

Бригінець Валентин, Подласов Сергій, Матвійчук Олексій

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

ОЦІНКА ЯКОСТІ ЗАВДАНЬ У ТЕСТОВІЙ ФОРМІ

Анотація. Стаття присвячена аналізу ускладнень, які виникають при створенні і практичному застосуванню тестів та аналізі якості завдань у тестовій формі. Для запобігання передачі інформації про зміст тесту від одних студентів до інших викладач змушений створювати паралельні форми окремих завдань для їх застосування в різних групах. Це суттєво уповільнює накопичення даних, необхідних для одержання достовірних результатів статистичного аналізу. Показано, що оцінювання рівня трудності завдань для малих виборок не може дати надійні значення цього параметра. При проведенні комп'ютерного тестування на базі LMS Moodle вбудований пакет статистичної обробки даних тестування дозволяє одержати параметри як окремих завдань, так і сукупності паралельних форм завдань (тестових категорій). Застосування результатів такого аналізу дозволило скорегувати зміст тестів поточного контролю.

***Ключові слова:** тест, орієнтований на критерій; модель Дж. Раша; LMS Moodle; тестові категорії; індекс легкості; стандартне відхилення.*

Постановка проблеми. Ефективним інструментом оцінювання навчальних досягнень студенті є тестування. Можливість адекватного оцінювання визначається якістю завдань, з яких складається тест, та умовами проведення тестування.

Оцінити якість тестових матеріалів можна тільки за результатами статистичної обробки даних тестування на репрезентативній вибірці тестованих за умови, що тестування проводилося в однакових умовах. Ця вимога створює певні ускладнення для пересічного викладача, який повинен не тільки створювати завдання для формування тесту, але й організувати їх апробацію та статистичну обробку результатів.

Аналіз останніх досліджень і публікацій. Ефективність тестових методик контролю була обґрунтована в роботах С. Архангельського, Н. Тализіної, Д. Чернілевського, Ю. Бабанського, В. Безпалька, Т. Ільїної та інших. Питання створення та практичного застосування тестів обговорювалися в роботах вітчизняних та зарубіжних вчених В. Аванесова, П. Атаманчука, Р. Гласера, В. Ким, П. Клайна, А. Майорова, О. Ляшенка, М. Остапчука, В. Сергієнка, В. Хлебнікова та багатьох інших. Методика обробки результатів тестування висвітлювалися в роботах В. С. Аванесова, В.С. Кім, А.Н. Майоров, М.Б. Челишкової, В. D. Wright, G. N. Masters, T. G. Bond та інших вчених.

Мета роботи полягає в аналізі ускладнень, які виникають при створенні завдань у тестовій формі, їх апробації та статистичній обробці одержаних результатів.

Методи досліджень. Відповідно до мети роботи використовувалися такі методи дослідження: загальнонаукові, зокрема, вивчення й узагальнення педагогічного досвіду щодо засобів контролю знань учнів та студентів; методи інших наук, у т. ч. соціологічні в опитування викладачів; методи обробки статистичних результатів тестування.

Виклад основного матеріалу. Тестовий контроль навчальних досягнень студентів може проводитися на різних етапах і відповідно до цього виокремлюють поточний, тематичний, рубіжний та підсумковий. Зміст завдань тесту визначається педагогічними цілями, а організаційна форма – технічними можливостями його проведення.

Педагогічними цілями тестування можуть бути або ранжування студентів (учнів) за рівнем навчальних досягнень (визначення рейтингу), або ж виявлення рівня оволодіння ними знаннями у певній предметній області. У першому випадку застосовують тести, орієнтовані на норму, у другому – на критерій. Нормативно орієнтований тест має складатися із завдань рівномірно зростаючої трудності [1]. При застосуванні критеріально орієнтованого тесту (КОрТ) можна визначати тільки відсоток правильно виконаних завдань, однак і в цьому випадку бажано знати рівень трудності завдань для формування еквівалентних паралельних варіантів тесту, а також вибору оптимального критеріального балу (визначення «зараховано», чи «не зараховано»), або ж оцінювання результатів роботи студента при рубіжному, чи підсумковому контролі на основі таксономії Б. Блума [8], або її розвитку у роботах L.W. Anderson [6], та D. R. Krathwohl [5, 6].

Достовірні статистичні характеристики завдань у тестовій формі можуть бути одержані на репрезентативній вибірці тестованих, тобто кожне завдання повинно бути запропоновано не менше ніж 150 респондентам [3], котрі виконують його в однакових умовах. Необхідні і достатні умови для визначення статистичних характеристик завдань створюються при проведенні зовнішнього незалежного оцінювання, коли всі абітурієнти виконують однакові завдання в однакових умовах. Що ж стосується студентів ВНЗ, то подібні до ЗНО умови створити складно, оскільки пересічний викладач найчастіше проводить тестування в окремих студентських групах, які складаються з 15 – 25 осіб. При цьому тестування в різних групах проводиться неодноразово, відтак студенти різних груп знаходяться в різних умовах. Крім того, як добре відомо, психологія наших студентів є здебільше колективістською і дати списати, або підказати сусіду в більшості випадків вважається звитягою і при виконанні однакових завдань слабо підготовлені студенти мають шанс одержати

навіть кращі результати, ніж більш сильно підготовлені. Треба також урахувати, що при неодночасному проведенні тестування в різних групах при використанні одних і тих самих завдань при їх невеликій кількості зміст завдань і правильні відповідь швидко стають відомими студентам заздалегідь і тестування як метод контролю втрачає сенс [4].

З метою запобігання списуванню, взаємним підказкам та передачі інформації від одних студентських груп до інших, викладач мусить скласти декілька варіантів завдань, які є близькими за своєю сутністю, але відрізняються звучанням, тобто створювати паралельні форми, які рекомендує В. С. Аванесов [1]. Для забезпечення однакових умов для всіх студентів рівень складності завдань повинен бути близьким.

Початкове значення рівня складності завдання у тестовій формі встановлюється укладачем тесту. Однак, як показує практика, рівень складності є поняттям доволі суб'єктивним, оскільки він залежить не тільки від змісту завдань та рівня підготовленості студентів (учнів) з даної та суміжних дисциплін, але й від їхнього тезаурусу, відтак можливості розуміння змісту завдання. Наприклад, для виконання завдань з фізики необхідні знання з математики, та знання значення загальнонаукових термінів. На наш подив непоодинокі випадки, коли студенти першого курсу та учні випускних класів не знають що таке «шарнір», дехто під блоком розуміє прямокутний будівельний блок замість блока, який мається на увазі у задачах фізики, і т. д. Таких прикладів, нажаль, можна навести ще чимало і все це визначає можливість студента дати правильну відповідь на завдання, відтак і рівень його складності. Саме тому складність окремих завдань може бути встановленим тільки за даними статистичної обробки результатів тестування, яку може проводитися за класичною, або сучасною методикою в моделі IRT – Item Response Theory (IRT), запропонованої Г. Рашем [10, 11] і розвиненої в роботах А. Бірнбаума [7] та інших вчених.

Трудність завдань для окремих студентських груп визначалася нами за результатами вхідного тестування з фізики студентів першого курсу. Кількість студентів у групах коливалася від 15 до 25 осіб. Тестування проводилося режимі віддаленого доступу за завданнями, які пропонувалися абітурієнтам при зовнішньому незалежному оцінюванні в 2008 – 2017 роках. Завдання розміщені на сайті кафедри загальної фізики та фізики твердого тіла Київського політехнічного інституту ім. Ігоря Сікорського (<http://physics.kpi.ua>).

Обробка результатів тестування здійснювалася за допомогою програми WinSteps, в якій реалізується алгоритм IRT.

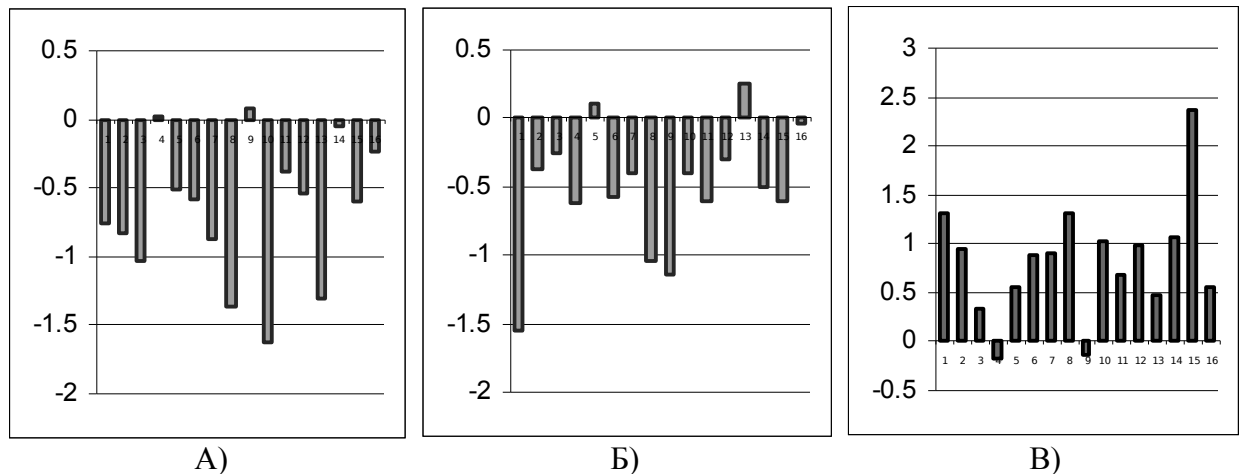


Рис. 1. Результати обчислення рівня трудності завдань закритої А), Б) та закритої В) форм за даними тестування різних студентських груп

Результати обчислень (рис. 1) показали, що для різних студентських груп одні й ті самі завдання мають суттєво відмінні значення трудності, що не дозволяє використовувати ці дані навіть для наближеної оцінки цього параметра.

Необхідність індивідуалізації включених в тест завдань ускладнює оцінювання трудності кожного з них, оскільки зростає час, необхідний для одержання статистично достовірної кількості даних випробовування. За таких умов, на початковому етапі, можна визначати трудність та інші статистичні характеристики групи завдань. Саме така можливість

реалізована в підсистемі тестування LMS Moodle. В ній можна створювати тестові категорії, з яких випадковим чином вибираються завдання при формуванні тесту для кожного студента. Загальна кількість завдань тесту для всіх студентів однакова і визначається запланованою викладачем кількістю категорій.

Таке формування тесту, з одного боку, дозволяє запобігати передачі інформації від одних студентських груп до інших, але з іншого – значно ускладнює процедуру визначення статистичних параметрів окремих завдань, оскільки одержання репрезентативної виборки відповідей потребує значного часу.

Вбудований в Moodle пакет статистичної обробки результатів тестування надає можливість користувачу одержувати як параметри тесту як цілого, так і параметри окремих завдань і тестових категорій.

Для тесту в цілому визначаються такі параметри як кількість спроб, середня оцінка, медіана, стандартне відхилення, асиметрія та ексцес розподілу одержаних оцінок, коефіцієнт внутрішньої узгодженості та деякі інші.

Для тестових категорій та окремих завдань визначаються: загальна кількість проб, успішність виконання завдання (індекс легкості), стандартне відхилення, оцінка навмання, призначена вага, ефективна вага, розрізнення (індекс дискримінації) та ефективність розрізнення (дискримінації).

Загальна кількість спроб дозволяє робити висновки про репрезентативність виборки, для якої виконані обчислення статистичних параметрів.

Успішність виконання завдання (індекс легкості) дає відношення середньої кількості балів, набраних студентами за дане завдання, до максимально можливої кількості балів. Якщо оцінка за завдання може бути тільки 1 – виконано, 0 – не виконано, то цей параметр дорівнюватиме

відношенню кількості студентів, котрі дали правильну відповідь, до загальної кількості відповідей на дане запитання. Збалансований тест повинен включати завдання з різним індексом легкості, однак завдання із значенням цього параметра близьким до нуля, або одиниці бажано виключати з тесту.

Стандартне відхилення показує розкид значень оцінок. Якщо цей параметр менший ніж 0,3, то переважна більшість студентів дали однакову відповідь на це завдання, отже за його допомогою не вдається ефективно відокремити слабо і добре підготовлених студентів.

Оцінка навмання визначається тільки для завдань закритого типу і показує ймовірність угадування правильної відповіді.

Призначена вага – визначає призначений укладачем тесту відсотковий внесок правильної відповіді на це завдання в загальну оцінку. Цей параметр визначається тільки для тестової категорії, тобто для позиції завдання в тесті.

Ефективна вага характеризує фактичний відсотковий внесок завдань певної категорії в загальну оцінку. В ідеалі призначена та ефективна вага повинні бути однаковими. Враховавши значення ефективної ваги, можна скорегувати призначену вагу.

Розрізнення (індекс дискримінації) наближено показує здатність завдання відокремити сильно і слабо підготовлених студентів.

Ефективність розрізнення (дискримінації) є коефіцієнтом кореляції між відповіддю на дане завдання і відповіддю на всі завдання тесту.

Для тестів досягнень, які орієнтовані на критерій і визначають рівень оволодіння студентами запланованим обсягом знань, найбільш важливими є індекс легкості, стандартне відхилення, призначена та ефективна вага. Параметри, що характеризують дискримінативність завдань, для критеріально орієнтованого, тесту особливого значення не мають.

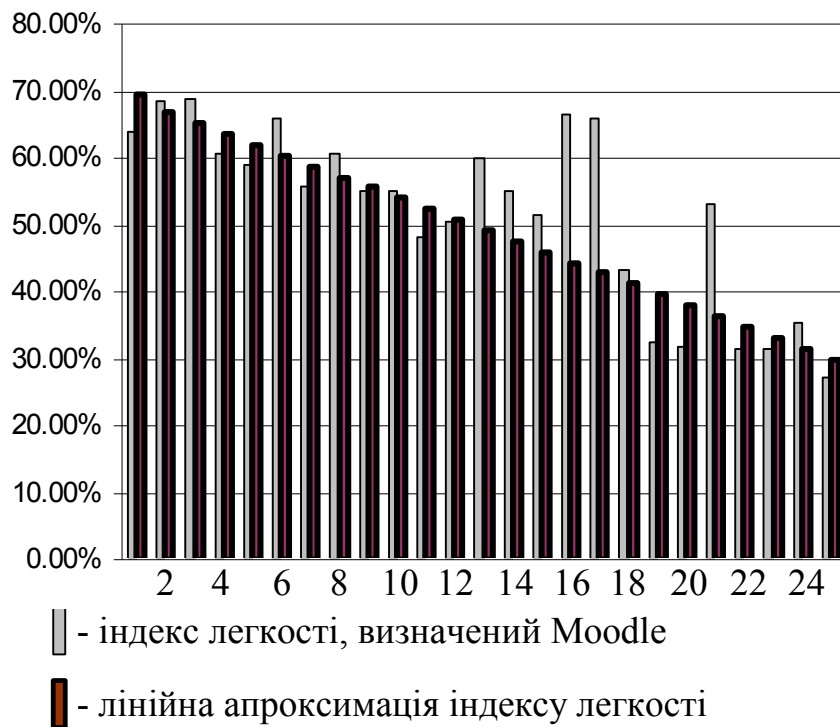


Рис. 2. Індекс легості завдань комплексного тесту з розділу «Механіка»

На кафедрі загальної фізики та фізики твердого тіла КПІ ім. І Сікорського LMS Moodle використовується для розміщення навчальних матеріалів та моніторингу результатів навчальної діяльності студентів та проведення профорієнтаційної роботи з фізики з абітурієнтами [2]

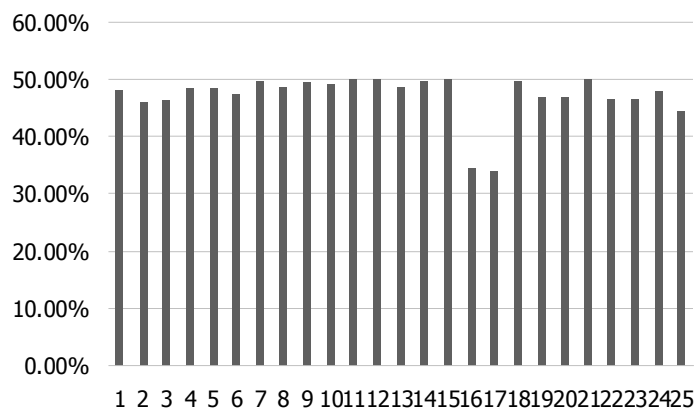


Рис. 3. Стандартне відхилення для тестових завдань комплексного тесту з розділу «Механіка»

Для оцінки результатів навчальної діяльності студентів складені тестові завдання з курсу фізики за програмою підготовки бакалаврів. Аналіз статистики виконання завдань тестів поточного контролю тесту, які

ми вважаємо критеріально орієнтованими, дозволив виявити недоліки складених тестів. Наприклад, завдання комплексного тесту з розділу «Механіка» мають монотонно зростаючий рівень складності – інакше практично лінійно спадаючий індекс легкості (рис. 2). Виняток складають завдання з категорій 16, 17 і 21, які потребують корекції. Про це свідчить низьке значення стандартного відхилення (рис. 3) для цих завдань.

Висновки.

1) Статистичний аналіз результатів тестування показав, що трудність завдань, яка визначається на малих виборках тестованих, не можна використовувати навіть для наближеної оцінки цього параметра.

2) При застосуванні системі підтримки навчального процесу Moodle, можна створювати тестові категорії, які містять паралельні форми тестових завдань, і автоматично одержувати статистичні характеристики як окремих завдань, так і відповідних категорій. Це дозволяє виявляти невдалі завдання, або тестові категорії і вносити відповідні корективи.

У подальшому планується розширення бази тестових завдань і накопичення даних, що дозволить одержати їх достовірні статистичні характеристики

БІБЛІОГРАФІЯ

1. Аванесов В.С. Композиция тестовых заданий. / Аванесов В.С. – М. Центр тестирования, 3 изд, 2002 г. – 239 с.
2. Долянiвська О.В. Тестування учнів з фізики при використанні програмної платформи Moodle / Долянiвська О.В., Матвійчук О.В., Подласов С.О. // Вісник Чернігiвського національного педагогічного університету [Текст]. Вип. 89. / Чернігiвський національний педагогічний університет імені Т.Г. Шевченка; гол. ред. Носко М.О. – Чернігiв: ЧНПУ, 2011. – С. 242-245.

3. Подласов С. Статистичний аналіз тестових завдань. Інноваційні комп'ютерні технології у вищій школі / Подласов С., Бригінець В., Матвійчук О. // Матеріали 7-ї науково-практичної конференції. м. Львів, 17-19 листопада 2015 року./ [Текст]. Відп. за випуск Л.Д. Озірковський – Львів: Видавництво Львівської політехніки, 2015. – 162 с.
4. Подласов С.О. Аналіз системи тестових завдань для поточного контролю навчальної діяльності студентів / Подласов С.О., Матвійчук О.В. // Збірник наукових праць. Педагогічні науки. [Текст]. Випуск LXI. – Херсон: Видавництво ХДУ, 2012. – С.287- 291.
5. A model of learning objectives based on a taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives [Електронний ресурс]. – Режим доступу : <http://www.celt.iastate.edu/teaching-resources/effective-practice/revised-blooms-taxonomy/>
6. Anderson L. W. Rethinking Bloom's Taxonomy: Implication for testing and assessment / Anderson L. W. - [Електронний ресурс]. – Режим доступу : <https://files.eric.ed.gov/fulltext/ED435630.pdf>
7. Birnbaum A. Some latent trait models and their use in inferring an examinee's ability / Birnbaum A. – In F.M. Lord & M.R. Novick (Eds.), Statistical theories of mental test scores. Addison-Wesley Publ. Co. Reading, Mass, 1968. – P.397-479.
8. Bloom, B.S. Taxonomy of educational objectives: the classification of educational goals. [Text]. Handbook 1: Cognitive domain. / Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. – New York: David McKay, 1956.
9. Krathwohl D. R. A Revision of Bloom's Taxonomy: An Overview / [Електронний ресурс]. – Режим доступу : http://www.unco.edu/cetl/sir/stating_outcome/documents/Krathwohl.pdf

10. Rasch, G. An item analysis which takes individual differences into account. / Rasch G. // British Journal of Mathematical and Statistical Psychology, 1966. – V.19, p.49-57.

11. Wright B. D., Rasch G. Probabilistic Models: Foreword and Preface, 1960.

Bryhinets Valentyn, Podlasov Serhii, Matviichuk Oleksii

National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine

ASSESSMENT OF QUALITY OF TASKS IN THE TEST FORM

Abstract: The article is devoted to the analysis of complications that arise when creating, using and analyzing the quality of tasks in a test form. Using the same tasks to carry out testing in small student groups leads to the fact that information about the content of the test and the correct answers becomes known to a large number of students and the test as a means of evaluation loses its meaning. In order to prevent the dissemination of information about the content of the test, the teacher must create close in meaning, but formulated in different ways, i.e. parallel forms. Since the current and final control provides for testing the quality of learning the student's material, the test is critically oriented. For parallel forms of tasks for such a test, the main indicator is the level of difficulty. As the results of our study the statistical parameters of the tasks in the test form can be considered valid, if the number of responses to each of them is not less than 150. In the case of the small student's groups it occurred that the calculated level of difficulty for them is not stable and can't be used even for approximate estimation of this parameter.

When using LMS Moodle for testing, parallel forms of items are placed in test categories and the statistical parameters of individual items and categories are automatically calculated using the built-in mathematical package. For tests of achievements are criterion-oriented and allow to assess the level of mastery of students in the study material. The most important for such tests are the

facility index, the intended weight, the effective weight and the standard deviation. Other parameters defined in Moodle, such as the discrimination index and the discriminative efficiency are not important parameters for criterion-oriented tests.

The results of the analysis in LMS Moodle of our complex test on the "Mechanics" chapter of the physics course allowed us to establish that the facility index of assignments of some categories falls out of the established regularity. For the same categories, the standard deviation values are close to the threshold value. This indicates that most students give the same answers to the questions of this category. The value of facility index indicates an insufficient average level of difficulty for these tasks and determines the necessity of its correction.

Keywords: *test; criterion-oriented test; G. Rasch model; LMS Moodle; test category; facility index; standard deviation.*

Бригинец Валентин, Подласов Сергей, Матвийчук Алексей

*Национальный технический университет Украины «Киевский
политехнический институт имени Игоря Сикорского»*

ОЦЕНКА КАЧЕСТВА ЗАДАНИЙ В ТЕСТОВОЙ ФОРМЕ

Аннотация. Статья посвящена анализу трудностей, возникающих при создании, практическом применении тестов и анализе качества заданий в тестовой форме. Для предотвращения передачи информации о содержании теста от одних студентов к другим преподаватель должен создавать параллельные формы отдельных заданий для использования в различных группах. Это существенно замедляет накопление данных, необходимых для расчета значений статистических параметров отдельных заданий, которые можно было бы считать достоверными. Показано, что оценка уровня трудности заданий для малых выборок не

может дать надежные значения этого параметра. При проведении компьютерного тестирования на базе LMS Moodle встроенный пакет статистической обработки данных тестирования позволяет получить параметры как отдельных заданий, так и совокупности параллельных форм заданий, объединенных в тестовых категориях. Применение результатов такого анализа позволило скорректировать содержание тестов текущего контроля.

***Ключевые слова:** тест, ориентированный на критерий; модель Дж. Раша; LMS Moodle; тестовые категории; индекс легкости; стандартное отклонение.*

ВІДОМОСТІ ПРО АВТОРІВ

Бригінець Валентин Петрович – кандидат фізико-математичних наук, доцент кафедри загальної та теоретичної фізики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Коло наукових інтересів: дистанційна освіта, застосування сучасних інформаційних технологій у навчальному процесі.

Подласов Сергій Олександрович – старший викладач кафедри загальної фізики та фізики твердого тіла Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Коло наукових інтересів: дистанційна освіта, застосування сучасних інформаційних технологій у навчальному процесі, наступність навчання фізики в школі та ВНЗ.

Матвійчук Олексій Васильович – кандидат педагогічних наук, старший викладач кафедри загальної фізики та фізики твердого тіла Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»

Коло наукових інтересів: дистанційна освіта, застосування сучасних інформаційних технологій у навчальному процесі, міжпредметні зв'язки фізики та інформатики, наступність навчання фізики в школі та ВНЗ.