

Додаток М

до Звіту про результати першого циклу загальнодержавного моніторингового дослідження якості початкової освіти

«Стан сформованості читацької та математичної компетентностей випускників початкової школи закладів загальної середньої освіти». Частина I. Методологія та технологія. 2018 р.

Словник термінів

Асиметрія (Skewness (Sk)) – характеристика «перекосу» чи несиметричності «крил» розподілу тестових балів, набраних учнями-учасниками тестування, які виконували певну тестову частину. $Sk=0$ свідчить про відсутність асиметрії, що дає змогу робити висновок про нормальність розподілу тестових балів. $Sk>0$ свідчить, що асиметрія є додатною, або лівобічною, тобто в розподілі найчастіше трапляються значення менші M . $Sk<0$ свідчить, що асиметрія є від'ємною, або правобічною, тобто в розподілі найчастіше трапляються значення більші M . Помітним відхиленням значення Sk від 0 вважають $\pm 0,1$ і більше.

Бал тестовий (*Test Score*) – це частина інформації, зазвичай у числовому вигляді, яка відображає результати тестування. Одним з офіційних визначень тестового бала є «зведення доказів, що містяться у відповідях випробовуваного на тестові завдання, які стосуються вимірюваного конструкту чи конструктів» (*Thissen, D., & Wainer, H. (2001)*).

Бал у шкалі 100–300 – лінійно перетворений бал у шкалі логітів із фіксацією середнього значення шкали на рівні 200 балів і середнім відхиленням, рівним 30 балам. Середнє значення відповідає значенню нуля логітів (значення, яке відповідає ймовірності 50 % розв'язати тестове завдання зі складністю нуля логітів).

Бал у шкалі логітів – перетворений тестовий бал у бал шкали логітів, розрахований на основі одної з моделей IRT.

Вибіркова сукупність, або **вибірка** (*Sample*) – певна кількість елементів генеральної сукупності, які спеціальним чином відібрані для участі в моніторинговому дослідженні та є мікромоделлю генеральної сукупності. Структура вибіркової сукупності повинна максимально збігатися зі структурою генеральної та відображати всі її основні характеристики й ознаки. Вибіркова сукупність учнів формується різними методами (*sampling methods*) шляхом випадкового відбору учнів із генеральної сукупності. Вибіркова сукупність – це фактична кількість учасників, які брали участь у моніторинговому дослідженні.

Вибірковий метод (*Sampling Method*) – це сукупність математичних засобів і обґрунтувань, які використовують при застосуванні вибіркового методу дослідження, за якого закономірності й характеристики, властиві якійсь генеральній сукупності, визначають дослідженням деякої її частини. Він дає змогу дістати потрібні дані з меншими затратами часу й коштів, а також швидше підбивати підсумки й робити відповідні висновки, тобто є економічнішим та оперативнішим.

Генеральна сукупність (*Population*) – уся сукупність елементів, що підлягає дослідженню. Генеральною сукупністю учнів початкової школи, тобто цільовою групою учасників моніторингового дослідження (*national target population*), були всі учні, які

завершують здобуття початкової освіти в ЗЗСО незалежно від типу ЗЗСО, місця проживання (місця розташування ЗЗСО), розміру ЗЗСО тощо.

Дизайн вибірки (*Sample Design*) – організаційно-логічна модель структури вибіркової сукупності.

Ексцес (*Kurtosis (K)*) – характеристика «гостроверхості» графіка розподілу тестових балів, набраних учнями-учасниками тестування, які виконували певну тестову частину, порівняно з графіком нормального розподілу (розподілу Гаусса), що має такі ж значення середнього та стандартного відхилення. $K=0$ свідчить, що ексцес відсутній, а крива розподілу балів не «гостріша» і не «плоскіша» за гауссівську, тобто відповідає нормальному розподілу. $K>0$ свідчить, що ексцес додатний, а крива розподілу має вищу та «гострішу» вершину, ніж крива нормального (гауссівського) розподілу. $K<0$ свідчить, що ексцес від'ємний, а крива розподілу має нижчу та «плоскішу» вершину, ніж крива нормального (гауссівського) розподілу.

Індивідуальна крива підготовленості учасника (*Person Response Curve*) – графічне відображення умовної ймовірності того, що учасник із певним рівнем підготовки буде виконувати тестове завдання з різним рівнем складності.

Книжка для читання – у межах моніторингового дослідження умовна назва для буклету з кольоровим друком, що містить текст, із яким має працювати учень-учасник під час виконання тесту з читання. За своєю природою книжка для читання – відокремлене представлення стимулу, необхідного для виконання тестових завдань (тестової частини).

Компетентність – динамічна комбінація знань, вмінь і практичних навичок, способів мислення, професійних, світоглядних і громадянських якостей, морально-етичних цінностей, яка визначає здатність особи успішно здійснювати професійну та подальшу навчальну діяльність і є результатом навчання на певному рівні вищої освіти (*стаття 1 Закон України «Про вищу освіту»*).

Компетентність математична – здатність особи бачити математику в житті, створювати математичні моделі об'єктів, явищ, процесів навколишнього світу, застосовувати досвід математичної діяльності під час розв'язування навчально-пізнавальних і практично зорієнтованих завдань (*Програма загальнодержавного моніторингового дослідження якості початкової освіти «Стан сформованості читацької та математичної компетентностей випускників початкової школи загальноосвітніх навчальних закладів», затверджена наказом МОНУ від 20.03.2018 р. № 256*).

Компетентність читацька – здатність особи широко розуміти текст як частину повсякденного життя й навчальної діяльності, шукати нову інформацію, відтворювати та використовувати її, інтерпретувати зміст і формулювати умовиводи, осмислювати й оцінювати зміст і форму тексту тощо (*Програма загальнодержавного моніторингового дослідження якості початкової освіти «Стан сформованості читацької та математичної компетентностей випускників початкової школи загальноосвітніх навчальних закладів», затверджена наказом МОНУ від 20.03.2018 р. № 256*).

Кореляція тестового завдання (*Correlation*) – зв'язок між результатами виконання певного тестового завдання тестової частини учнем-учасником тестування та загальним тестовим балом, отриманим ним за всю тестову частину, – обчислюють у вигляді коефіцієнта:

– Rit (*Item-Test Correlation*) – зв'язок між результатом виконання певного тестового завдання тестової частини учнем-учасником тестування та загальним тестовим балом, отриманим ним за всю тестову частину, **ураховуючи тестовий бал** за виконання цього завдання;

– *Rir* (Item-Rest Correlation) – зв'язок між результатом виконання певного тестового завдання тестової частини учнем-учасником тестування та загальним тестовим балом, отриманим ним за всю тестову частину, **не враховуючи тестовий бал** за виконання цього завдання.

Що сильніший зв'язок, то вищим є коефіцієнт кореляції, і тим більшою є розподільна здатність тестового завдання. Значення коефіцієнта кореляції належить інтервалу $[-1;1]$, тобто варіюється від 0 (випадковий зв'язок) до 1 (абсолютний лінійний зв'язок) або -1 (абсолютний негативний лінійний зв'язок).

У таблиці наведено інтервали значень коефіцієнта кореляції та інтерпретація розподільної здатності тестового завдання.

Інтервали значень показника <i>Rit</i>	Інтерпретація розподільної здатності тестового завдання
$> 0,25 - 1$	Висока
$> 0,2 - 0,24$	Прийнятна
$-1 - 0,19$	Низька

Крива розподілу (*Distribution Curve*) – графік частот різних значень змінної в статистичному розподілі. Описує закономірність співвідношення варіант і частот.

Крива розподілу тестових балів (*Distribution Curve of Test Scores*) – графічне відображення частот набраних балів учасниками тестування.

Максимально можливий бал (*Max*) – бал, який можуть набрати учні-учасники моніторингового дослідження, правильно виконавши / надавши повну відповідь на всі завдання певної тестової частини, та обчислений згідно зі Схемами нарахування тестових балів за виконання тестових завдань.

Максимально набраний бал (*max*) – найбільший бал, який фактично набрали учні-учасники тестування, правильно виконавши/надавши повну відповідь на всі завдання певної тестової частини та обчислений згідно зі Схемами нарахування тестових балів за виконання тестових завдань ($max \leq Max$).

Медіана (*Median (Me)*) – значення, яке вказує, що кількість учнів-учасників тестування, які виконували певну тестову частину й отримали більше вказаного бала, дорівнює кількості учнів-учасників, які отримали менше вказаного бала. Медіана розподілу балів – це бал, який розподіляє всіх учасників тестування, які виконували певну тестову частину на дві рівні групи (за кількістю учасників у кожній з них): тих, хто отримав тестовий бал більший, ніж указаний, і тих, хто отримав тестовий бал менший, ніж указаний.

Мода (*Mode (Mo)*) – значення, яке вказує, що найчастіше в учнів-учасників тестування трапляється певний бал, що відповідає найвищому стовпчику гістограми розподілу учнів-учасників тестування, які виконували певну тестову частину, за кількістю набраних тестових балів, тобто її вершині.

Моніторинг в освіті (лат. *Monitor* – той, що нагадує, наглядає, застерігає) – спеціальна система збору, обробки, зберігання й поширення інформації про стан освіти, прогнозування на підставі об'єктивних даних динаміки і основних тенденцій її розвитку та розроблення науково обґрунтованих рекомендацій для прийняття управлінських рішень стосовно підвищення ефективності функціонування освітньої галузі. У сфері освіти, зокрема управління якістю освіти, моніторинг використовується порівняно недавно та формується як інформаційна база системи управління освітою. Для України моніторинг в освіті є новим явищем (*Енциклопедія освіти / Академія пед. наук України; голов. ред. В. Г. Кремень. Київ: Юрінком Інтер, 2008*).

Моніторинг якості освіти – це система послідовних і систематичних заходів, що здійснюються з метою виявлення та відстеження тенденцій у розвитку якості освіти в країні, на окремих територіях, у закладах освіти (інших суб'єктах освітньої діяльності), встановлення відповідності фактичних результатів освітньої діяльності її заявленим цілям, а також оцінювання ступеня, напряду і причин відхилень від цілей. Моніторинг якості освіти може бути внутрішній та зовнішній. Внутрішній моніторинг якості освіти проводиться закладами освіти (іншими суб'єктами освітньої діяльності). Зовнішній моніторинг якості освіти може проводитися будь-якими органами, підприємствами, установами, організаціями, іншими юридичними особами, що здійснюють незалежне оцінювання якості освіти та освітньої діяльності. Участь закладів освіти (інших суб'єктів освітньої діяльності) та учасників освітнього процесу у зовнішньому моніторингу якості освіти є добровільною, крім випадків, встановлених законодавством (*стаття 48 Закону України «Про освіту»*).

Моніторингове дослідження – моніторинг, що проводиться з метою вивчення стану функціонування об'єкта дослідження, процесів, що характеризують його, і ґрунтується на методології наукового дослідження з чітким визначенням мети, завдань, предмета, об'єкта, дослідження, концептуальних засад і гіпотез та ін.

Основний етап моніторингового дослідження – період збору даних моніторингового дослідження, на основі обробки й аналізу яких проводиться визначення результатів й оцінювання стану сформованості читацької й математичної компетентностей випускників початкової школи.

Основний тест – тестова частина, відповідно до якої відбувається вирівнювання всіх інших тестових частин, що використовувалися в поточному циклі дослідження.

Пілотний етап моніторингового дослідження – період збору даних з метою валідації та оптимізації матеріалів і процедур моніторингового дослідження. Після проведення пілотного етапу проводяться необхідні адаптації й зміни для забезпечення якісних матеріалів і процедур для основного етапу моніторингового дослідження.

Поріг математичної/читацької компетентності – чисельний вимір оволодіння випускником початкової школи набором знань, умінь, ставлень та ін. у межах відповідної предметної галузі. Визначається за результатом виконання учасниками тестування тестових завдань тестових частин через установа експертами відповідної кількості балів на шкалі тестових балів основного тесту, що відповідає мінімальним/значним досягненням учасника у визначених галузях, шляхом переведення тестових балів у шкальовані (шкала 100–300). У межах моніторингового дослідження визначено два основні пороги – *базовий* і *високий* пороги математичної/читацької компетентності, які відповідають балам 170 і 230 на шкалі 100–300 відповідно. Також визначено середній поріг (200 балів на шкалі 100–300), однак його значення не можна співвідносити із будь-яким середнім значенням когнітивних здобутків учасників. Середнє значення відповідає значенню нуля логітів (значення, яке відповідає ймовірності 50 % розв'язати тестове завдання зі складністю нуля логітів). У першому циклі опис цієї характеристики немає легкопояснюваного змісту, однак у наступних циклах моніторингового дослідження зміщення результатів від цієї точки демонструватиме динаміку змін результатів.

Програма моніторингового дослідження – Програма загальнодержавного моніторингового дослідження якості початкової освіти «Стан сформованості читацької та математичної компетентностей випускників початкової школи загальноосвітніх навчальних закладів», затверджена наказом МОНУ від 20.03.2018 р. № 256.

Процентиль (Percentile) – указує на відсоткову частку учнів, які набрали балів менше за відповідне процентильне значення. Процентиль указує на 1/100 частку розподілу, упорядкованого за значеннями. Процентиль – це показник того, який відсоток значень знаходиться нижче певного рівня. Наприклад, значення 50-го процентиля вказує, що 50 % значень розташовані нижче цього рівня.

Рівень математичної / читацької компетентності - чисельний вимір оволодіння випускником початкової школи набором знань, умінь, ставлень та ін. у межах предметної галузі, що визначається на шкалі 100–300 за результатом виконання учасникам тестування тестових завдань тестових частин. У межах дослідження визначено п'ять основних рівнів досягнень учасників, а саме:

передбазовий рівень – результати від 100 до 170 балів на шкалі 100–300, на цьому рівні знаходяться всі учасники, які не досягли базового порогу;

базовий рівень – результати від 170 до 300 балів на шкалі 100–300, на цьому рівні знаходяться всі учасники, які досягли базового порогу;

середній рівень – результати від 200 до 300 балів на шкалі 100–300, на цьому рівні знаходяться всі учасники, які досягли середнього порогу;

високий рівень – результати від 230 до 300 балів на шкалі 100–300, на цьому рівні знаходяться всі учасники, які досягли високого порогу.

Розподільна здатність (дискримінативність) тестового завдання (*D-index*) – здатність тестового завдання відділяти учнів-учасників тестування з різним рівнем сформованості компетентності. Дискримінативність тестового завдання визначають як різницю складності тестового завдання для сильної та слабкої (добре й погано підготовленої) груп учнів-учасників тестування. У таблиці наведено інтервали значень розподільної здатності та характеристики розподільної здатності (дискримінативності) тестового завдання.

Інтервал значення <i>D-index</i>		Характеристика розподільної здатності (дискримінативності) тестового завдання
на шкалі 0 –100 %	на шкалі 0–1	
41–100	0,41–1,0	Дуже хороша
31–40	0,31–0,40	Хороша
21–30	0,21–0,30	Середня
≤ 20	≤ 0,20	Низька

Розподільна здатність тестової частини – середнє арифметичне значень показників розподільної здатності (*D-index*) усіх тестових завдань відповідної тестової частини. У таблиці наведено інтервали значень розподільної здатності та характеристики тестової частини.

Інтервал значення <i>D-index</i>		Характеристика тестової частини
на шкалі 0 –100 %	на шкалі 0–1	
41–100	0,41–1,0	З дуже хорошою розподільною здатністю
31–40	0,31–0,40	З хорошою розподільною здатністю
21–30	0,21–0,30	Із середньою розподільною здатністю
≤ 20	≤ 0,20	З низькою розподільною здатністю

Середнє (Mean (M)) – значення, яке вказує, що середній набраний бал для всіх учнів-учасників тестування, які виконували певну тестову частину, становить певний бал – M .

Складність тестового завдання (P -value) – показник успішності виконання цього тестового завдання учнями-учасниками тестування. Визначають як відношення (у відсотках (шкала 0–100 %) або в шкалі 0–1,0) кількості балів, набраних усіма учнями-учасниками за виконання цього тестового завдання, до максимальної кількості тестових балів, яку вони могли б отримати за його виконання. У таблиці наведено інтервали значень складності та характеристику тестового завдання.

Інтервал значення P -value		Характеристика тестового завдання
на шкалі 0 –100 %	на шкалі 0–1	
> 80	> 0,80	Дуже легке
60–79	0,60–0,79	Легке
40–59	0,40–0,59	Оптимальне
21–39	0,21–0,39	Складне
≤ 20	$\leq 0,20$	Дуже складне

Складність тестової частини – середнє арифметичне значень показників складності (P -value) усіх завдань тестової частини. У таблиці наведено інтервали значень складності та характеристику тестової частини.

Інтервал значення P -value		Характеристика тестової частини
на шкалі 0 –100 %	на шкалі 0–1	
> 80	> 0,80	Дуже легка
60 – 79	0,60–0,79	Легка
40 – 59	0,40–0,59	Оптимальна
21 – 39	0,21–0,39	Складна
≤ 20	$\leq 0,20$	Дуже складна

Стандартна похибка вимірювання (Standard Error of Measurement (SEM)) – статистичний показник, що відображає ступінь точності вимірювання та виражається в балах і залежить від надійності тесту та стандартного відхилення. Задовільною є якість тесту, у якого величина $SEM < 1/2$ від величини $St.Dev.$

На основі стандартної похибки вимірювання будують довірчий інтервал (*Confidence Interval*):

– для близько 68 % упевненості, що оцінка учасника тестування знаходиться всередині цього інтервалу, він становить $\pm 1 \cdot SEM$;

– для близько 95 % упевненості, що оцінка учасника тестування знаходиться всередині цього інтервалу, він становить $\pm 2 \cdot SEM$.

Що вужчим є діапазон SEM , то більшою є впевненість, що тестовий бал представляє реальний стан успішності окремого учасника тестування в галузі, стосовно якої проводиться тестування.

Стандартне відхилення (Standard Deviation ($St.Dev.$)) – значення, яке вказує, що в середньому величина відхилення тестового бала, отриманого кожним учнем-учасником тестування, який виконував певну тестову частину, від середнього (M) становить певний

тестовий бал. Величина *St.Dev.* показує те, наскільки широко тест розподіляє учнів-учасників тестування за набраними ними тестовими балами. Мале значення *St.Dev.* свідчить, що більшість учнів-учасників тестування отримують практично однакову кількість тестових балів. Таке трапляється або коли тестування проходить група учасників, що дійсно має дуже близький один до одного рівень підготовки, або коли тест не є якісним і має низьку розподільну здатність. Що більша величина *St.Dev.*, то краще тест розподіляє учасників тестування. Для оцінки якості тесту значення *St.Dev.* порівнюють зі значенням *M*, а також зі стандартною похибкою вимірювання. Тест вважають таким, що має достатню розподільну здатність, якщо значення *St.Dev.* становить $>15\%$ від значення *M*.

Тест – у межах моніторингового дослідження назва для позначення сукупності тестових завдань, уміщених у двох окремих тестових зошитах – у Частині 1 та Частині 2.

Тестова частина – у межах моніторингового дослідження, зокрема у звітних документах, назва кожної з двох структурних одиниць тесту – Частини 1 та Частини 2.

Тестове завдання – елемент тесту з чітко визначеними психометричними характеристиками, що орієнтований на вимірювання певної властивості учасника моніторингового дослідження.

Тестове завдання на вибір відповіді – тип тестових завдань, що передбачають вибір учасником тестування одного або кількох варіантів відповіді з-поміж запропонованих. У межах моніторингового дослідження використано тестові завдання на вибір однієї правильної відповіді із чотирьох запропонованих варіантів відповіді. Іншими назвами цього типу тестових завдань є «закриті тестові завдання», «тестові завдання множинного вибору» тощо

Тестове завдання на надання відповіді – тип тестових завдань, що передбачають самостійне створення учасником тестування відповіді. Видами цього типу тестових завдань, які використані в межах моніторингового дослідження, є тестові завдання на надання короткої відповіді, на надання розгорнутої відповіді, на побудову. Іншими назвами цього типу тестових завдань є «відкриті тестові завдання», «тестові завдання відкритої форми» тощо.

Тестовий зошит – у межах моніторингового дослідження назва друкованого буклету (брошури) з чорно-білим друком з уміщеними в ньому (ній) всіма тестовими завданнями однієї тестової частини (як структурної одиниці тесту), які має виконати учень-учасник моніторингового дослідження, позначивши у визначених умовами тестових завдань місцях тестового зошита відповіді на відповідні тестові завдання.

Учасники моніторингового дослідження – учні 4-х класів (випускники початкової школи) та вчителі, які їх навчали.

Характеристична крива тестового завдання (*Item Characteristic Curve*) – графічне відображення функціональної залежності ймовірності правильної відповіді учасником тестування на тестове завдання залежно від латентної характеристик, яка лежить в основі виконання тестових завдань тестів.

Шкала 100–300 – стандартизована шкала із середнім значенням 200 балів і середнім відхиленням 30 балів, що визначена для звітування про стан сформованості читацької/математичної компетентності випускників початкової школи в межах моніторингового дослідження.

Шкала логітів – метрична шкала, яка відображає складність тестових завдань і латентних характеристик (індивідуальні здібності, успішність, рівні підготовленості) учасників тестування в єдиній метриці. Значення в шкалі логітів латентних характеристик учасника тестування відображає ймовірність скласти тест нульової складності цим

учасником. Значення в шкалі логітів складності тестового завдання відображає ймовірність вирішити це тестове завдання учасників з нульовим значенням латентної характеристики. Параметри тестових завдань і латентні характеристики учасників тестування можуть змінюватися в межах $(-\infty, +\infty)$, але фактично їхні значення приблизно змінюється в межах $(-4, +4)$.

Якість освіти – відповідність результатів навчання вимогам, установленим законодавством, відповідним стандартом освіти та/або договором про надання освітніх послуг (*стаття 1 Закону України «Про освіту»*).