

# Bygga stratifierade Lichess benchmark-kurvor för Elo+Chess

Elo+Chess Forskningsnoteringar

31 maj 2026

## Abstract

Elo+Chess benchmark-rapporter är byggda från stora stratifierade urval av Lichess.org-spel, med benchmark-exemplet som används i denna anteckning som innehåller 483 974 1+0 kulspel, 479 974 3+0 blixspel, 925 636 10+0 snabbspel, och 660 540 längre snabbspel. För varje större speltyp utökas pipelinen varje spel i före-drag och efter-drag styrelse stater, beräknar hundratals mellanliggande och spelar-spelvariabler, aggregerar dessa variabler till coachningsstatistik och sedan genomsnittet av dessa mätvärden med Lichess Elo-segment. Den provet är avsiktligt stratifierat snarare än proportionellt mot den råa spelaren pool: syftet är att uppskatta stabila beteendekurvor på varje ratingnivå, inklusive hinkar som annars skulle vara glesa, inte för att återge det naturliga betygsfördelning. Resultatet är en familj av empiriska benchmarkkurvor: för varje mätvärde kan vi fråga hur ofta spelare på olika Elo-nivåer visar en given vana eller princip i riktiga spel. Denna not beskriver metoden, omfattningen av beräkningen och flera öppningsprincipexempel som involverar castling, utveckling av mindre stycken, upprepade öppningsspel och tidig drottning. Tyngdpunkten är metodologisk: Elo+Chess förutsätter inte att schackprinciper är universella linjär över alla färdighetsnivåer. Istället inspekteras benchmark-kurvor av betyg och produkten fokuserar på nybörjar- till tidig-avancerad region där relationerna är starkast och mest tolkbara.

## 1 Ändamål

Målet med benchmarksystemet är att omvandla breda schackråd till mätbara mängder. Ett uttalande som "utveckla dina mindre delar" är användbar, men det blir mer handlingskraftigt när vi kan säga hur många mindre bitar a spelare utvecklas vanligtvis genom drag 10, hur det kan jämföras med jämnåriga i samma betyggruppen och hur den vanan förändras över hela betygsspektrumet.

Elo+Chess bygger därför ett metriskt bibliotek från faktiska spel. Varje måttenhet är beräknas för varje spelare-spelrad, sammanfattad efter speltyp och Elo-hink, och används som riktmärke i rapporten. Rapporten kan sedan placera en användare på samma kurva och visa om användaren är över, under eller nära beteendet observeras bland jämförbara spelare.

## 2 Källdata och stratifiering

Benchmarkkällan är den offentliga speldatabasen Lichess.org. Elo+Chess använder stora stratifierade prover efter speltyp och klassificeringshink snarare än en liten bekvämlighetsprov. De nuvarande produktionsartefakterna täcker fyra stora spelkategorier:

- 1+0 kula,
- 3+0 blitz,

- 10+0 snabbt,
- snabba spel längre än 10 minuter.

Stratifieringsmålet är bredd över betygsnivåer. En proportionell slumpmässig prov skulle vara effektivt för att beskriva den totala spelarpoolen, men det skulle det vara ineffektivt för att bygga benchmarkkurvor. De vanligaste betygshinkarna skulle dominera provet, medan lågstödsskopor skulle ha för få spel för stabila metriska medelvärden. Elo+Chess samplar därför efter genomsnittligt spel Elo hink. I den nuvarande renprovskonstruktionen är råspel först filtreras till giltiga betygsatta spel med kända spelare, kända betyg, giltiga resultat, och kända tidskontroller. Spel filtreras sedan till spelare med tillräckligt många spel i den hastigheten och tillräckligt stabila betyg för att representera en sammanhängande färdighetsnivå: för varje spelare beräknar Elo+Chess spelarens klassificeringsspann inom hastigheten ( $\max(\text{Elo}) - \min(\text{Elo})$ ) över källpoolen och behåller spel endast när båda spelarna har minst det konfigurerade minsta antalet spel och inte mer än 300 Elo-punkter av drift inom hastighet. Driftfiltret är inte avsett för att ta bort normal förbättring eller varians; det minskar fall där en spelares spel skulle blanda väsentligt olika färdighetsnivåer, såsom preliminära konton, återkommande spelare, eller konton som rörde sig snabbt genom flera hinkar under provperioden. Slutligen, kvalificerade spel rankas inom varje genomsnittlig Elo hink av en deterministisk hash av spel-id och seed, och ett begränsat antal spel behålls från varje hink.

Denna design gör benchmarkkurvorna mycket mer användbara. Varje Elo-hink bidrar med tillräckligt många observationer för att uppskatta medelvärdet för varje mätvärde, och sällsynta skopor med hög eller låg betyg dränks inte av den naturliga formen av distributionen för aktiva spelare. Benchmarkkurvorna som används av Elo+Chess är sedan begränsad till det sortiment som är mest relevant för produkten, ungefär nybörjare genom tidigt avancerat spel. I Lichess-skalan rapporterar det primära intervallet är 600–1600. När en användare tittar på rapporten på Chess.com-skalan, visas motsvarande klassificeringssegment mappas till samma underliggande Lichess benchmark-skopor med den separata plattformsberoende ratingkartläggningsmetoden beskrivs i den kompletterande uppsatsen: [Estimating Cross-Platform Chess Rating Mappings with Modal regression](#).

### 3 Betygsdistributioner för aktiva spelare

Om vi tittar på fördelningen av Lichess.org-spelarens Elo-betyg i slutet i mars 2026, efter större speltyp, ser vi följande räkningar och percentiler. För denna distributionsvy räknas varje spelare en gång per exakt speltyp, använder spelarens senaste observerade betyg i motsvarande råa tidskontrollrader, efter att ha krävt minst fem spel i det exakta spelet typ.<sup>1</sup>

Denna fördelning ska inte förväxlas med det stratifierade riktmärkesurvalet. Fördelningen beskriver den aktiva Lichess spelarpool som observerats i rådata skanna. Benchmark-provet stratifieras sedan medvetet så att Elo hinkar har tillräckligt med stöd för stabila metriska kurvor.

Att mappa från ett Lichess.org Elo-betyg för en viss speltyp till ett motsvarande Chess.com-betyg, Elo+Chess använder den speltypsspecifika modalen regressionsekvationer uppskattade i

<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 visa distributionerna med och utan detta filter.

| Game type  | Players   | Median | 75th pct. | 80th pct. | % $\leq 1600$ | 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: Senast observerade aktiva spelare Lichess betygsfördelning efter exakt speltyp efter att ha krävt minst fem spel i just den speltypen.

följeslagaren [Elo+Chess rating-mapping paper](#):

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

Dessa ekvationer kartlägger betygsvärden, inte percentilrankningar. Percentile ranger är definieras i en viss spelarpool, och Chess.com- och Lichess-spelaren pooler har olika former. I tabell 2, för 10-minuters snabbspelstyp, den första kolumnen ger ett litet urval av möjliga Chess.com Elo betygsvärden. Den andra kolumnen ger motsvarande Chess.com percentilrankning visas på Chess.com den 31 maj 2026. A 71,7

ranking på Lichess.org för 10-minuters fort motsvarar en Lichess Elo-betyg på 1,644<sup>2</sup> och det värdet visas i tredje kolumnen. Om vi tillämpar Chess.com–Lichess mappningsekvationen, 1 644 på Lichess motsvarar cirka 1 271 på Chess.com. Om man antar det uppskattade mappningsekvationen är rimlig, +522 gapet mellan Chess.com-betyget i kolumn 1 och det mappade betyget i kolumn 4 kan förklaras av ett stort antal av tillfälliga spelare i Chess.com spelarpool som blåser upp Chess.com percentil rang. Med andra ord, att vara i 72:a percentilen på Lichess.org är en större prestation än att vara i 72:a percentilen på Chess.com, eftersom percentilrangerna kartläggs inte rent från ett system till ett annat. Faktiska Elo värden kan kartläggas med hjälp av ekvationer som uppskattas från samma spelares poäng de två systemen.

| 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: Illustrativ jämförelse mellan Chess.com 10-minuters snabb percentil poäng och samma percentilpoäng i den aktiva Lichess 10+0-snabben distribution efter att ha krävt minst fem 10+0-spel. “Kartlagt Chess.com equiv.” tillämpar nuvarande Elo+Chess 10+0 snabb omvandling av betygsskala till samma procentuella Lichess-betyg. Den rader med högst svans ingår för att visa skalans skillnad, men den monterade betygslinjen är mest tillförlitlig inom det huvudsakliga coachningssortimentet.

<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.

Anledningen till luckorna i Tabell 2 är att mappningen uppskattar motsvarande spelstyrka, medan en percentil beskriver position inom en plattforms egen befolkning. En 71,7:e-percentilspelare på Chess.com är därför inte automatiskt en 71,7:e-percentilspelare på Lichess, och vice versa. De observerade siffrorna överensstämmer med Chess.coms publik rapid pool har en mycket större lägre rankad eller casual-spelare massa, medan aktiv exakt-kontroll Lichess pool är mer koncentrerad bland erfarna upprepa spelare. Det viktiga är inte att en percentilskalan är "rätt" och att annat är "fel"; det är att percentilen inte är portabel över plattformar såvida inte de underliggande spelarfördelningarna har samma form.

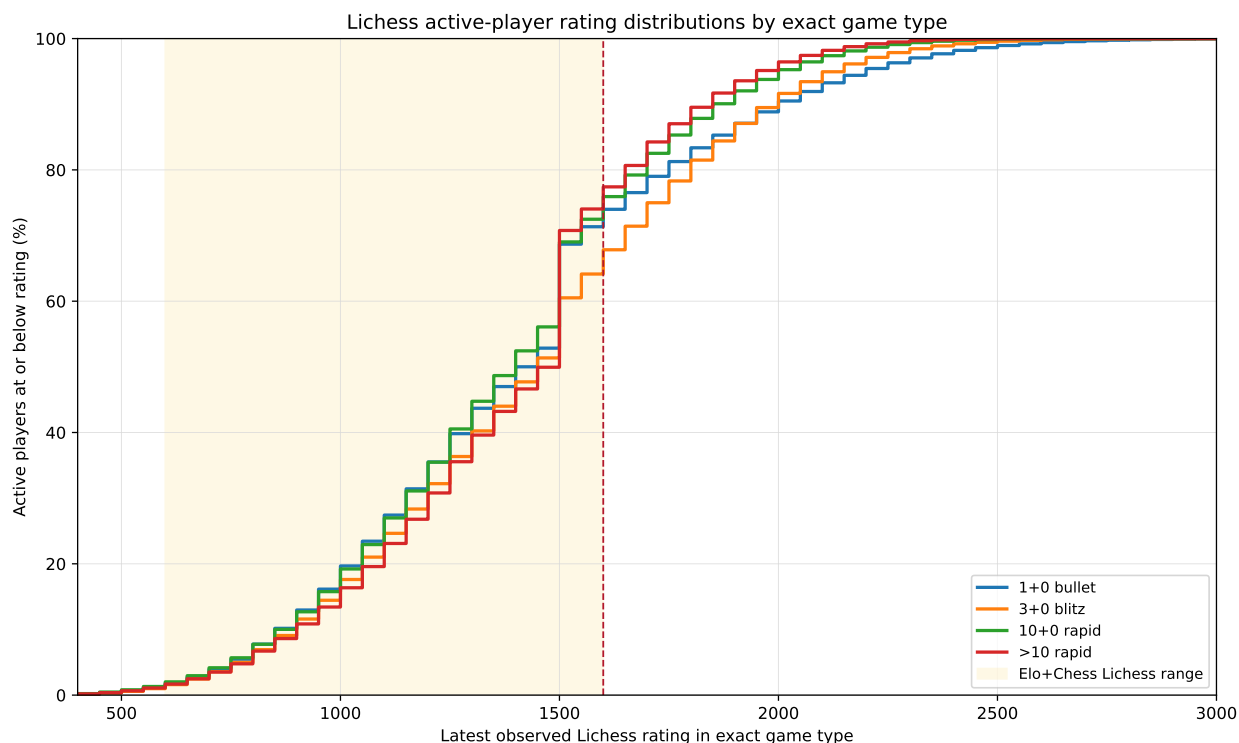


Figure 1: Kumulativa fördelningar av aktiva spelares betyg efter exakt speltyp. Den gula band markerar 600–1600 Lichess-serien som används av huvud Elo+Chess riktmärken för coaching.

Note on the 1500 spike. The visible jump near 1500 finns i data snarare än skapas av utjämning. I det senaste observerad betygsfördelning, exakta 1500 betyg står för 13,1 bullet-spelare, 5,6 % av 3+0 blitzspelare, 9,4 % av 10+0 snabba spelare, och 17,6 start eller provisoriskt ankare i klassificeringssystemet.

Den exakta 1500 spiken är till stor del en artefakt av lätt aktiva konton. Till kolla detta, vi räknade om samma distributioner med senaste betyg efter att ha krävt att varje spelare ska ha minst fem spel i den exakta speltypen. Piggan nästan försvinner under detta filter: exakta 1500 betyg faller från 5,6–17,6 spelare i fördelningen för alla spelare till 0,2–1,2 fördelning rapporterad i tabell 1.

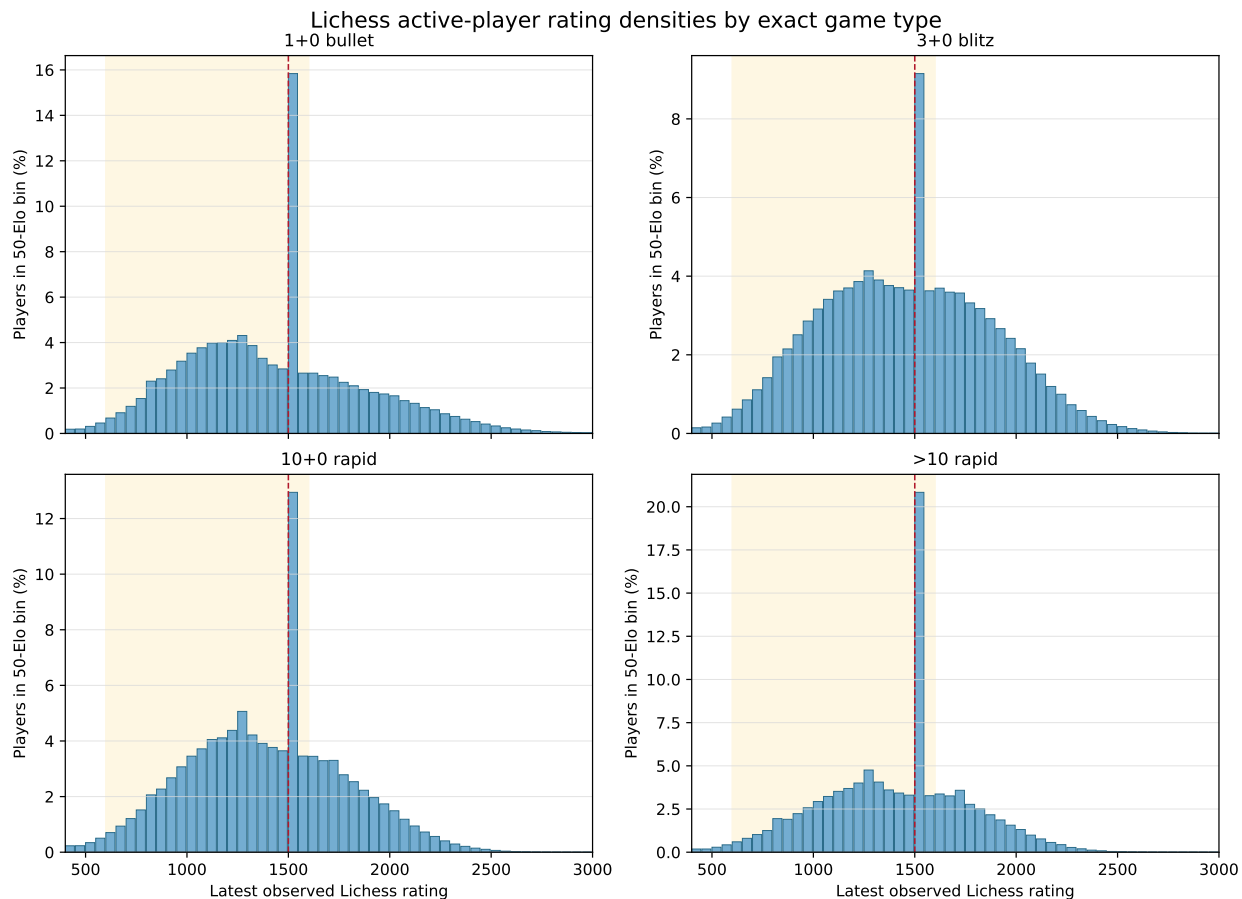


Figure 2: Densitetsvy av samma betygsfördelningar för aktiva spelare. Den streckad linje markerar 1500 Lichess, där de råa senaste betygsdistributionerna har en stor massapunkt med exakt värdering. Det gula bandet markerar 600–1600 Lichess intervall som används av Elo+Chess huvudstatistikrapporter. Piggen är mest synlig i tätheten för alla spelare eftersom många lätt aktiva konton förblir exakt på 1500; Figur 4 upprepar densitetsdiagrammet efter att ha krävt minst fem spel i den exakta speltypen, vilket till stor del tar bort denna artefakt.

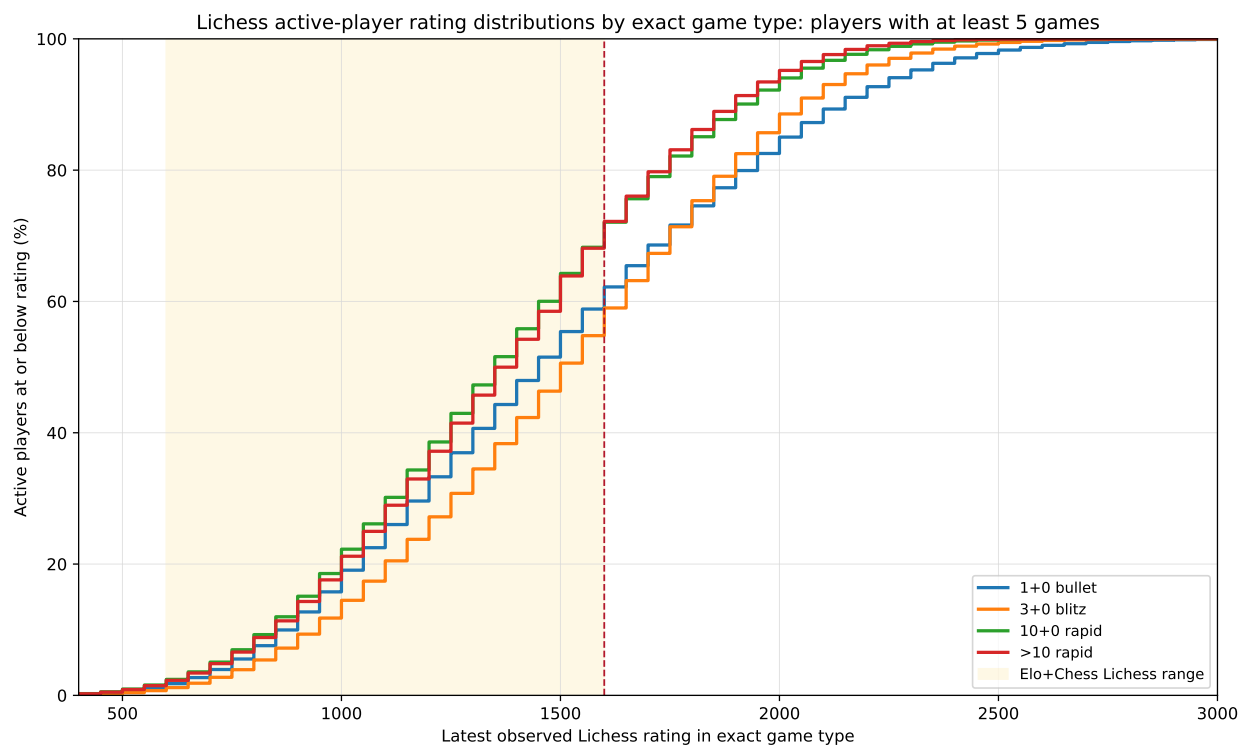


Figure 3: Kumulativa utdelningar av aktiva spelares betyg efter krav på kl minst fem spel i den exakta speltypen. 1500-hoppet reduceras kraftigt.

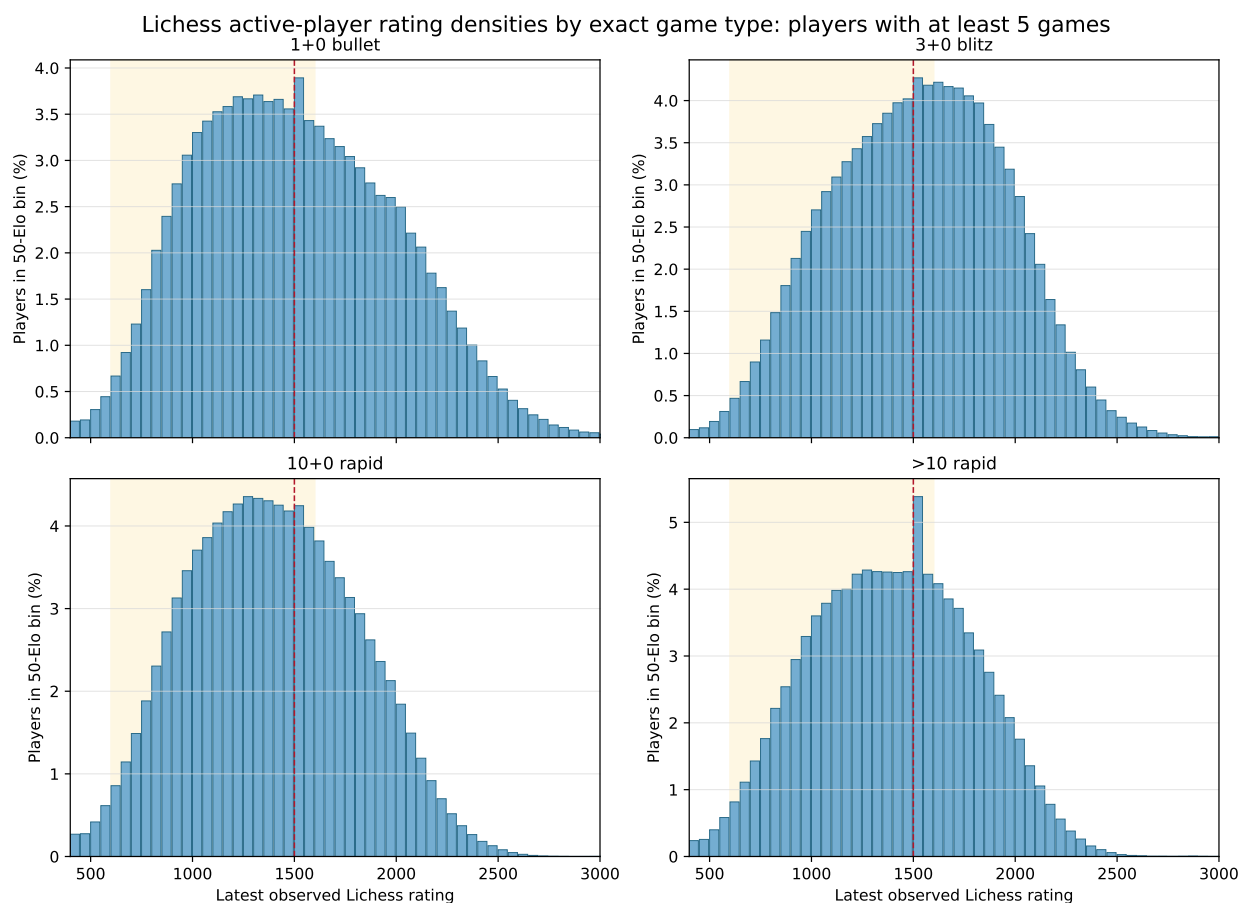


Figure 4: Densitetsvy efter att ha krävt minst fem spel i det exakta spelet typ. Denna uppfattning bekräftar att den stora 1500 massan poäng i all-player distributionen drivs till största delen av svagt aktiva konton. Det gula bandet markerar intervallet 600–1600 Lichess som används av Elo+Chess huvudmetriska rapporter.

## 4 Skala av nuvarande benchmark-artefakter

Beräkningen är avsiktligt stor. Det är inte en profilskrapa, en liten uppsättning nedladdade PGNs, eller en handkurerad samling lärorika exempel. Det första Lichess-passet täckte kompletta månatliga filer från januari 2025 till mars 2026. Den ostratifierade skanningstabellen innehåller 1 382 178 087 spel och 2 764 356 174 betygsrader för spelarspel. Över dessa rader, 6 623 438 unika spelarens användarnamn visas minst en gång.

| 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: Ostratifierad Lichess skanningsskala före benchmarksampling. Skanningen täcker fullständiga månatliga spelhistorikfiler från januari 2025 till mars 2026.

Produktionsriktmärket artefakter är mycket mindre än råskanningen eftersom de är valda för metrisk konstruktion. De är fortfarande stora: riktmärket stadium är en rörledning för expansion och aggregering av flytt tillstånd över hundratals tusentals spel per speltyp.

| 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: Skalan för de stratifierade benchmark-artefakterna efter anpassning till vanligt 277-kolumns runtime-funktionsschema. A “player-game row” betyder ett spel från en spelares perspektiv, så de flesta spel producera två spelare-spelrader.

Enbart för 3+0 blitz använder det nuvarande produktionsriktmärket 479 974 spel involverar 129 398 unika spelare och 28 138 904 distinkta rader i rörelsetillstånd. Den aligned blitz player-game feature table innehåller 277 härledda variabler. Den snabba funktionsbyggande artefakter är ännu bredare i rörelsestadiet: hela snabb rörelse-tillståndstabell som används under funktionskonstruktion innehåller 168 flytta-tillstånd kolumner innan aggregering till en 277-kolumner spelar-spel funktion bord. Frontend runtime benchmark artefakter för alla fyra speltyper är sedan anpassad till samma 277-kolumner spelar-spel funktionsschema, inklusive utökade taktik-taxonomi breakout-variabler. Från dessa funktionstabeller, rapporten väljer och visar de coachningsmått som är mest tolkbara.

Dessa räkningar är viktiga för att tolka resultaten. Benchmark-kurvorna är inte uppskattningar från några hundra manuellt granskade spel. Istället, de kommer från tiotals miljoner utvärderade styrelsestater och mer än fyra miljoner spelare-spelrader över de fyra nuvarande produktionsspelkategorierna.

## 5 Pipeline översikt

Benchmark-konstruktionspipelinen har sex huvudsteg:

1. Välj klassade Lichess-spel i målspelestyp och klassificeringsskikt.
2. Analysera varje spel drag för drag och lagra före-draget och efter-draget styrelsestat.
3. Beräkna mellanliggande rörelse-tillståndsvariabler från tavlan, draget, spelarperspektivet och motståndarperspektivet.
4. Aggregera rörelse-tillståndsvariabler till spelare-spelfunktionsvariabler.
5. Konvertera spelar-spelvariabler till rapportstatistik med en tydlig tolkningsriktning: högre är vanligtvis bättre, lägre är vanligtvis bättre eller beskrivande.
6. Sätt ett genomsnitt för varje mätvärde med Elo-hink och inspektera resultatet manuellt relation.

Denna struktur är viktig eftersom många rapportmått inte är direkt synlig i ett PGN-huvud. De kräver att rekonstruera positionen, bestämma vilka pjäser flyttade från vilka startrutor, jämför före-drag och efterflyttning material och struktur, och sedan aggregera resultatet från flytta nivå till spelar-spel nivå.

## 6 Mellanvariabler

De mellanliggande variablerna är utformade för att beskriva vad som förändrades på tavlan och varför det är viktigt ur reporterns perspektiv. Exempel inkluderar:

- dragnummer, spelares rörelsenummer, sida till rörelse, rörelsens färg och spelare färg;
- från ruta, till ruta, rörlig pjästyp, infångad pjästyp och om flytten var ett tillfångatagande, check, slott, befordran eller kompis;
- materialbalans före rörelse och efter rörelse;
- om en mindre pjäs lämnade sin startruta med ett givet dragnummer;
- om samma pjäs flyttas på på varandra följande spelare flyttar i öppning;
- om drottningen flyttade före drag 10;
- castlings sida, casting move nummer och om torn blev sammankopplade efter rokering;
- lösa pjäser, hängande pjäser, dubblerade bönder, isolerade bönder, godkänd bönder och baklänges bönder;
- tornfilstatus, öppna filer, halvöppna filer och king-safety drag;
- taktiska möjligheter, genomförd taktik, materialtryckstaktik, garanterade uppföljningsvinster och realiserade vinster.

Den exakta mellansatsen varierar beroende på byggstadiet. Runtime rapportdatabaser behålla endast de kolumner som behövs för att visa rapporter snabbt, medan hela funktionsbygget artefakter inkluderar bredare tabeller för rörelsetillstånd och funktioner som används för att konstruera de sista spelar-spelfunktionsraderna.

## 7 Exempel på öppningsprincipen

Öppningsprinciper är användbara exempel eftersom de är lätta för användarna förstå och lätt att ansluta till flyttlägesberäkningar.

### 7.1 Minor Pieces Developed by Move 10

Detta mått räknar hur många av en spelares fyra riddare och biskopar som har kvar sina hemmarutor efter spelarens 10:e drag. Mellanräkningen använder flyttarfärgen, rörlig pjästyp, startruta och spelarrörelse nummer. För vit är hemrutorna b1, g1, c1 och f1. För Black är de de b8, g8, c8 och f8. För varje spelare-spel räknar Elo+Chess det distinkta hemmet rutor bland riddare och biskopar som flyttade efter spelares drag 10.

Detta är ett högre-är-vanligtvis-bättre mått. Det hävdar inte att varje minderårig pjäs måste alltid flyttas för drag 10, men bland utvecklande spelare den empiriska relationen är tydlig: spelare som når högre Elo-hinkar tenderar att utvecklas fler mindre bitar tidigare.

### 7.2 Repeated Piece Moves in the Opening

Detta mått räknar extra tidiga drag som går åt till att flytta samma pjäs igen istället att utveckla något nytt. Den mellanliggande beräkningen använder den föregående destinationsruta, den aktuella ursprungsrutan och spelarens dragnummer. Om samma pjäs flyttades på föregående spelardrag och flyttas igen inom 10 första spelarna flyttar, antalet upprepade drag ökar.

Detta är ett lägre-är-vanligtvis-bättre mått. Den mäter bortkastade öppningstempi i en enkelt sätt. Vissa upprepade drag är taktiskt motiverade, men i skala bucket curve fångar den allmänna tendensen: starkare spelare vid målet nivåer spenderar färre tidiga drag på att flytta samma pjäs upprepade gånger.

### 7.3 Queen Touched Before Move 10

Detta mått räknar damdrag före spelardrag 10. Mellanspelet beräkningen använder rörlig pjästyp och spelarens dragnummer. Om den rörliga pjäsen är drottningen och spelarens dragnummer är högst 10, räkningen ökar.

Detta är också lägre-är-vanligtvis-bättre för målområdet Elo+Chess. Tidigt drottningdrag kan vinna material eller tvinga eftergifter, men många utvecklande spelare ta ut drottningen för tidigt, tappa tempi till attacker mot drottningen och fördröja mindre bit utveckling och kung säkerhet.

### 7.4 Castling and King Safety

Castling-mått använder castling-flaggor, castling-sida, casting-rörelsenummer och styrelsestaten efter kastning. Exempel inkluderar om spelaren kastade, om spelaren kastade av drag 10, om spelaren kastade kung eller queenside, om torn var kopplade efter kastning och om tornfiler var öppna eller halvöppna efter roking.

Alla dessa mätvärden tolkas inte på samma sätt. Slottsavslutning och kastning med drag 10 är i allmänhet högre-är-vanligtvis-bättre i målområdet. Queenside casting eller casting i öppna filer kan vara mer kontextberoende. Det är därför Elo+Chess inte bara hårdkodar en moralisk regel; det bygger hink kurvor och inspekterar vilka relationer som är tillräckligt stabila för coachning.

## 8 Skokkurvor och manuell inspektion

Efter att spelar-spelsstatistiken har beräknats, grupperar Elo+Chess dem efter Lichess Elo hink och beräknar hinkmedelvärden. Bucket-genomsnittet är den råa empirin riktmärke för det måttet. En trendlinje används sedan i rapporten till tilldela betyg i Elo-stil till användarvärden.

Kurvorna inspekteras manuellt innan de används. Ett mått är mer användbart när hinken visar ett sammanhängande förhållande över betygsintervallet betjänas av produkten. Ett mått är mindre användbart när det är bullrigt, platt, mycket icke-monotona, eller drivs av speciella fall som är svåra att förklara för en användare.

Detta manuella inspektionssteg är särskilt viktigt ovanför Elo+Chess-överdelen cutoff. Starkare spelare vet när de ska bryta öppningsprinciperna och gör det effektivt. Över intervallet nybörjare-till-tidigt avancerade, förhållandet mellan enkla vanor och betyg plattar ofta till eller ändrar skepnad. Produkten betonar därför intervallet där de empiriska sambanden är starka och pedagogiskt användbar.

## 9 Exempel på öppningskurvor för fullt intervall

Figur 5 visar fyra 3+0 blitzöppningsprincipkurvor. Grå punkter visar det bredare betygsintervallet för Lichess. Blå punkter och den blå trenden linje visar 600–1600 Lichess serien som används av den huvudsakliga Elo+Chess coaching riktmärken. Den streckade röda linjen markerar den övre avskärningen.

Samma mönster är synligt i hela intervallet 10+0 snabb och längre snabb artefakter. Det centrala modelleringsbeslutet är oförändrat: använd det empiriska stabil räckvidd för coaching påståenden och undvika att låtsas som en enkel vanekurva förblir lika informativt för alltid.

Samma mönster visas numeriskt i Tabell 5. Backar är rapporteras som metrisk värdeförändring per 100 Elo-poäng. Slott, mollstycke utveckling, upprepad pjäsrörelse och tidig drottningrörelse visar alla tydliga rörelse över 600–1600. Över 1600 är backarna mycket mindre i dessa exempel, vilket stämmer överens med tanken att högre rankade spelare bryter grundläggande principer mer selektivt och mer framgångsrikt. En andra mekanism är mättnad. För vissa vanor når måtten ett praktiskt tak eller söt plats: starka spelare har redan lärt sig vanan, så det finns ingen anledning till kurvan för att fortsätta röra sig kraftigt uppåt. Mer är inte heller alltid bättre efter den punkten. Uppgifterna plattas därför till både eftersom starkare spelare kan gå sönder regler i konkreta positioner och eftersom många grundläggande vanor redan har nått deras användbara arbetsområde.

## 10 Varför den övre cutoff spelar roll

Målanvändaren Elo+Chess är en nybörjare-till-tidigt avancerad spelare. För dessa spelare, grundläggande vanor är starkt förknippade med betyg: utveckla pjäser, undvik upprepade tidiga pjäsdrag, undvik onödiga tidiga drottningsdrag, slott i tid, koppla ihop torn, hantera material och undvik att lämna pjäser lösa eller hängande.

At higher ratings, those same surface habits become less diagnostic. A strong player may delay castling because the center is closed, move the queen early because the position contains a concrete tactic, or move the same piece twice because it wins a tempo or forces a structural concession. The raw habit still has meaning, but the relationship between the habit and rating becomes more conditional. This is why Elo+Chess uses an upper cutoff for its main coaching

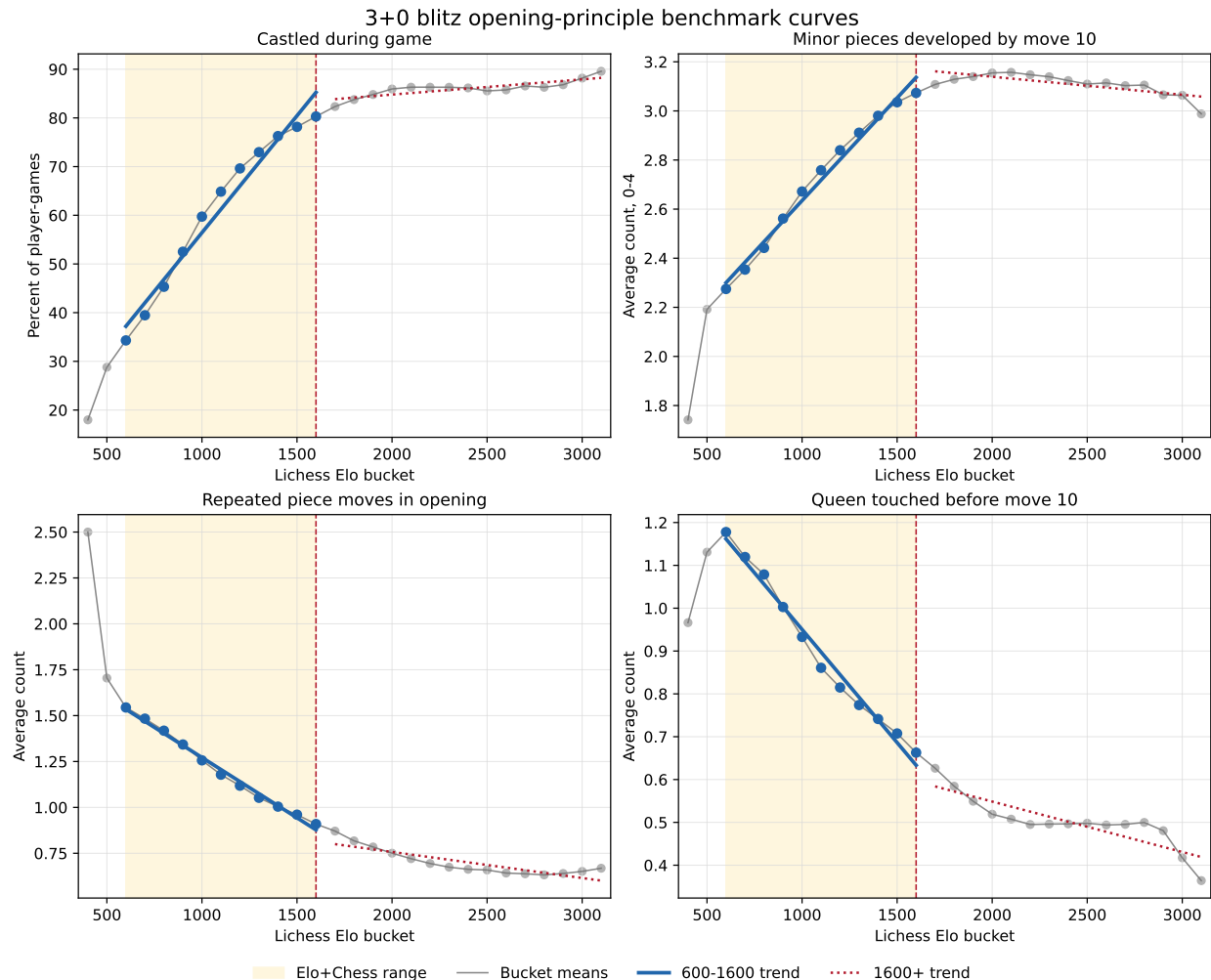


Figure 5: Utvalda benchmarkkurvor för öppningsprincipen för 3+0 blitz. Den relationer är starkast och mest användbara i Elo+Chess coachingsortimentet. Ovanför den övre avskärningen plattar flera förhållanden avsevärt ut.

claims and why users above that range are warned that the tight relationship between principled habits and rating breaks down.

## 11 Tvåsidiga taktiska mått och procentmål

Vissa mått har motsatt form från exemplen med öppningsprincipen. Gafflar är ett användbart exempel eftersom händelsen till sin natur är tvåsidig. En spelare som hittar och konverterar mer användbara gafflar än kamrater gör något bra, men befolkningsgraden av konverterade gafflar faller med betyg eftersom starkare motståndare tillåter färre gaffelmöjligheter och försvarar sig mot taktiska hot mer aktivt.

Figur 8 visar denna effekt. Båda konverterade gafflarna av rapportspelaren och konverterade gafflar av motståndaren blir mindre frekventa som betyg stiger. Det betyder inte att det är dåligt att utföra gafflar. Det betyder den råa händelsen frekvensen styrs delvis av motståndarens fel-

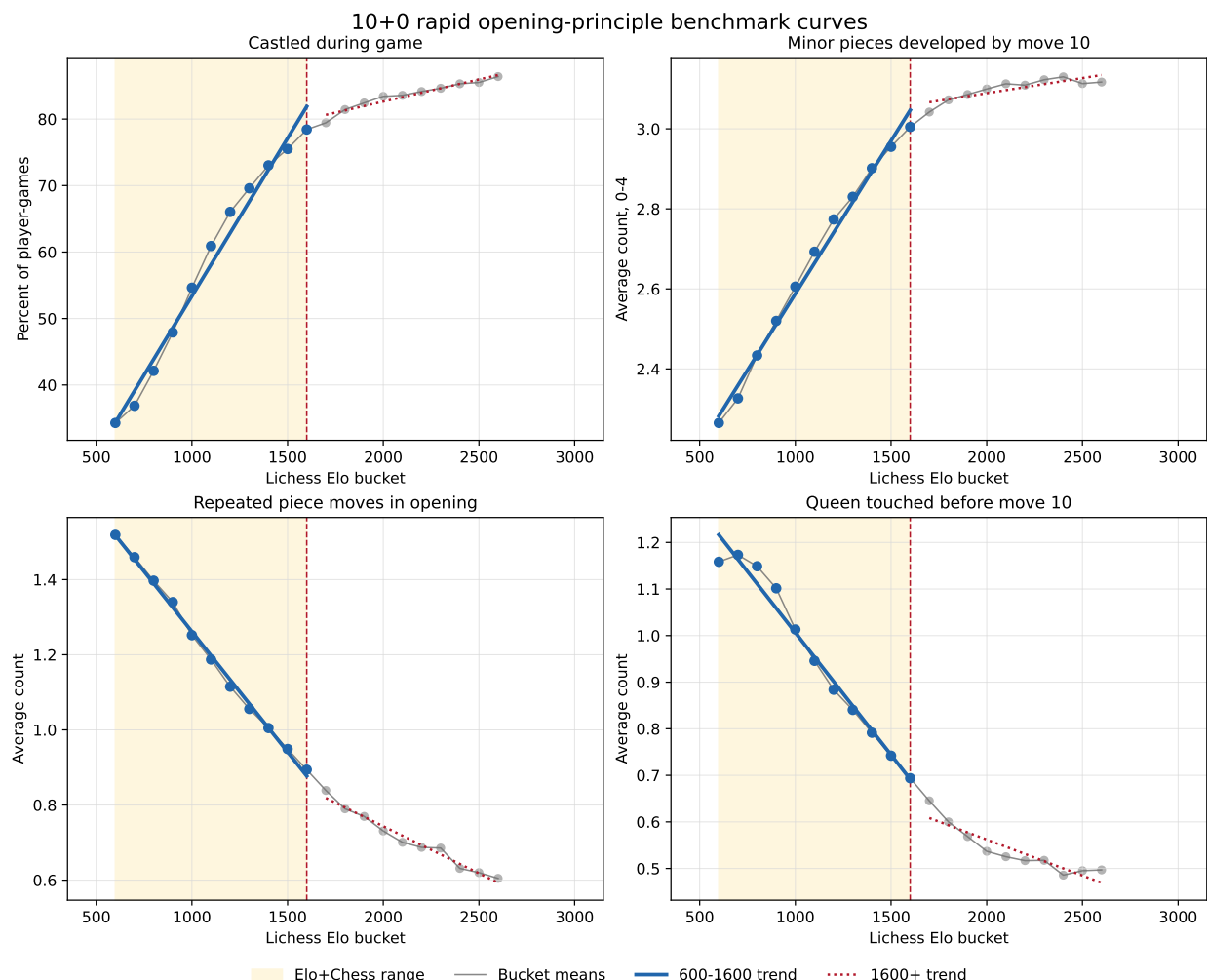


Figure 6: Utvalda benchmarkkurvor för öppningsprincipen för 10+0 snabba spel. Den relationer är starkast och mest användbara i Elo+Chess coachingsortimentet. Ovanför den övre avskärningen plattar flera förhållanden avsevärt ut.

frekvens och spelstruktur. För mått så här använder Elo+Chess därför peer-history-percentiler inuti spelarens egen Elo-hink istället för att bara förlita sig på en global trend.

För att bygga dessa percentilmål samplar Elo+Chess spelare efter speltyp och Elo-bucket, använder sedan de senaste kvalificerande spelen för varje samplade spelare. Den nuvarande måluppsättning använder 11 hinkar från 600 till 1600, 200 spelare per bucket och upp till 50 senaste kvalificerande spel per spelare för varje produktion speltyp. Detta ger 2 200 spelarprover per speltyp och ungefär 110 000 spelare-splrader per speltyp för percentil inom hink distributioner.

Figur 9 illustrerar det resulterande målet distribution för samplade spelares konverterade gaffel-hastighet i 3+0 blitz. Varje låda är en fördelning inom segmentet mellan spelare i urvalet, inte ett medelvärde för segment över individuella spel. I rapportens användargränssnitt jämförs en användares värde med det relevanta box för användarens mappade Elo-hink. Detta ligger till grund för uttalanden som t.ex om en spelares konverterade gaffelhastighet är hög eller låg i förhållande till jämnåriga samma nivå, även när den totala populationshändelsefrekvensen minskar med betyget.

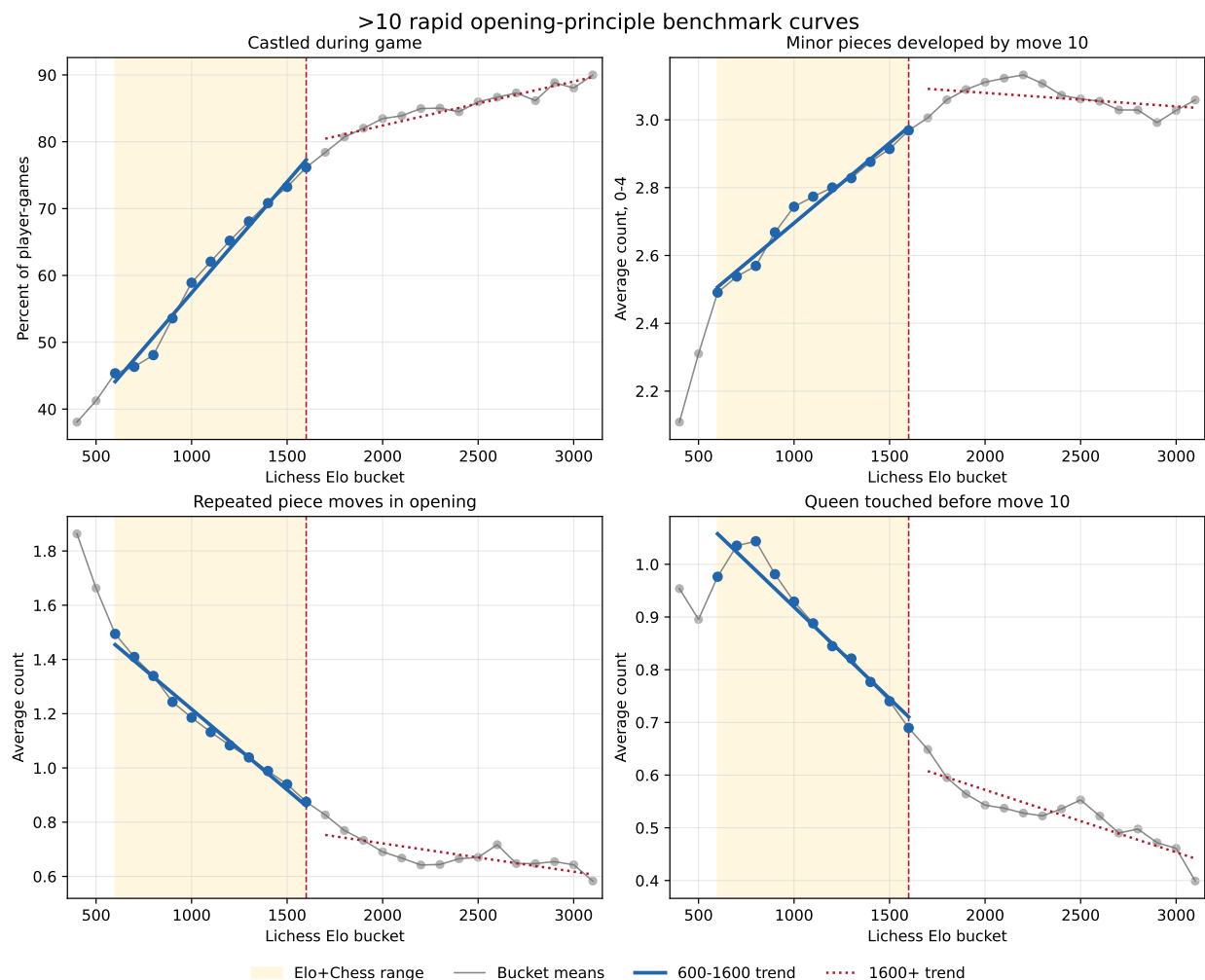


Figure 7: Utvalda benchmarkkurvor med öppningsprincip för längre snabba spel.

## 12 Replikeringskostnad

The method is reproducible in principle, but it is not lightweight. Ombyggnad the benchmark curves requires:

- tillgång till stora Lichess månatliga spelhistorikfiler;
- stratified filtering by game type, rating bucket, and eligibility regler;
- en legal move parser och board-state evaluator som kan expandera tiotals miljoner flyttstater;
- har logik för öppning, material, bondstruktur, kungsäkerhet, torn aktivitet, taktik, tidshantering och spelresultat;
- aggregering från flyttläge till spelare-spelfunktioner;
- Sammanfattningar på hinknivå för varje rapportmått;

| 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: Ungefärliga skopkurvlutningar för de fyra exemplen på öppningsprincipen mätvärden över produktionsspeltypen.

| 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: Aktuella måluppsättningar för peer-historia-percentiler. Dessa är spelare-historikprover som används för att uppskatta distributioner inom hink för percentilliknande mått.

- manuell granskning av varje benchmarkkurva för kandidater innan den används i en tränar-rapport.

För 3+0 blitz betyder det nästan en halv miljon spel, nästan en miljon spelare-spel rader, och mer än 28 miljoner flytta-stat rader. Tvärs över de fyra benchmark-speltypen innehåller de nuvarande benchmark-artefakterna mer än 2,55 miljoner spel, mer än 5,10 miljoner spelare-spelrader och mer än 147 miljoner flytta-stat-rader. Skalan är central för värdet av benchmark: det låter rapporten visa användarna vad som faktiskt händer i spel som spelas av spelare nära deras nivå snarare än att bara förlita sig på allmänna schackslogans.

## 13 Begränsningar

Benchmarkkurvorna är empiriska beskrivningar, inte kausala bevis. En spelare får inte betyg bara genom att matcha ett hinkmedelvärde. Mätvärden är också ofullkomliga sammanfattningar av komplexa positioner: ett tidigt drottningssdrag kan vara utmärkt i en position och hänsynslös i en annan. Syftet med kurvorna är att identifiera återkommande vanor och jämföra dem mot kamraters beteende, inte att ersätta beräkning eller coachande omdöme.

Kurvorna är också kurvor i Lichess-skala. När användare tar med Chess.com-spelet historik, mappar Elo+Chess den relevanta betygsgruppen till Lichess benchmark skala med den separata

Two-sided converted-fork rates by Elo bucket

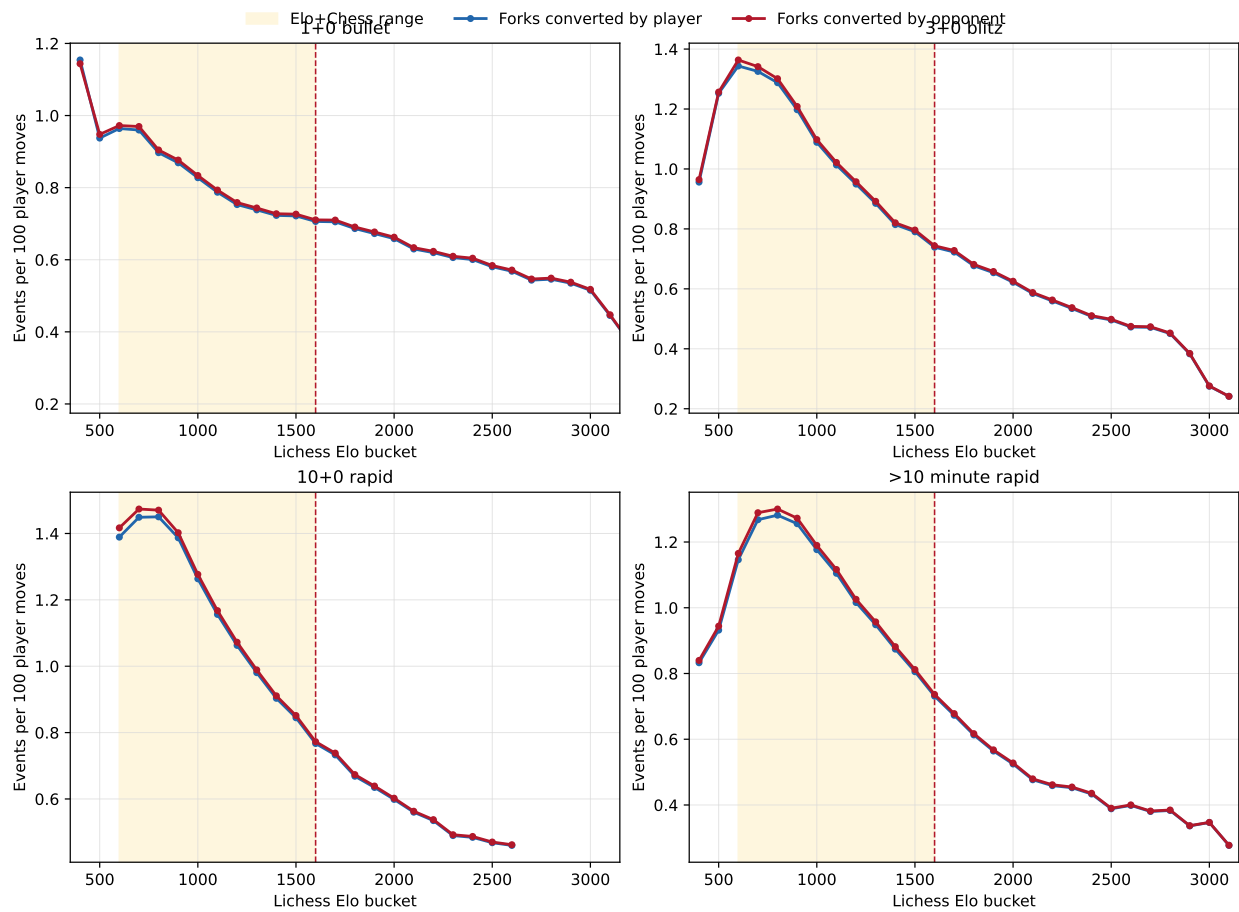


Figure 8: Tvåsidiga konverterade gaffelpriser med Elo-skopa. Konverterade gaffelhändelser falla på högre betyg eftersom båda sidor försvarar bättre och färre rena gaffel konverteringar är tillgängliga.

plattformsoberoende kartläggningsmodellen. Det kartläggningslagret är beskrivs i den kompletterande forskningsanteckningen Elo+Chess om Lichess-to-Chess.com betygskonvertering: [Estimating Cross-Platform Chess Rating Mappings with Modal regression](#). Denna mappning ska inte läsas som bevarande percentil rang. Den bevarar en uppskattad ratingekvivalens mellan de två betygen system; percentilrankningen förblir plattformsspecifik eftersom den aktiva spelaren fördelningarna är olika.

## 14 Produktionsrapport grafik

Produktionsrapporten förvandlar benchmark-artefakterna till paneler på metrisk nivå som kan läsas utan att känna till hela byggpipelinen. Siffror 10 and 11 visa tre exempel från Elo+Chess-rapportdesignen.

Öppningsprincippanelerna i figur 10 använd den vanliga benchmark-kurvans presentation. Varje kort börjar med metrisk familj och metriskt namn, ger sedan en kort tolkningsregel. Den stora betyg i det övre högra hörnet är avsiktligt separerat från det råa metriska värdet: det svarar på den

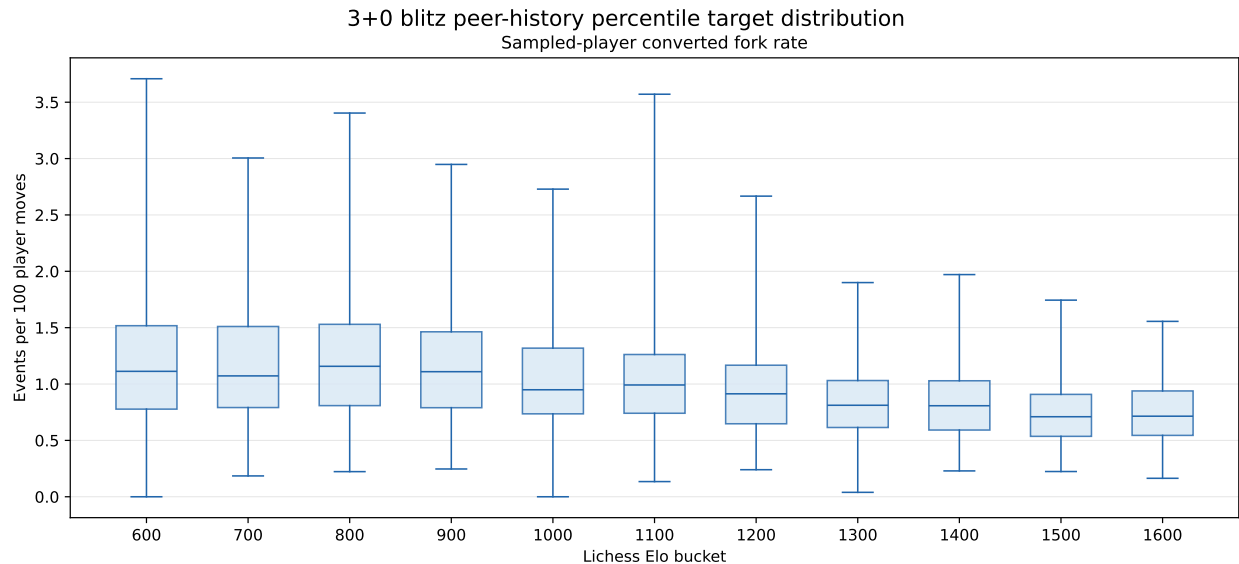


Figure 9: Exempel 3+0 blitz peer-historia målfördelning för samplade spelare konverterad gaffelkurs. Percentiler beräknas inom Elo bucket så rapporten jämför en spelare med jämnåriga som står inför i stort sett liknande taktiska möjligheter priser.

användarvårdande frågan "hur betygsätts denna vana mot riktmärket?" medan raderna nedan bevarar de uppmätta värdena. De fyra sammanfattningsraderna visar Livstid, Senaste 10, Vinster och Förluster. Detta gör att paneldiagnostik snarare än bara dekorativ. Om Lifetime och Last 10 överensstämmer, vanan är förmodligen stabil. Om vinster och förluster skiljer sig åt kan måttet vara det resultatberoende eller knuten till speltillstånd snarare än en enkel vana.

Diagramområdet använder en återhållen mörk bakgrund, guldtrendlinje och färgad markörer för de fyra skivorna. Trendanteckningen anger uttryckligen om benchmarkförhållandet stiger eller faller med Elo. Detta är viktigt eftersom vissa mätvärden är högre-är-bättre, såsom utveckling av mindre delar, medan andra är lägre-är-bättre, till exempel tidiga drotningsrörelser. Kortet alltså ber inte användaren att sluta sig till riktning enbart från slutningen. Finalen Raderna "Vad detta betyder" och "Coachingtips" översätter benchmarkresultatet till en konkret schackinstruktion samtidigt som det empiriska sammanhanget bevaras.

Figur 11 visar percentilstilen presentation som används för mätvärden vars råhastighet inte representeras väl av en enda global trend. Vinnande gafflar är ett taktiskt exempel: befolkningsgraden av konverterade gafflar förändringar med både spelarens skicklighet och motståndarens felfrekvens. Den kortet jämför därför användarens värde med peer-spelares historia i mappad Elo-hink. Boxplotten visar peer-fördelningen, inklusive interkvartilintervall, median och morrhår. Användarens markör är ritad på samma skala och märkt med både råvärde och percentilrankning. Intelligande jämförelsehinkar förblir synliga i dämpad form så att användaren kan se att peer hink är lokal, inte en universell skala.

Denna design gör att percentilmått kan granskas. Användaren ser råvärdet, percentilen, peer-bucket som används för jämförelse och den översatta Chess.com-till-Lichess-skopmappning. Samma livstid, senaste 10, vinster och Förlustrader behålls så att taktiska percentilpoäng fortfarande kan kontrolleras för nyligen genomförda effekter och resultatberoende.

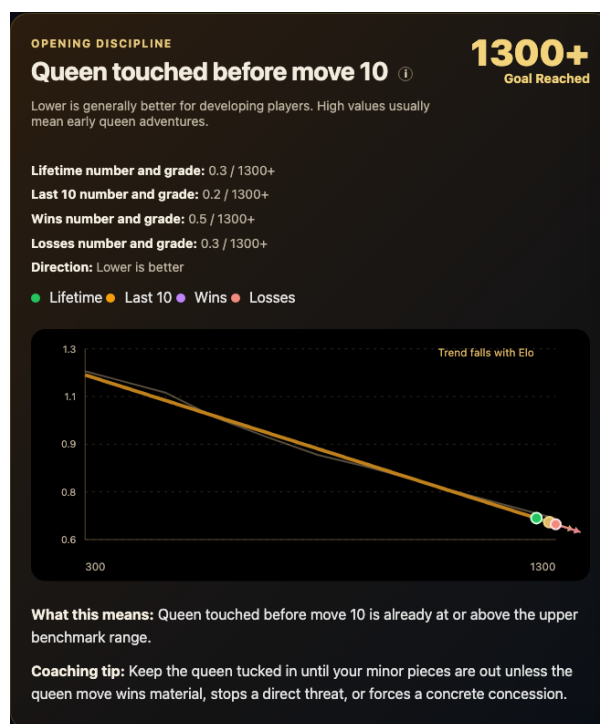
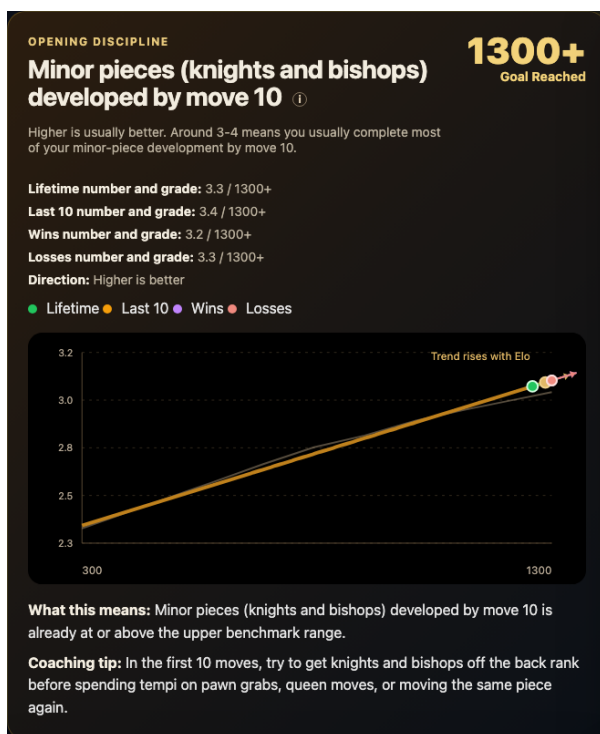


Figure 10: Paneler med produktionsöppningsprincip. Samma visuella grammatik hanterar a högre-är-bättre mått, utveckling i mindre delar och ett lägre-är-bättre mått, tidig drottningrörelse. Båda panelerna visar Lifetime, Senaste 10, Vinster och Förluster så rapporten kan skilja stabila vanor från nya eller resultatberoende mönster.

## 15 Slutsats

Elo+Chess benchmark-rapporter är byggda från en stor flytt-tillståndsberäkning över stratifierade Lichess-prover. Varje mått börjar som en uppsättning av betongbrädetillstånd och flytta-tillstånd-variabler, blir en spelare-spel-funktion, och medelvärde beräknas sedan av Elo bucket för att bilda en empirisk benchmarkkurva. Öppningsprincipen exempel visar varför detta tillvägagångssätt är användbart: enkla schackvanor har starka, synliga samband med betyg i målintervall, medan de relationer plattar ofta ut eller blir mer villkorade ovanför den övre cutoff. Det är den centrala anledningen till att Elo+Chess fokuserar sina tränarkrav på betyget område där data stöder dem tydligast.

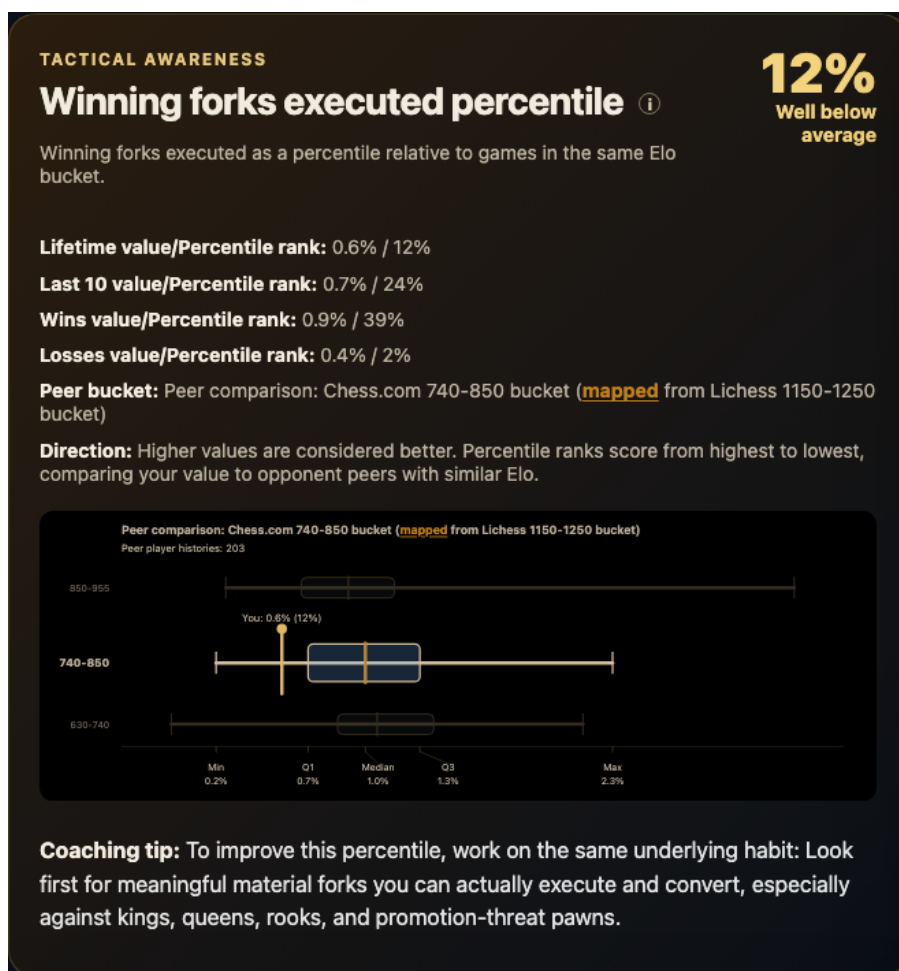


Figure 11: Produktionspercentilpanel för vinnande gafflar utförd. Boxplotten visar samma jämbördiga fördelning som används av rapporten, medan guldets markör visar användarens råvärde och percentilrankning.