

# Gestratificeerde Lichess-benchmarkcurven bouwen voor Elo+Chess

Elo+Chess Onderzoeksnotities

31 mei 2026

## Abstract

Elo+Chess-benchmarkrapporten zijn opgebouwd uit grote gestratificeerde steekproeven van Lichess.org-games, waarbij het benchmarkvoorbeeld dat in deze notitie wordt gebruikt, bevat 483.974 1+0 kogelspellen, 479.974 3+0 blitzspellen, 925.636 10+0 snelle spellen, en 660.540 langere snelle spellen. Voor elk belangrijk speltype breidt de pijplijn zich uit elk spel in de bordstatussen vóór en na de zet, berekent er honderden intermediaire en speler-spelvariabelen, voegt deze variabelen samen in coachingstatistieken, en berekent vervolgens het gemiddelde van deze statistieken per Lichess Elo-bucket. De steekproef is opzettelijk gestratificeerd in plaats van proportioneel aan de ruwe speler pool: het doel is om stabiele gedragscurves op elk ratingniveau te schatten, inclusief emmers die anders schaars zouden zijn, niet om het natuurlijke te reproduceren beoordelingsverdeling. Het resultaat is een familie van empirische benchmarkcurven: voor elke statistiek kunnen we ons afvragen hoe vaak spelers op verschillende Elo-niveaus een gegeven laten zien gewoonte of principe in echte games. Deze nota beschrijft de werkwijze, de omvang van de berekening, en verschillende voorbeelden van openingsprincipes met betrekking tot rokade, ontwikkeling van kleine stukken, herhaalde bewegingen van het openingsstuk en vroege damebeweging. De nadruk ligt methodologisch: Elo+Chess gaat er niet van uit dat schaakprincipes universeel zijn lineair over alle vaardigheidsniveaus. In plaats daarvan worden benchmarkcurves geïnspecteerd door beoordelingsbereik, en het product richt zich op de regio van beginner tot vroeg-gevorderd waar de relaties het sterkst en het meest interpreteerbaar zijn.

## 1 Doel

Het doel van het benchmarksysteem is om brede schaakadviezen om te zetten in meetbare hoeveelheden. Een uitspraak als “ontwikkel je kleine stukken” is dat wel nuttig, maar het wordt praktischer als we kunnen zeggen hoeveel kleine stukjes a De speler ontwikkelt zich gewoonlijk in zet 10, hoe dat zich verhoudt tot zijn collega's in dezelfde zet ratingbucket, en hoe die gewoonte over het hele ratingspectrum verandert.

Elo+Chess bouwt daarom een metrische bibliotheek op van echte games. Elke metriek is berekend voor elke speler-spelrij, samengevat per speltype en Elo-bucket, en gebruikt als benchmarkcurve in het rapport. Het rapport kan vervolgens een gebruiker op plaatsen dezelfde curve en laat zien of de gebruiker zich boven, onder of dichtbij het gedrag bevindt waargenomen bij vergelijkbare spelers.

## 2 Brongegevens en stratificatie

De benchmarkbron is de openbare Lichess.org-speldatabase. Elo+Chess gebruikt grote gestratificeerde steekproeven per speltype en ratingbucket in plaats van een kleine gemaksmonster. De huidige productieruntime-artefacten bestrijken vier grote spelcategorieën:

- 1+0 kogel,
- 3+0 blitz,
- 10+0 snel,
- snelle spellen langer dan 10 minuten.

Het stratificatiedoel is de breedte van alle ratingniveaus. Een proportionele willekeurige Een steekproef zou efficiënt zijn voor het beschrijven van de totale spelerspool, maar dat is wel zo inefficiënt zijn voor het bouwen van benchmarkcurves. De meest voorkomende beoordelingsbuckets zouden het monster domineren, terwijl staartbakken met lage ondersteuning er te weinig zouden hebben games voor stabiele metrische gemiddelden. Elo+Chess samplet daarom op gemiddeld spel Elo-emmer. In de huidige clean-sample-constructie staan rauwe games voorop gefilterd op geldig beoordeelde games met bekende spelers, bekende beoordelingen, geldige resultaten, en bekende tijdcontroles. Games worden vervolgens gefilterd op spelers met voldoende games in die snelheid en stabiel genoeg beoordelingen om een samenhangend vaardigheidsniveau te vertegenwoordigen: voor elke speler berekent Elo+Chess de beoordelingsreeks binnen de snelheid van de speler ( $\max(\text{Elo}) - \min(\text{Elo})$ ) in de bronpool en behoudt alleen games als beide spelers ten minste het geconfigureerde minimale aantal games hebben en niet meer dan 300 Elo-punten met drift binnen de snelheid. Het driftfilter is niet bedoeld om normale verbeteringen of afwijkingen te verwijderen; het vermindert het aantal gevallen waarin een speler betrokken is games zouden wezenlijk verschillende vaardigheidsniveaus combineren, zoals voorlopige accounts, terugkerende spelers, of accounts die tijdens de periode snel door verschillende buckets zijn gegaan de bemonsterde periode. Ten slotte worden in aanmerking komende games gerangschikt binnen elke gemiddelde Elo bucket door een deterministische hash van de game-ID en het zaad, en een beperkt aantal games worden uit elke emmer bewaard.

Dit ontwerp maakt de benchmarkcurven veel nuttiger. Elke Elo-emmer draagt voldoende observaties bij om de gemiddelde waarde van elke metriek te schatten, en zeldzame emmers met een hoge of lage rating worden niet overstemd door de natuurlijke vorm van de actieve-spelerdistributie. De benchmarkcurven die door Elo+Chess worden gebruikt, zijn vervolgens beperkt tot het bereik dat het meest relevant is voor het product, grofweg beginners door vroeg geavanceerd spel. In de Lichess-schaal rapporteert dat primaire bereik ligt tussen 600 en 1600. Wanneer een gebruiker het rapport op de Chess.com-weegschaal bekijkt, wordt de bijbehorende beoordelingssegmenten worden toegewezen aan dezelfde onderliggende Lichess benchmark-buckets met behulp van de afzonderlijke platformafhankelijke rating mapping-methode beschreven in het begeleidende document: [Estimating Cross-Platform Chess Rating Mappings with Modale regressie](#).

## 3 Waarderingsverdelingen voor actieve spelers

Als we kijken naar de verdeling van de Elo-beoordelingen van Lichess.org-spelers vanaf het einde van maart 2026 zien we per belangrijk speltype de volgende aantallen en percentielen. Voor deze

verdelingsweergave wordt elke speler één keer geteld per exact speltype, gebruikmakend van de laatst waargenomen beoordeling van die speler in de overeenkomstige ruwe versie tijdcontrolelijsten, nadat er in dat exacte spel minstens vijf games nodig zijn geweest type.<sup>1</sup>

| 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: Laatst waargenomen Lichess-beoordelingsverdeling voor actieve spelers op exacte schaal speltype nadat u minimaal vijf spellen in dat exacte speltype nodig heeft.

Deze verdeling mag niet worden verward met de gestratificeerde benchmarksteekproef. De distributie beschrijft de actieve Lichess-spelerspool die in de ruwe versie wordt waargenomen scannen. De benchmarksteekproef wordt vervolgens opzettelijk gestratificeerd, zodat Elo-buckets voldoende ondersteuning hebben voor stabiele metrische curven.

Om een Lichess.org Elo-beoordeling voor een bepaald speltype toe te wijzen aan een gelijkwaardige Chess.com-beoordeling, Elo+Chess gebruikt de speltypespecifieke modale regressievergelijkingen geschat in de begeleidende tekst [Elo+Chess rating-mapping paper](#):

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

Deze vergelijkingen brengen beoordelingswaarden in kaart, geen percentielrangen. Percentielrangen zijn dat wel gedefinieerd binnen een bepaalde spelerspool, en de Chess.com- en Lichess-speler zwembaden hebben verschillende vormen. In Tabel 2, voor de Snel speltype van 10 minuten, de eerste kolom geeft een klein voorbeeld van de mogelijke spellen Chess.com Elo-beoordelingswaarden. De tweede kolom geeft het overeenkomstige weer Chess.com percentielrangschikking weergegeven op Chess.com vanaf 31 mei 2026. Een 71,7 rang op Lichess.org voor een snelle 10 minuten komt overeen met een Lichess Elo-beoordeling van 1,644<sup>2</sup> en die waarde wordt weergegeven in de derde kolom. Als we de kaartvergelijking Chess.com–Lichess toepassen, 1.644 op Lichess komt overeen met ongeveer 1.271 op Chess.com. Uitgaande van de geschatte mapping-vergelijking is redelijk, de +522 kloof tussen de Chess.com rating in kolom 1 en de in kaart gebrachte beoordeling in kolom 4 kunnen worden verklaard door een groot aantal van casual spelers in de Chess.com-spelerspool die de Chess.com opblaast percentiel rang. Met andere woorden: in het 72e percentiel op Lichess.org staan is dat wel een grotere prestatie dan in het 72e percentiel op Chess.com te staan, omdat percentielrangen worden niet duidelijk van het ene systeem naar het andere weergegeven. Echt Elo waarden kunnen in kaart worden gebracht met behulp van vergelijkingen die zijn geschat op basis van scores van dezelfde speler de twee systemen.

<sup>1</sup>We 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 toon de verdelingen met en zonder dit filter.

<sup>2</sup>Calculated 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: Illustratieve vergelijking tussen Chess.com 10 minuten snel percentiel punten en dezelfde percentiepunten in de actieve Lichess 10+0 rapid distributie na minimaal vijf 10+0-spellen. “Chess.com in kaart gebracht equiv.” geldt voor de huidige Elo+Chess 10+0 snelle conversie op beoordelingsschaal naar Lichess-beoordeling met hetzelfde percentiel. De rijen met de hoogste staart zijn opgenomen om het schaalonderscheid te tonen, maar de gepaste De beoordelingslijn is het meest betrouwbaar in het belangrijkste coachingsbereik.

De reden voor de gaten in Tabel 2 is dat de mapping schat gelijkwaardige speelsterktebeoordelingen, terwijl een percentiel beschrijft positie binnen de eigen populatie van een platform. Een speler met het 71,7e percentiel Chess.com is daarom niet automatisch een speler met het 71,7e percentiel op Lichess, en omgekeerd. De waargenomen cijfers komen overeen met het publiek van Chess.com snelle pool met een veel grotere massa met een lagere rating of gewone spelers, terwijl de actieve exacte controle Lichess-pool is meer geconcentreerd onder ervaren herhalingen spelers. Het belangrijkste punt is niet dat één percentielschaal ‘goed’ is en de ander is “fout”; het is dat de percentielrangschikking niet overdraagbaar is tussen platforms tenzij de onderliggende spelersverdelingen dezelfde vorm hebben.

Note on the 1500 spike. The visible jump near 1500 is aanwezig in de gegevens en niet gecreëerd door afvlakking. In de laatste waargenomen beoordelingsverdeling, exact 1500 beoordelingen zijn goed voor 13,1

kogelspelers, 5,6

17,6

startend of voorlopig anker in het ratingsysteem.

De piek van exact 1500 is grotendeels een artefact van licht actieve accounts. Aan controleer dit, we hebben dezelfde laatste beoordelingsverdelingen opnieuw berekend nadat we dit hadden vereist elke speler moet minimaal vijf spellen in het exacte speltype hebben. De piek bijna verdwijnt onder dit filter: exacte 1500 beoordelingen vallen van 5,6–17,6

spelers in de verdeling voor alle spelers naar 0,2–1,2

distributie gerapporteerd in Tabel 1.

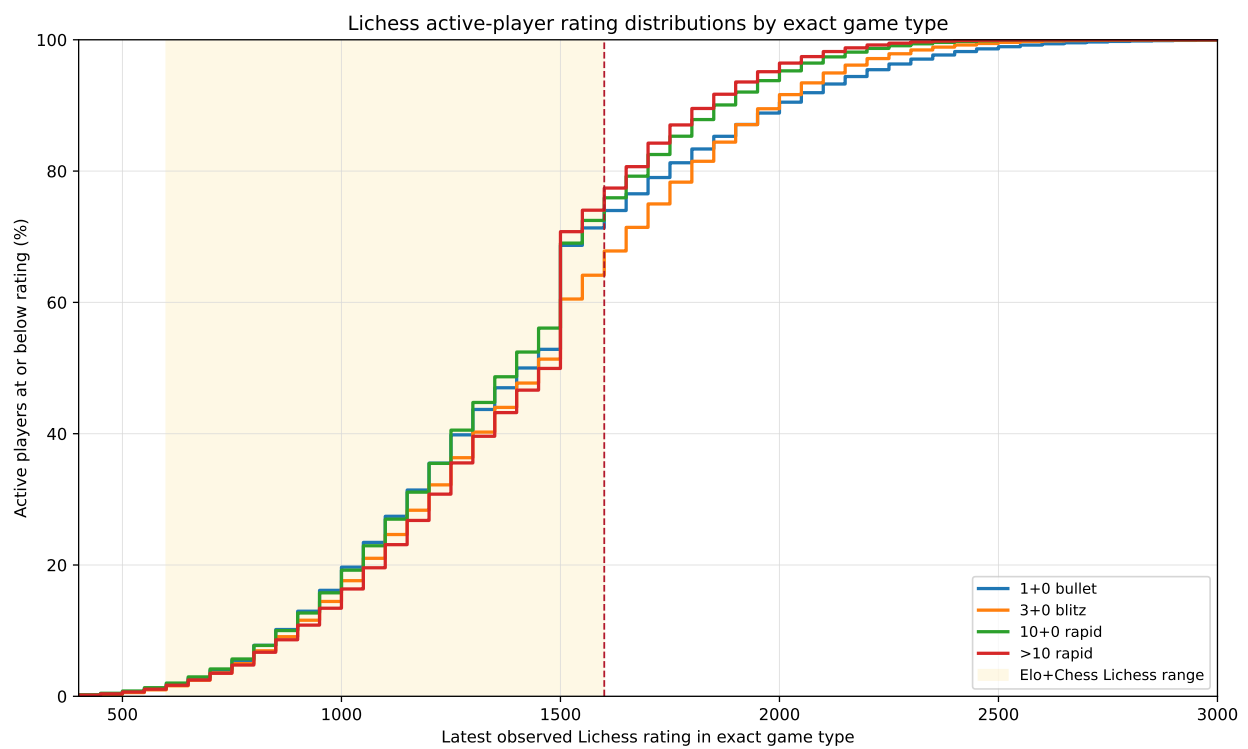


Figure 1: Cumulatieve classificatieverdelingen voor actieve spelers per exact speltype. De gele band markeert de 600–1600 Lichess-reeks die wordt gebruikt door de belangrijkste Elo+Chess benchmarks voor coaching.

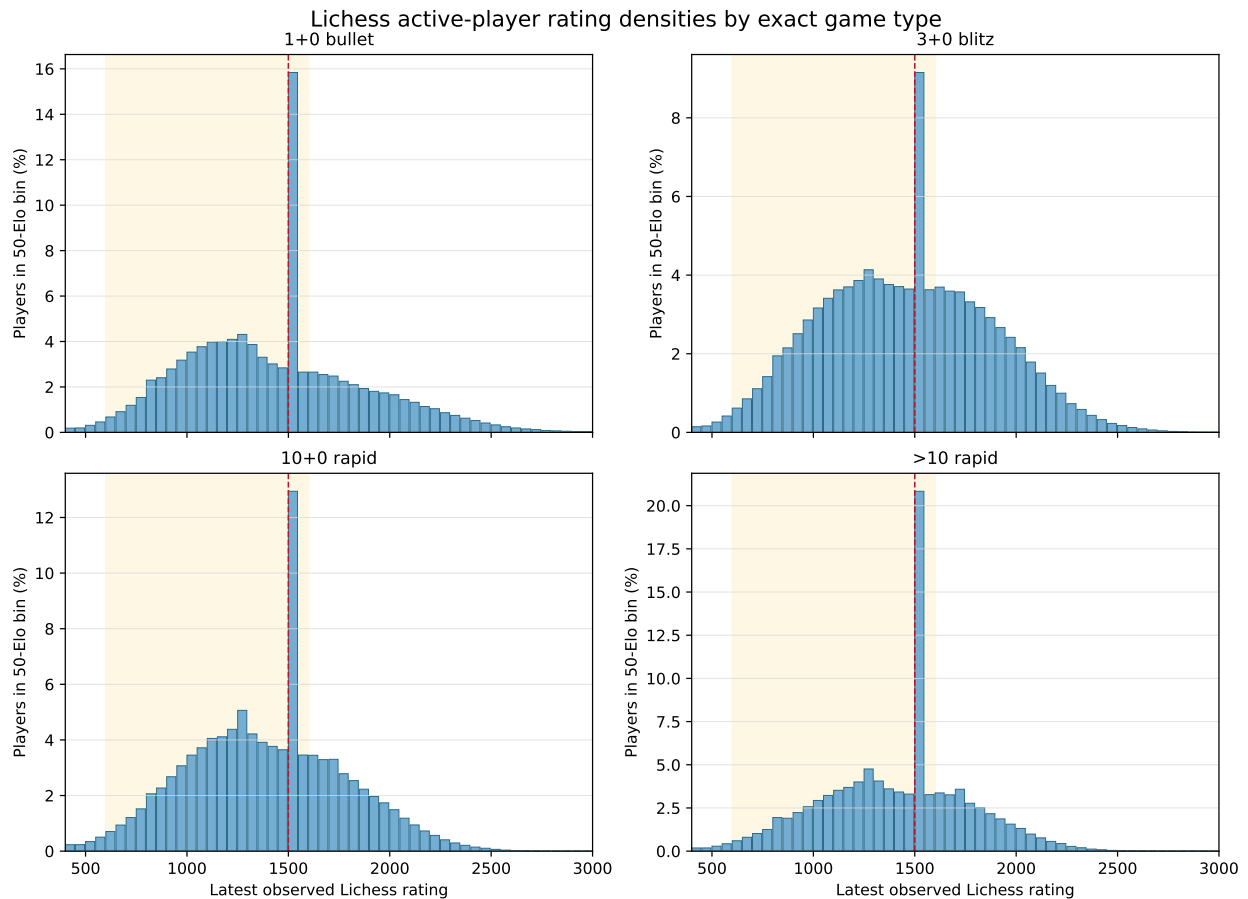


Figure 2: Dichtheidsweergave van dezelfde classificatieverdelingen voor actieve spelers. De stip-pellijn markeert 1500 Lichess, waar de onbewerkte nieuwste beoordelingsverdelingen zijn een groot massapunt met exacte beoordeling. De gele band markeert de 600–1600 Lichess bereik dat wordt gebruikt door de belangrijkste Elo+Chess-metrische rapporten. De piek is het meest zichtbaar in de dichtheid van alle spelers, omdat veel licht actieve accounts precies op hetzelfde niveau blijven 1500; Figuur 4 herhaalt de dichtheidsplot na het vereisen van ten minste vijf games in het exacte speltype, wat grotendeels wordt verwijderd dit artefact.

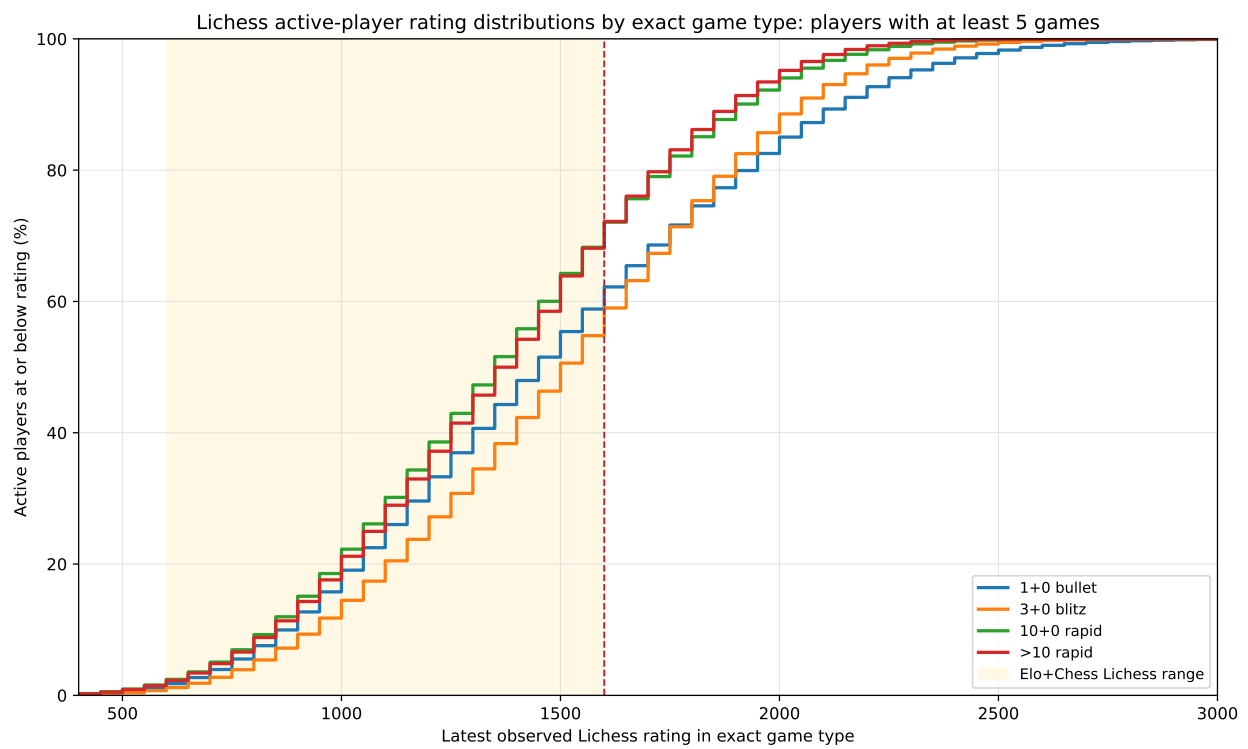


Figure 3: Cumulatieve classificatieverdelingen voor actieve spelers na vereiste at minimaal vijf games in het exacte speltype. De sprong van 1500 is sterk verminderd.

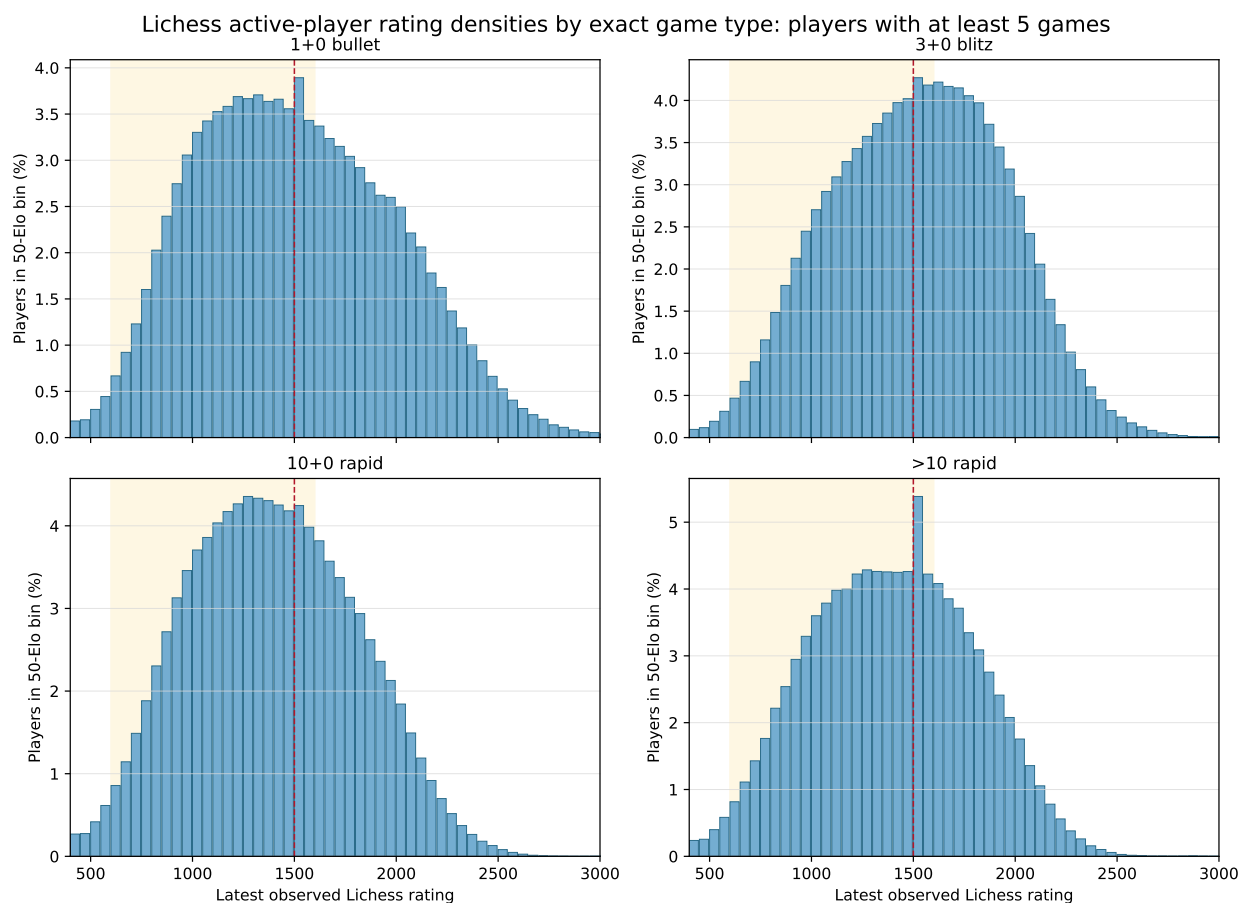


Figure 4: Dichtheidsweergave na minimaal vijf games in het exacte spel typ. Deze visie bevestigt dat het grote massapunt van 1500 in de all-player de distributie wordt grotendeels bepaald door licht actieve accounts. De gele band markeert de 600–1600 Lichess-reeks die wordt gebruikt door de belangrijkste Elo+Chess-metrische rapporten.

## 4 Schaal van de huidige benchmarkartefacten

De berekening is opzettelijk groot gemaakt. Het is geen profielschraap, een klein set gedownloade PGN's, of een met de hand samengestelde verzameling leerzame voorbeelden. De initiële Lichess-pas dekte volledige maandbestanden vanaf januari 2025 tot en met maart 2026. De niet-gestratificeerde scantabel bevat 1.382.178.087 spellen en 2.764.356.174 rijen met speler-gamebeoordelingen. In deze rijen zijn er 6.623.438 uniek spelersgebruikersnamen verschijnen minstens één keer.

| 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: Niet-gestratificeerde Lichess-scanschaal vóór benchmark-sampling. De scan omvat volledige maandelijkse gamesgeschiedenisbestanden van januari 2025 tot en met maart 2026.

De productiebenchmarkartefacten zijn veel kleiner dan de onbewerkte scan, omdat ze zijn geselecteerd op metrische constructie. Ze zijn nog steeds groot: de maatstaf fase is een uitbreidings- en aggregatiepijplijn van honderden toestanden duizenden spellen per speltype.

| 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: Schaal van de gestratificeerde benchmarkartefacten na afstemming op de algemeen runtime-functieschema met 277 kolommen. EEN 'speler-spelrij' betekent één spel vanuit het perspectief van één speler, dus de meeste spellen twee speler-spelrijen produceren.

Alleen al voor 3+0 blitz gebruikt de huidige productiebenchmark 479.974 games met 129.398 unieke spelers en 28.138.904 verschillende bewegingsstatusrijen. De De uitgelijnde blitzspeler-spelfunctietabel bevat 277 afgeleide variabelen. De snelle feature-build-artefacten zijn zelfs nog breder in de move-state-fase: het volledige rapid move-state tabel gebruikt tijdens de feature-constructie bevat 168 move-state-kolommen vóór aggregatie in een speler-spelfunctie met 277 kolommen tafel. De frontend runtime-benchmarkartefacten voor alle vier de speltypen zijn dat wel vervolgens uitgelijnd met hetzelfde speler-game-functieschema met 277 kolommen, inclusief de uitgebreide tactiek-taxonomie breakout-variabelen. Van deze functietabellen wordt de rapport selecteert en toont de coachingstatistieken die het meest interpreteerbaar zijn.

These counts are important for interpreting the results. The benchmark curves are not estimates from a few hundred manually reviewed games. Instead, they come from tens of millions of evaluated board states and more than four million player-game rows across the four current production game categories.

## 5 Pijpleidingoverzicht

De benchmark-bouwpijplijn bestaat uit zes belangrijke stappen:

1. Selecteer beoordeelde Lichess-spellen in het doelspeltype en de beoordelingstrata.
2. Analyseer elke partij, zet voor zet, en sla de voor- en nazet op board state.
3. Bereken tussenliggende bewegingsvariabelen van het bord, de zet, het perspectief van de speler en het perspectief van de tegenstander.
4. Voeg bewegingsstatusvariabelen samen in variabelen voor spelerspelfuncties.
5. Converteer speler-gamevariabelen naar rapportstatistieken met een duidelijke interpretatierichting: hoger is meestal beter, lager is meestal beter of beschrijvend.
6. Bereken het gemiddelde van elke statistiek per Elo-bucket en inspecteer het resultaat handmatig relatie.

Deze structuur is belangrijk omdat veel rapportstatistieken niet direct zijn zichtbaar in een PGN-header. Ze vereisen het reconstrueren van de positie, het bepalen ervan welke stukken vanaf welke startvelden zijn verplaatst, vergelijking vóór zet en materiaal en structuur na de verplaatsing, en vervolgens het resultaat van de aggregatie verplaats niveau naar het speler-speelniveau.

## 6 Tussenliggende variabelen

De tussenliggende variabelen zijn bedoeld om te beschrijven wat er op het bord is veranderd en waarom het belangrijk is vanuit het perspectief van de verslaggever. Voorbeelden zijn onder meer:

- zetnummer, zetnummer van de speler, te zetten kant, kleur van de zet en speler kleur;
- van-vierkant, naar-vierkant, bewegend stuktype, vastgelegd stuktype, en of de zet een verovering, schaak, kasteel, promotie of mat was;
- materiaalbalans vóór en na de zet;
- of een klein stuk zijn startveld met een bepaald zetnummer heeft verlaten;
- of hetzelfde stuk beweegt bij opeenvolgende spelersbewegingen in de opening;
- of de dame vóór zet 10 heeft bewogen;
- rokadezijde, rokadezetnummer en of torens verbonden raakten na rokade;
- losse stukken, hangende stukken, dubbele pionnen, geïsoleerde pionnen, geslaagd pionnen en achterwaartse pionnen;
- torenbestandsstatus, geopende bestanden, halfopen bestanden en koningsveiligheid functies;
- tactische kansen, uitgevoerde tactieken, materiële druktactieken, gegarandeerde vervolgwinsten en gerealiseerde winsten.

De exacte tussenset verschilt per bouwfase. Runtime-rapportdatabases blijven behouden alleen de kolommen die nodig zijn om rapporten snel weer te geven, terwijl de volledige functionaliteit is opgebouwd artefacten omvatten bredere move-state-tabellen en feature-patches die worden gebruikt om te construeren de laatste speler-spelfunctierijen.

## 7 Openingsprincipe metrische voorbeelden

Openingsprincipes zijn nuttige voorbeelden omdat ze voor gebruikers gemakkelijk te gebruiken zijn begrijpen en gemakkelijk te verbinden met move-state berekeningen.

### 7.1 Minor Pieces Developed by Move 10

Deze metriek telt hoeveel van de vier ridders en lopers van een speler er nog over zijn hun thuisveld door de 10e zet van de speler. De tussenberekening gebruikt de kleur van de verplaatsing, het type verplaatsingsstuk, het startveld en de verplaatsing van de speler nummer. Voor wit zijn de thuisvelden b1, g1, c1 en f1. Voor zwart zijn ze dat wel b8, g8, c8 en f8. Voor elk spelersspel telt Elo+Chess het afzonderlijke huis vierkanten tussen ridders en lopers die bewogen zijn door speler zet 10.

Dit is een hogere-is-meestal-beter-statistiek. Er wordt niet beweerd dat elke minderjarige het stuk moet altijd per zet 10 worden verplaatst, maar bij zich ontwikkelende spelers is dit empirisch De relatie is duidelijk: spelers die hogere Elo-buckets bereiken, hebben de neiging zich te ontwikkelen meer kleine stukken eerder.

### 7.2 Repeated Piece Moves in the Opening

Deze statistiek telt extra vroege zetten die zijn besteed aan het opnieuw verplaatsen van hetzelfde stuk van het ontwikkelen van iets nieuws. De tussentijdse berekening maakt gebruik van de voorgaande bestemmingsveld, het huidige oorsprongsveld en het zetnummer van de speler. Als hetzelfde stuk werd verplaatst tijdens de zet van de vorige speler en wordt opnieuw verplaatst binnen de eerste 10 zetten van de speler, neemt het aantal herhaalde zetten toe.

This is a lower-is-usually-better metric. It measures wasted opening tempi in a simple way. Some repeated moves are tactically justified, but at scale the bucket curve captures the general tendency: stronger players at the target levels spend fewer early moves moving the same piece repeatedly.

### 7.3 Queen Touched Before Move 10

Deze metriek telt de zetten van de dame vóór speler 10. De tussenliggende De berekening maakt gebruik van het type bewegend stuk en het zetnummer van de speler. Als het bewegende stuk de dame is en het zetnummer van de speler maximaal 10 is, neemt het aantal toe.

Dit is ook lager-is-meestal-beter voor het Elo+Chess-doelbereik. Vroeg koninginzetten kunnen materiaal winnen of concessies afdwingen, maar veel ontwikkelende spelers breng de koningin te vroeg naar buiten, verlies tempi door aanvallen op de koningin en stel uit ontwikkeling van kleine stukken en koningsveiligheid.

## 7.4 Castling and King Safety

De rokadestatistieken maken gebruik van rokadevlaggen, rokadezijde, rokadezetnummer en de bordstatus na de rokade. Voorbeelden hiervan zijn onder meer of de speler heeft gecastreed, of de speler rokeerde op zet 10, of de speler de koningsvleugel rokeerde of damevleugel, of torens verbonden waren na rokade, en of torens fileren waren na het rokeren open of halfopen.

Deze statistieken worden niet allemaal op dezelfde manier geïnterpreteerd. Voltooiing van rokeren en Rokeren met zet 10 ligt over het algemeen hoger-is-meestal-beter in het doelbereik. Rokeren aan de koninginzijde of rokeren in open bestanden kan meer contextafhankelijk zijn. Dat is de reden waarom Elo+Chess niet simpelweg een morele regel hardcodeert; het bouwt een emmer rondt en inspecteert welke relaties stabiel genoeg zijn voor coaching.

## 8 Bakcurven en handmatige inspectie

Nadat de speler-gamestatistieken zijn berekend, groepeerde Elo+Chess ze op Lichess Elo bucket en berekent bucketgemiddelden. Het bucketgemiddelde is het ruwe empirische benchmarkpunt voor die statistiek. In het rapport wordt vervolgens een trendlijn gebruikt wijs cijfers in Elo-stijl toe aan gebruikerswaarden.

De bochten worden handmatig geïnspecteerd voordat ze worden gebruikt. Een statistiek is nuttiger wanneer de bucket-middelen een samenhangend verband vertonen over het beoordelingsbereik geserveerd door het product. Een metriek is minder nuttig als deze luidruchtig, vlak of hoog is niet-monotoon, of gedreven door speciale gevallen die moeilijk uit te leggen zijn aan een gebruiker.

Deze handmatige inspectiestap is vooral belangrijk bij het bovenwerk Elo+Chess afsluiting. Sterkere spelers weten wanneer ze de openingsprincipes moeten overtreden en doen dat ook effectief. Boven het bereik van beginner tot vroeg-gevorderd, de relatie tussen eenvoudige gewoonten en waardering vlakkt vaak af of verandert van vorm. Het product benadrukt daarom het bereik waarin de empirische relaties sterk zijn en pedagogisch bruikbaar.

## 9 Voorbeeld van openingscurven over het volledige bereik

Figuur 5 toont vier 3+0 blitz-openingsprincipecurven. Grijs punten tonen het bredere Lichess-beoordelingsbereik. Blauwe punten en de blauwe trend lijn toont de 600–1600 Lichess-reeks die wordt gebruikt door de belangrijkste Elo+Chess-coaching benchmarks. De rode stippellijn markeert de bovenste grenswaarde.

Hetzelfde patroon is zichtbaar in de 10+0 rapid en de langere-rapid met volledig bereik artefacten. De centrale modelleringsbeslissing is ongewijzigd: gebruik empirisch stabiel bereik voor coachingclaims en vermijd te doen alsof het een eenvoudige gewoontecurve is blijft voor altijd even informatief.

Hetzelfde patroon verschijnt numeriek in Tabel 5. Hellingen zijn gerapporteerd als verandering van de metrische waarde per 100 Elo-punten. Rokade, klein stuk ontwikkeling, herhaalde stukbewegingen en vroege damebewegingen zijn allemaal duidelijk te zien beweging tussen 600 en 1600. Boven 1600 zijn de hellingen hier veel kleiner voorbeelden, wat consistent is met het idee dat spelers met een hogere rating breken basisprincipes selectiever en met meer succes. Een tweede mechanisme is verzadiging. Voor sommige gewoonten bereikt de metriek een praktisch of zoet plafond spot: sterke spelers hebben de gewoonte al aangeleerd, dus daar is geen reden

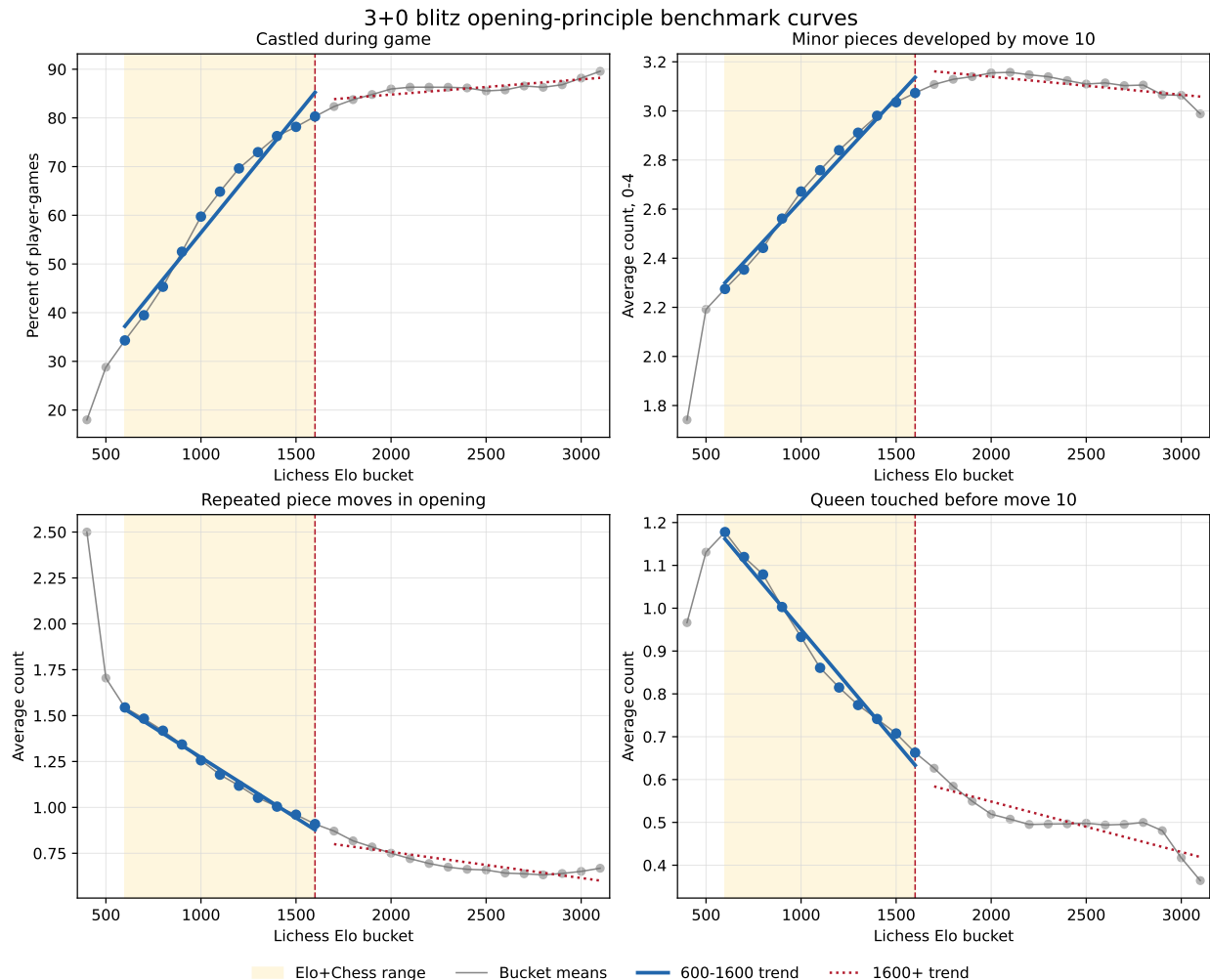


Figure 5: Geselecteerde benchmarkcurves volgens het openingsprincipe voor 3+0 blitz. De relaties zijn het sterkst en het nuttigst in het Elo+Chess coachingassortiment. Boven de bovengrens worden verschillende relaties aanzienlijk vlakker.

voor de curve scherp omhoog blijft bewegen. Meer is achteraf ook niet altijd beter dat punt. De gegevens vervlakken dus beide, omdat sterkere spelers kunnen breken regels in concrete posities en omdat veel basisingewoonten al bereikt zijn hun bruikbare bereik.

## 10 Waarom de bovenste grenswaarde belangrijk is

De doelgebruiker van de Elo+Chess is een beginnende tot vroeg-gevorderde speler. Voor deze spelers zijn basisingewoonten sterk geassocieerd met beoordeling: stukken ontwikkelen, vermijd herhaalde vroege stukzetten, vermijd onnodige vroege damezetten, kasteel Verbind op tijd torens, beheer materiaal en voorkom dat stukken los of los blijven hangend.

Bij hogere beoordelingen worden diezelfde oppervlakkige gewoonten minder diagnostisch. Een sterke de speler mag het rokeren uitstellen omdat het centrum gesloten is, de dame eerder verplaatsen omdat de stelling een concrete tactiek bevat, of hetzelfde stuk twee keer verplaatsen

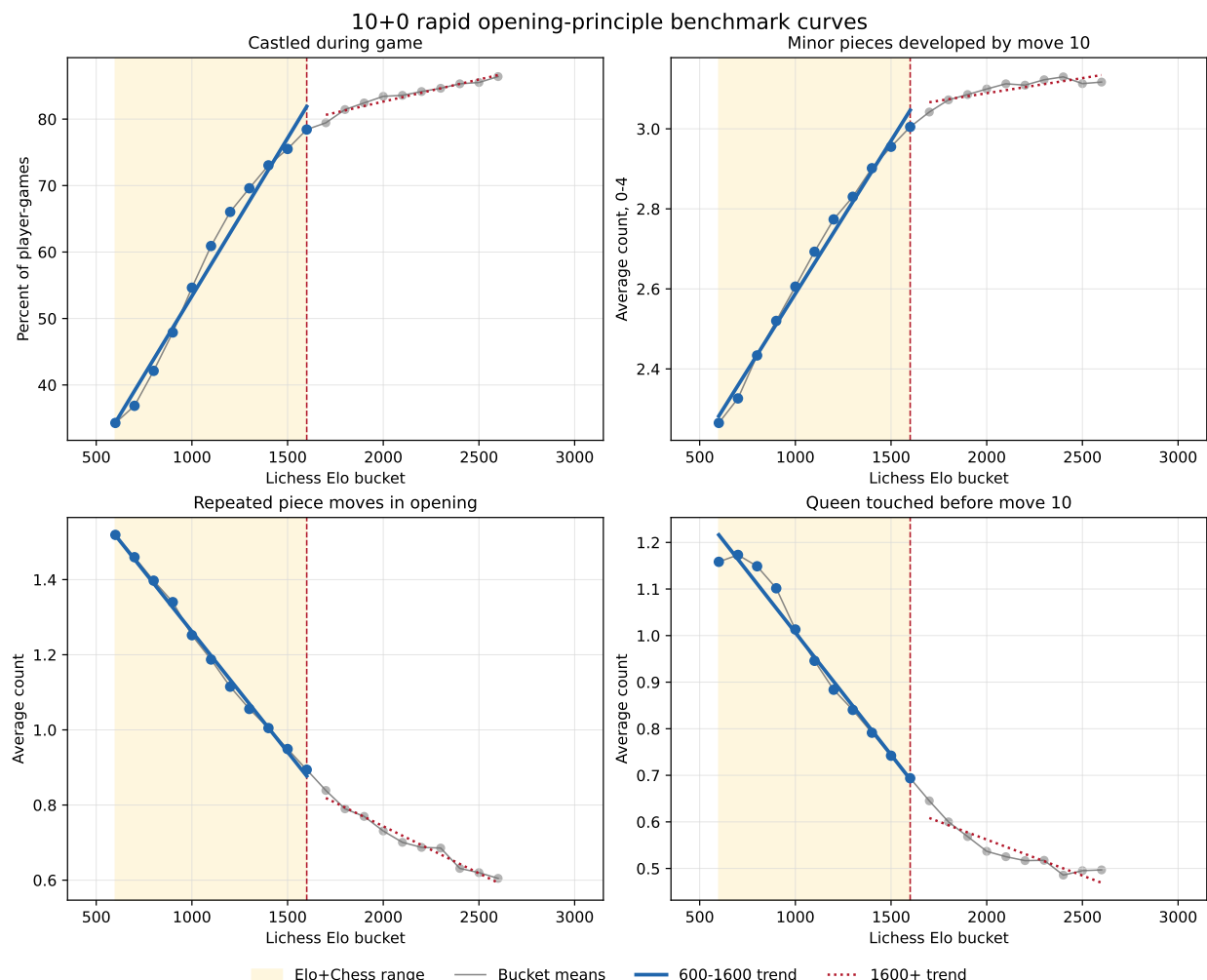


Figure 6: Geselecteerde benchmarkcurves volgens het openingsprincipe voor 10+0 snelle games. De relaties zijn het sterkst en het nuttigst in het Elo+Chess coachingassortiment. Boven de bovengrens worden verschillende relaties aanzienlijk vlakker.

omdat het een tempo wint of een structurele concessie afdwingt. Nog steeds de rauwe gewoonte heeft betekenis, maar de relatie tussen de gewoonte en de beoordeling wordt groter voorwaardelijk. Dit is de reden waarom Elo+Chess een hogere grenswaarde gebruikt voor de belangrijkste coaching claims en waarom gebruikers boven dat bereik worden gewaarschuwd voor de nauwe relatie tussen principiële gewoonten en waardering verdwijnt.

## 11 Tweezijdige tactische statistieken en percentieldoelstellingen

Sommige metriekeken hebben de tegenovergestelde vorm van de voorbeelden van het openingsprincipe. Vorken zijn een nuttig voorbeeld omdat de gebeurtenis inherent tweezijdig is. Een speler die nuttiger vorken vindt en converteert dan collega's doet iets goeds, maar het populatiepercentage van omgebouwde vorken daalt met de beoordeling omdat ze sterker zijn tegenstanders laten minder vorkkansen toe en verdedigen zich tegen tactische bedreigingen actiever.

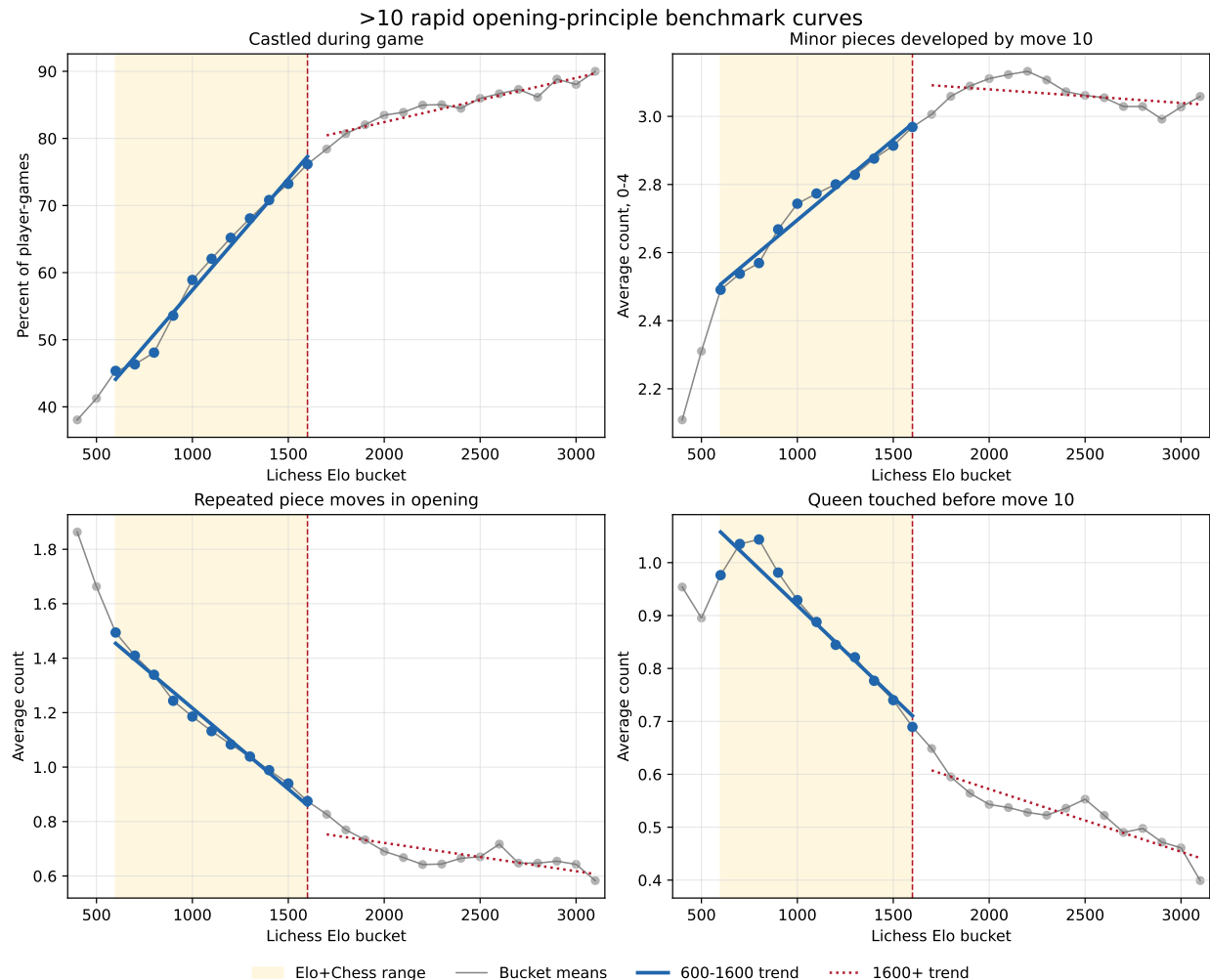


Figure 7: Geselecteerde benchmarkcurves volgens het openingsprincipe voor langere, snelle games.

Figuur 8 toont dit effect. Beide omgebouwde vorken door de rapportspeler en geconverteerde vorken door de tegenstander komen minder vaak voor kijkcijfers stijgen. Dat betekent niet dat het uitvoeren van vorken slecht is. Het betekent de rauwe gebeurtenis Het percentage wordt gedeeltelijk bepaald door het foutenpercentage van de tegenstander en de speltextuur. Voor statistieken op deze manier gebruikt Elo+Chess daarom peer-geschiedenis percentielen binnen de de eigen Elo-bucket van de speler in plaats van alleen te vertrouwen op een mondiale bucket-mean-trend.

Om deze percentiëldoelstellingen te bepalen, bemonstert Elo+Chess spelers op speltype en Elo bucket, gebruikt vervolgens recente kwalificatiewedstrijden voor elke in de steekproef opgenomen speler. De huidige doelset gebruikt 11 buckets van 600 tot en met 1600, 200 spelers per bucket en maximaal 50 recente kwalificatiewedstrijden per speler voor elke productie speltype. Dit levert ongeveer 2.200 spelersvoorbeelden per speltype op 110.000 speler-spelrijen per speltype voor binnen-bucket-percentiel distributies.

Figuur 9 illustreert het resulterende doel distributie voor gesamplede speler geconverteerde vorksnelsnelheid in 3+0 blitz. Elke doos is een verdeling binnen de bucket over de spelers in de steekproef, en niet een gemiddelde over de hele bucket individuele spellen. In de rapport-UI wordt

| 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: Geschatte hellingen van de bakcurve voor het vier voorbeeldopeningsprincipe statistieken voor productiegametypen.

| 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: Huidige percentieldoelstellingen voor de peer-geschiedenis. Dit zijn voorbeelden van spelersgeschiedenis die worden gebruikt om distributies binnen buckets te schatten statistieken in percentielstijl.

de waarde van een gebruiker vergeleken met de relevante waarde vak voor de in kaart gebrachte Elo-bucket van de gebruiker. Dit is de basis voor uitspraken als of de geconverteerde fork-rate van een speler hoog of laag is in vergelijking met zijn collega's op de op hetzelfde niveau, zelfs als het totale aantal gebeurtenissen in de populatie afneemt met de beoordeling.

## 12 Replicatiekosten

De methode is in principe reproduceerbaar, maar niet licht van gewicht. Wederopbouw de benchmarkcurven vereisen:

- toegang tot grote Lichess maandelijkse gamegeschiedenisbestanden;
- gestratificeerde filtering op speltype, ratingbucket en geschiktheid regels;
- een legal move parser en board-state evaluator die in staat is uit te breiden tientallen miljoenen bewegingsstaten;
- voorzien van logica voor openen, materiaal, pionnenstructuur, koningsveiligheid, toren activiteit, tactiek, tijdmanagement en spelresultaten;

Two-sided converted-fork rates by Elo bucket

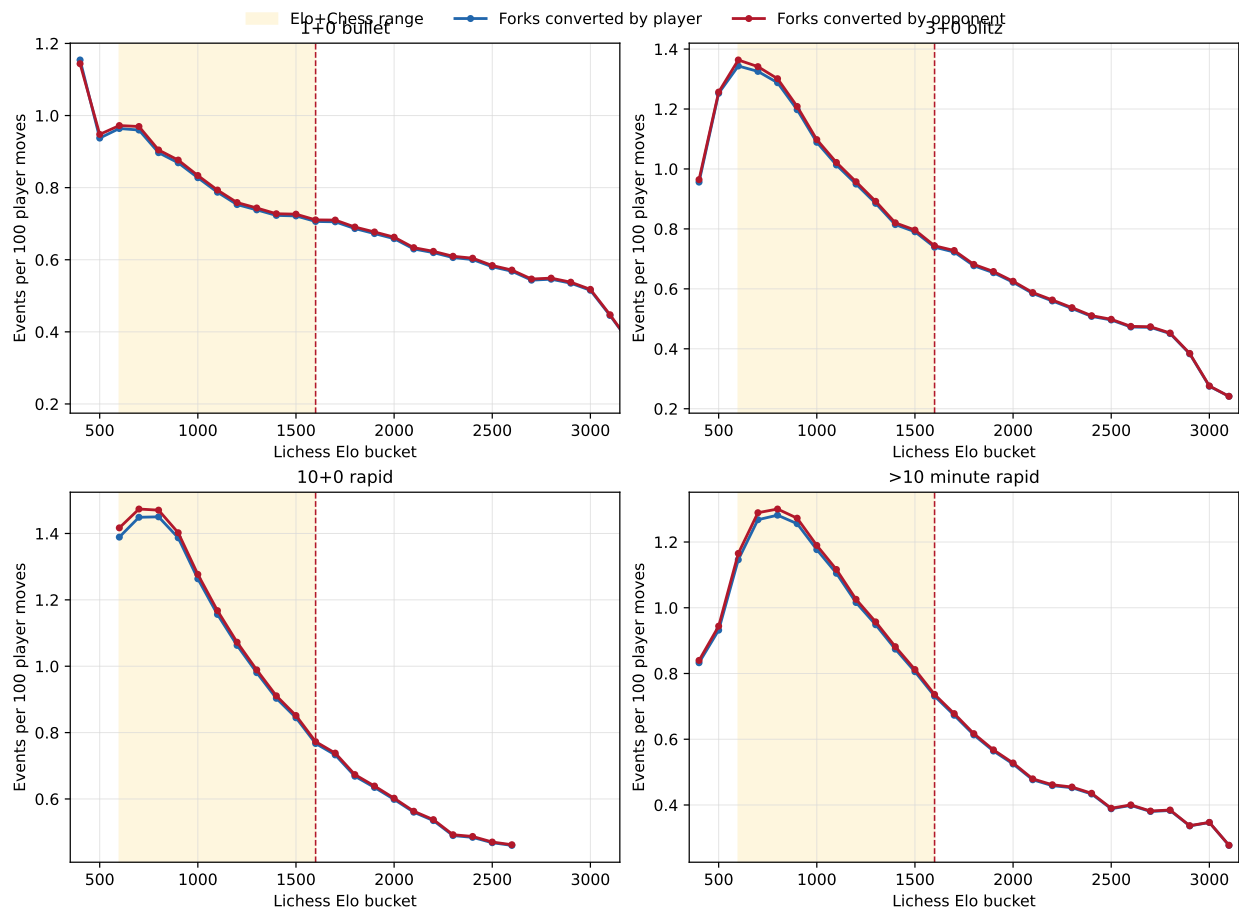


Figure 8: Dubbelzijdige omgezette vorken door Elo-bak. Geconverteerde vorkgebeurtenissen dalen bij hogere beoordelingen omdat beide partijen beter verdedigen en minder schone vorken conversies zijn beschikbaar.

- aggregatie van verplaatsingsstatus naar speler-spelfuncties;
- samenvattingen op bucketniveau voor elke rapportstatistiek;
- handmatige beoordeling van elke kandidaat-benchmarkcurve voordat deze wordt gebruikt een coachrapport.

Voor 3+0 blitz betekent dit bijna een half miljoen games, bijna een miljoen speler-game-rijen en meer dan 28 miljoen move-state-rijen. Over de vier benchmark-speltypen bevatten de huidige benchmark-artefacten meer dan 2,55 miljoen spellen, meer dan 5,10 miljoen speler-spelrijen en meer dan 147 miljoen bewegingsstatusrijen. De schaal staat centraal in de waarde van de benchmark: hiermee kan het rapport gebruikers laten zien wat er daadwerkelijk gebeurt in gespeelde games door spelers die dichtbij hun niveau liggen, in plaats van alleen te vertrouwen op algemene schaakslogans.

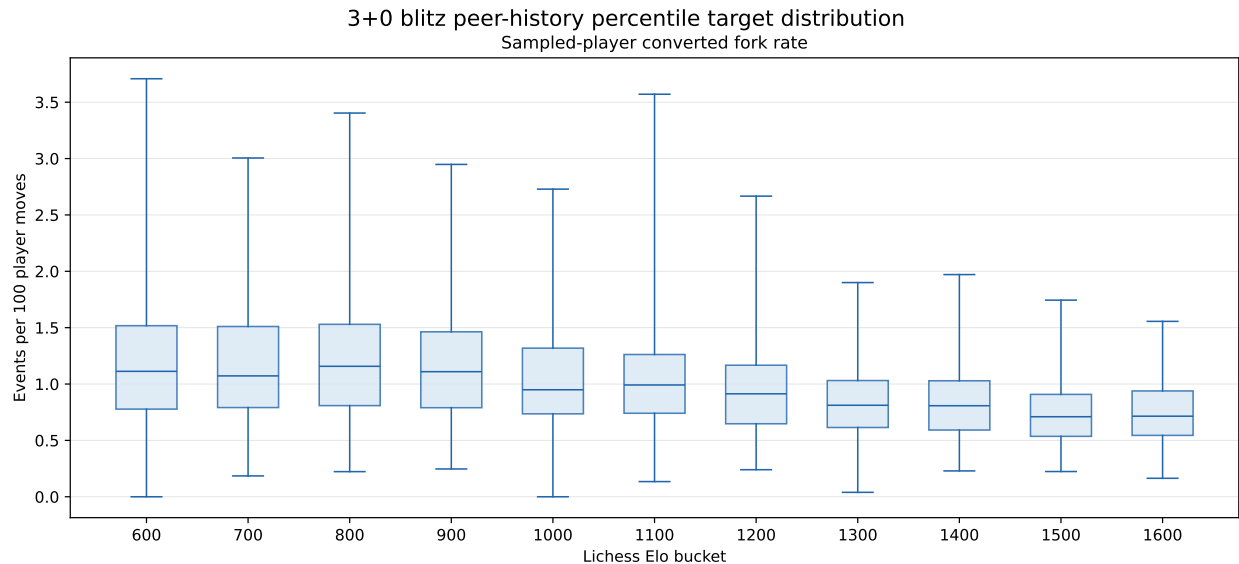


Figure 9: Voorbeeld 3+0 blitz peer-geschiedenis doeldistributie voor gesampled speler omgerekende vorksnelheid. Percentielen worden berekend binnen de Elo-bucket, aldus het rapport vergelijkt een speler met leeftijdsgenoten die in grote lijnen dezelfde tactische kansen hebben tarieven.

## 13 Beperkingen

De benchmarkcurves zijn empirische beschrijvingen en geen causaal bewijs. Een speler krijgt geen rating alleen door het matchen van een bucket-gemiddelde. Metrieken zijn dat ook onvolmaakte samenvattingen van complexe stellingen: een vroege damezet kan uitstekend zijn in de ene positie en roekeloos in de andere. Het doel van de curven is om te identificeren terugkerende gewoonten en vergelijk ze met het gedrag van leeftijdsgenoten, en niet met dat van anderen vervanging van berekening of coachend oordeel.

De curven zijn ook curven op Lichess-schaal. Wanneer gebruikers het Chess.com-spel meenemen geschiedenis brengt Elo+Chess de relevante ratingbucket in kaart op de Lichess-benchmark schaal met behulp van het afzonderlijke platformafhankelijke kaartmodel. Die kaartlaag is dat wel beschreven in de begeleidende Elo+Chess-onderzoeksnotitie over Lichess-naar-Chess.com beoordeling conversie: [Estimating Cross-Platform Chess Rating Mappings with Modale regressie](#). Deze mapping mag niet worden gelezen als behoud van het percentiel rang. Het behoudt een geschatte gelijkwaardigheid tussen de twee beoordelingen systemen; percentielrang blijft platformspecifiek omdat het de actieve speler is distributies zijn verschillend.

## 14 Grafische productierapporten

Het productierapport zet de benchmarkartefacten om in panelen op metrisch niveau die kunnen worden gelezen zonder de volledige build-pijplijn te kennen. Cijfers 10 and 11 laat drie voorbeelden zien van het live Elo+Chess-rapportontwerp.

De panelen met het openingsprincipe in Figuur 10 gebruik de gewone benchmark-curvepresentatie. Elke kaart begint met de metrische familie en metrische naam, en geeft vervolgens een korte interpretatieregel. De grote kwaliteit in de rechterbovenhoek is opzettelijk gescheiden van de onbewerkte versie metrische waarde: het beantwoordt de vraag van de gebruiker: 'Hoe scoort deze

gewoonte?’ ten opzichte van de benchmark?’ terwijl de rijen eronder de gemeten waarden behouden. De vier overzichtsrijen tonen levensduur, laatste 10, overwinningen en verliezen. Dit maakt de paneeldiagnostisch in plaats van louter decoratief. Als Lifetime en Last 10 het eens zijn, de gewoonte is waarschijnlijk stabiel. Als winsten en verliezen uiteenlopen, kan de maatstaf dat zijn resultaatafhankelijk of gebonden aan de spelstatus in plaats van een eenvoudige gewoonte.

Het grafiekgebied heeft een ingetogen donkere achtergrond, een gouden trendlijn en is gekleurd markeringen voor de vier plakjes. In de trendannotatie wordt expliciet aangegeven of de benchmarkrelatie stijgt of daalt met Elo. Dit is belangrijk omdat sommige statistieken zijn hoger-is-beter, zoals de ontwikkeling van kleine stukjes, terwijl andere zijn lager-is-beter, zoals vroege koninginnenbeweging. De kaart dus vraagt de gebruiker niet om de richting alleen uit de helling af te leiden. De finale De lijnen ‘Wat dit betekent’ en ‘Coachingtip’ vertalen het benchmarkresultaat tot een concrete schaakinstructie met behoud van de empirische context.

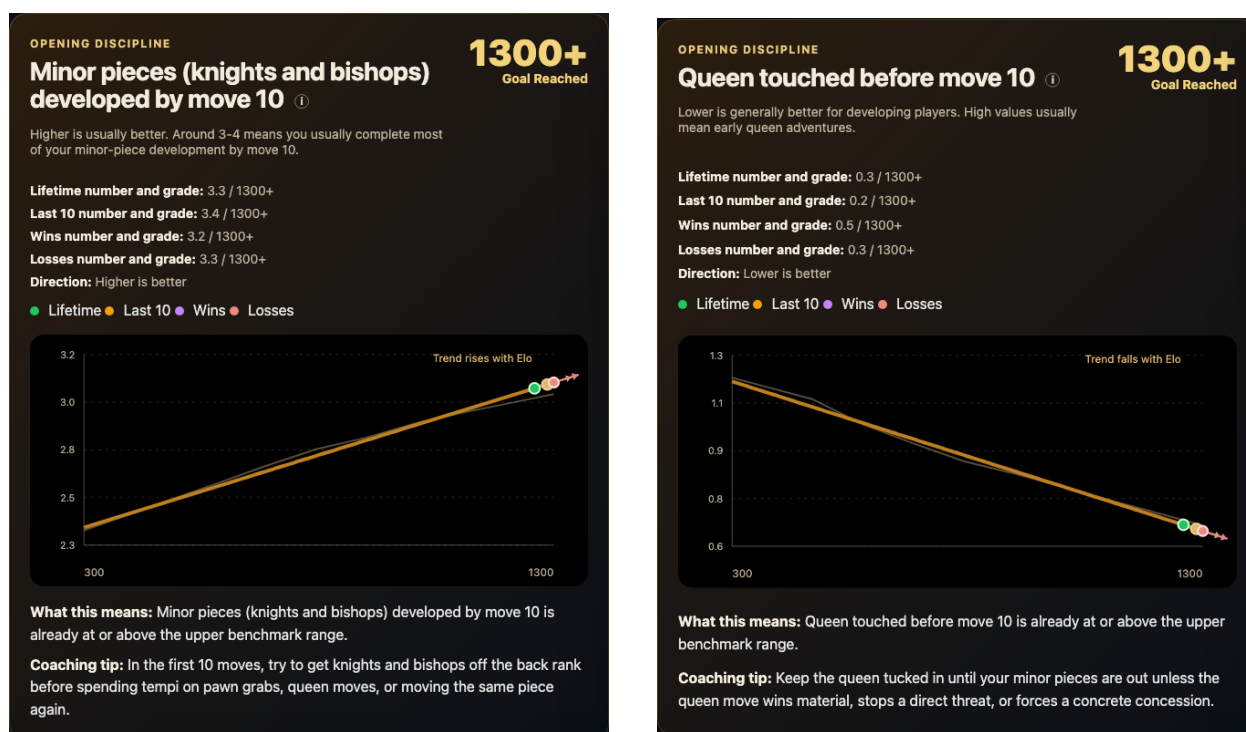


Figure 10: Productie-openingsprincipepanelen. Dezelfde visuele grammatica behandelt a hogere-is-beter-metrick, ontwikkeling van kleine stukjes en een lagere-is-beter-metrick, vroege koninginnenbeweging. Beide panelen tonen levensduur, laatste 10, overwinningen en verliezen zodat het rapport stabiele gewoonten kan scheiden van recente of resultaatafhankelijke gewoonten patronen.

Figuur 11 toont de percentielstijl presentatie die wordt gebruikt voor statistieken waarvan de ruwe waarde niet goed wordt weergegeven door a enkele mondiale trend. Winnende vorken zijn een tactisch voorbeeld: het bevolkingscijfer van geconverteerde vorken verandert met zowel de vaardigheid van de speler als het foutenpercentage van de tegenstander. De card vergelijkt daarom de waarde van de gebruiker met de geschiedenis van peer-spelers in de Elo-emmer in kaart gebracht. De boxplot geeft de peerverdeling weer, inclusief de interkwartielafstand, mediaan en snorharen. De markering van de gebruiker wordt uitgezet op het dezelfde schaal en gelabeld met zowel de ruwe waarde als de percentielrang. Aangrenzend vergelijkingsbuckets bli-

jven in gedempte vorm zichtbaar, zodat de gebruiker kan zien dat de peer-bucket is lokaal, geen universele schaal.

Dit ontwerp maakt percentielstatistieken controleerbaar. De gebruiker ziet de ruwe waarde, het percentiel, de peer-bucket die ter vergelijking wordt gebruikt, en de vertaalde waarde Chess.com-naar-Lichess-baktoewijzing. Dezelfde levensduur, laatste 10, overwinningen en Verliezenrijen blijven behouden, zodat tactische percentielscores nog steeds kunnen worden gecontroleerd voor recentheidseffecten en resultaatafhankelijkheid.

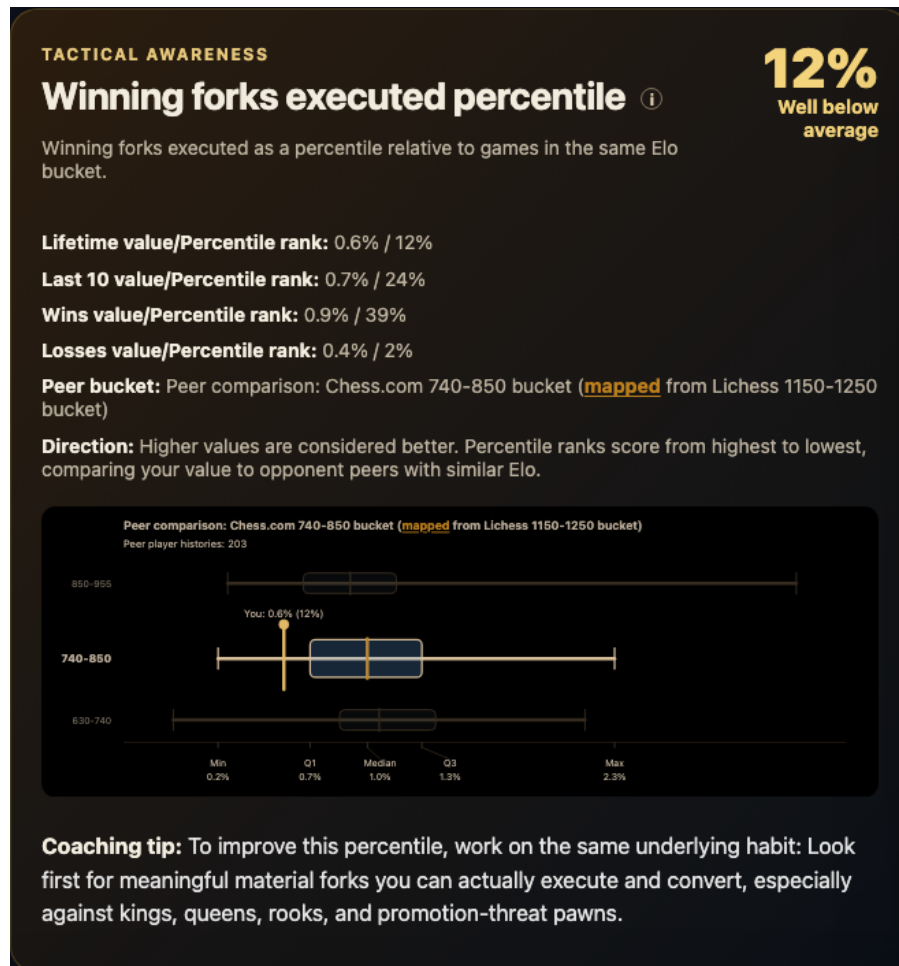


Figure 11: Productiepercentielpaneel voor uitgevoerde winnende vorken. De boxplot toont de same-bucket peer-distributie die in het rapport wordt gebruikt, terwijl het goud De markering toont de ruwe waarde en percentielrangschikking van de gebruiker.

## 15 Conclusie

Elo+Chess-benchmarkrapporten zijn opgebouwd op basis van een grote move-state-berekening gestratificeerde Lichess-monsters. Elke metriek begint als een reeks concrete bordstatussen en verplaatsingsvariabelen, wordt een speler-spelfunctie en wordt vervolgens gemiddeld door Elo bucket om een empirische benchmarkcurve te vormen. Het openingsprincipe Voorbeelden laten

zien waarom deze aanpak nuttig is: eenvoudige schaakgewoonten hebben sterke, zichtbare relaties met beoordelingen in het doelbereik, terwijl die Relaties vlakten vaak af of worden conditioneler boven de bovenste grens. Dat is de centrale reden waarom Elo+Chess zijn coachingclaims richt op de rating bereik waar de gegevens dit het duidelijkst ondersteunen.