

Побудова стратифікованих еталонних кривих Lichess для Elo+Chess

Elo+Chess Нотатки дослідження

31 травня 2026 р

Abstract

Elo+Chess benchmark reports are built from large stratified samples of Lichess.org games, with the benchmark sample used in this note containing 483,974 1+0 bullet games, 479,974 3+0 blitz games, 925,636 10+0 rapid games, and 660,540 longer rapid games. For each major game type, the pipeline expands every game into before-move and after-move board states, calculates hundreds of intermediate and player-game variables, aggregates those variables into coaching metrics, and then averages those metrics by Lichess Elo bucket. The sample is deliberately stratified rather than proportional to the raw player pool: the purpose is to estimate stable behavior curves at each rating level, including buckets that would otherwise be sparse, not to reproduce the natural rating distribution. The result is a family of empirical benchmark curves: for each metric, we can ask how often players at different Elo levels show a given habit or principle in real games. This note describes the method, the scale of the computation, and several opening-principle examples involving castling, minor-piece development, repeated opening piece moves, and early queen movement. The emphasis is methodological: Elo+Chess does not assume that chess principles are universally linear across all skill levels. Instead, benchmark curves are inspected by rating range, and the product focuses on the beginner-to-early-advanced region where the relationships are strongest and most interpretable.

1 призначення

Метою системи порівняння є перетворення широких шахових порад у вимірювані величини. Таке твердження, як «розвивайте свої другорядні частини». корисно, але це стає більш ефективним, коли ми можемо сказати, скільки другорядних фрагментів а гравець зазвичай розвивається на 10 хід, як це порівнюється з однолітками в тому самому рейтинговий сегмент і як ця звичка змінюється в діапазоні рейтингів.

Таким чином, Elo+Chess створює бібліотеку показників із реальних ігор. Кожен показник є обчислюється для кожного рядка гри гравця, узагальнюється за типом гри та сегментом Elo, а також використовується як еталонна крива у звіті. Після цього звіт може розмістити користувача ту саму криву та показує, чи знаходиться користувач вище, нижче чи поблизу поведінки спостерігається серед порівнянних гравців.

2 Вихідні дані та стратифікація

Джерелом тесту є публічна база даних ігор Lichess.org. Elo+Chess використовує великі стратифіковані вибірки за типом гри та рейтинговим сегментом, а не маленькі зручний зразок. Артефакти поточного робочого середовища охоплюють чотири основні категорії гри:

- 1+0 куля,
- 3+0 бліц,
- 10+0 рапід,
- швидкі ігри тривалістю більше 10 хвилин.

Метою стратифікації є охоплення рейтингових рівнів. Пропорційний випадковий зразок був би ефективним для опису загального пулу гравців, але він буде бути неефективним для побудови еталонних кривих. Найпоширеніші рейтингові бакети домінували б у зразку, тоді як хвостові ковші з низькою опорою мали б занадто мало ігри для стабільних середніх показників. Elo+Chess тому зразки середньої гри Ело відро. У поточній структурі чистого зразка першими є необроблені ігри відфільтровано до дійсних рейтингових ігор із відомими гравцями, відомими рейтингами, дійсними результатами, і відомі засоби керування часом. Потім ігри фільтруються до гравців із достатньою кількістю ігор з такою швидкістю та досить стабільними рейтингами, щоб відобразити узгоджений рівень навичок: для кожного гравця Elo+Chess обчислює рейтинг гравця в межах швидкості ($\max(\text{Elo}) - \min(\text{Elo})$) через вихідний пул і зберігає ігри лише тоді, коли обидва гравці мають принаймні налаштовану мінімальну кількість ігор і не більше 300 балів Ело внутрішньошвидкісного дрейфу. Дрейфовий фільтр не має на увазі для усунення нормального поліпшення або відхилення; це зменшує випадки, коли гравець ігри змішували б суттєво різні рівні навичок, такі як тимчасові облікові записи, гравці, що повернулися, або облікові записи, які швидко переміщувалися через кілька сегментів протягом вибіркового періоду. Нарешті, відповідні ігри ранжуються в межах кожного середнього Ело відро за допомогою детермінованого хешу ідентифікатора гри та початкового числа, а також обмеження кількості ігор зберігається з кожного відра.

Ця конструкція робить еталонні криві набагато кориснішими. Кожне відро Ело вносить достатньо спостережень, щоб оцінити середнє значення кожного показника, а рідкісні ковші з високим або низьким рейтингом не затьмарюються природною формою розподілу активних гравців. Еталонні криві, які використовує Elo+Chess, є потім обмежується діапазоном, який найбільше відповідає продукту, приблизно для початківців через ранню просунуту гру. У звітах за шкалою Lichess цей первинний діапазон становить 600–1600. Коли користувач переглядає звіт за шкалою Chess.com, відповідні рейтингові сегменти відображаються на той самий базовий Lichess порівняльні сегменти за допомогою окремого міжплатформного методу відображення рейтингів описано в супровідному документі: [Estimating Cross-Platform Chess Rating Mappings with Модальна регресія](#).

3 Розподіл рейтингів активних гравців

Якщо ми подивимося на розподіл рейтингів Elo гравця Lichess.org станом на кінець березня 2026 року за основними типами ігор ми бачимо такі підрахунки та проценти. У цьому перегляді розподілу кожен гравець зараховується один раз для конкретного типу гри, використовуючи останній спостережуваний рейтинг цього гравця у відповідній сировині рядки з контролем часу, після того, як потрібно щонайменше п'ять ігор у цій грі типу.¹

¹We apply a five-game minimum because lightly active accounts disproportionately remain at the 1500 starting/provisional anchor, creating an artificial spike in all-player latest-rating distributions. Figures 1–4 показати дистрибутиви з цим фільтром і без нього.

Game type	Players	Median	75th pct.	80th pct.	% ≤ 1600	Exact 1500	1500 share
1+0 bullet	1,221,146	1,479	1,858	1,951	58.9%	4,756	0.4%
3+0 blitz	1,144,045	1,542	1,845	1,912	54.9%	2,315	0.2%
10+0 rapid	1,679,401	1,381	1,691	1,765	68.3%	3,472	0.2%
>10 rapid	668,255	1,400	1,686	1,753	68.2%	7,969	1.2%

Table 1: Останній спостережуваний точний розподіл рейтингу активних гравців Lichess типу гри після того, як буде потрібно щонайменше п'ять ігор саме цього типу.

Цей розподіл не слід плутати зі стратифікованою контрольною вибіркою. Розподіл описує пул активних гравців Lichess, який спостерігається в необробленому вигляді сканування. Потім еталонний зразок навмисно стратифікується, щоб отримати Elo мають достатню підтримку для стабільних метричних кривих.

Щоб відобразити рейтинг Lichess.org Elo для певного типу гри на еквівалентний рейтинг Chess.com, Elo+Chess використовує модальний тип гри рівняння регресії, оцінені в компаньоні [Elo+Chess rating-mapping paper](#):

$$\begin{aligned}
 \hat{R}_{\text{Chess.com,bullet}} &= -531.71 + 0.9871R_{\text{Lichess}}, \\
 \hat{R}_{\text{Chess.com,blitz}} &= -551.54 + 1.0853R_{\text{Lichess}}, \\
 \hat{R}_{\text{Chess.com,10minrapid}} &= -495.04 + 1.0741R_{\text{Lichess}}, \\
 \hat{R}_{\text{Chess.com,>10rapid}} &= -369.02 + 0.9374R_{\text{Lichess}}.
 \end{aligned}$$

Ці рівняння відображають рейтингові значення, а не процентильні ранги. Процентильні ранги є визначені всередині конкретного пулу гравців, і гравці Chess.com і Lichess басейни мають різну форму. У таблиці 2, для 10-хвилинний швидкий тип гри, перший стовпець містить невелику вибірку можливих Chess.com Значення рейтингу Elo. Друга колонка дає відповідні Процентиль рейтингу Chess.com, що відображається на Chess.com станом на 31 травня 2026 р. А 71,7

рейтинг на Lichess.org для 10-хвилинного рапиду відповідає рейтингу Elo Lichess 1,644², і це значення показано в третій стовпець. Якщо застосувати рівняння відображення Chess.com–Lichess, 1644 на Lichess еквівалентний приблизно 1271 на Chess.com. Припускаючи оціночне рівняння відображення розумне, розрив +522 між рейтингом Chess.com у стовпець 1 і відображений рейтинг у стовпці 4 можна пояснити великою кількістю випадкових гравців у пулі гравців Chess.com, що надуває Chess.com процентний ранг. Іншими словами, перебування в 72-му процентилі на Lichess.org є більшим досягненням, ніж перебування в 72-му процентилі на Chess.com, тому що Процентильні ранги не відповідають чітко від однієї системи до іншої. Справжнє Ело значення можна зіставити за допомогою рівнянь, оцінених на основі результатів одного гравця дві системи.

Причина прогалин у таблиці 2 полягає в тому, що відображення оцінює еквівалентні рейтинги ігрової сили, тоді як процентиль описує позиція всередині населення однієї платформи. Гравець із 71,7-процентилем Таким чином, Chess.com не є автоматично 71,7-процентильним гравцем на Lichess, і навпаки. Спостережувані цифри відповідають публічним Chess.com швидкий пул, що має набагато більшу масу з нижчим рейтингом або випадкових гравців, тоді як пул активного точного контролю Lichess більш сконцентрований серед досвідчених повторів

²Calculated from the cumulative distribution function of the ZXQPROT3ZXQ 10+0 rapid latest-rating sample after requiring at least five games in that exact game type; see Table 1 and Figure 3.

Chess.com rating	Chess.com pct.	Same-pct. Lichess	Mapped equiv.	Gap
749	71.7%	1,644	1,271	+522
1,012	86.2%	1,871	1,515	+503
1,358	95.4%	2,095	1,755	+397
2,383	99.9%	2,557	2,251	-132

Table 2: Ілюстративне порівняння між Chess.com 10-хвилинним швидким процентилям і ті самі процентильні пункти в активному Lichess 10+0 rapid розповсюдження після принаймні п'яти ігор 10+0. “Відображено Chess.com екв.” застосовує поточний Elo+Chess Швидке перетворення рейтингової шкали 10+0 у той самий процентиль рейтингу Lichess. The рядки найвищого хвоста включені, щоб показати різницю в масштабі, але підігнані Рейтингова лінія найбільш надійна в основному тренерському діапазоні.

гравців. Ключовим моментом є не те, що одна процентильна шкала є «правильною», а інший є «неправильним»; він полягає в тому, що процентний ранг не переноситься між платформами якщо базові розподіли гравців не мають однакової форми.

Note on the 1500 spike. The visible jump near 1500 присутній у даних, а не створений шляхом згладжування. В останньому спостерігається розподіл рейтингів, точні 1500 рейтингів становлять 13,1

гравців у кульку, 5,6

17,6

початковий або тимчасовий якір у рейтинговій системі.

Точний сплеск у 1500 є здебільшого артефактом малоактивних облікових записів. до перевірте це, ми повторно обчислили ті самі розподіли останнього рейтингу після запиту кожен гравець повинен мати принаймні п'ять ігор певного типу. Шип майже зникає під цим фільтром: точні 1500 оцінок падають з 5,6–17,6

гравців у розподілі всіх гравців до 0,2–1,2

розподіл наведено в таблиці 1.

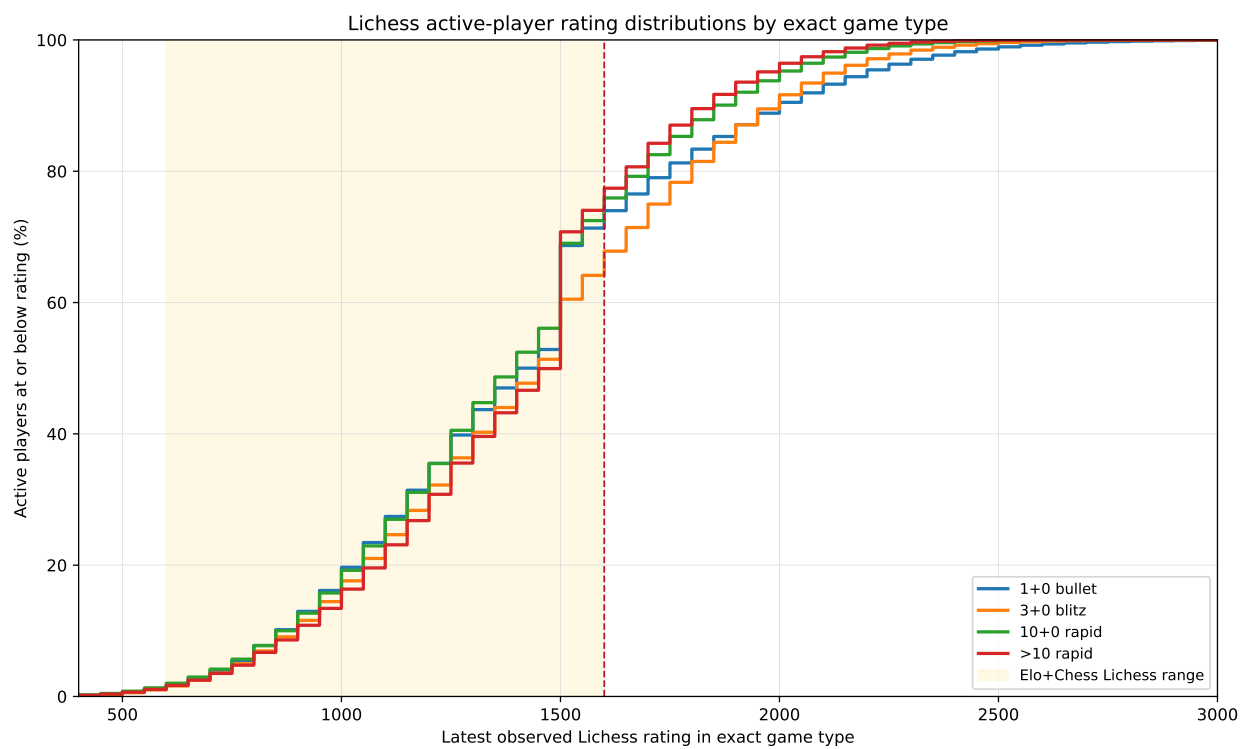


Figure 1: Розподіл сукупного рейтингу активних гравців за типом гри. The жовта смуга позначає діапазон 600–1600 Lichess, який використовується основним Elo+Chess тренінгові орієнтири.

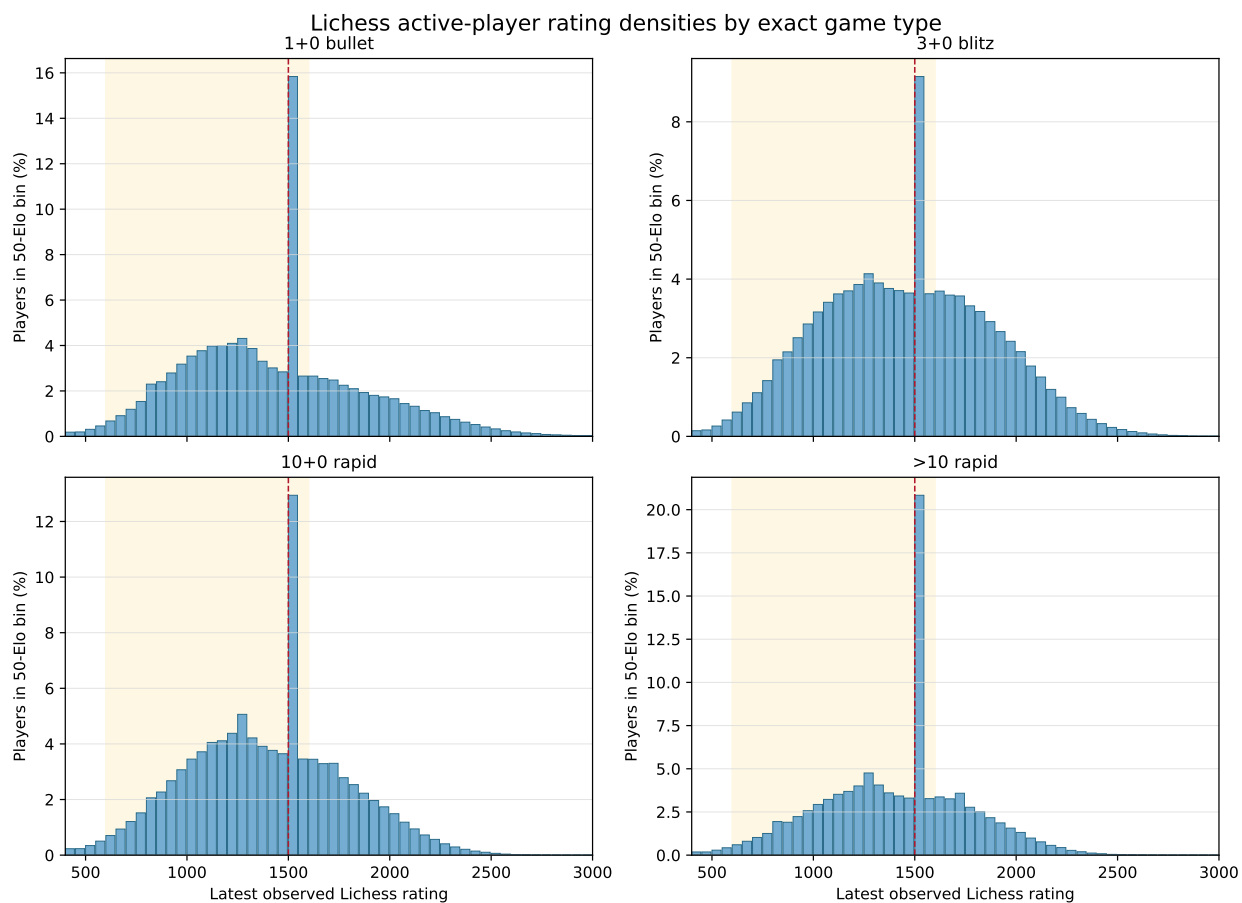


Figure 2: Перегляд щільності тих самих розподілів рейтингів активних гравців. The пунктирна лінія позначає 1500 Lichess, де необроблені останні рейтингові розподіли мають велика точка точної номінальної маси. Жовта смуга позначає 600–1600 Lichess діапазон, який використовується основними звітами про показники Elo+Chess. Шип найбільше видно в щільність усіх гравців, оскільки багато малоактивних облікових записів залишаються рівно на рівні 1500; Рисунок 4 повторює графік щільності після того, як потрібно принаймні п'ять ігор у точному типі гри, що значною мірою видаляє цей артефакт.

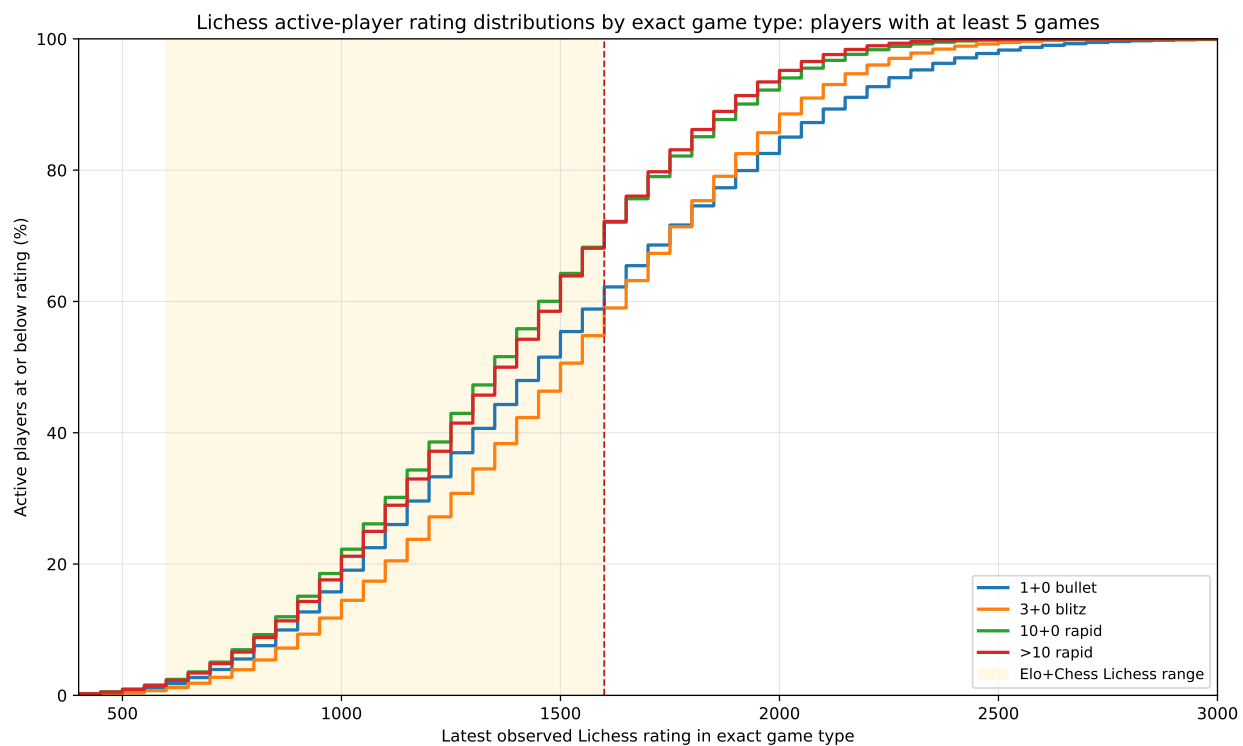


Figure 3: Розподіл сукупного рейтингу активних гравців після вимоги ат принаймні п'ять ігор у точному типі гри. Стрибок 1500 значно зменшений.

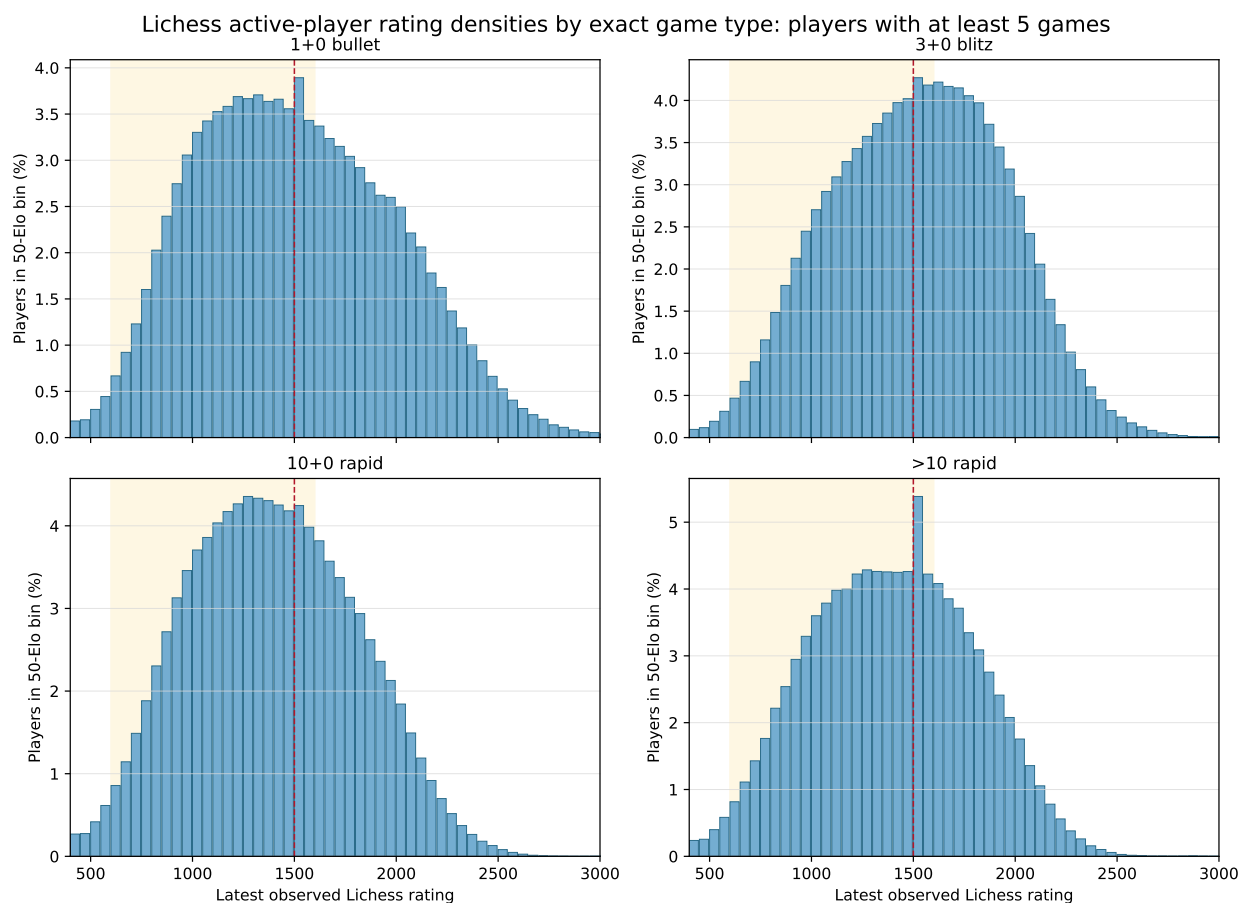


Figure 4: Перегляд щільності після того, як потрібно принаймні п'ять ігор у точній грі типу. Ця точка зору підтверджує, що велика точка маси 1500 у всіх гравців Розповсюдження здебільшого відбувається через малоактивні облікові записи. Жовта смуга позначає діапазон 600–1600 Lichess, який використовується в основних звітах метрики Elo+Chess.

4 Масштаб артефактів поточного тесту

Обчислення навмисно велике. Це не профільний шарп, маленький набір завантажених PGN або підібрана вручну колекція повчальних прикладів. Початковий абонемент Lichess охоплював усі щомісячні файли з січня 2025 року до березня 2026. Таблиця нестратифікованого сканування містить 1 382 178 087 ігор і 2 764 356 174 рейтингових рядків гри. У цих рядках 6 623 438 унікальних імена користувачів гравців з'являються принаймні один раз.

Speed	Games in raw scan	Player-game rows	Unique players
Bullet	519,375,423	1,038,750,846	3,077,635
Blitz	839,823,863	1,679,647,726	5,987,634
Rapid	22,978,801	45,957,602	1,535,865
All scanned speeds	1,382,178,087	2,764,356,174	6,623,438

Table 3: Нестратифікована шкала сканування Lichess перед контрольною вибіркою. Сканування охоплює повні місячні файли історії гри з січня 2025 року по березень 2026 рік.

Артефакти контрольного тесту виробництва значно менші, ніж необроблене сканування, тому що їх вибирають для метричної побудови. Вони все ще великі: еталон стадію є розширення стану переміщення та конвеєр агрегації на сотнях тисячі ігор на тип гри.

Game type	Games	Player-game rows	Unique players	Move states	Feature columns
1+0 bullet	483,974	967,948	119,732	22,985,272	277
3+0 blitz	479,974	959,948	129,398	28,138,904	277
10+0 rapid	925,636	1,851,272	229,235	57,554,339	277
>10 rapid	660,540	1,321,080	133,661	38,580,214	277

Table 4: Шкала стратифікованих еталонних артефактів після вирівнювання загальна схема виконання з 277 стовпцями. А "Ряд гравця-ігри" означає одну гру з точки зору одного гравця, тому більшість ігор створити два рядки гравця-ігри.

Лише для близу 3+0 поточний контрольний тест використовує 479 974 гри за участю 129 398 унікальних гравців і 28 138 904 різних рядків стану руху. The вирівняна таблиця характеристик бліц-гра містить 277 похідних змінних. The артефакти швидкого створення функцій ще ширші на етапі стану переміщення: повний Таблиця станів швидкого переміщення, яка використовується під час створення функції, містить 168 стовпці стану переміщення перед агрегацією в функцію гравця-гри з 277 стовпцями стіл. Порівняльні артефакти зовнішнього середовища для всіх чотирьох типів ігор є потім узгоджено з тією самою схемою функцій гравця з 277 стовпцями, включаючи розширені тактико-таксономічні змінні прориву. З цих таблиць характеристик, звіт вибирає та відображає коучингові показники, які найкраще інтерпретуються.

Ці підрахунки важливі для інтерпретації результатів. Еталонні криві не є оцінками кількох сотень перевірених вручну ігор. Натомість вони походять з десятків мільйонів оцінюваних штатів і більше чотирьох мільйон рядків ігор для гравців у чотирьох поточних категоріях ігор.

5 Огляд трубопроводу

Еталонний конвеєр має шість основних етапів:

1. Виберіть ігри з рейтингом Lichess у цільовому типі гри та рейтингових класах.
2. Розбирайте кожну гру хід за ходом і зберігайте хід до та хід після правління держ.
3. Обчисліть проміжні змінні стану ходу з дошки, ходу, точки зору гравця та точки зору суперника.
4. Агрегувати змінні стану руху в змінні функцій гравця.
5. Перетворюйте змінні гри гравця на показники звіту з кліком напрям інтерпретації: вище зазвичай краще, нижче звичайно краще, або описовий.
6. Усередніть кожну метрику за сегментом Elo та вручну перевірте результат відносини.

Ця структура важлива, оскільки багато показників звіту не є прямими видимий у заголовку PGN. Вони вимагають перебудови позиції, визначення які фігури перемістилися з яких початкових полів, порівнюючи попередній хід і після переміщення матеріалу та структури, а потім агрегування результату перейти до рівня гравця-гри.

6 Проміжні змінні

Проміжні змінні призначені для опису того, що змінилося на дошці і чому це важливо з точки зору гравця звіту. Приклади:

- номер ходу, номер ходу гравця, сторона руху, колір гравця та гравець колір;
- від квадрата, до квадрата, тип рухомої фігури, тип захопленої фігури та чи був цей хід захопленням, шахом, замком, підвищенням або матом;
- передперевізний і післяперевізний матеріальний баланс;
- чи залишила другорядна фігура своє початкове поле на заданий номер ходу;
- чи рухалася та сама фігура під час послідовних ходів гравця в відкриття;
- чи ходив ферзь перед ходом 10;
- сторона рокіровки, номер ходу рокіровки та чи з'єдналися тури після рокіровки;
- вільні фігури, фігури, що висять, здвоєні пішаки, окремі пішаки, пас пішаки, і задні пішаки;
- статус файлу тури, відкриті файли, напіввідкриті файли та безпека короля особливості;
- тактичні можливості, виконана тактика, тактика матеріального тиску, гарантовані подальші прибутки та реалізовані прибутки.

Точний проміжний набір залежить від етапу будівництва. Зберігаються бази даних звітів про виконання лише стовпці, необхідні для швидкого обслуговування звітів, у той час як повна збірка функцій артефакти включають ширші таблиці стану руху та патчі функцій, які використовуються для побудови Останні рядки для гравців.

7 Приклади метричних принципів відкриття

Принципи відкриття є корисними прикладами, тому що користувачам їх легко зрозуміти зрозумілі та легкі для підключення до обчислень стану руху.

7.1 Minor Pieces Developed by Move 10

Цей показник підраховує, скільки з чотирьох коней і єпископів у гравця залишилося їхні домашні поля до 10-го ходу гравця. Проміжний розрахунок використовує колір рушія, тип рухомої фігури, початкове поле та хід гравця номер. Для білих домашніми полями є b1, g1, c1 і f1. Для Чорних вони є b8, g8, c8 і f8. Для кожної гри гравця Elo+Chess зараховує окремий дім поля серед лицарів і єпископів, які переміщувалися за хід гравця 10.

Це метрика «вище – зазвичай краще». Воно не стверджує, що кожен неповнолітній фігура завжди повинна рухатися за ходом 10, але серед гравців, що розвиваються, емпіричний зв'язок зрозумілий: гравці, які досягають вищих показників Ело, мають тенденцію розвиватися більш дрібні частини раніше.

7.2 Repeated Piece Moves in the Opening

Цей показник підраховує додаткові ранні ходи, витрачені на повторне переміщення тієї самої фігури розробки чогось нового. Проміжний розрахунок використовує попередній поле призначення, поточне початкове поле та номер ходу гравця. Якщо ту саму фігуру, яка була переміщена під час попереднього ходу гравця, і знову переміщується в межах перші 10 ходів гравця, кількість повторних ходів збільшується.

Це показник «нижче – зазвичай краще». Він вимірює втрачений темп відкриття в а простий спосіб. Деякі повторні кроки тактично виправдані, але в масштабі крива ковша відображає загальну тенденцію: сильніші гравці в цілі рівні витрачають менше ранніх ходів, багаторазово переміщаючи ту саму фігуру.

7.3 Queen Touched Before Move 10

Цей показник підраховує ходи ферзя до того, як гравець переміститься на 10. Проміжний показник обчислення використовує тип рухомої фігури та номер ходу гравця. Якщо рухомий фрагмент є ферзем, а число ходів гравця не більше 10, кількість збільшується.

Це також менше – зазвичай краще для цільового діапазону Elo+Chess. Рано Ходи ферзя можуть виграти матеріал або змусити поступитися, але багато гравців розвиваються виведіть ферзя занадто рано, втрачайте темп через атаки на ферзя та затримуйте другорядний розвиток і безпека короля.

7.4 Castling and King Safety

Показники рокировки використовують прапорці рокировки, сторону рокировки, номер ходу рокировки та стан дошки після рокировки. Приклади: чи рокував гравець, чи рокував гравець ходом 10, чи рокував гравець королівським флангом або ферзевий фланг, чи були з'єднані тури після рокировки, і чи були тури були відкриті або напіввідкриті після рокировки.

Ці показники не всі інтерпретуються однаково. Довершення рокировки і рокировка 10-м ходом, як правило, вище, зазвичай, краще в цільовому діапазоні. Рокировка ферзеві сторони або рокировка у відкриті файли може бути більш залежною від контексту. Ось чому

Elo+Chess не просто закріплює моральне правило; він будує відро криві та перевіряє, які стосунки достатньо стабільні для коучинга.

8 Криві ковша та ручна перевірка

Після обчислення показників гравця-ігри Elo+Chess групує їх за Lichess Elo відро та обчислює середні значення для відра. Середнє відро є необробленим емпіричним контрольна точка для цього показника. Потім у звіті використовується лінія тренду призначати оцінки в стилі Elo значенням користувача.

Криві перевіряються вручну перед використанням. Метрика більш корисна коли середні значення групи показують узгоджений зв'язок у всьому діапазоні оцінок обслуговується продуктом. Метрика менш корисна, якщо вона шумна, рівна, висока немонотонні або керовані особливими випадками, які важко пояснити користувачеві.

Цей етап перевірки вручну особливо важливий над верхньою частиною Elo+Chess відсікання. Сильніші гравці знають, коли порушувати принципи відкриття, і роблять це ефективно. Вище діапазону від початківця до раннього просунутого, відносини між простими звичками і рейтингом часто вирівнюється або змінює форму. Продукт тому підкреслює діапазон, де сильні емпіричні зв'язки і педагогічно корисний.

9 Приклад кривих відкриття повного діапазону

На малюнку 5 показано чотири бліц криві принципу відкриття 3+0. Сірі точки показують ширший діапазон рейтингу Lichess. Сині точки і синій тренд лінія показує діапазон 600–1600 Lichess, який використовується основним коучингом Elo+Chess контрольні показники. Пунктирна червона лінія позначає верхню границю.

Така ж картина спостерігається в рапіді 10+0 на повному діапазоні та в довшому рапіді артефакти. Основне рішення моделювання залишається незмінним: використовувати емпірично стабільний діапазон для тренерських претензій і уникайте вдавати, що це проста крива звички залишається однаково інформативним назавжди.

Той самий шаблон відображається числово в таблиці 5. Схили є повідомляється як зміна метричного значення на 100 балів Elo. Ротировка, мінор розвиток, повторювані рухи фігур і ранній рух ферзя – все це чітко видно рух через 600–1600. Вище 1600 схили тут значно менші приклади, що узгоджується з ідеєю, що гравці з вищим рейтингом ламаються основні принципи більш вибірково та успішніше. Другий механізм є насиченість. Для деяких звичок метрика досягає практичної стелі або солодкого пляма: сильні гравці вже вивчили цю звичку, тому для цього немає причин крива продовжувати рухатися різко вгору. Більше також не завжди означає краще після ця точка. Тому дані зрівнюють обидва, оскільки сильніші гравці можуть зламати правил у конкретних позиціях і тому, що багато основних звичок уже досягли їх корисний робочий діапазон.

10 Чому верхня границя має значення

Цільовий користувач Elo+Chess — це гравець від початківця до раннього просунутого рівня. Для цих гравці, основні звички тісно пов'язані з рейтингом: розробляти фігури, уникайте

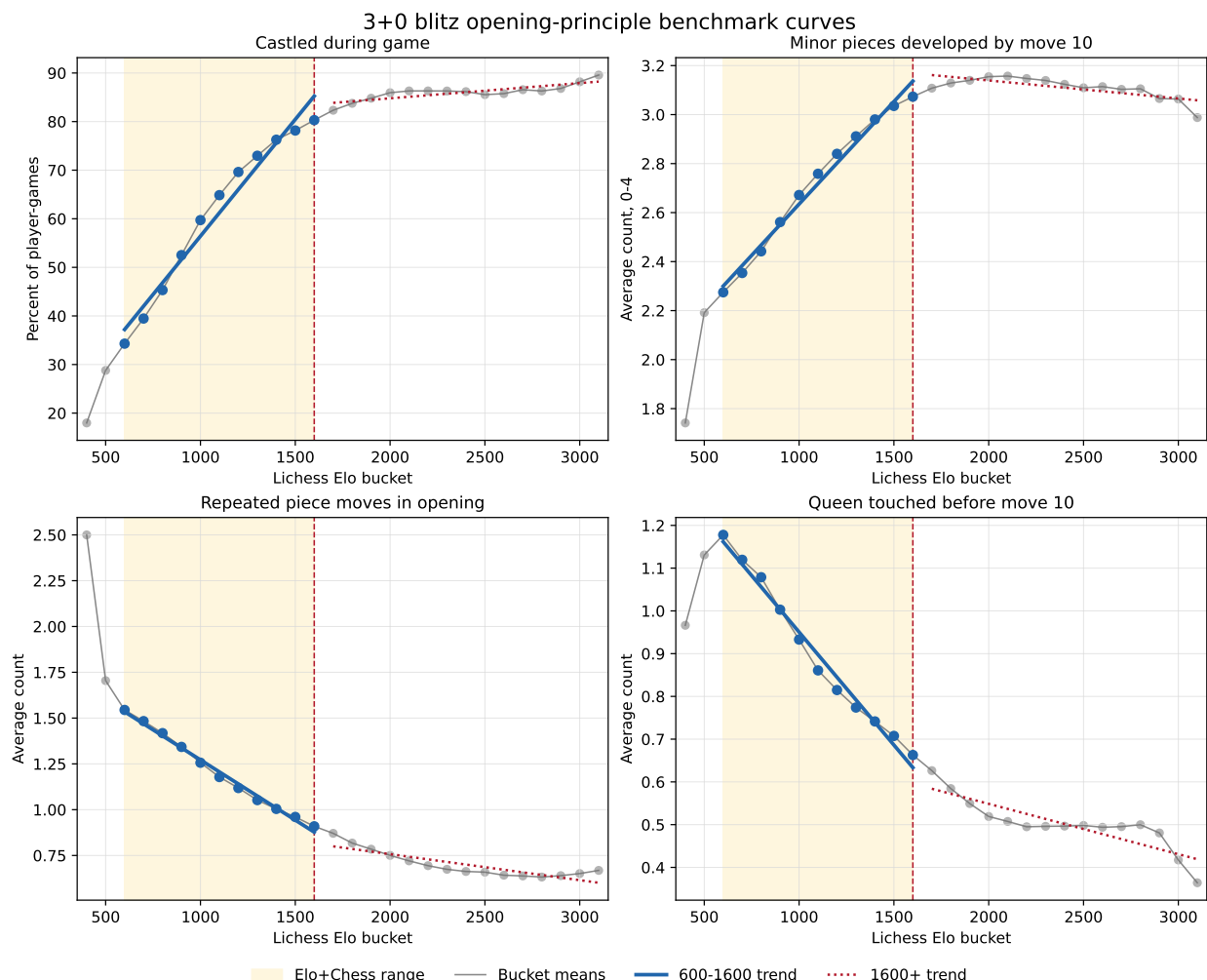


Figure 5: Вибрані базові криві початкового рівня для близу 3+0. The стосунки є найміцнішими та найкориснішими в серії коучингів Elo+Chess. Вище верхньої межі кілька зв'язків суттєво згладжуються.

повторних ранніх ходів фігур, уникайте непотрібних ранніх ходів ферзя, замку вчасно з'єднайте тури, керуйте матеріалом і не залишайте фігури незакріпленими або висить.

При вищих оцінках ті самі зовнішні звички стають менш діагностичними. Сильний гравець може відкласти рокировку, тому що центр закрито, перемістіть ферзя раніше оскільки позиція містить конкретну тактику, або перемістіть ту саму фігуру двічі тому що він виграє в темпі або змушує структурну поступку. Сира звичка все ще має значення, але зв'язок між звичкою та оцінкою стає більшим умовний. Ось чому Elo+Chess використовує верхнє відсічення для свого основного тренування претензій і чому користувачів вище цього діапазону попереджають про тісний зв'язок між принциповими звичками і рейтингом розбивається.

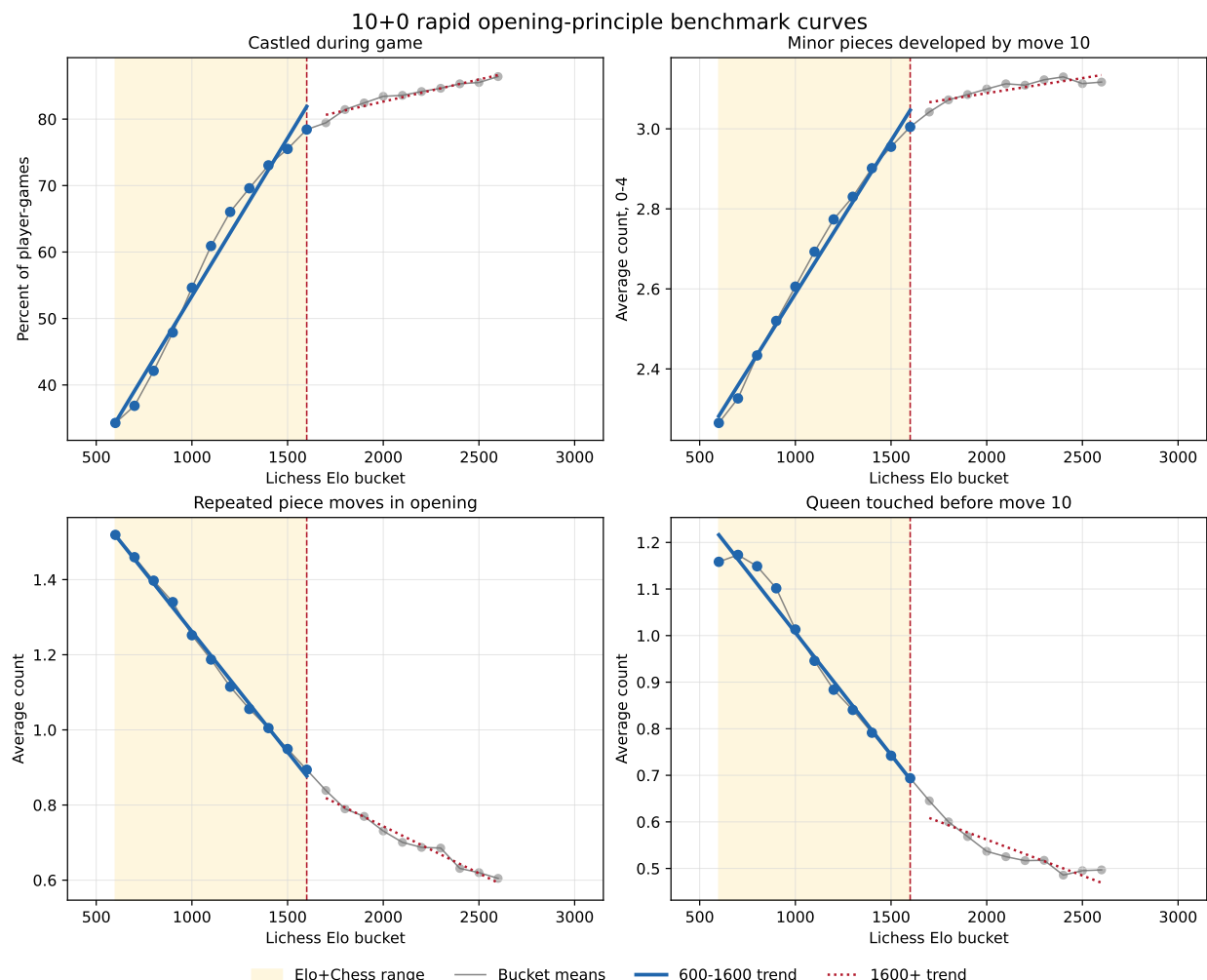


Figure 6: Вибрані криві тесту початкового принципу для стрімких ігор 10+0. The стосунки є найміцнішими та найкориснішими в серії коучингів Elo+Chess. Вище верхньої межі кілька зв'язків суттєво згладжуються.

11 Двосторонні тактичні показники та процентильні цілі

Деякі метрики мають протилежну форму від прикладів початкового принципу. Вилки є корисним прикладом, оскільки подія за своєю суттю є двосторонньою. Гравець, який знаходить і перетворює більш корисні форки, ніж аналоги, робить щось хороше, але рівень популяції перетворених форків падає з рейтингом, оскільки сильніше опоненти дають менше можливостей для форків і захищаються від тактичних загроз більш активно.

Рисунок 8 показує цей ефект. Обидва перероблені вилки гравець звіту та конвертовані вилки суперником стають рідшими, оскільки рейтинги підвищуються. Це не означає, що виконання форків погане. Це означає сиру подію частота частково контролюється частотою помилок супротивника та текстурою гри. Для метрики таким чином, Elo+Chess використовує перцентилі історії однорангових всередині власне відро гравця, а не покладатися лише на глобальну тенденцію середнього відра.

Щоб створити ці цільові процентилі, Elo+Chess відбирає гравців за типом гри та Elo bucket,

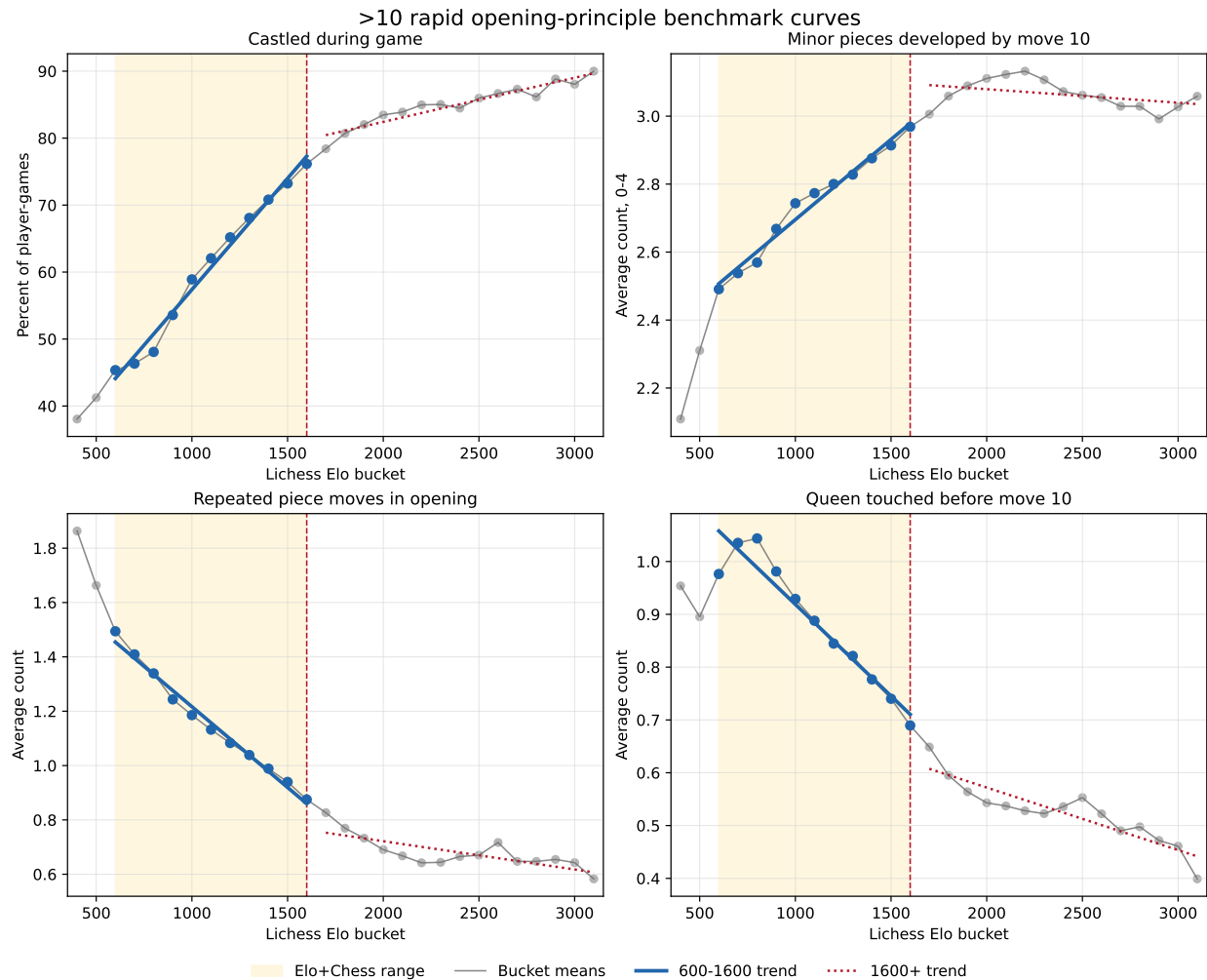


Figure 7: Вибрані криві початкового принципу тестування для тривалих швидких ігор.

а потім використовує останні кваліфікаційні ігри для кожного обраного гравця. The поточний цільовий набір використовує 11 сегментів від 600 до 1600, 200 гравців на відро, і до 50 останніх кваліфікаційних ігор на гравця для кожного виробництва тип гри. Це дає приблизно 2200 зразків гравців для кожного типу гри 110 000 рядків ігор гравця на тип гри для перцентиля всередині сегмента розподіли.

Рисунок 9 ілюструє результуючу ціль розподіл для гравців із вибіркою, конвертований коефіцієнт вилки в 3+0 бліц. Кожна коробка є розподіл між гравцями всередині групи, а не середнє значення за групою індивідуальні ігри. В інтерфейсі користувача звіту значення користувача порівнюється з відповідним вікно для призначеного користувача відра Elo. Це є основою для таких тверджень, як чи високий або низький конвертований форк гравця порівняно з аналогами на тому самому рівні, навіть якщо загальна частота подій у популяції знижується разом із рейтингом.

Game type	Metric	600–1600 slope per 100 Elo	1600+ slope per 100 Elo
1+0 bullet	Castled during game	4.137	0.550
	Minor pieces developed by move 10	0.063	0.009
	Repeated piece moves in opening	-0.051	-0.026
	Queen touched before move 10	-0.040	-0.011
3+0 blitz	Castled during game	4.804	0.313
	Minor pieces developed by move 10	0.084	-0.007
	Repeated piece moves in opening	-0.066	-0.014
	Queen touched before move 10	-0.053	-0.012
10+0 rapid	Castled during game	4.753	0.665
	Minor pieces developed by move 10	0.076	0.008
	Repeated piece moves in opening	-0.064	-0.025
	Queen touched before move 10	-0.052	-0.015
>10 minute rapid	Castled during game	3.320	0.659
	Minor pieces developed by move 10	0.047	-0.004
	Repeated piece moves in opening	-0.059	-0.010
	Queen touched before move 10	-0.035	-0.012

Table 5: Приблизні нахили ковшової кривої для чотирьох прикладів принципу відкриття показники для типів продуктивних ігор.

Game type	Elo buckets	Players per bucket	Games per player	Player samples
1+0 bullet	11	200	49–50	2200
3+0 blitz	11	200	27–50	2200
10+0 rapid	11	200	37–50	2200
>10 minute rapid	11	200	47–50	2200

Table 6: Поточні набори цільових процентилів історії аналогів. Це є зразки історії гравців, які використовуються для оцінки розподілу всередині сегмента для метрики в процентильному стилі.

12 Вартість реплікації

Метод в принципі відтворюваний, але він не легкий. Перебудова контрольні криві вимагають:

- доступ до великих щомісячних файлів історії ігор Lichess;
- стратифіковане фільтрування за типом гри, рейтинговим сегментом і правомочністю правила;
- законний аналізатор ходів і оцінювач стану правління, здатний розширюватися десятки мільйонів станів руху;
- особливість логіки відкриття, матеріал, пішакова структура, захист короля, тура діяльність, тактика, тайм-менеджмент і результати гри;
- агрегування від стану переміщення до функцій гравця-гри;
- підсумки на рівні сегментів для кожного показника звіту;
- ручний перегляд кожної потенційної кривої бенчмарку перед її використанням тренерський звіт.

Two-sided converted-fork rates by Elo bucket

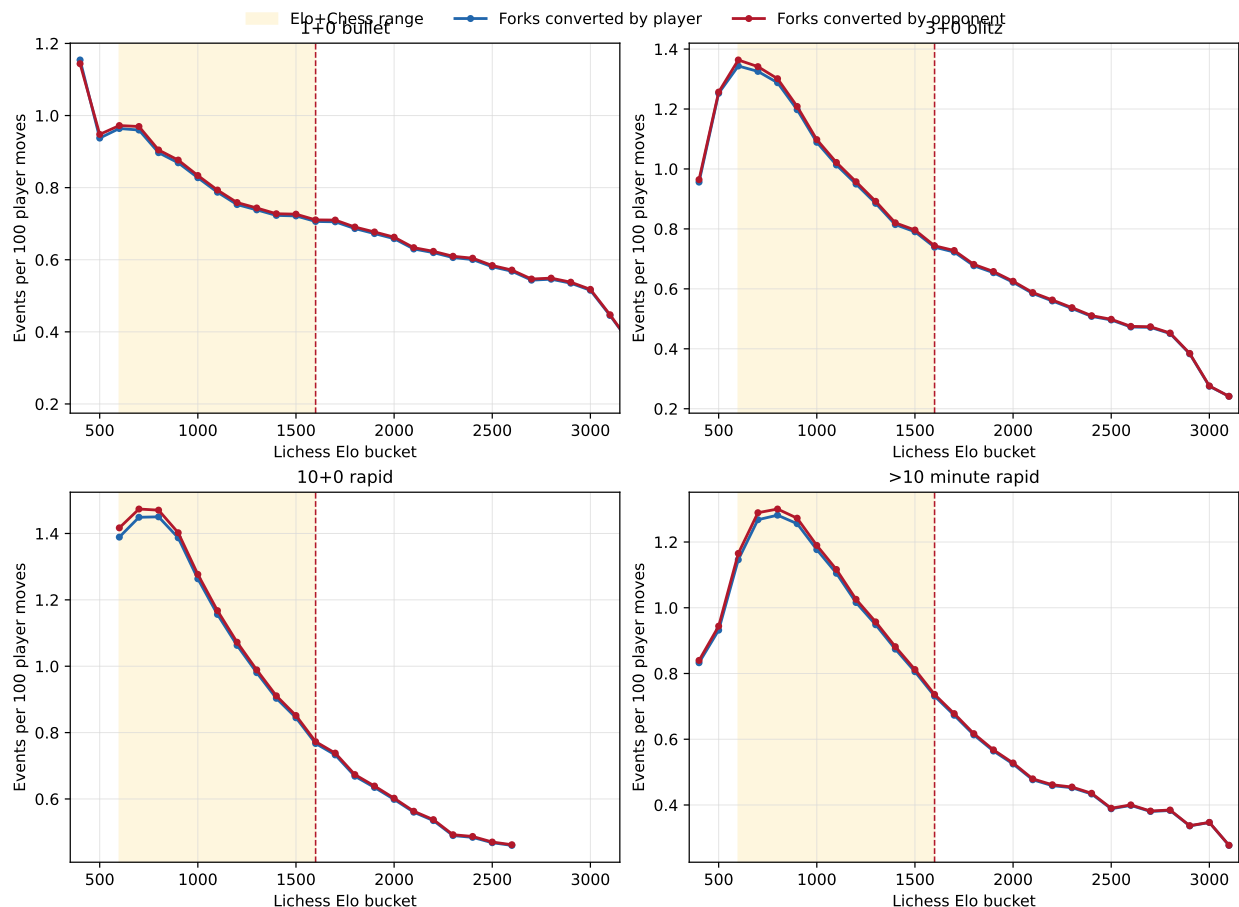


Figure 8: Двосторонні перетворені ставки вилки відром Ело. Перетворені події розгалуження падають на вищі рейтинги, тому що обидві сторони краще захищають і менше чистих форків конверсії доступні.

Для бліцу 3+0 це означає майже півмільйона ігор, майже мільйон рядки гравця-ігри та понад 28 мільйонів рядків стану руху. Через чотири типів еталонних ігор, поточні артефакти еталонного тесту містять понад 2,55 мільйонів ігор, понад 5,10 мільйонів рядків ігор гравця та понад 147 мільйонів рядків стану переміщення. Масштаб є центральним у значенні тест: він дозволяє звіту показувати користувачам, що насправді відбувається в іграх гравцями близького рівня, а не покладатися лише на загальні шахові гасла.

13 Обмеження

Еталонні криві є емпіричними описами, а не причинно-наслідковими доказами. Гравець не отримує рейтинг лише шляхом відповідності середнього значення. Метрики також недосконалі підсумки складних позицій: ранній хід ферзя може бути чудовим в одній позиції і безрозсудно в іншій. Призначення кривих полягає в тому, щоб визначити звички, що повторюються, і порівняти їх із поведінкою однолітків, а не з замінити розрахунок або тренерське судження.

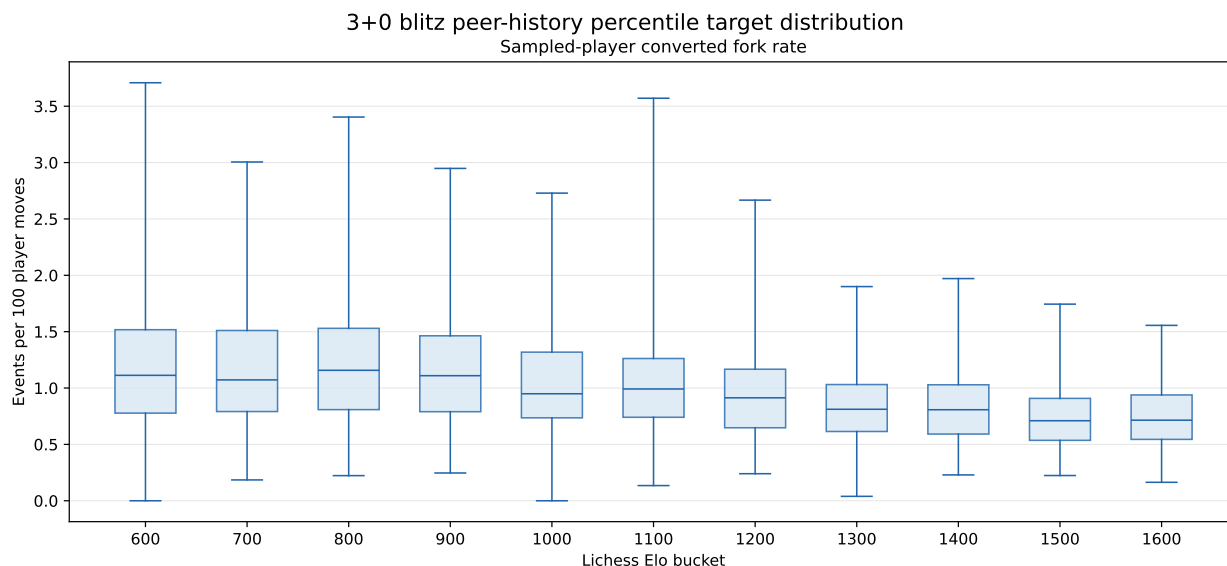


Figure 9: Приклад 3+0 бліц-історія однорангового розподілу цілей для гравця з вибірки конвертований форк курс. Процентилі обчислюються в сегменті Elo, тому звіт порівнює гравця з однолітками, які стикаються із загальом подібними тактичними можливостями ставки.

Криві також є кривими масштабу Lichess. Коли користувачі приносять гру Chess.com історій, Elo+Chess відображає відповідну групу рейтингів на тест Lichess масштабувати за допомогою окремої міжплатформної моделі відображення. Це шар відображення описано в супутній дослідницькій записці Elo+Chess на Lichess-to-Chess.com перетворення рейтингу: [Estimating Cross-Platform Chess Rating Mappings with Модальна перспекія](#). Це відображення не слід читати як збереження процентилів ранг. Він зберігає приблизний рейтинговий еквівалент між двома рейтингами системи; процентний рейтинг залишається залежним від платформи, оскільки активний гравець розподіли різні.

14 Графіка звіту про виробництво

Звіт про виробництво перетворює еталонні артефакти на панелі метричного рівня які можна прочитати, не знаючи повного конвеєра збірки. Цифри 10 and 11 показати три приклади з дизайну звіту Elo+Chess.

Панелі принципу відкриття на малюнку 10 використовувати звичайне представлення еталонної кривої. Кожна картка починається з родина метрики та назва метрики, а потім наводиться коротке правило інтерпретації. The великий сорт у правому верхньому куті навмисно відокремлений від сирого значення метрики: воно відповідає на запитання користувача «як оцінюється ця звичка». порівняно з еталонним показником?”, а рядки нижче зберігають виміряні значення. Чотири підсумкові рядки показують тривалість, останні 10, перемоги та поразки. Це робить панель діагностична, а не просто декоративна. Якщо тривалість життя та останні 10 погоджуються, звичка мабуть стійка. Якщо виграші та поразки відрізняються, показник може бути таким залежить від результату або пов'язано зі станом гри, а не простою звичкою.

Область діаграми використовує стриманий темний фон, золоту лінію тренду та кольоровий колір маркери для чотирьох шматочків. В анотації тенденції чітко зазначено, чи є еталонний

зв'язок зростає або падає разом з Elo. Це важливо, оскільки деякі показники є вищими, тим краще, як-от розробка незначних частин інші – нижче – краще, наприклад ранній рух ферзя. Тому карта не просить користувача визначити напрямок лише за нахилом. Фінал Рядки “Що це означає” і “Порада тренера” перекладають результат тестування в конкретну шахову інструкцію, зберігаючи емпіричний контекст.

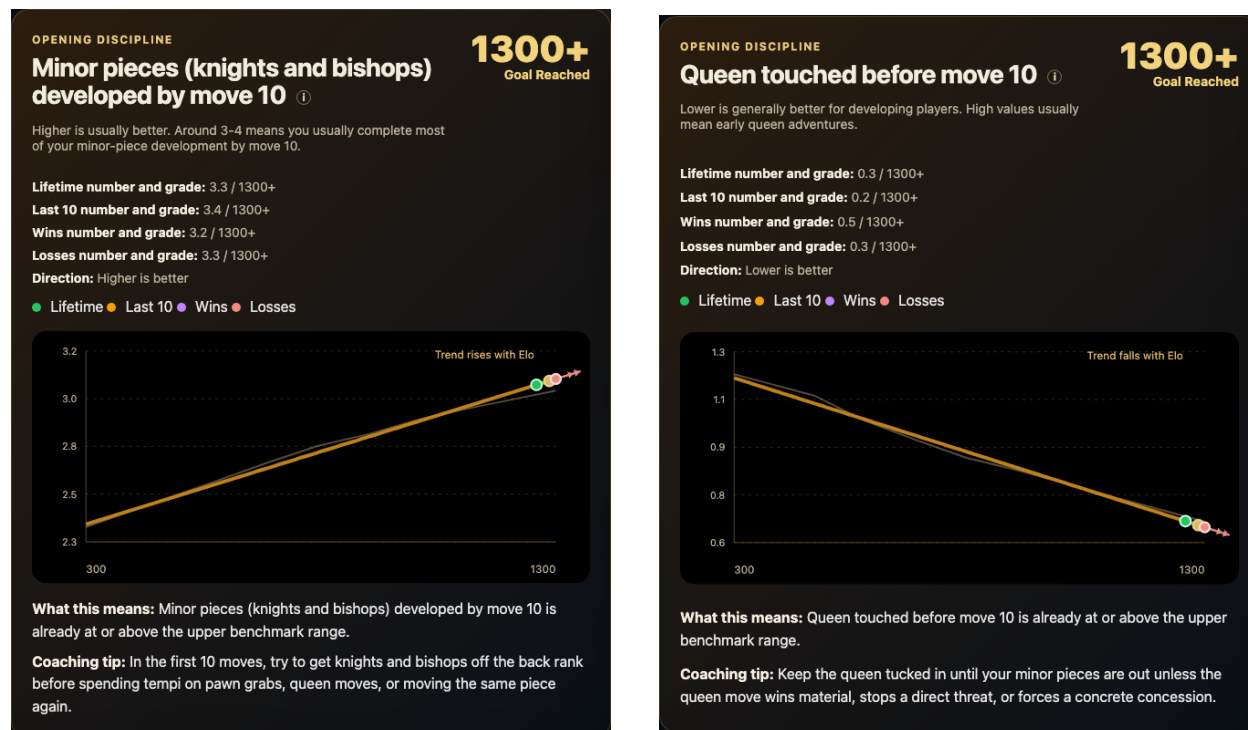


Figure 10: Виготовлення панелей принципу відкриття. Така ж візуальна граматика обробляє метрика «вище — це краще», розвиток менших частин і метрика «нижче — краще», ранній рух ферзя. Обидві панелі показують тривалість життя, останні 10, перемоги та поразки тому у звіті можна відокремити стабільні звички від останніх або залежних від результату візерунки.

Рисунок 11 показує процентний стиль представлення, що використовується для метрик, необроблена швидкість яких погано представлена а єдиний світовий тренд. Виграшні вилки є тактичним прикладом: швидкість населення перетворених вилок змінюється як з навичками гравця, так і з частотою помилок суперника. The таким чином картка порівнює значення користувача з історією інших гравців у зіставлене відро Elo. Коробковий графік відображає розподіл аналогів, у тому числі інтерквартильний діапазон, медіана та вискери. Маркер користувача наноситься на той самий масштаб і позначений необробленим значенням і процентним рангом. Прилеглий відра порівняння залишаються видимими в приглушеній формі, щоб користувач міг побачити, що однорангове відро є локальним, а не універсальним масштабом.

Ця конструкція робить процентильні показники доступними для перевірки. Користувач бачить необроблене значення, процентиль, однорангове відро, що використовується для порівняння, і перекладений Відображення сегментів Chess.com-Lichess. Той самий Lifetime, Last 10, Wins і Рядки втрат зберігаються, тому тактичні процентильні бали все ще можна перевірити для ефектів нещодавності та залежності результату.

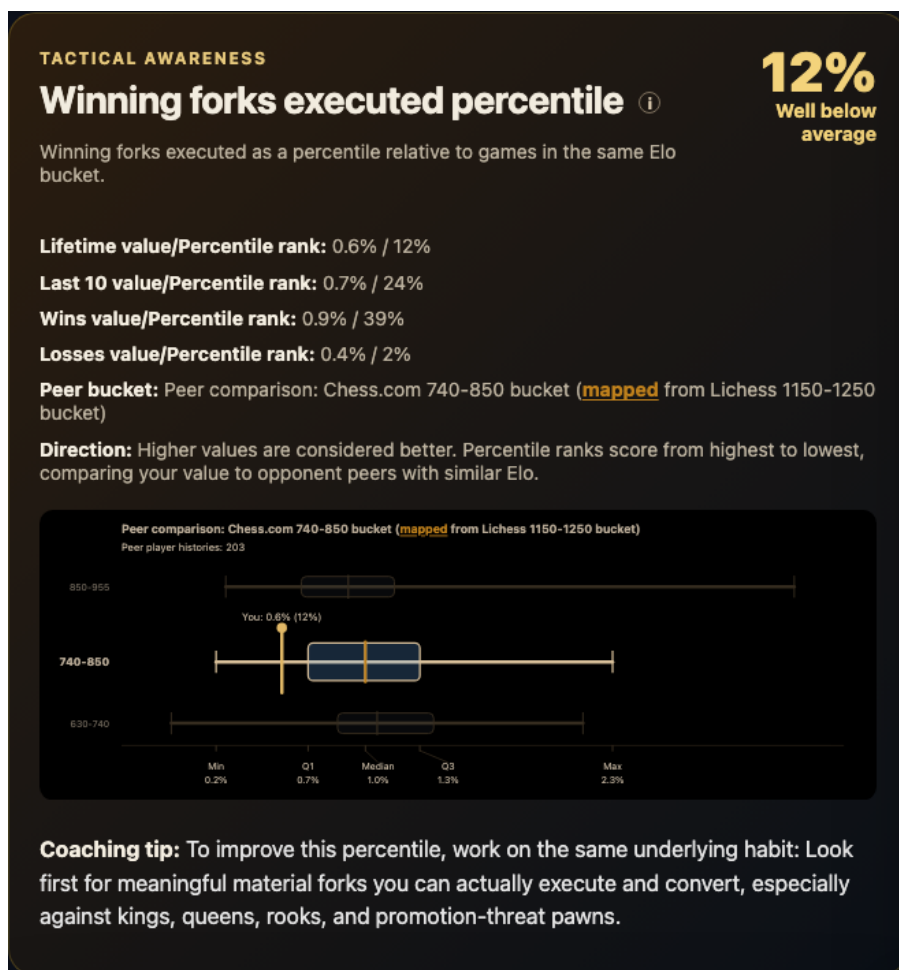


Figure 11: Панель виробничих процентилів для виграшних вилок виконано. Коробковий сюжет показує аналогічний розподіл, який використовується у звіті, тоді як золоте маркер показує необроблене значення користувача та процентний ранг.

15 Висновок

Звіти про контрольні показники Elo+Chess побудовані на основі великого обчислення стану руху стратифіковані зразки Lichess. Кожна метрика починається як набір конкретних станів дошки і змінні стану руху, стає функцією гравця-гри, а потім усереднюється відром Elo, щоб сформувати емпіричну еталонну криву. Принцип відкриття приклади показують, чому цей підхід корисний: прості шахові звички мають сильні, видимі зв'язки з рейтингом у цільовому діапазоні, тоді як ті відносини часто згладжуються або стають більш умовними вище верхньої межі. Це головна причина, чому Elo+Chess зосереджує свої тренерські претензії на рейтингу діапазон, де дані найбільш чітко підтверджують їх.