18-1500>.
- Le Pelley, M. E., Griffiths, O., & Beesley, T. (2017). Associative accounts of causal cognition. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 13–28). Oxford University Press. <https://psycnet.apa.org/record/2017-30850-002>.
- McMorris, R. F., DeMer, L. P., & Schwarz, Sh. P. (1987). Attitudes, Behaviors, and Reasons for Changing Responses Following Answer-Changing Instruction. *Journal of Educational Measurement*, 24(2), 131–143. <https://doi.org/10.1111/j.1745-3984.1987.tb00269.x>.
- Moss, S. (2018). *Probabilistic Knowledge*. Oxford University Press.
- Mustakim, M., Mansyur, J., Hatibe, A., Rizal, M., & Kaharu, S. N. (2020). Analysis of Students' Causal Reasoning in Physics Problem Solving. *Journal of Physics: Conference Series*, 1521, 022058. <https://doi.org/10.1088/1742-6596/1521/2/022058>.
- Naimnule, L. & Corebima, A. D. (2018). The correlation between metacognitive skills and critical thinking skills toward students' process skills in biology learning. *Journal of Pedagogical Research*, 2(2), 122–134. <https://eric.ed.gov/?id=EJ1300666>.
- Rivers, M. L., Dunlosky, J., & Persky, A. M. (2020). Measuring Metacognitive Knowledge, Monitoring, and Control in the Pharmacy Classroom and Experiential Settings. *American journal of pharmaceutical education*, 84(5), 7730. <https://doi.org/10.5688/ajpe7730>.
- Russell, D. (1982). The Causal Dimension Scale: A measure of how individuals perceive causes. *Journal of Personality and Social Psychology*, 42(6), 1137–1145. <https://doi.org/10.1037/0022-3514.42.6.1137>.
- Padilla, M. J. (1990). *The science process skills*. NARST. A global organization for improving science teaching and learning through research. <https://narst.org/research-matters/science-process-skills>.
- Pazzani, M., Oyer, M., & Flowers, M. (1987, August). Using prior learning to facilitate the learning of new causal theories. In *IJCAI'87: Proceedings of the 10th international joint conference on Artificial intelligence (Volume 1)* (pp. 277–279). <https://dl.acm.org/doi/abs/10.5555/1625015.1625071>.
- Prinsell, C. P., Ramsey, Ph. H., & Ramsey P. P. (1994). Score Gains, Attitudes, and Behavior Changes Due to Answer-Changing Instruction. *Journal of Educational Measurement*, 31(4), 327–337. <https://doi.org/10.1111/j.1745-3984.1994.tb00450.x>.
- Shou, Yi., & Smithson, M. (2015) Effects of question formats on causal judgments and model evaluation. *Frontiers in Psychology*, (6). <https://doi.org/10.3389/fpsyg.2015.00467>.

- Taib, F., & Yusoff, M. S. B. (2014). Difficulty index, discrimination index, sensitivity and specificity of long case and multiple choice questions to predict medical students' examination performance. *Journal of Taibah University Medical Sciences*, 9(2), 110–114. <https://doi.org/10.1016/j.jtumed.2013.12.002>.
- Tiffin-Richards, S. P., Lenhart, J., & Marx, P. (2022). When do examinees change their initial answers? The influence of task instruction, response confidence, and subjective task difficulty. *Learning and instruction*, 82, 101654. <https://doi.org/10.1016/j.learninstruc.2022.101654>.

References

- Hryvko, A. V. (2020). Vplyv poslidovnosti vykonannia testovykh zavdan na rezultaty testuvannia: ohliad i uzahalnennia aktualnykh doslidzhen. U I. B. Vasheniak (Hol. red.), *Otsiniuvannia dosiahnen zdozvachiv zahalnoi serednoi osvity: suchasni vyklyky ta perspektyvy: zbirnyk naukovykh prats za materialamy Vseukrainskoi naukovopraktychnoi konferentsii* (s. 32–38). Vydavnytstvo KhOIPPO. <https://lib.iitta.gov.ua/726253/>. (in Ukrainian).
- Hryvko, A. V., & Zhuk, Yu. O. (2019). Vykorystannia zasobiv IKT u protsesi eksperymentalnoho doslidzhennia emotyvno-otsinnoho stavlennia uchniv do riznykh form testovykh zavdan z ukrainskoi movy. *Informatsiini tekhnologii i zasoby navchannia*, 70(2), 285–297. <https://doi.org/10.33407/itlt.v70i2.2621>. <https://journal.iitta.gov.ua/index.php/itlt/article/view/2621>. (in Ukrainian).
- Zhuk, Yu. O. (1998). Vykorystannia elementiv formalnoi lohiky u protsesi rozviazuvannia navchalnoi fizychnoi zadachi. U *Materialy naukovopraktychnoi konferentsii «Dydaktychni problemy fizychnoi osvity v Ukraini»* (s. 63–65). Chernihivskiy DPI. (in Ukrainian).
- Zhuk, Yu. O., & Vashchenko, L. S. (2019). Vplyv formy testovykh zavdan ta profilnoi spetsializatsii klaviv na rezultaty vykonannia zdozvachamy serednoi osvity testu z biolohii. *Ukrainskyi pedahohichnyi zhurnal*, (2), 90–99. <https://doi.org/10.32405/2411-1317-2019-2-90-99>. <https://uej.undip.org.ua/index.php/journal/article/view/428>. (in Ukrainian).
- Zhuk, Yu. O., & Vashchenko, L. S. (2020). Osoblyvosti vplyvu kompozytsii testu z biolohii na rezultatyvni pokaznyky testuvannia starshoklasnykiv. *Ukrainskyi pedahohichnyi zhurnal*, (1), 20–31. <https://doi.org/10.32405/2411-1317-2020-1-20-31>. <https://uej.undip.org.ua/index.php/journal/article/view/369>. (in Ukrainian).
- Postanova Kabinetu Ministriv Ukrainy № 898 «Pro deaki pytannia derzhavnykh standartiv povnoi zahalnoi serednoi osvity» (2020, Veresen 30). *Verkhovna Rada Ukrainy. Zakonodavstvo Ukrainy*. <https://zakon.rada.gov.ua/laws/show/898-2020-%D0%BF#Text> (in Ukrainian).
- Bauer, D., Kopp, V., & Fischer, M. R. (2007). Answer changing in multiple choice assessment change that answer when in doubt – and spread the word!. *BMC Medical Education*, 7, 28 (2007), <https://doi.org/10.1186/1472-6920-7-28>. (in English).
- Berkant, H. G. (2009). An Investigation of Students' Meaningful Causal Thinking Abilities in Terms of Academic Achievement, Reading Comprehension and Gender. *Educational Sciences: Theory and Practice*, 9(3), 1149–1165. <https://eric.ed.gov/?id=EJ858922>. (in English).
- Bratfisch, O., Borg, G., & Dornič, S. (1972). Perceived Item-Difficulty in Three Tests of Intellectual Performance Capacity. *Reports from the Institute of Applied Psychology, the University of Stockholm*, 29. <https://eric.ed.gov/?id=ED080552>. (in English).
- Buchsbaum, D., Griffiths, T. L., Plunkett, D., Gopnik, A., & Baldwin, D. (2015). Inferring action structure and causal relationships in continuous sequences of human action. *Cognitive Psychology*, 76, 30–77. <https://doi.org/10.1016/j.cogpsych.2014.10.001>. (in English).
- Buehner, M. J., & Cheng, P. W. (2005). Causal Learning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 143–168). Cambridge University Press. <https://psycnet.apa.org/record/2005-09680-007>. (in English).
- Casteel, C. A. (2014). Answer Changing on Multiple-Choice Test Items Among Eighth-Grade Reader. *The Journal of Experiential Education*, 59(4), 300–309. <https://doi.org/10.1080/00220973.1991.10806568>. (in English).
- Chang, S. L. (2015). Students' Perceived Test Difficulty, Perceived Performance and Actual Performance of Oral Tests. 23. 1225–1242. https://www.researchgate.net/publication/285206775_Students'_perceived_test_difficulty_perceived_performance_and_actual_performance_of_oral_tests. (in English).
- Chen, Ch., Bao, L., Fritchman, J. C., & Ma, H. (2021). Causal reasoning in understanding Newton's third law. *Physical Review Physics Education Research*, 17, 010128. <https://doi.org/10.1103/PhysRevPhysEducRes.17.010128>. (in English).

- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10), 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>. (in English).
- Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological review*, 116(4), 661–716. <https://doi.org/10.1037/a0017201>. <https://psycnet.apa.org/record/2009-18254-002>. (in English).
- Hanna, G. S. (2010). To Change Answers or Not to Change Answers: That is the Question. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 62(9), 414–416. <https://doi.org/10.1080/00098655.1989.10114108>. (in English).
- Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, 62, 135–163. <https://doi.org/10.1146/annurev.psych.121208.131634>. (in English).
- Hryvko, A., & Zhuk, Yu. (2020). The research of the effect of different question forms and their combination variability on the results of testing. *SHS Web Conf.*, 75, 04004. <https://doi.org/10.1051/shsconf/20207504004>. (in English).
- Laleye, S. A. (2020). The logic of language and the counterintuitiveness of the logical tool of material implication. *Sophia: An African journal of Philosophy*, 19(2), 220–228. <https://www.ajol.info/index.php/sophia/article/view/213691>. (in English).
- Lalor, J. P., Wu, H., Munkhdalai, T., & Yu, H. (2018). Understanding Deep Learning Performance through an Examination of Test Set Difficulty: A Psychometric Case Study. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 4711–4716). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D18-1500>. (in English).
- Le Pelley, M. E., Griffiths, O., & Beesley, T. (2017). Associative accounts of causal cognition. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 13–28). Oxford University Press. <https://psycnet.apa.org/record/2017-30850-002>. (in English).
- McMorris, R. F., DeMer, L. P., & Schwarz, Sh. P. (1987). Attitudes, Behaviors, and Reasons for Changing Responses Following Answer-Changing Instruction. *Journal of Educational Measurement*, 24(2), 131–143. <https://doi.org/10.1111/j.1745-3984.1987.tb00269.x>. (in English).
- Moss, S. (2018). *Probabilistic Knowledge*. Oxford University Press. (in English).
- Mustakim, M., Mansyur, J., Hatibe, A., Rizal, M., & Kaharu, S. N. (2020). Analysis of Students' Causal Reasoning in Physics Problem Solving. *Journal of Physics: Conference Series*, 1521, 022058. <https://doi.org/10.1088/1742-6596/1521/2/022058>. (in English).
- Naimnule, L. & Corebima, A. D. (2018). The correlation between metacognitive skills and critical thinking skills toward students' process skills in biology learning. *Journal of Pedagogical Research*, 2(2), 122–134. <https://eric.ed.gov/?id=EJ1300666>. (in English).
- Rivers, M. L., Dunlosky, J., & Persky, A. M. (2020). Measuring Metacognitive Knowledge, Monitoring, and Control in the Pharmacy Classroom and Experiential Settings. *American journal of pharmaceutical education*, 84(5), 7730. <https://doi.org/10.5688/ajpe7730>. (in English).
- Russell, D. (1982). The Causal Dimension Scale: A measure of how individuals perceive causes. *Journal of Personality and Social Psychology*, 42(6), 1137–1145. <https://doi.org/10.1037/0022-3514.42.6.1137>. (in English).
- Padilla, M. J. (1990). *The science process skills*. NARST. A global organization for improving science teaching and learning through research. <https://narst.org/research-matters/science-process-skills> (in English).
- Pazzani, M., Oyer, M., & Flowers, M. (1987, August). Using prior learning to facilitate the learning of new causal theories. In *IJCAI'87: Proceedings of the 10th international joint conference on Artificial intelligence (Volume 1)* (pp. 277–279). <https://dl.acm.org/doi/abs/10.5555/1625015.1625071>. (in English).
- Prinsell, C. P., Ramsey, Ph. H., & Ramsey P. P. (1994). Score Gains, Attitudes, and Behavior Changes Due to Answer-Changing Instruction. *Journal of Educational Measurement*, 31(4), 327–337. <https://doi.org/10.1111/j.1745-3984.1994.tb00450.x>. (in English).
- Shou, Yi., & Smithson, M. (2015) Effects of question formats on causal judgments and model evaluation. *Frontiers in Psychology*, (6). <https://doi.org/10.3389/fpsyg.2015.00467>. (in English).
- Taib, F., & Yusoff, M. S. B. (2014). Difficulty index, discrimination index, sensitivity and specificity of long case and multiple choice questions to predict medical students' examination performance. *Journal of Taibah University Medical Sciences*, 9(2), 110–114. <https://doi.org/10.1016/j.jtumed.2013.12.002>. (in English).

Tiffin-Richards, S. P., Lenhart, J., & Marx, P. (2022). When do examinees change their initial answers? The influence of task instruction, response confidence, and subjective task difficulty. *Learning and instruction, 82*, 101654. <https://doi.org/10.1016/j.learninstruc.2022.101654>. (in English).

Yurii Zhuk, D. Sc. (Pedagogy), Associate Professor, Head of the Department of Monitoring and Assessment of the Quality of General Secondary Education, Institute of Pedagogy of the National Academy of Educational Sciences of Ukraine, Kyiv, Ukraine.

Research interests: pedagogical quality metrics; learning environment.

Lidiia Vashchenko, PhD (Pedagogy), Senior Researcher, Senior Researcher of the Department of Monitoring and Assessment of the Quality of General Secondary Education, Institute of Pedagogy of the National Academy of Educational Sciences of Ukraine, Kyiv, Ukraine.

Research interests: peculiarities of internal quality control of general secondary education; technologies for evaluating of students' educational achievements in biology of general secondary education institutions; test assessment technologies.

Svitlana Naumenko, PhD (Pedagogy), Senior Researcher, Senior Researcher of the Department of Monitoring and Assessment of the Quality of General Secondary Education, Institute of Pedagogy of the National Academy of Educational Sciences of Ukraine, Kyiv, Ukraine.

Research interests: quality of general secondary education; monitoring the quality of education; international comparative studies of the quality of general secondary education; test technologies.

AN EXPERIMENTAL STUDY OF THE FORMATION OF CAUSAL THINKING OF APPLICANTS FOR GENERAL SECONDARY EDUCATION

Abstract. The authors proposed and tested a methodology for assessing the level of formation of causal thinking, based on test technologies. The methodology rests upon the using of a criterion-referenced test containing two groups of questions that have the wording which initiates different mental activities of students. There are options in one group of questions, for possible cause-and-effect relationships between processes, phenomena or events studied in this subject area are offered for selection (causally oriented group of questions). Another group of questions is aimed at knowledge of previously learned concepts and clearly does not contain a request to indicate cause-and-effect relationships between them (not a causally oriented group of questions). Both groups of questions are represented by test tasks with the choice of an answer from a number of proposed options.

In order to study the properties of the test, the “difficulty” and “hesitating” options were used in its structure. The data obtained as a result of the experiment allow us to conclude that the proposed structure of the test showed a sufficient ability to differentiate the results of pedagogical measurements with regard to causally oriented and non-causally oriented questions, which, in turn, allows us to compare the students' levels of assimilation of educational material and the formation of causal thinking throughout of using one test. The short format of the test makes it possible to use it directly during the real educational process. Compilation of test tasks is quite simple for the teacher to implement independently in accordance with the goals of contextual learning control.

Keywords: applicants for education; causal thinking; test tools; criterion-referenced test.