

Основними показниками якості тестів є складність тесту, його розподільна здатність, надійність тесту тощо.

Пропонуємо ознайомитися зі статистичними та психометричними характеристиками на прикладі тесту ЗНО з української мови і літератури 2017 року.

Назва характеристики	Кількісне значення
Кількість учасників	229 958
Максимально можливий бал	104
Максимально набраний бал	104*
Середнє	53,41
Стандартне відхилення	21,86
Асиметрія	0,19
Ексцес	2,05
Середня складність тесту	52,30
Середня розподільна здатність тесту	50,22
Надійність тесту	0,94

Кількість учасників – фактична кількість осіб, які взяли участь у зовнішньому незалежному оцінюванні з цього навчального предмета.

Максимально можливий бал – бал, який могли набрати учасники тестування з цього навчального предмета правильно виконавши всі завдання тесту.

Максимально набраний бал – бал, який фактично набрали учасники тестування, та обчислений відповідно до Схем нарахування балів за виконання завдань різних форм, уключених до тесту з цього навчального предмета.

* – отримало **п'ять** учасників.

Середнє (Mean (M)) – означає, що середній набраний бал по всіх учасниках тестування з цього навчального предмета становить 53,41.

Стандартне відхилення (Standard Deviation ($St.Dev.$)) – означає, що в середньому величина відхилення бала, отриманого кожним учасником тестування з цього навчального предмета, від середнього (M) становить 21,86 бала. Величина $St.Dev.$ показує те, наскільки широко тест розподіляє учасників тестування за набраним ними балами. Мале значення $St.Dev.$ свідчить, що більшість учасників тестування отримують практично однакову кількість балів. Таке трапляється або коли тестування складає група учасників, що дійсно мають дуже близький один до одного рівень підготовки, або коли тест не є якісним і має низьку розподільну здатність. Чим більша величина $St.Dev.$, тим краще тест розрізняє учасників тестування. Для оцінки якості тесту значення $St.Dev.$ порівнюють зі значенням M , а також зі стандартною похибкою вимірювання. Тест вважається таким, що має недостатню розподільну здатність, якщо значення $St.Dev.$ становить <15 % від значення M .

Асиметрія (Skewness (A_s)) – характеристика «перекосу» чи несиметричності «крил» розподілу балів, набраних учасниками тестування з цього навчального предмета. $A_s > 0$ свідчить що асиметрія є позитивною або лівосторонньою, тобто в розподілі найчастіше зустрічаються значення менше M . $A_s < 0$ свідчить що асиметрія є від'ємною або правосторонньою, тобто в розподілі найчастіше зустрічаються значення більше M . Помітним відхиленням значення A_s від 0 вважається $\pm 0,1$ і більше.

Ексцес (Kurtosis (K)) – характеристика «гостроверхості» графіка розподілу балів, набраних учасниками тестування з цього навчального предмета, в порівнянні з графіком «нормального розподілу» (розподілу Гаусса), що має такі ж значення M і $St.Dev.$ Якщо значення $K > 0$, то ексцес додатній, а крива розподілу має вищу та «гострішу» вершину ніж відповідна крива «нормального (Гауссовського) розподілу». Якщо значення $K < 0$, то ексцес

від'ємний, а крива розподілу має нижчу та «плоскішу» вершину ніж крива «нормального (Гауссовського) розподілу».

Середня складність тесту – середнє арифметичне значень показників складності (*P-value*) усіх завдань тесту з цього навчального предмета.

Складність тестового завдання (*P-value*) – показник успішності виконання цього завдання учасниками тестування. Визначається як відношення (у відсотках) кількості балів, набраних усіма учасниками за виконання цього завдання, до максимальної кількості балів, яку вони могли б отримати за його виконання.

У таблиці наведено інтервали значень складності тестового завдання та характеристику тестового завдання.

Інтервал значення	Характеристика завдання
понад 80%	дуже легке
60–79%	легке
40–59%	оптимальне
20–39%	складне
менше 20%	дуже складне

Середня розподільна здатність тесту – середнє арифметичне значень показників розподільної здатності (*D-index*) усіх завдань тесту з цього навчального предмета.

Розподільна здатність (дискримінативність) тестового завдання (*D-index*) – здатність тестового завдання відділяти учасників тестування з різним рівнем навчальних досягнень. Дискримінативність завдання визначається як різниця складності завдання для сильної та слабкої (добре і погано підготовленої) груп учасників тестування.

У таблиці наведено інтервали значень розподільної здатності та характеристику дискримінативності завдання.

Інтервал значення	Характеристика дискримінативності завдання
41 – 100%	дуже хороша
31 – 40%	хороша
21 – 30%	середня
20% і менше	низька

Надійність тесту – показник, як правило, внутрішньої узгодженості завдань тесту з цього навчального предмета. Внутрішня узгодженість відображає, наскільки похибки, що вносяться в процес вимірювання окремими завданнями компенсують одна одну, чи навпаки, підсилюють одна одну.

Загальноживаним показником надійності тесту є коефіцієнт Альфа Кронбаха (*Cronbach's alpha*). Референсними є значення Альфи $\geq 0,75$. Тестування, у яких Альфа Кронбаха виявилася більшою від 0,75, вважаються надійними. Чим більшим є значення Альфи, тим надійнішим, стійкішим є результат тестування стосовно впливу сторонніх випадкових факторів, які не є об'єктами вимірювання, і тим меншою є похибка вимірювання.

У таблиці наведено інтервали значень коефіцієнта Альфа Кронбаха та характеристика надійності тесту як інструменту вимірювання.

Інтервал значення	Характеристика надійності тесту
0,90 – 0,99	відмінна
0,85 – 0,89	дуже хороша
0,80 – 0,84	хороша
0,75 – 0,79	задовільна
менше 0,75	малозадовільна