

Building Stratified Lichess Benchmark Curves para sa Elo+Chess

Elo+Chess Mga Tala sa Pananaliksik

Mayo 31, 2026

Abstract

Elo+Chess benchmark na mga ulat ay binuo mula sa malalaking stratified sample ng Lichess.org laro, na may benchmark na sample na ginamit sa talang ito na naglalaman ng 483,974 1+0 bullet games, 479,974 3+0 blitz games, 925,636 10+0 rapid games, at 660,540 na mas mabibilis na laro. Para sa bawat pangunahing uri ng laro, lumalawak ang pipeline bawat laro sa before-move at after-move board states, kinakalkula ang daan-daang intermediate at player-game variable, pinagsasama-sama ang mga variable na iyon mga sukatan ng pagtuturo, at pagkatapos ay i-average ang mga sukatan na iyon sa pamamagitan ng Lichess Elo bucket. Ang ang sample ay sadyang stratified sa halip na proporsyonal sa raw player pool: ang layunin ay upang tantyahin ang mga matatag na curve ng pag-uugali sa bawat antas ng rating, kabilang ang mga balde na kung hindi man ay kalat-kalat, hindi para magparami ng natural pamamahagi ng rating. Ang resulta ay isang pamilya ng empirical benchmark curves: para sa bawat sukatan, maaari nating tanungin kung gaano kadalas nagpapakita ng ibinigay ang mga manlalaro sa iba't ibang antas ng Elo ugali o prinsipyo sa totoong laro. Inilalarawan ng talang ito ang pamamaraan, ang sukat ng ang pagkalkula, at ilan mga halimbawa ng pambungad na prinsipyo na kinasasangkutan ng castling, minor-piece development, paulit-ulit na opening piece moves, at early queen movement. Ang diin ay metodolohikal: Hindi ipinapalagay ng Elo+Chess na ang mga prinsipyo ng chess ay pangkalahatan linear sa lahat ng antas ng kasanayan. Sa halip, ang mga benchmark na curve ay sinusuri ng saklaw ng rating, at ang produkto ay nakatuon sa beginner-to-early-advanced na rehiyon kung saan ang mga relasyon ay pinakamatibay at pinakanaiintindihan.

1 Layunin

Ang layunin ng sistema ng benchmark ay i-convert ang malawak na payo sa chess masusukat na dami. Ang isang pahayag tulad ng “develop your minor pieces” ay kapaki-pakinabang, ngunit nagiging mas naaaksyunan kapag masasabi natin kung ilang menor de edad na piraso a karaniwang nabubuo ang manlalaro sa pamamagitan ng paglipat 10, kung paano ito maihahambing sa mga kapantay sa parehong rating bucket, at kung paano nagbabago ang ugali na iyon sa buong spectrum ng rating.

Samakatuwid, ang Elo+Chess ay bumubuo ng isang panukat na library mula sa mga aktwal na laro. Ang bawat sukatan ay na-compute para sa bawat hilera ng player-game, na buod ayon sa uri ng laro at Elo bucket, at ginamit bilang benchmark curve sa ulat. Ang ulat ay maaaring maglagay ng user sa ang parehong curve at ipakita kung ang user ay nasa itaas, ibaba, o malapit sa gawi naobserbahan sa mga maihahambing na manlalaro.

2 Pinagmulan ng Data at Stratification

Ang benchmark na pinagmulan ay ang pampublikong Lichess.org database ng laro. Elo+Chess ay gumagamit malalaking stratified sample ayon sa uri ng laro at rating bucket sa halip na maliit sample ng kaginhawaan. Ang kasalukuyang production runtime artifacts ay sumasaklaw sa apat na major mga kategorya ng laro:

- 1+0 bala,
- 3+0 blitz,
- 10+0 mabilis,
- mabilis na laro na mas mahaba sa 10 minuto.

Ang stratification na layunin ay ang lawak sa mga antas ng rating. Isang proporsyonal na random sample ay magiging mahusay para sa paglalarawan ng kabuuang player pool, ngunit ito ay maging hindi mahusay para sa pagbuo ng mga benchmark na kurba. Ang pinakakaraniwang mga bucket ng rating ay mangibabaw sa sample, habang ang mga low-support tail bucket ay magkakaroon ng masyadong kakaunti mga laro para sa mga stable na average na sukatan. Samakatuwid, ang Elo+Chess ay nagsa-sample ayon sa karaniwang laro Elo balde. Sa kasalukuyang clean-sample construction, ang mga raw na laro ang una na-filter sa mga wastong na-rate na laro na may mga kilalang manlalaro, kilalang mga rating, wastong resulta, at kilalang mga kontrol sa oras. Ang mga laro ay sinasala sa mga manlalaro na may sapat na mga laro sa bilis na iyon at sapat na matatag na mga rating upang kumatawan sa magkakaugnay na antas ng kasanayan: para sa bawat manlalaro, kinukuwenta ng Elo+Chess ang tagal ng rating ng player sa loob ng bilis ($\max(\text{Elo}) - \min(\text{Elo})$) sa kabuuan ng source pool at pinapanatili laro lamang kapag ang parehong mga manlalaro ay may hindi bababa sa naka-configure na minimum na bilang ng laro at hindi higit sa 300 Elo point ng within-speed drift. Ang drift filter ay hindi sinadya upang alisin ang normal na pagpapabuti o pagkakaiba; binabawasan nito ang mga kaso kung saan ang isang manlalaro. Ang mga laro ay maghahalo ng materyal na magkakaibang antas ng kasanayan, tulad ng mga pansamantalang account, bumabalik na mga manlalaro, o mga account na mabilis na lumipat sa ilang mga bucket habang ang sample na panahon. Panghuli, ang mga karapat-dapat na laro ay niraranggo sa loob ng bawat average-Elo bucket sa pamamagitan ng isang tiyak na hash ng game id at seed, at isang nalimitahan na bilang ng ang mga laro ay pinananatili mula sa bawat balde.

Ginagawang mas kapaki-pakinabang ng disenyong ito ang mga benchmark na curve. Ang bawat balde ng Elo nag-aambag ng sapat na mga obserbasyon upang matantya ang average na halaga ng bawat sukatan, at ang mga bihirang bucket na may mataas o mababang rating ay hindi nalunod ng natural na hugis ng pamamahagi ng aktibong manlalaro. Ang mga benchmark na curve na ginagamit ng Elo+Chess ay pagkatapos ay pinaghihigpitan sa hanay na pinakanauugnay sa produkto, halos baguhan sa pamamagitan ng maagang advanced na paglalaro. Sa Lichess-scale na mga ulat, ang pangunahing hanay na iyon ay 600–1600. Kapag tiningnan ng isang user ang ulat sa Chess.com scale, ang ang mga katumbas na bucket ng rating ay namamapa sa parehong pinagbabatayan na Lichess mga benchmark na bucket gamit ang hiwalay na paraan ng pagmamapa ng rating ng cross-platform inilarawan sa kasamang papel: [Estimating Cross-Platform Chess Rating Mappings with Modal Regression](#).

3 Mga Pamamahagi ng Rating ng Aktibong Manlalaro

Kung titingnan natin ang pamamahagi ng Lichess.org player Elo ratings sa pagtatapos ng Marso 2026, ayon sa pangunahing uri ng laro, nakikita namin ang mga sumusunod na bilang at porsyento. Para sa view ng pamamahagi na ito, ang bawat manlalaro ay binibilang nang isang beses sa bawat eksaktong uri ng laro, gamit ang pinakabagong naobserbahang rating ng manlalaro sa kaukulang raw time-control row, pagkatapos mangailangan ng hindi bababa sa limang laro sa eksaktong larong iyon uri.¹

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: Pinakabagong naobserbahang pamamahagi ng rating ng aktibong manlalaro na Lichess ayon sa eksaktong uri ng laro pagkatapos mangailangan ng hindi bababa sa limang laro sa eksaktong uri ng larong iyon.

Ang distribusyon na ito ay hindi dapat malito sa stratified benchmark sample. Inilalarawan ng pamamahagi ang aktibong Lichess player pool na naobserbahan sa raw i-scan. Ang benchmark na sample ay sinadyang stratified upang Elo bucket may sapat na suporta para sa mga matatag na metric curve.

Upang i-map mula sa isang Lichess.org Elo rating para sa isang partikular na uri ng laro sa isang katumbas ng Chess.com rating, Elo+Chess ay gumagamit ng game-type-specific modal regression equation na tinatantya sa kasama [Elo+Chess rating-mapping paper](#):

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

Ang mga equation na iyon ay nagmamapa ng mga halaga ng rating, hindi percentile rank. Ang mga ranggo ng porsyento ay tinukoy sa loob ng isang partikular na pool ng manlalaro, at ang manlalaro ng Chess.com at Lichess ang mga pool ay may iba't ibang hugis. Sa Talahanayan 2, para sa 10 minutong mabilis na uri ng laro, ang unang column ay nagbibigay ng maliit na sample ng posible Chess.com Elo na mga halaga ng rating. Ang ikalawang hanay ay nagbibigay ng katumbas Chess.com percentile rank na ipinapakita sa Chess.com noong Mayo 31, 2026. A 71.7 ranggo sa Lichess.org para sa 10 minutong mabilis na tumutugma sa isang Lichess Elo rating ng 1,644² at ang halagang iyon ay ipinapakita sa ikatlong hanay. Kung ilalapat natin ang Chess.com–Lichess mapping equation, 1,644 sa Ang Lichess ay katumbas ng humigit-kumulang 1,271 sa Chess.com. Ipagpalagay ang tinantyang makatwiran ang mapping equation, ang +522 gap sa

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 ipakita ang mga distribusyon na mayroon at walang filter na ito.

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.

pagitan ng Chess.com rating sa column 1 at ang naka-map na rating sa column 4 ay maaaring ipaliwanag ng malaking bilang ng mga kaswal na manlalaro sa Chess.com player pool na nagpapalaki sa Chess.com porsiyentong ranggo. Sa madaling salita, ang pagiging nasa 72nd percentile sa Lichess.org ay isang mas malaking tagumpay kaysa sa pagiging nasa 72nd percentile sa Chess.com, dahil Ang mga percentile rank ay hindi malinis na nagmamapa mula sa isang sistema patungo sa isa pa. Talagang Elo ang mga halaga ay maaaring imapa gamit ang mga equation na tinantiya mula sa mga marka ng parehong manlalaro sa kabuuan ang dalawang sistema.

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: Mapaglarawang paghahambing sa pagitan ng Chess.com 10 minutong mabilis na percentile mga puntos at ang parehong mga percentile na puntos sa aktibong Lichess 10+0 mabilis pamamahagi pagkatapos mangailangan ng hindi bababa sa limang 10+0 laro. “Nakamapang Chess.com equiv.” nalalapat ang kasalukuyang Elo+Chess 10+0 mabilis na pagbabago sa scale ng rating sa parehong-percentile na rating ng Lichess. Ang Ang mga hilera ng pinakamataas na buntot ay kasama upang ipakita ang pagkakaiba ng sukat, ngunit ang angkop ang linya ng rating ay pinaka maaasahan sa pangunahing hanay ng pagtuturo.

Ang dahilan ng mga gaps sa Table 2 ay ang pagmamapa tinatantya ang katumbas na mga rating ng lakas ng paglalaro, samantalang inilalarawan ng isang percentile posisyon sa loob ng sariling populasyon ng isang platform. Isang 71.7th-percentile player sa Samakatuwid, ang Chess.com ay hindi awtomatikong isang 71.7th-percentile player sa Lichess, at vice versa. Ang mga naobserbahang numero ay pare-pareho sa publiko ng Chess.com mabilis na pool na may mas malaking lower-rated o casual-player mass, habang ang Ang aktibong eksaktong-kontrol na Lichess pool ay mas puro sa mga karanasang umuulit mga manlalaro. Ang pangunahing punto ay hindi ang isang percentile scale ay “tama” at ang ang iba ay “ mali ”; ito ay ang percentile rank ay hindi portable sa mga platform maliban kung ang pinagbabatayan na mga pamamahagi ng manlalaro ay may parehong hugis.

Note on the 1500 spike. The visible jump near Ang 1500 ay naroroon sa data sa halip na nilikha sa pamamagitan ng pagpapakinis. Sa pinakahuling naobserbahang pamamahagi ng rating, eksaktong 1500 rating ang account para sa 13.1

bullet player, 5.6

17.6

panimula o pansamantalang anchor sa rating system.

Ang eksaktong-1500 spike ay higit sa lahat ay isang artifact ng mga hindi gaanong aktibong account. Upang suriin ito, ni-compute namin ang parehong mga pamamahagi ng pinakabagong rating pagkatapos na kailanganin bawat manlalaro ay magkaroon ng hindi bababa sa limang laro sa eksaktong uri ng laro. Ang spike halos nawawala sa ilalim ng filter na ito: ang eksaktong 1500 rating ay bumaba mula sa 5.6–17.6

mga manlalaro sa all-player distribution sa 0.2–1.2

pamamahagi na iniulat sa Talahanayan 1.

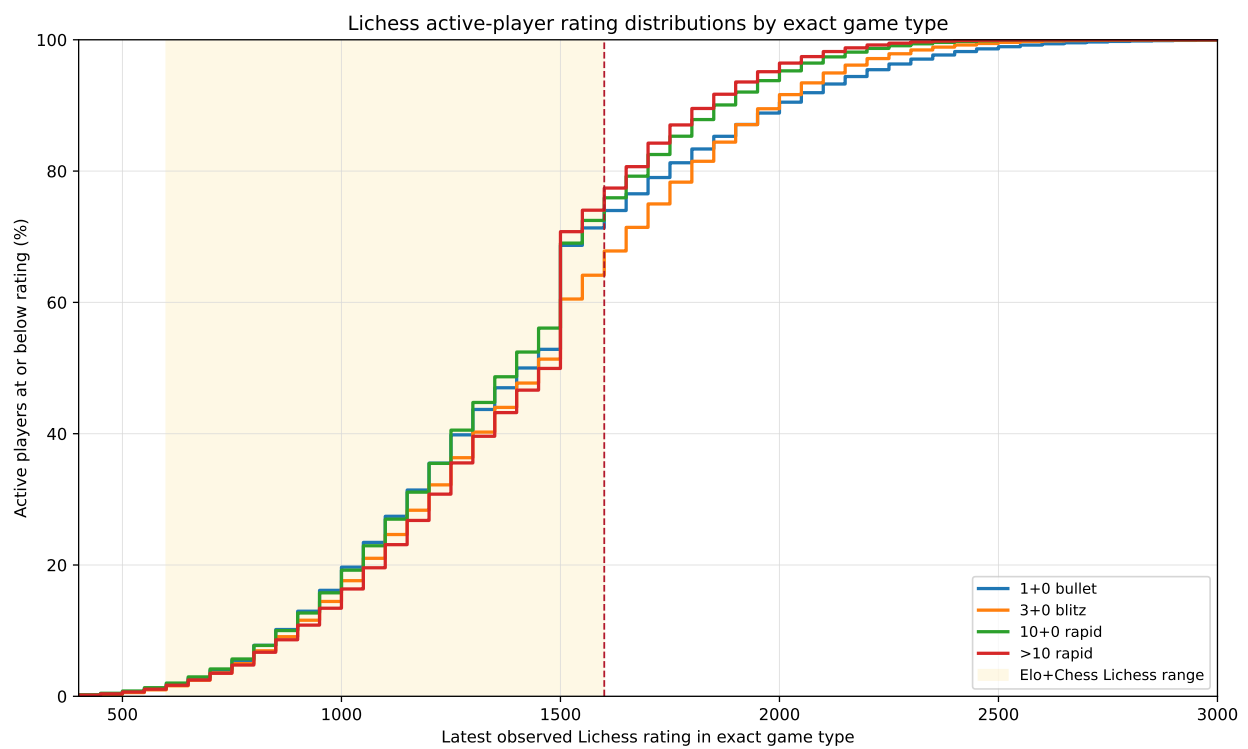


Figure 1: Pinagsama-samang pamamahagi ng rating ng aktibong manlalaro ayon sa eksaktong uri ng laro. Ang ang dilaw na banda ay nagmamarka ng 600–1600 Lichess range na ginagamit ng pangunahing Elo+Chess mga benchmark sa pagtuturo.

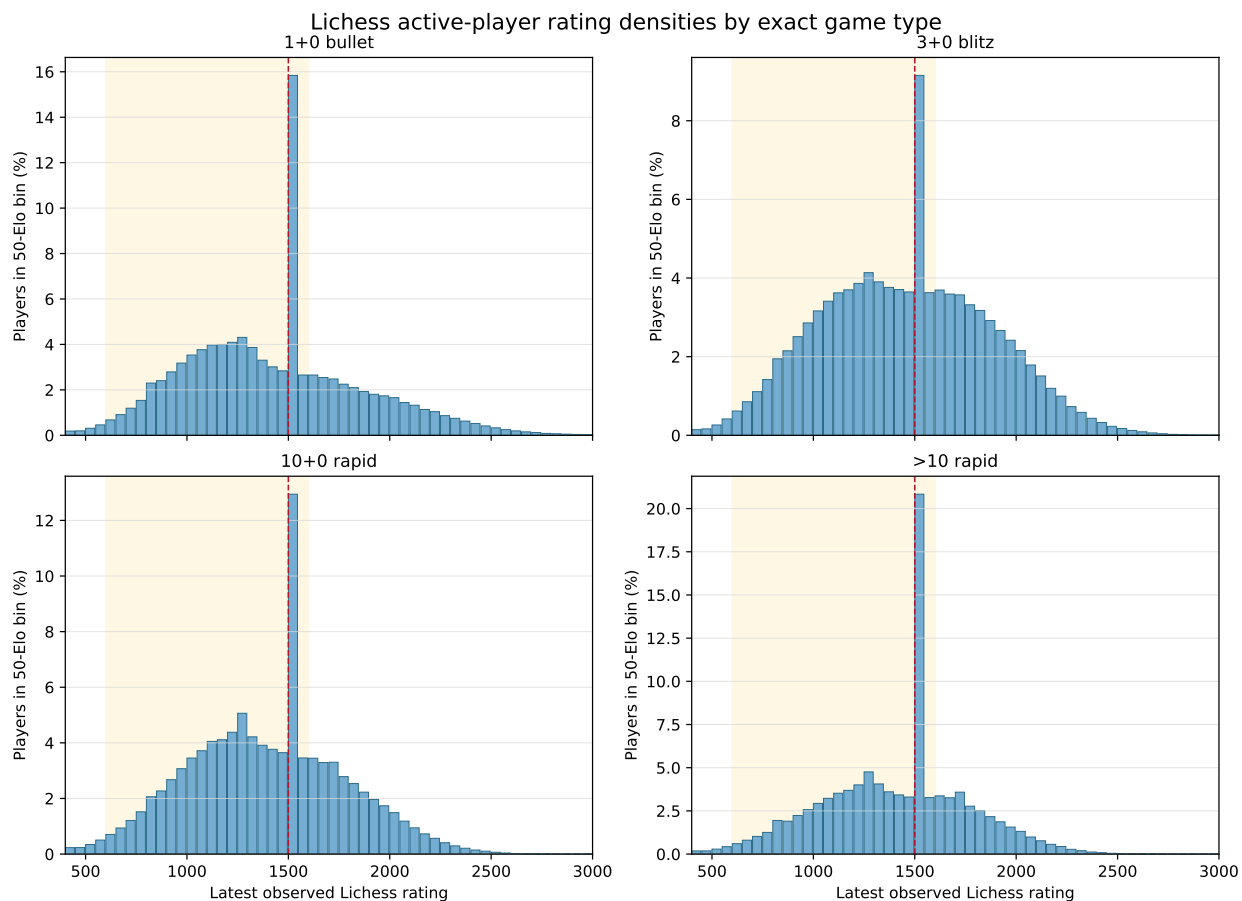


Figure 2: Density view ng parehong mga pamamahagi ng rating ng aktibong manlalaro. Ang ang putol-putol na linya ay nagmarka ng 1500 Lichess, kung saan ang mga pamamahagi raw ng pinakabagong rating ay mayroong isang malaking eksaktong-rating na mass point. Ang dilaw na banda ay nagmamarka ng 600–1600 Lichess saklaw na ginagamit ng mga pangunahing ulat ng sukatan ng Elo+Chess. Ang spike ay pinaka nakikita sa ang densidad ng lahat ng manlalaro dahil maraming mga hindi gaanong aktibong account ang nananatiling eksakto sa 1500; Figure 4 inuulit ang density plot pagkatapos mangailangan ng hindi bababa sa limang laro sa eksaktong uri ng laro, na higit na nag-aalis artifact na ito.

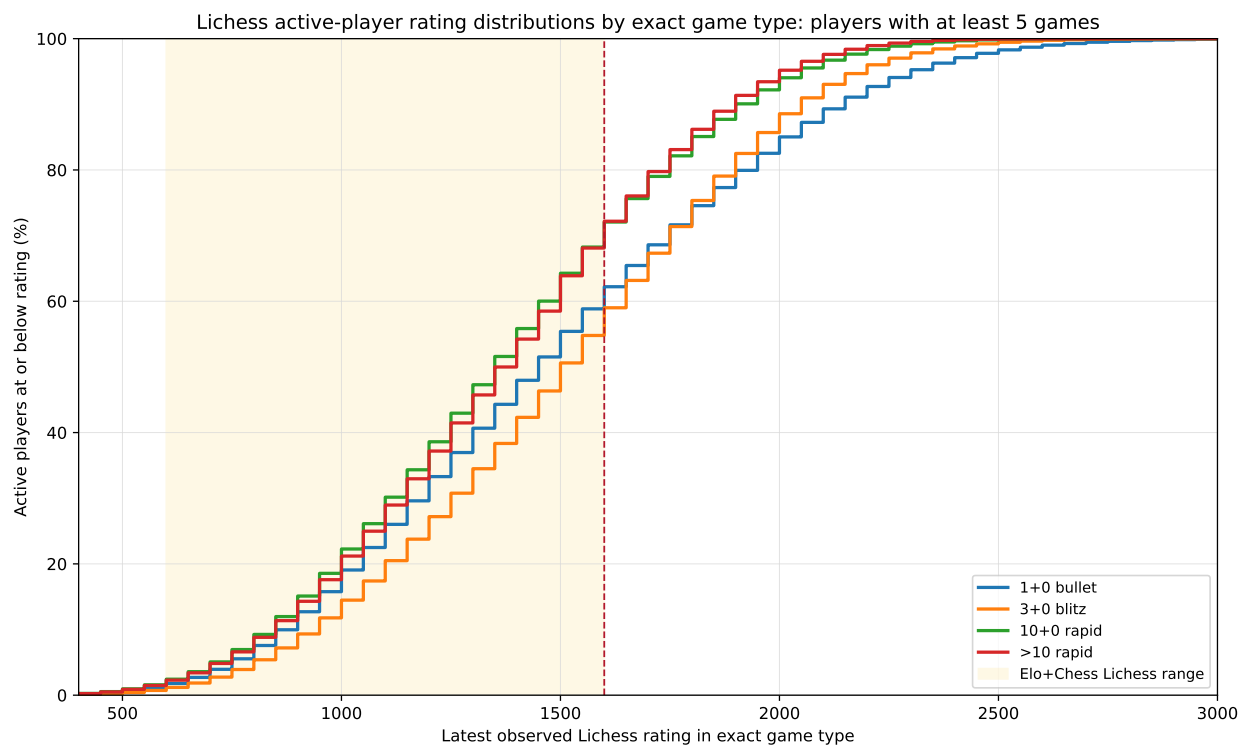


Figure 3: Pinagsama-samang mga pamamahagi ng rating ng aktibong manlalaro pagkatapos na kailanganin sa hindi bababa sa limang laro sa eksaktong uri ng laro. Ang 1500 jump ay lubhang nabawasan.

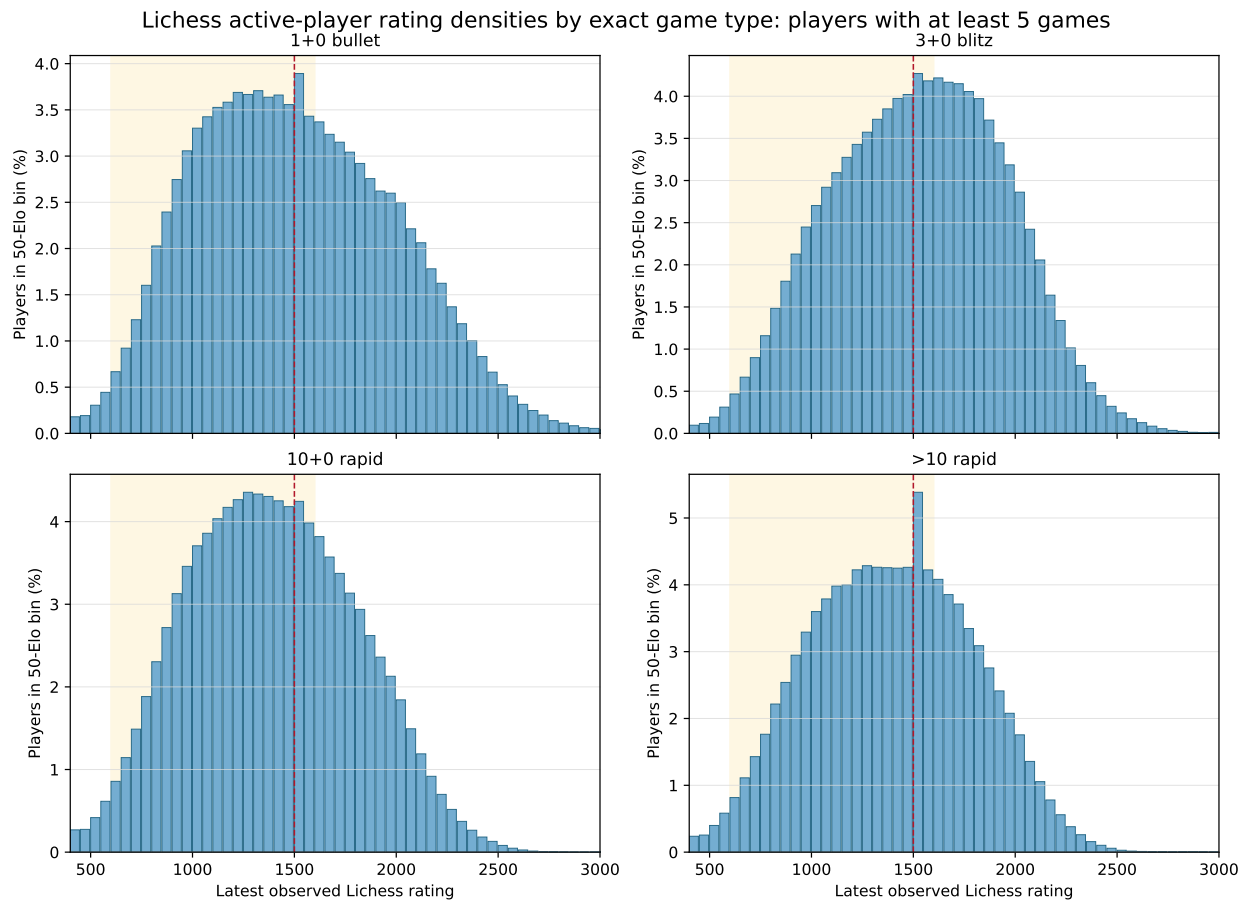


Figure 4: Density view pagkatapos mangailangan ng hindi bababa sa limang laro sa eksaktong laro uri. Kinukumpirma ng view na ito na ang malaking 1500 mass point sa all-player ang pamamahagi ay kadalasang hinihimok ng mga hindi gaanong aktibong account. Ang mga marka ng dilaw na banda ang hanay na 600–1600 Lichess na ginagamit ng mga pangunahing ulat ng sukatan ng Elo+Chess.

4 Scale ng Kasalukuyang Benchmark Artifact

Ang pagkalkula ay sadyang malaki. Ito ay hindi isang profile scrape, isang maliit hanay ng mga na-download na PGN, o isang koleksyon ng mga halimbawang nakapagtuturo na ginawa ng kamay. Ang paunang Lichess pass ay sumasaklaw sa kumpletong buwanang mga file mula Enero 2025 hanggang Marso 2026. Ang unstratified scan table ay naglalaman ng 1,382,178,087 laro at 2,764,356,174 na hilera ng rating ng player-game. Sa mga row na iyon, 6,623,438 ang natatangi ang mga username ng manlalaro ay lilitaw nang hindi bababa sa isang beses.

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: Hindi na-stratified na Lichess scan scale bago ang benchmark sampling. Ang pag-scan sumasaklaw sa kumpletong buwanang mga file ng kasaysayan ng laro mula Enero 2025 hanggang Marso 2026.

Ang mga artifact ng benchmark ng produksyon ay mas maliit kaysa sa raw scan dahil ang mga ito ay pinili para sa metric construction. Malalaki pa rin sila: ang benchmark stage ay isang move-state expansion at aggregation pipeline sa daan-daang libu-libong laro sa bawat uri ng laro.

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: Scale ng mga stratified benchmark artifact pagkatapos ng alignment sa karaniwang 277-column runtime feature schema. A Ang “player-game row” ay nangangahulugang isang laro mula sa pananaw ng isang manlalaro, kaya karamihan sa mga laro gumawa ng dalawang hilera ng player-game.

Para sa 3+0 blitz lang, ang kasalukuyang production benchmark ay gumagamit ng 479,974 na laro kinasasangkutan ng 129,398 natatanging manlalaro at 28,138,904 natatanging mga hilera ng paglipat-estado. Ang aligned blitz player-game feature table ay naglalaman ng 277 derived variables. Ang mas malawak pa ang mga artifact ng mabilis na feature-build sa move-state stage: ang buo Ang talahanayan ng mabilis na paglipat-estado na ginamit sa panahon ng pagbuo ng tampok ay naglalaman ng 168 ilipat ang estado ng mga column bago ang pagsasama-sama sa isang 277-column player-game feature mesa. Ang frontend runtime benchmark artifact para sa lahat ng apat na uri ng laro ay pagkatapos ay nakahanay sa parehong 277-column player-game feature schema, kasama ang pinalawak na tactic-taxonomy breakout variable. Mula sa mga feature table na ito, ang ang ulat ay pumipili at nagpapakita ng mga sukatan ng pagtuturo na pinakanaiintindihan.

Ang mga bilang na ito ay mahalaga para sa pagbibigay-kahulugan sa mga resulta. Ang benchmark ay kurba ay hindi mga pagtatantya mula sa ilang daang manu-manong sinuri na mga laro. Sa halip, sila mula sa sampu-sampung milyong mga nasuri na estado ng board at higit sa apat

milyong hilera ng player-game sa apat na kasalukuyang kategorya ng production game.

5 Pangkalahatang-ideya ng Pipeline

Ang benchmark construction pipeline ay may anim na pangunahing hakbang:

1. Piliin ang na-rate na Lichess na mga laro sa target na uri ng laro at strata ng rating.
2. I-parse ang bawat galaw ng laro sa pamamagitan ng galaw at iimbak ang before-move at after-move estado ng board.
3. Kalkulahin ang mga intermediate na variable na estado ng paglipat mula sa board, ang paglipat, ang pananaw ng manlalaro, at ang pananaw ng kalaban.
4. Pagsama-samahin ang mga variable ng move-state sa mga variable na feature ng player-game.
5. I-convert ang mga variable ng player-game sa mga sukatan ng ulat na may malinaw direksyon ng interpretasyon: mas mataas ay karaniwang mas mahusay, mas mababa ay karaniwang mas mahusay, o naglalarawan.
6. I-average ang bawat sukatan sa pamamagitan ng Elo bucket at manu-manong suriin ang resulta relasyon.

Ang istrukturang ito ay mahalaga dahil maraming mga sukatan ng ulat ay hindi direkta makikita sa isang PGN header. Nangangailangan sila ng muling pagtatayo ng posisyon, pagtukoy kung aling mga piraso ang lumipat mula sa kung saan nagsisimula ang mga parisukat, paghahambing ng bago-galaw at after-move material at structure, at pagkatapos ay pagsasama-samahin ang resulta mula sa ilipat ang antas sa antas ng player-game.

6 Mga Intermediate Variable

Ang mga intermediate variable ay idinisenyo upang ilarawan kung ano ang nagbago sa pisara at kung bakit ito mahalaga mula sa pananaw ng manlalaro ng ulat. Kasama sa mga halimbawa ang:

- numero ng paglipat, numero ng paglipat ng manlalaro, bahagi upang ilipat, kulay ng mover, at player kulay;
- from-square, to-square, moving piece type, captured piece type, at kung ang paglipat ay isang pagkuha, tseke, kastilyo, promosyon, o kapareha;
- balanse ng materyal bago ilipat at pagkatapos ilipat;
- kung ang isang menor de edad na piraso ay umalis sa panimulang parisukat sa pamamagitan ng isang ibinigay na numero ng paglipat;
- kung ang parehong piraso ay inilipat sa magkakasunod na manlalaro na gumagalaw sa pagbubukas;
- kung lumipat ang reyna bago lumipat 10;

- castling side, castling move number, at kung ang mga rook ay naging konektado pagkatapos ng castling;
- maluwag na piraso, nakabitin na piraso, dobleng mga pawn, nakahiwalay na mga pawn, naipasa pawns, at paatras pawns;
- katayuan ng rook file, bukas na mga file, kalahating bukas na mga file, at king-safety mga tampok;
- mga taktikal na pagkakataon, mga taktika na isinagawa, mga taktika sa materyal-presyon, garantisadong follow-up na mga pakinabang, at natanto na mga pakinabang.

Ang eksaktong intermediate set ay nag-iiba ayon sa yugto ng pagbuo. Pinapanatili ang mga database ng ulat ng runtime ang mga column lang ang kailangan para makapaghatid ng mga ulat nang mabilis, habang ang buong feature-build Kasama sa mga artifact ang mas malawak na move-state na mga talahanayan at mga tampok na patch na ginagamit upang bumuo ang huling mga hilera ng tampok na laro ng manlalaro.

7 Mga Halimbawang Sukatan ng Pambungad na Prinsipyo

Ang mga pambungad na prinsipyo ay mga kapaki-pakinabang na halimbawa dahil madali ang mga ito para sa mga gumagamit maunawaan at madaling kumonekta sa mga kalkulasyon ng paglipat-estado.

7.1 Minor Pieces Developed by Move 10

Binibilang ng panukat na ito kung ilan sa apat na kabalyero at obispo ng manlalaro ang natira kanilang home squares sa ika-10 galaw ng manlalaro. Ang intermediate na pagkalkula gumagamit ng kulay ng mover, uri ng gumagalaw na piraso, panimulang parisukat, at paglipat ng manlalaro numero. Para sa White, ang mga parisukat sa bahay ay b1, g1, c1, at f1. Para kay Black, sila b8, g8, c8, at f8. Para sa bawat laro ng manlalaro, binibilang ng Elo+Chess ang natatanging tahanan mga parisukat sa mga kabalyero at obispo na ginalaw ng manlalaro ay gumagalaw 10.

Ito ay isang mas mataas-ay-karaniwang-mas mahusay na sukat. Hindi nito inaangkin na ang bawat menor de edad piraso ay dapat palaging ilipat sa pamamagitan ng paglipat 10, ngunit sa pagbuo ng mga manlalaro ang empirical malinaw ang relasyon: ang mga manlalaro na nakakaabot ng mas mataas na Elo bucket ay may posibilidad na bumuo mas menor de edad na piraso kanina.

7.2 Repeated Piece Moves in the Opening

Ang sukatang ito ay nagbibilang ng mga karagdagang maagang galaw na ginugol sa paglipat muli sa parehong piraso sa halip ng pagbuo ng bago. Ang intermediate na pagkalkula ay gumagamit ng nauna destination square, ang kasalukuyang origin square, at ang player move number. Kung ang parehong piraso ay inilipat sa nakaraang paglipat ng manlalaro at inilipat muli sa loob ng unang 10 galaw ng manlalaro, tataas ang bilang ng paulit-ulit na galaw.

Ito ay isang mas mababa-ay-karaniwang-mas mahusay na sukat. Sinusukat nito ang wasted opening tempi sa isang simpleng paraan. Ang ilang mga paulit-ulit na galaw ay taktikal na makatwiran, ngunit sa sukat ang kinukuha ng bucket curve ang pangkalahatang tendensya: mas malakas na mga manlalaro sa target ang mga antas ay gumugugol ng mas kaunting mga maagang galaw sa paglipat ng parehong piraso nang paulit-ulit.

7.3 Queen Touched Before Move 10

Binibilang ng panukat na ito ang paglipat ng reyna bago lumipat ang manlalaro ng 10. Ang intermediate Ang pagkalkula ay gumagamit ng gumagalaw na uri ng piraso at numero ng paglipat ng manlalaro. Kung ang gumagalaw na piraso ay ang reyna at ang numero ng paglipat ng manlalaro ay hindi hihigit sa 10, tataas ang bilang.

Ito rin ay mas mababa-ay-karaniwang-mas mahusay para sa hanay ng target na Elo+Chess. Maaga ang mga queen moves ay maaaring manalo ng materyal o puwersahang mga konsesyon, ngunit maraming umuunlad na mga manlalaro ilabas ang reyna ng masyadong maaga, mawala ang tempiyo sa mga pag-atake sa reyna, at antalahin minor-piece development at king safety.

7.4 Castling and King Safety

Gumagamit ang mga sukatan ng castling ng mga castling flag, castling side, castling move number, at ang board state pagkatapos ng castling. Kasama sa mga halimbawa kung nag-cast ang manlalaro, kung ang player ay na-castled sa pamamagitan ng move 10, kung ang player ay nag-castled sa kingside o queenside, kung ang mga rook ay konektado pagkatapos ng castling, at kung ang mga rook file ay bukas o kalahating bukas pagkatapos ng castling.

Ang mga sukatang ito ay hindi lahat ay binibigyang kahulugan sa parehong paraan. Pagkumpleto ng castling at Ang castling sa pamamagitan ng paglipat 10 ay karaniwang mas mataas-ay-karaniwang-mas mahusay sa target na hanay. Ang Queenside castling o castling sa mga bukas na file ay maaaring mas nakadepende sa konteksto. Iyon ang dahilan kung bakit ang Elo+Chess ay hindi lamang hard-code ng isang moral na tuntunin; ito ay gumagawa ng balde curves at sinisiyasat kung aling mga relasyon ang sapat na matatag para sa pagtuturo.

8 Bucket Curves at Manu-manong Inspeksyon

Pagkatapos kalkulahan ang mga sukatan ng laro ng manlalaro, ipapangkat ng Elo+Chess ang mga ito ayon sa Lichess Elo bucket at nag-compute ng mga average ng bucket. Ang bucket average ay ang raw empirical benchmark na punto para sa sukatan na iyon. Ang isang trend line ay gagamitin sa ulat sa magtalaga ng mga Elo-style na marka sa mga value ng user.

Ang mga kurba ay manu-manong siniyasat bago gamitin. Mas kapaki-pakinabang ang isang sukatan kapag ang ibig sabihin ng bucket ay nagpapakita ng magkakaugnay na relasyon sa buong saklaw ng rating inihain ng produkto. Ang isang sukatan ay hindi gaanong kapaki-pakinabang kapag ito ay maingay, patag, mataas hindi monotonic, o hinihimok ng mga espesyal na kaso na mahirap ipaliwanag sa isang user.

Ang hakbang na ito ng manu-manong inspeksyon ay lalong mahalaga sa itaas ng Elo+Chess sa itaas cutoff. Alam ng mas malalakas na manlalaro kung kailan nilalabag ang mga simulain sa pagbubukas at gawin ito mabisa. Sa itaas ng beginner-to-early-advanced range, ang relasyon sa pagitan ng mga simpleng gawi at rating ay kadalasang nagiging flattens o nagbabago ng hugis. Ang produkto samakatuwid ay binibigyang-diin ang saklaw kung saan ang mga empirikal na relasyon ay malakas at pedagogically kapaki-pakinabang.

9 Halimbawa Full-Range Opening Curves

Ang Figure 5 ay nagpapakita ng apat na 3+0 blitz na opening-principle curve. Ipinapakita ng mga gray na puntos ang mas malawak na hanay ng rating ng Lichess. Mga asul na puntos at ang

asul na trend line show ang 600–1600 Lichess range na ginagamit ng pangunahing Elo+Chess coaching mga benchmark. Ang putol-putol na pulang linya ay nagmamarka sa itaas na cutoff.

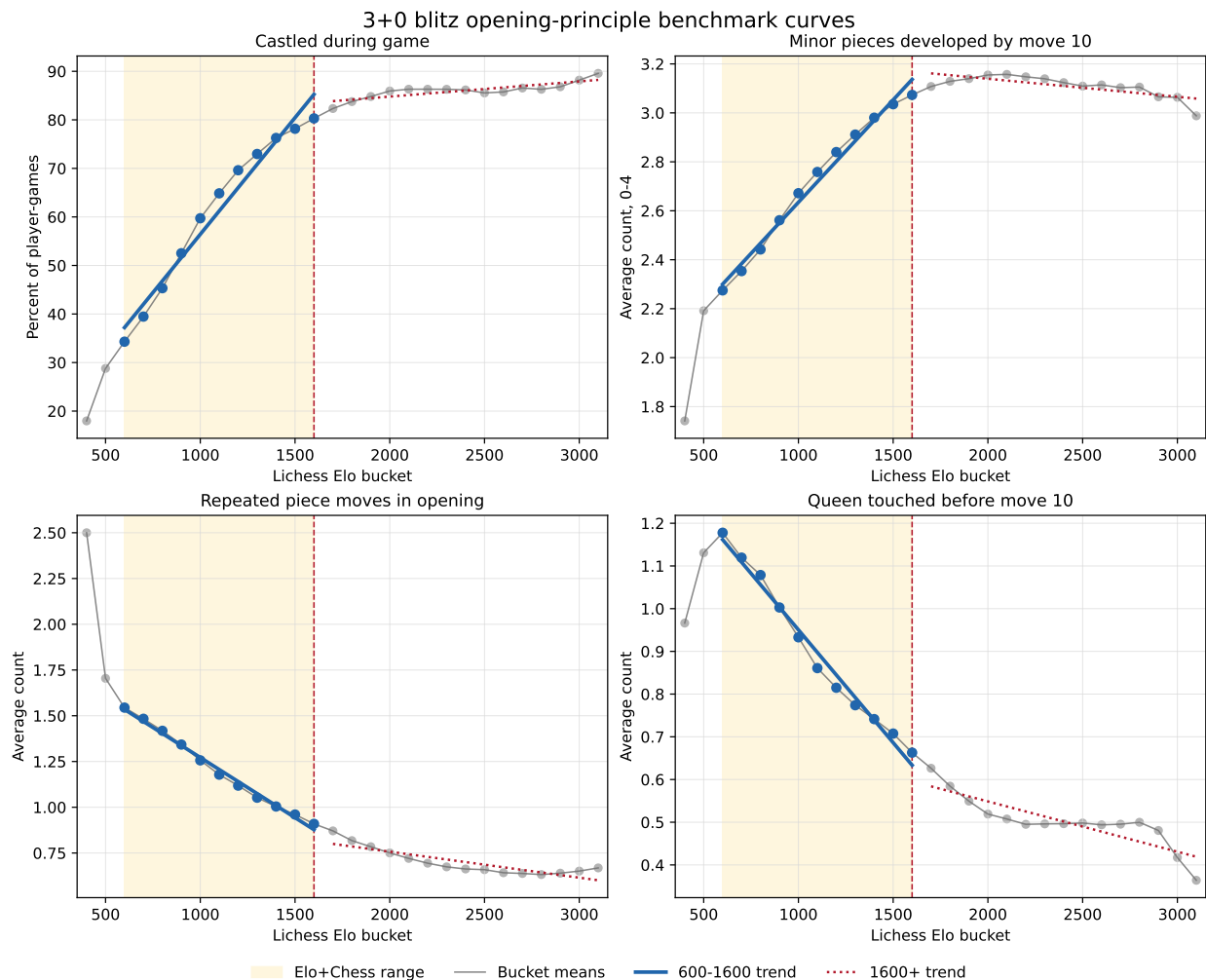


Figure 5: Mga napiling opening-principle benchmark curves para sa 3+0 blitz. Ang ang mga relasyon ay pinakamatibay at pinakakapaki-pakinabang sa hanay ng pagtuturo ng Elo+Chess. Sa itaas ng itaas na cutoff, ang ilang mga relasyon ay na-flat nang malaki.

Ang parehong pattern ay makikita sa full-range na 10+0 na mabilis at mas mahabang mabilis mga artifact. Ang sentral na desisyon sa pagmomodelo ay hindi nagbabago: gamitin ang empirically matatag na hanay para sa mga claim sa pagtuturo at iwasang magpanggap na isang simpleng kurba ng ugali nananatiling pantay na nagbibigay-kaalaman magpakailanman.

Ang parehong pattern ay lilitaw ayon sa numero sa Talahanayan 5. Ang mga slope ay iniulat bilang pagbabago sa metric-value sa bawat 100 Elo point. Castling, minor-piece pag-unlad, paulit-ulit na paggalaw ng piraso, at maagang paggalaw ng reyna lahat ay malinaw paggalaw sa 600–1600. Sa itaas ng 1600, ang mga slope ay mas maliit sa mga ito mga halimbawa, na naaayon sa ideya na masira ang mga manlalarong may mataas na rating mga pangunahing prinsipyo nang mas pinipili at mas matagumpay. Ang pangalawang mekanismo ay saturation. Para sa ilang mga gawi, ang sukatan ay umaabot sa isang praktikal na kisame o matamis spot: natutunan na ng mga malalakas na manlalaro ang ugali, kaya walang dahilan para ang kurba upang patuloy

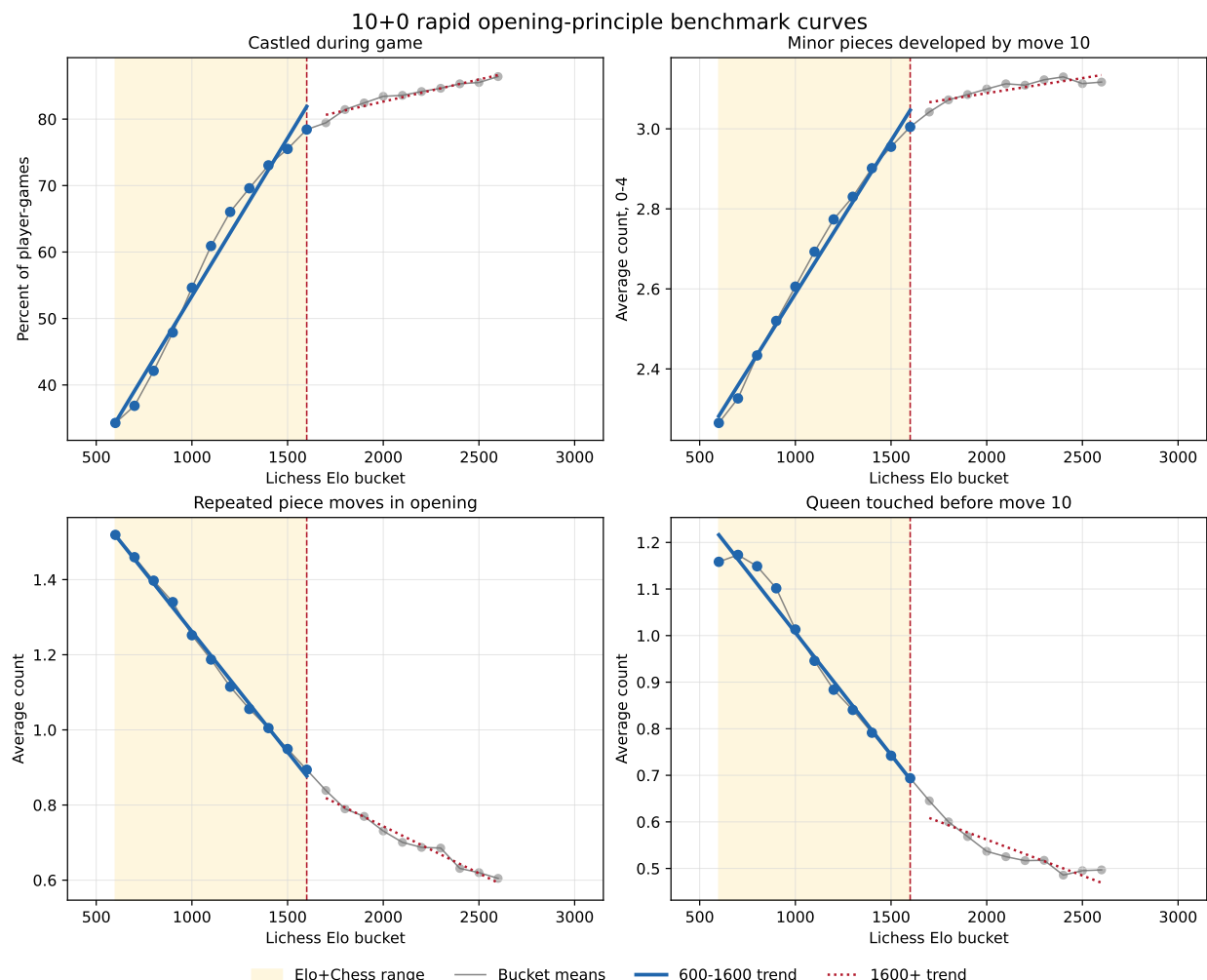


Figure 6: Mga napiling benchmark na curve ng pambungad na prinsipyo para sa 10+0 mabilis na laro. Ang ang mga relasyon ay pinakamatibay at pinakakapaki-pakinabang sa hanay ng pagtuturo ng Elo+Chess. Sa itaas ng itaas na cutoff, ang ilang mga relasyon ay na-flat nang malaki.

na gumagalaw nang husto pataas. Higit din ay hindi palaging mas mahusay pagkatapos puntong iyon. Ang data samakatuwid ay patagin pareho dahil ang mas malalakas na manlalaro ay maaaring masira mga alituntunin sa mga konkretong posisyon at dahil marami nang batayang gawi ang naabot na ang kanilang kapaki-pakinabang na saklaw ng pagpapatakbo.

10 Bakit Mahalaga ang Upper Cutoff

Ang Elo+Chess target na user ay isang beginner-to-early-advanced na player. Para sa mga ito mga manlalaro, ang mga pangunahing gawi ay malakas na nauugnay sa rating: bumuo ng mga piraso, iwasan ang paulit-ulit na maagang mga galaw ng piraso, iwasan ang mga hindi kinakailangang maagang paggalaw ng reyna, kastilyo pagdating ng panahon, ikonekta ang mga rook, pamahalaan ang materyal, at iwasang mag-iwan ng mga piraso na maluwag o nakabitin.

Sa mas mataas na mga rating, ang parehong mga gawi sa ibabaw ay nagiging hindi gaanong

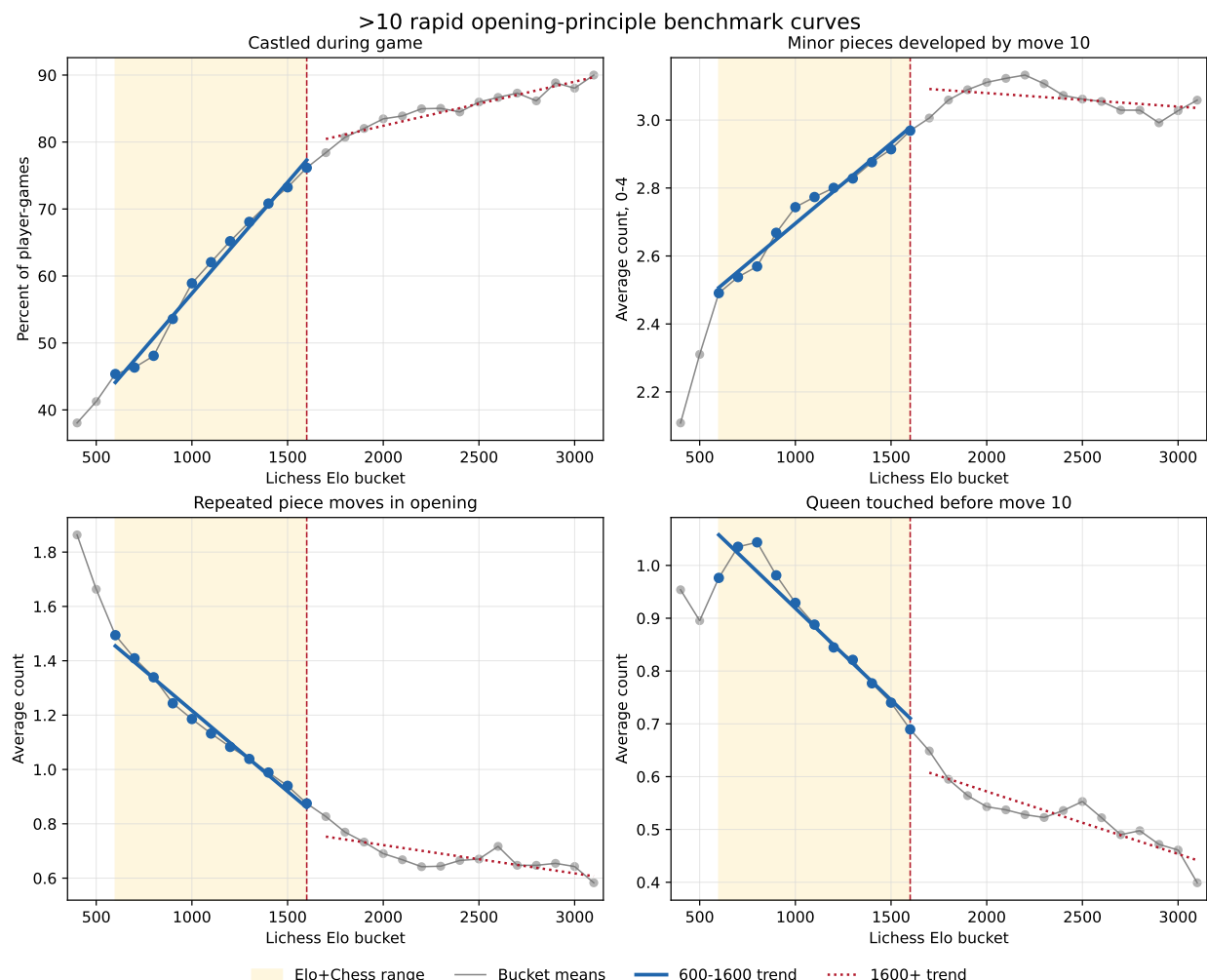


Figure 7: Mga napiling opening-principle benchmark curves para sa mas mahahabang mabilis na laro.

diagnostic. Isang malakas Maaaring maantala ng manlalaro ang castling dahil sarado ang center, ilipat ang reyna ng maaga dahil ang posisyon ay naglalaman ng isang kongkretong taktika, o ilipat ang parehong piraso nang dalawang beses dahil nanalo ito ng isang tempo o pinipilit ang isang structural concession. Ang hilaw na ugali pa ay may kahulugan, ngunit ang relasyon sa pagitan ng ugali at rating ay nagiging higit pa may kondisyon. Ito ang dahilan kung bakit gumagamit ang Elo+Chess ng upper cutoff para sa pangunahing coaching nito mga claim at kung bakit ang mga user sa itaas ng saklaw na iyon ay binabalaan na ang mahigpit na relasyon sa pagitan ng mga prinsipyong gawi at rating break down.

11 Two-Sided Tactical Sukatan at Percentile Target

Ang ilang sukatan ay may kabaligtaran na hugis mula sa mga halimbawa ng pambungad na prinsipyo. Mga tinidor ay isang kapaki-pakinabang na halimbawa dahil ang kaganapan ay likas na may dalawang panig. Isang manlalaro na nakakahanap at nagko-convert ng mas kapaki-pakinabang

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: Tinatayang bucket-curve slope para sa apat na halimbawa ng pambungad na prinsipyo mga sukatan sa mga uri ng laro ng produksyon.

na mga tinidor kaysa sa mga kapantay ay gumagawa ng isang bagay na mabuti, ngunit ang rate ng populasyon ng mga na-convert na tinidor ay bumaba sa rating dahil mas malakas pinapayagan ng mga kalaban ang mas kaunting pagkakataon sa tinidor at ipagtanggol laban sa mga taktikal na banta mas aktibo.

Ipinapakita ng Figure 8 ang epektong ito. Parehong na-convert na tinidor ni ang report player at na-convert na tinidor ng kalaban ay nagiging mas madalas tumaas ang ratings. Hindi iyon nangangahulugan na ang pag-execute ng mga tinidor ay masama. Nangangahulugan ito ng hilaw na kaganapan Ang rate ay bahagyang kinokontrol ng rate ng error ng kalaban at texture ng laro. Para sa mga sukatan tulad nito, Elo+Chess samakatuwid ay gumagamit ng peer-history percentiles sa loob ng sariling Elo bucket ng manlalaro sa halip na umasa lamang sa isang pandaigdigang bucket-mean na trend.

Upang bumuo ng mga percentile na target na ito, ang Elo+Chess ay nagsa-sample ng mga manlalaro ayon sa uri ng laro at Elo bucket, pagkatapos ay gumagamit ng kamakailang mga qualifying na laro para sa bawat sample na manlalaro. Ang kasalukuyang target na set ay gumagamit ng 11 bucket mula 600 hanggang 1600, 200 manlalaro bawat bucket, at hanggang 50 kamakailang qualifying na laro bawat manlalaro para sa bawat produksyon uri ng laro. Nagbibigay ito ng 2,200 sample ng manlalaro bawat uri ng laro at humigit-kumulang 110,000 player-game row sa bawat uri ng laro para sa within-bucket percentile mga pamamahagi.

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: Kasalukuyang peer-history percentile target set. Ang mga ito ay mga sample ng player-history na ginamit upang tantyahin ang mga within-bucket na pamamahagi para sa mga sukatan sa istilong percentile.

Two-sided converted-fork rates by Elo bucket

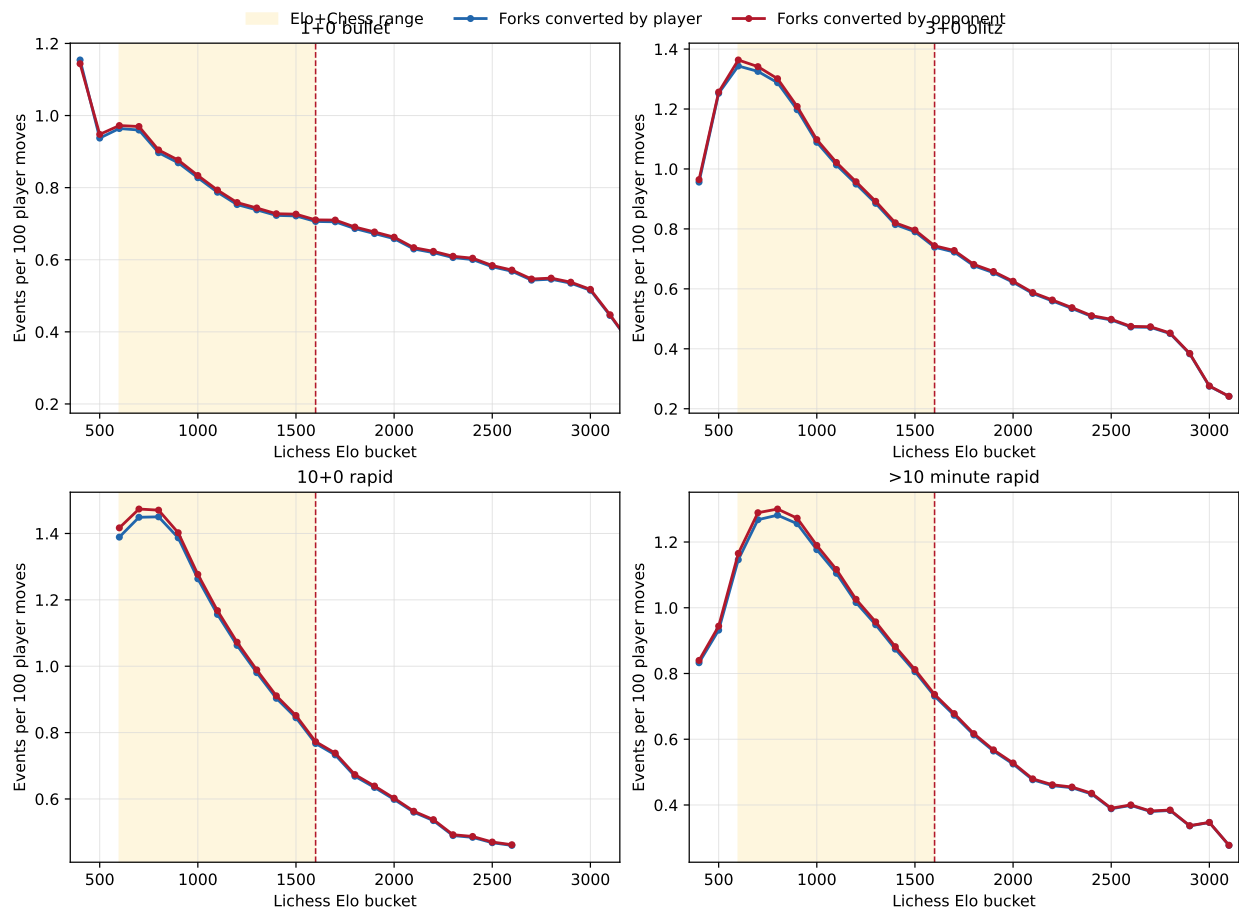


Figure 8: Two-sided converted-fork rates ng Elo bucket. Na-convert na mga kaganapan sa tinidor bumagsak sa mas mataas na mga rating dahil ang magkabilang panig ay nagtatanggol nang mas mahusay at mas kaunting malinis na tinidor magagamit ang mga conversion.

Ang Figure 9 ay naglalarawan ng resultang target pamamahagi para sa sampled-player na na-convert na fork rate sa 3+0 blitz. Ang bawat kahon ay isang pamamahagi sa loob ng bucket sa mga naka-sample na manlalaro, hindi isang bucket na ibig sabihin sa kabuuan mga indibidwal na laro. Sa UI ng ulat, inihahambing ang halaga ng isang user sa nauugnay kahon para sa nakamapang Elo bucket ng gumagamit. Ito ang batayan ng mga pahayag tulad ng kung mataas o mababa ang rate ng na-convert na fork ng manlalaro kumpara sa mga kapantay sa parehong antas, kahit na ang kabuuang rate ng kaganapan ng populasyon ay bumaba nang may rating.

12 Gastos sa Pagitiklop

Ang pamamaraan ay maaaring kopyahin sa prinsipyo, ngunit hindi ito magaan. Muling pagtatayo ang mga benchmark na curves ay nangangailangan ng:

- access sa malalaking Lichess buwanang mga file ng kasaysayan ng laro;

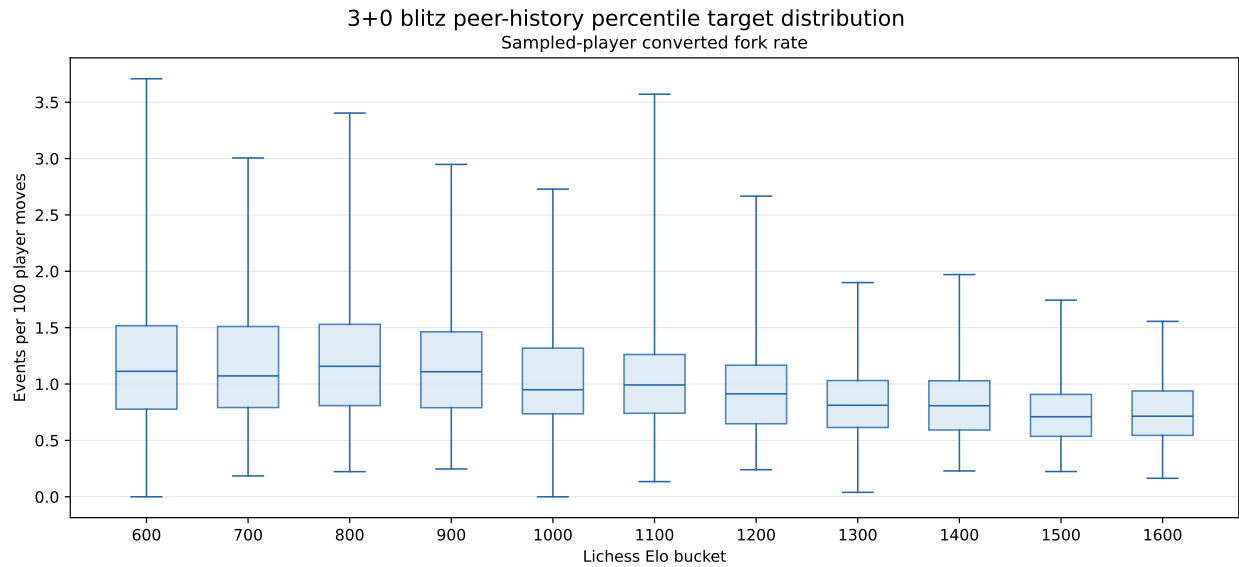


Figure 9: Halimbawa 3+0 blitz peer-history target distribution para sa sampled-player na-convert na rate ng tinidor. Ang mga porsyento ay kinukuwenta sa loob ng Elo bucket kaya ang ulat iniha-hambing ang isang manlalaro sa mga kapantay na nahaharap sa malawak na katulad na taktikal na pagkakataon mga rate.

- stratified na pag-filter ayon sa uri ng laro, bucket ng rating, at pagiging kwalipikado mga tuntunin;
- isang legal na move parser at board-state evaluator na may kakayahang palawakin sampu-sampung milyong estado ng paglipat;
- tampok na lohika para sa pagbubukas, materyal, istraktura ng pawn, kaligtasan ng hari, rook aktibidad, taktika, pamamahala ng oras, at mga resulta ng laro;
- pagsasama-sama mula sa estado ng paglipat patungo sa mga feature ng player-game;
- bucket-level na mga buod para sa bawat sukatan ng ulat;
- manu-manong pagsusuri ng bawat kurba ng benchmark ng kandidato bago ito gamitin sa isang ulat ng pagtuturo.

Para sa 3+0 blitz, nangangahulugan ito ng halos kalahating milyong laro, halos isang milyon mga hilera ng player-game, at higit sa 28 milyong mga hilera ng move-state. Sa kabila ng apat benchmark na mga uri ng laro, ang kasalukuyang benchmark na artifact ay naglalaman ng higit sa 2.55 milyong laro, higit sa 5.10 milyong hanay ng laro ng manlalaro, at higit sa 147 milyong mga hilera ng paglipat-estado. Ang iskala ay sentro ng halaga ng benchmark: hinahayaan nito ang ulat na ipakita sa mga user kung ano talaga ang nangyayari sa mga larong nilalaro ng mga manlalaro na malapit sa kanilang antas sa halip na umasa lamang sa mga pangkalahatang islogan ng chess.

13 Mga Limitasyon

Ang mga benchmark na curve ay mga empirical na paglalarawan, hindi sanhi ng patunay. Isang manlalaro ay hindi nakakakuha ng rating sa pamamagitan lamang ng pagtutugma ng ibig sabihin ng bucket. Ang mga sukatan ay din hindi perpektong buod ng mga kumplikadong posisyon: ang isang maagang paglipat ng reyna ay maaaring maging mahusay sa isang posisyon at walang ingat sa isa pa. Ang layunin ng mga kurba ay upang tukuyin ang mga umuulit na gawi at ihambing ang mga ito laban sa pag-uugali ng mga kasamahan, hindi sa palitan ang pagkalkula o paghuhusga sa pagtuturo.

Ang mga curve ay Lichess-scale curves din. Kapag ang mga gumagamit ay nagdala ng Chess.com na laro kasaysayan, ipinamapa ng Elo+Chess ang nauugnay na bucket ng rating papunta sa benchmark ng Lichess scale gamit ang hiwalay na cross-platform mapping model. Ang layer ng pagmamapa ay inilarawan sa kasamang Elo+Chess research note sa Lichess-to-Chess.com conversion ng rating: [Estimating Cross-Platform Chess Rating Mappings with Modal Regression](#). Ang pagmamapa na ito ay hindi dapat basahin bilang pagpepreserba ng percentile ranggo. Pinapanatili nito ang isang tinanyang katumbas ng rating sa pagitan ng dalawang rating mga sistema; nananatiling tukoy sa platform ang percentile rank dahil ang aktibong manlalaro iba-iba ang mga pamamahagi.

14 Mga Graphic ng Ulat sa Produksyon

Ginagawa ng ulat sa produksyon ang mga benchmark na artifact sa mga panel sa antas ng sukatan na maaaring basahin nang hindi alam ang buong build pipeline. Mga figure [10](#) and [11](#) magpakita ng tatlong halimbawa mula sa live na disenyo ng ulat ng Elo+Chess.

Ang mga panel ng pambungad na prinsipyo sa Figure [10](#) gamitin ang ordinaryong benchmark-curve presentation. Ang bawat card ay nagsisimula sa pamilya ng sukatan at pangalan ng sukatan, pagkatapos ay nagbibigay ng maikling panuntunan sa pagpapakahulugan. Ang malaking grado sa kanang sulok sa itaas ay sadyang ihiwalay sa hilaw halaga ng sukatan: sinasagot nito ang tanong na nakaharap sa gumagamit na “paano ang marka ng ugali na ito laban sa benchmark?” habang pinapanatili ng mga hilera sa ibaba ang mga nasusukat na halaga. Ang apat na buod na row ay nagpapakita ng Lifetime, Last 10, Wins, at Losses. Ginagawa nitong ang panel diagnostic sa halip na pandekorasyon lamang. Kung magkasundo ang Lifetime at Last 10, stable na siguro ang ugali. Kung magkaiba ang mga panalo at pagkatalo, ang sukatan ay maaaring nakadepende sa resulta o nakatali sa estado ng laro sa halip na isang simpleng ugali.

Gumagamit ang lugar ng tsart ng pinigilan na madilim na background, linya ng trend ng ginto, at may kulay mga marker para sa apat na hiwa. Ang anotasyon ng trend ay tahasang nagsasaad kung tumataas o bumababa ang benchmark na relasyon kay Elo. Mahalaga ito dahil ang ilang mga sukatan ay mas mataas-ay-mas mahusay, tulad ng minor-piece development, habang ang iba ay lower-is-better, gaya ng early queen movement. Ang card samakatuwid hindi hinihiling sa gumagamit na mag-isip ng direksyon mula sa slope nang mag-isa. Ang final Isinasalin ng mga linyang “Ano ang ibig sabihin nito” at “Coaching tip” ang benchmark na resulta sa isang konkretong pagtuturo ng chess habang pinapanatili ang empirical na konteksto.

Ipinapakita ng Figure [11](#) ang percentile-style presentation na ginagamit para sa mga sukatan na ang raw rate ay hindi mahusay na kinakatawan ng a iisang pandaigdigang kalakaran. Ang mga panalong tinidor ay isang taktikal na halimbawa: ang rate ng populasyon ng mga na-convert na tinidor ay nagbabago sa parehong kakayahan ng manlalaro at rate ng error ng kalaban. Ang samakatuwid, inihahambing ng card ang halaga ng user sa mga kasaysayan ng peer player sa

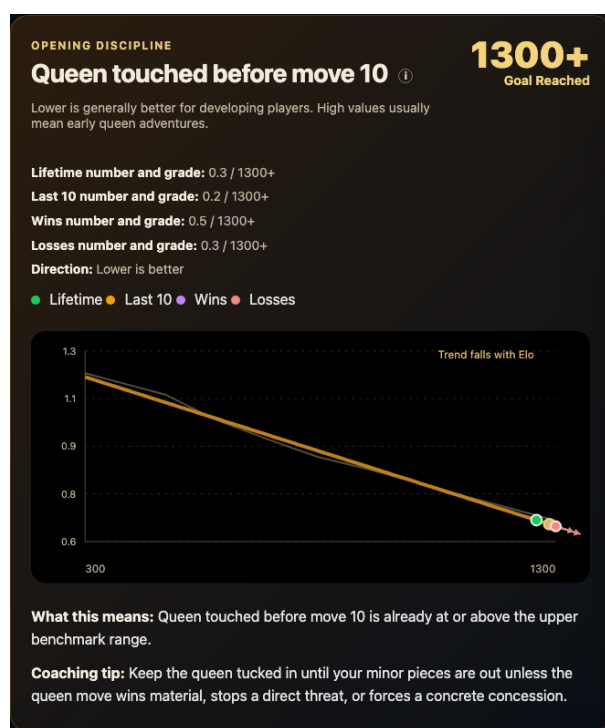
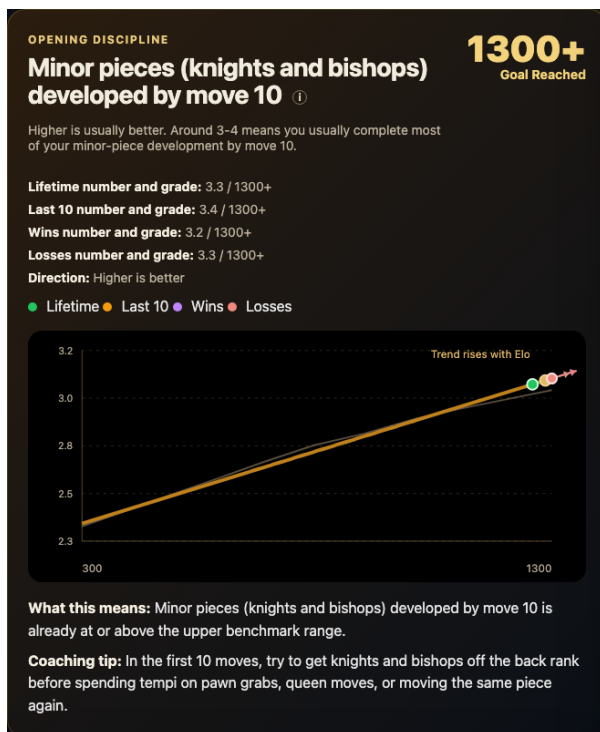


Figure 10: Mga panel ng pagbubukas-prinsipyo ng produksyon. Ang parehong visual grammar ay humahawak ng a mas mataas-ay-mas-mahusay na sukatan, maliit na piraso na pag-unlad, at mas mababa-ay-mas mahusay na sukatan, maagang kilusang reyna. Inilalantad ng parehong panel ang Panghabambuhay, Huling 10, Panalo, at Pagkatalo kaya't maaaring paghiwalayin ng ulat ang mga matatag na gawi mula sa kamakailan o umaasa sa resulta mga pattern.

nakamapang Elo bucket. Ang boxplot ay nagpapakita ng peer distribution, kasama ang interquartile range, median, at whisker. Ang marker ng user ay naka-plot sa parehong sukat at may label na may parehong raw value at percentile rank. Katabi Ang mga bucket ng paghahambing ay nana-natiling nakikita sa naka-mute na anyo upang makita ng user na ang Ang peer bucket ay lokal, hindi isang unibersal na sukat.

Ginagawang auditable ng disenyo ito ang percentile metrics. Nakikita ng user ang hilaw na halaga, ang percentile, ang peer bucket na ginamit para sa paghahambing, at ang isinalin Chess.com-to-Lichess bucket mapping. Ang parehong Lifetime, Last 10, Wins, at Pinapanatili ang mga row ng pagkalugi upang masuri pa rin ang mga marka ng taktikal na percentile para sa reency effect at result dependence.

15 Konklusyon

Ang mga ulat ng benchmark na Elo+Chess ay binuo mula sa isang malaking pagkalkula ng estado ng paglipat stratified Lichess sample. Ang bawat sukatan ay nagsisimula bilang isang set ng kongkretong board-state at mga variable ng move-state, nagiging feature ng player-game, at pagkatapos ay ina-average sa pamamagitan ng Elo bucket upang bumuo ng isang empirical benchmark curve. Ang pambungad na prinsipyo ipinapakita ng mga halimbawa kung bakit kapakipakinabang ang diskarteng ito: ang mga simpleng gawi sa chess ay may malakas, nakikitang mga relasyon sa rating sa target na hanay, habang ang mga ang mga relasyon ay madalas na patagin o

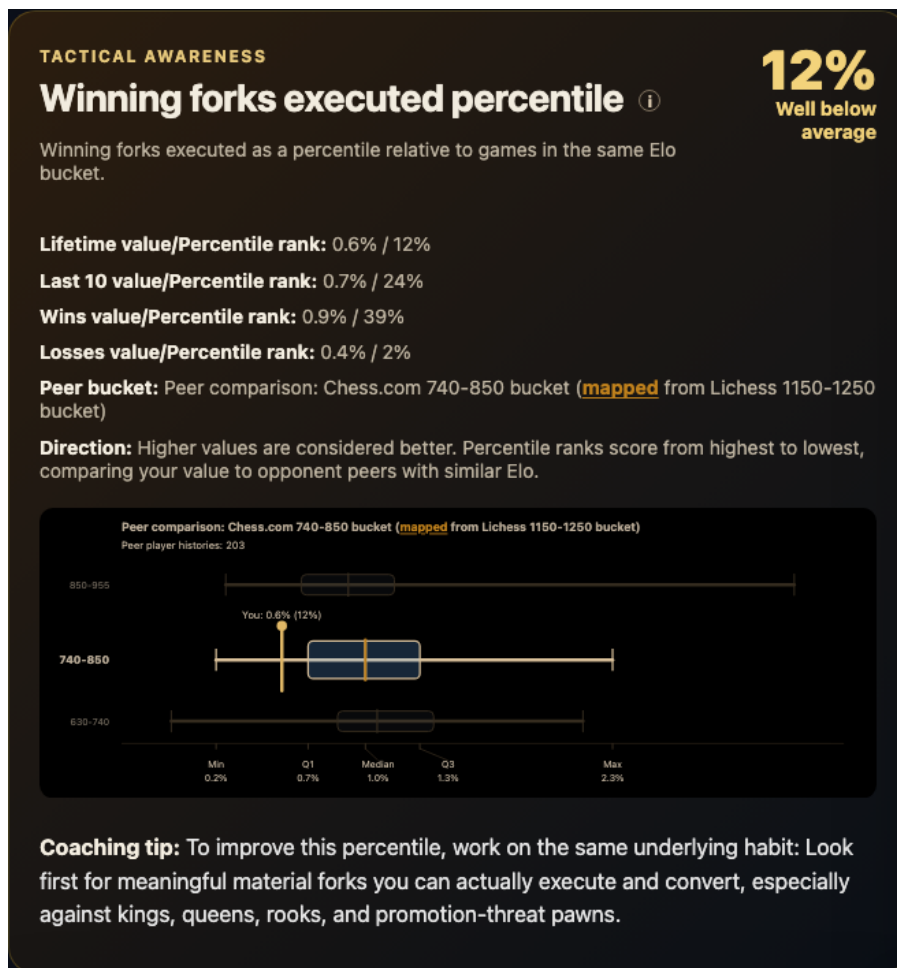


Figure 11: Naisagawa ang production percentile panel para sa mga nanalong tinidor. Ang boxplot ay nagpapakita ng parehong-bucket peer distribution na ginamit ng ulat, habang ang ginto ipinapakita ng marker ang raw value at percentile rank ng user.

nagiging mas may kondisyon sa itaas ng itaas na cutoff. Iyon ang pangunahing dahilan kung bakit itinuon ng Elo+Chess ang mga claim sa coaching nito sa rating saklaw kung saan mas malinaw na sinusuportahan ng data ang mga ito.