

为 Elo+Chess 构建分层 Lichess 基准曲线

Elo+Chess Research Notes

2026 年 5 月 31 日

Abstract

Elo+Chess 基准报告是根据大量分层样本构建的 Lichess.org 游戏，本说明中使用的基准示例包含 483,974 1+0 子弹游戏、479,974 3+0 闪电战、925,636 10+0 快速游戏、以及 660,540 个较长的快速游戏。对于每种主要游戏类型，管道都会扩展 每个游戏进入移动前和移动后的棋盘状态，计算数百个 中间变量和玩家游戏变量，将这些变量聚合成 指导指标，然后按 Lichess Elo 存储桶对这些指标进行平均。的 样本是故意分层的，而不是与原始玩家成比例 池：目的是估计每个评级级别的稳定行为曲线，包括否则会稀疏的桶，而不是再现自然 评级分布。结果是一系列经验基准曲线：对于每个指标，我们可以询问不同 Elo 级别的玩家展示给定的频率 真实游戏中的习惯或原则。本说明描述了方法、规模 计算，以及一些 涉及王位铸造、小件开发的开放原则示例，重复的开局棋子动作和早期的皇后动作。重点是 方法论：Elo+Chess 并不假设国际象棋原理是普遍适用的 所有技能水平呈线性。相反，基准曲线通过以下方式检查 评级范围，产品侧重于初级到早期高级区域 关系最强、最容易解释的地方。

1 目的

基准系统的目标是将广泛的国际象棋建议转化为 可测量的数量。诸如“发展你的小作品”之类的陈述是 很有用，但是当我们可以说出一个有多少个小部分时，它就变得更具有可操作性 玩家通常通过第 10 步进步，与同等水平的同龄人相比如何 评级范围，以及该习惯如何在评级范围内变化。

因此，Elo+Chess 根据实际游戏构建了一个指标库。每个指标是 计算每个玩家游戏行，按游戏类型和 Elo 桶进行汇总，以及 用作报告中的基准曲线。然后，该报告可以将用户置于 相同的曲线并显示用户是否高于、低于或接近该行为 在可比较的玩家中观察到。

2 源数据和分层

基准源是公共Lichess.org游戏数据库。 Elo+Chess用途 按游戏类型和评级桶划分的大分层样本，而不是小样本 方便样品。当前的生产运行时工件涵盖四个主要 游戏类别：

- 1+0 bullet,
- 3+0 blitz,
- 10+0快速，
- rapid games longer than 10 minutes.

分层目标是评级级别的广度。比例随机 样本对于描述整个玩家池是有效的，但它会 构建基准曲线效率低下。最常见的评级桶 将主导样本，而低支撑尾桶的数量太少 稳定指标平均值的游戏。Elo+Chess 因此按平均游戏进行采样 埃洛桶。在当前的干净样本构建中，原始游戏是第一位的

筛选出具有已知玩家、已知评级、有效结果的有效评级游戏，和已知的时间控制。然后将游戏过滤给拥有足够游戏的玩家 以足够的速度和稳定的评级来代表连贯的技能水平：对于每个玩家，Elo+Chess 计算玩家的速度内评级跨度 ($\max(\text{Elo}) - \min(\text{Elo})$) 跨源池并保持 仅当双方玩家都至少具有配置的最小游戏数时才进行游戏，并且 速度内漂移不超过 300 Elo 点。漂移滤波器并不是指消除正常的改进或差异；它减少了玩家的情况 游戏会混合不同的技能水平，例如临时账户，返回的玩家，或在多个存储桶中快速移动的帐户 采样周期。最后，符合条件的游戏在每个平均 Elo 内排名 存储桶由游戏 ID 和种子的确定性哈希以及上限数量组成 每个存储桶中都会保留游戏。

这种设计使基准曲线更加有用。每个 Elo 桶 提供足够的观察结果来估计每个指标的平均值，罕见的高评级或低评级桶不会被自然形状淹没 活跃玩家分布。Elo+Chess使用的基准曲线是然后限制在与产品最相关的范围内，大致是初学者 通过早期的高级游戏。在 Lichess 规模报告中，该主要范围 是600–1600。当用户查看 Chess.com 规模的报告时，相应的评级桶映射到相同的底层Lichess 使用单独的跨平台评级映射方法进行基准存储桶 配套论文中描述：[Estimating Cross-Platform Chess Rating Mappings with 模态回归](#)。

3 活跃玩家评分分布

如果我们看看截至最后 Lichess.org 玩家 Elo 评分的分布 截至 2026 年 3 月，按主要游戏类型，我们看到以下计数和百分位数。对于此分布视图，每个玩家按确切的游戏类型计数一次，使用该玩家在相应原始数据中最新观察到的评分 时间控制行，在该游戏中至少需要五场比赛之后 型号。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: 最新观察到的活跃玩家 Lichess 评分分布（按精确分布）需要至少五场该确切游戏类型的游戏。

该分布不应与分层基准样本相混淆。该分布描述了在原始数据中观察到的活跃 Lichess 玩家池扫描。然后故意对基准样本进行分层，以便 Elo 桶 对稳定的度量曲线有足够的支持。

将特定游戏类型的 Lichess.org Elo 评级映射到 等效 Chess.com 评级，Elo+Chess 使用游戏类型特定模式 同伴中估计的回归方程 [Elo+Chess rating-mapping paper](#):

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

这些方程映射的是评级值，而不是百分位数排名。百分位数排名是在特定玩家池中定义，以及 Chess.com 和 Lichess 玩家 水池有不同的形状。在表 2 中，对于 10 分钟快速游戏类型，第一栏给出了可能的小样本 Chess.com Elo 额定值。第二列给出了相应的 截至 2026 年 5 月 31 日，Chess.com 上显示的 Chess.com 百分位排名。A 71.7

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 显示有和没有此过滤器的分布。

Lichess.org 10 分钟快速排名对应于 Lichess Elo 评级 1,644²，该值显示在 第三栏。如果我们应用 Chess.com–Lichess 映射方程，则 1,644 Lichess 相当于 Chess.com 上的约 1,271。假设估计映射方程合理，+522与Chess.com评级之间的差距 第 1 列和第 4 列中的映射评级可以用大量数字来解释 Chess.com 玩家池中的休闲玩家使 Chess.com 膨胀 百分位数排名。换句话说，在 Lichess.org 上处于第 72 个百分位数是比 Chess.com 上的第 72 个百分位数更大的成就，因为百分位等级不能清晰地从一个系统映射到另一个系统。实际Elo 可以使用根据同一玩家得分估计的方程来映射值 两个系统。

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: Chess.com 10分钟快速百分位之间的说明性比较 积分与主动Lichess 10+0快速积分相同要求至少五场10+0比赛后分配。“映射的 Chess.com equiv.” applies the current Elo+Chess 10+0 快速评级量表转换为相同百分位数的 Lichess 评级。的 包括最高尾行以显示尺度区别，但拟合的评级线在主要教练范围内是最可靠的。

表 2 中存在间隙的原因是映射 估计同等的演奏强度等级，而百分位数描述 在一个平台自身群体中的地位。71.7% 的玩家 因此，Chess.com 不会自动成为 Lichess 上第 71.7% 的玩家，反之亦然。观察到的数字与Chess.com公开的一致 快速池拥有更大的低评级或休闲玩家群体，而 主动精确控制Lichess池更集中在经验丰富的重复中 玩家。关键不在于一个百分位数是“正确的”，而且其他的是“错误”；这是百分位数排名不可跨平台移植 除非底层玩家分布具有相同的形状。

Note on the 1500 spike. The visible jump near 1500 存在于数据中，而不是通过平滑创建的。在最新的 观察评分分布，精确的 1500 个评分占 1+0 的 13.1

子弹型玩家、3+0 闪电战型玩家中 5.6

17.6

评级系统中的起始或临时锚点。

确切的 1500 个峰值很大程度上是由轻度活跃账户造成的。至 检查这一点，我们在要求后重新计算了相同的最新评级分布 每个玩家至少有五场相同游戏类型的游戏。尖峰几乎 在此过滤器下消失：确切的 1500 个评级从 5.6–17.6

在五场或以上比赛中，所有球员分布中的球员达到 0.2–1.2

表 1 中报告了分布。

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.

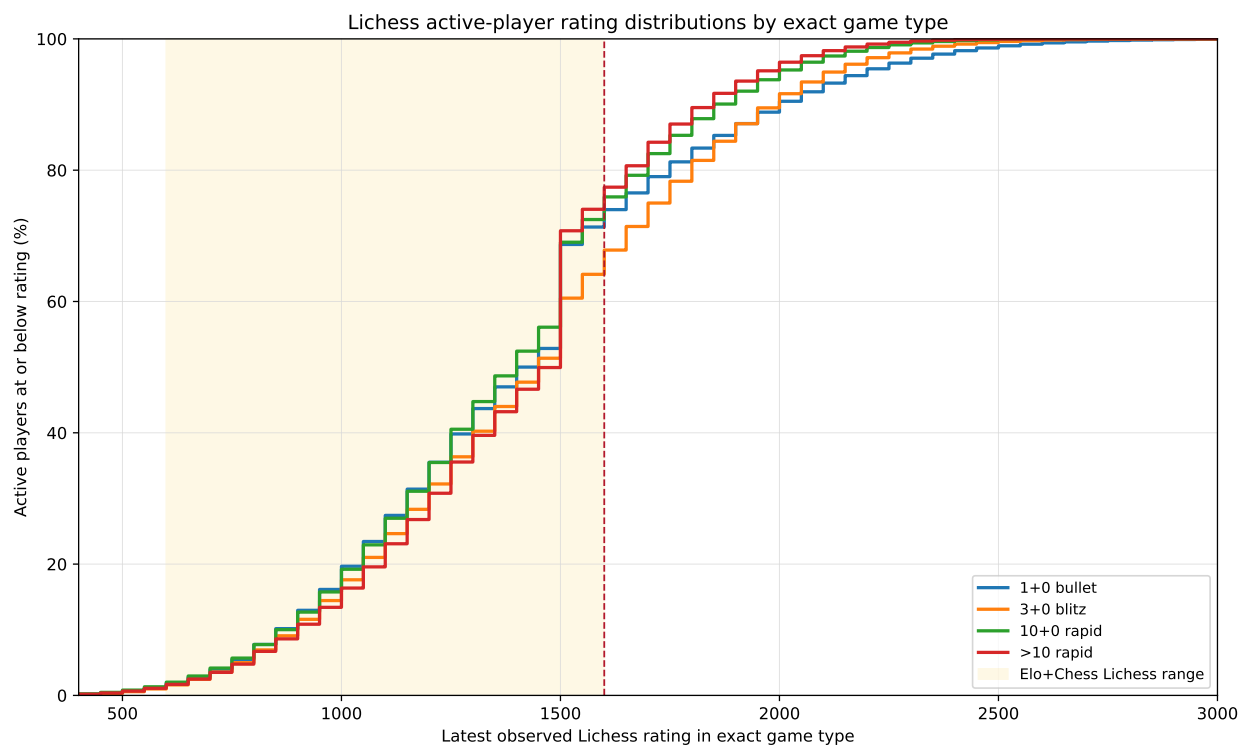


Figure 1: 按确切游戏类型划分的累计活跃玩家评分分布。的黄色带标记了主要Elo+Chess使用的600–1600 Lichess范围 教练基准。

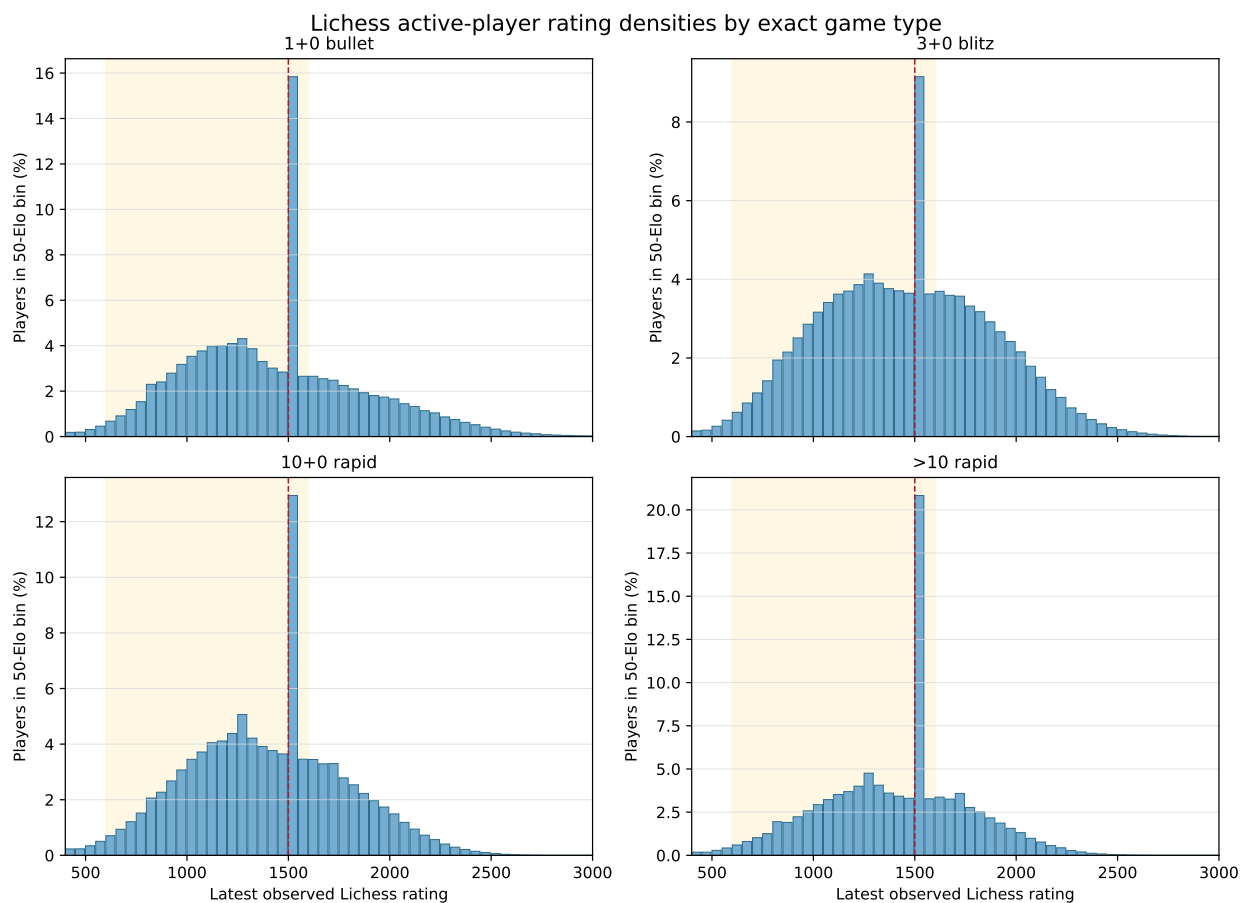


Figure 2: 相同活跃玩家评分分布的密度视图。的 虚线标记 1500 Lichess，其中原始最新评级分布一个大的精确额定质量点。黄色带标记600–1600 Lichess 主要 Elo+Chess 指标报告使用的范围。峰值最明显的是 所有玩家的密度，因为许多轻度活跃帐户仍然保持在 1500；图 4重复密度图 在要求至少五场特定游戏类型的游戏之后，这在很大程度上消除了 这个神器。

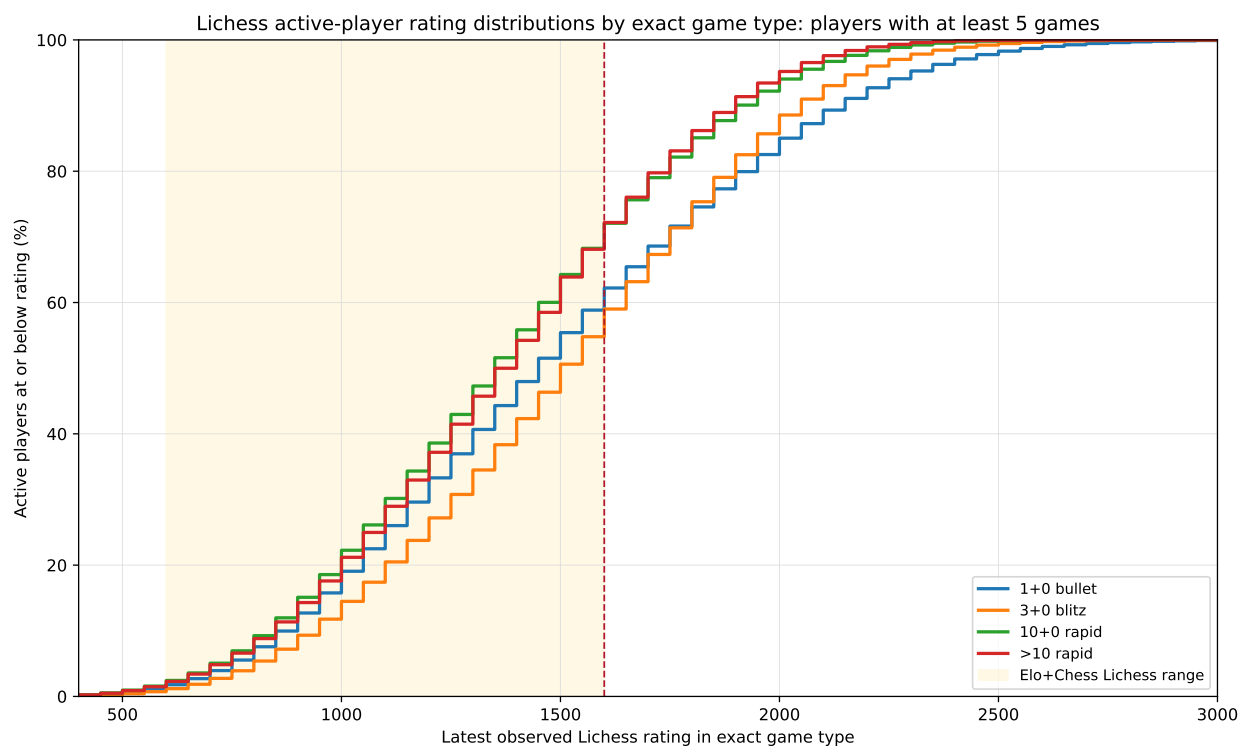


Figure 3: 要求后的累积活跃玩家评分分布 确切游戏类型中至少有五场比赛。 1500跳大大减少。

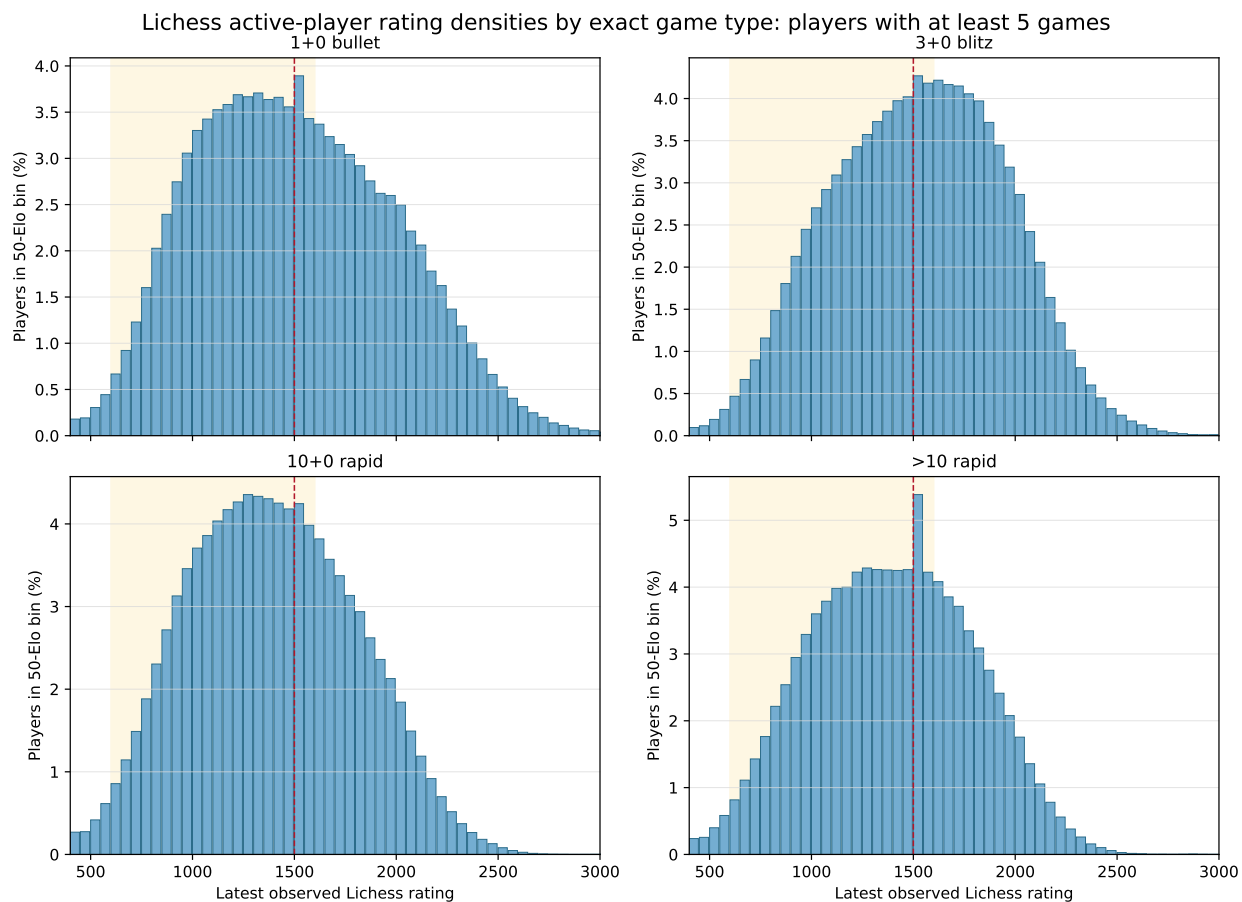


Figure 4: 要求在同一场比赛中至少进行五场比赛后的密度视图 类型。该视图证实了全玩家中的大 1500 个质点 分配主要是由活跃度较低的帐户驱动的。黄色带标记 主要 Elo+Chess 指标报告使用的 600–1600 Lichess 范围。

4 当前基准工件的规模

计算量故意很大。这不是轮廓刮伤，而是一个小 一组下载的 PGN，或手工策划的指导性示例集合。初始 Lichess 通行证涵盖 2025 年 1 月起的完整月度文件 截至 2026 年 3 月。非分层扫描表包含 1,382,178,087 场比赛 以及 2,764,356,174 个玩家游戏评分行。在这些行中，有 6,623,438 个唯一 玩家用户名至少出现一次。

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: 基准采样之前未分层的 Lichess 扫描比例。扫描 涵盖 2025 年 1 月至 3 月完整的每月游戏历史文件 2026 年。

生产基准测试工件比原始扫描小得多，因为 它们被选择用于公制结构。它们仍然很大：基准 stage 是一个超过数百个的移动状态扩展和聚合管道 每种游戏类型有数千个游戏。

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: 对齐后分层基准工件的规模 常见的 277 列运行时特征模式。一个“玩家-游戏行”是指从一个玩家的角度来看一场游戏，因此大多数游戏 产生两个玩家游戏行。

仅就 3+0 闪电战而言，当前的生产基准就使用了 479,974 场比赛 涉及 129,398 个独特的玩家和 28,138,904 个不同的移动状态行。的 对齐的闪电战玩家游戏特征表包含 277 个派生变量。的快速功能构建工件在移动状态阶段甚至更广泛：完整的 特征构建期间使用的快速移动状态表包含 168 聚合成 277 列玩家游戏功能之前的移动状态列 表。所有四种游戏类型的前端运行时基准工件是 然后与相同的 277 列玩家游戏特征模式对齐，包括 扩展的战术分类突破变量。从这些特征表中， 报告选择并显示最容易解释的辅导指标。

这些计数对于解释结果很重要。基准曲线 不是根据几百个人工审核的游戏来估计的。相反，他们 来自数以千万计的评估委员会国家和超过四个 当前四个制作游戏类别中的百万玩家游戏行。

5 管道概览

标杆建设流程有六大步骤：

1. 在目标游戏类型和评级层中选择评级为Lichess的游戏。
2. 逐步解析游戏的每一步棋并存储棋步前和棋步后的信息 董事会状态。
3. 从棋盘计算中间移动状态变量，移动， 玩家视角和对手视角。
4. 将移动状态变量聚合为玩家游戏特征变量。

5. 将玩家游戏变量转换为清晰的报告指标 解释方向：通常越高越好，通常越低 更好，或者描述性的。
6. 按 Elo 存储桶平均每个指标并手动检查结果 关系。

这种结构很重要，因为许多报告指标并不直接 在 PGN 标头中可见。他们需要重建位置，确定哪些棋子从哪些起始方格移动，比较移动前和 移动后的材料和结构，然后汇总结果 将级别移至玩家游戏级别。

6 中间变量

中间变量旨在描述板上发生的变化 以及为什么从报道者的角度来看它很重要。示例包括：

- 移动编号、玩家移动编号、移动方向、移动颜色和玩家 颜色；
- 从方格、到方格、移动棋子类型、捕获棋子类型以及 此举是否为占领、过牌、城堡、晋升或交配；
- 搬迁前和搬迁后的物料平衡；
- 小棋子是否离开起始方格给定的步数；
- 同一棋子是否在连续的玩家移动中移动 开幕；
- 皇后是否在第 10 步之前移动；
- 易车侧、易车移动次数以及车是否已连接 易位后；
- 松散棋子、悬挂棋子、双棋子、孤立棋子、通过 兵和后兵；
- rook 文件状态、打开文件、半打开文件和 king-safety 特征；
- 战术机会、执行战术、物质压力战术、 保证后续收益，实现收益。

确切的中间集因构建阶段而异。运行时报告数据库保存 只有快速提供报告所需的列，而完整的功能构建 工件包括更广泛的移动状态表和用于构建的功能补丁 最终的玩家游戏特征行。

7 开放原则度量示例

开放原则是有用的例子，因为它们很容易让用户理解 理解并易于连接到移动状态计算。

7.1 Minor Pieces Developed by Move 10

该指标计算玩家的四个骑士和主教中还剩下多少个 玩家第 10 步移动时，他们的主场方格。中间计算 使用移动者颜色、移动棋子类型、起始方格和玩家移动 数量。对于白方，本方格为 b1、g1、c1 和 f1。对于黑方来说，他们是 b8、g8、c8 和 f8。对于每个玩家游戏，Elo+Chess 计算不同的主场 由玩家移动 10 移动的骑士和主教之间的方格。

这是一个通常越高越好的指标。它并不声称每个未成年人 棋子必须始终按第 10 步移动，但在发展中的玩家中，经验值 关系很明确：达到更高 Elo 等级的玩家往往会发展 更早的小作品。

7.2 Repeated Piece Moves in the Opening

该指标计算了再次移动同一棋子所花费的额外早期移动 开发新东西。中间计算使用之前的 目的地方格、当前起始方格和玩家移动编号。如果 前一个玩家移动时移动的同一棋子，在该棋子内再次移动 玩家前 10 步移动后，重复移动次数会增加。

这是一个通常越低越好的指标。它测量浪费的打开速度 简单的方法。一些重复的举动在战术上是合理的，但从规模上看 桶曲线捕捉了总体趋势：目标上更强的玩家 关卡早期移动重复移动同一块的时间更少。

7.3 Queen Touched Before Move 10

该指标计算玩家移动 10 之前的皇后移动。计算使用移动棋子类型和玩家移动数量。如果移动件是皇后且玩家的移动次数最多为 10 次，则计数增加。

对于 Elo+Chess 目标范围来说，这通常也是越低越好。早期 皇后动作可以赢得物质或力量让步，但许多发展中的玩家 太早带出女王，失去攻击女王的节奏，并拖延 小件开发，安全为王。

7.4 Castling and King Safety

易位指标使用易位标志、易位方位、易位步数和 易位后的棋盘状态。例子包括玩家是否城堡，玩家是否通过第 10 步攻势，无论玩家在王翼还是攻势 后侧，车易位后车是否连接，以及车是否文件易位后呈开放式或半开放式。

这些指标的解释方式并不相同。易位完成及 第 10 步的易位通常在目标范围内越高，通常越好。后侧易位或易位至开放文件可能更依赖于上下文。这就是为什么 Elo+Chess 不是简单地硬编码道德规则；而是简单地硬编码道德规则。它建造了水桶 曲线并检查哪些关系对于指导来说足够稳定。

8 铲斗曲线和手动检查

计算玩家游戏指标后，Elo+Chess 按 Lichess Elo 对它们进行分组 桶并计算桶平均值。桶平均值是原始经验值 该指标的基准点。然后在报告中使用时趋势线来 将 Elo 风格的等级分配给用户值。

曲线在使用前经过手动检查。指标更有用 当桶均值在整个评级范围内显示出一致的关系时 由产品服务。当一个指标有噪音、平坦、高度集中时，它的用处就较小。非单调的，或者由难以向用户解释的特殊情况驱动的。

这个手动检查步骤在Elo+Chess鞋面上方尤其重要 截止。更强的玩家知道何时打破开局原则并这样做 有效地。在初级到早期高级范围之上，关系 简单的习惯和评级之间经常会变平或改变形状。产品 因此强调经验关系较强的范围 并且在教学上有用。

9 全范围开度曲线示例

图 5展示了4个3+0 blitz开启原理曲线。灰点表示更宽的 Lichess 额定范围。蓝点和蓝色趋势 线路显示主要Elo+Chess教练使用的600–1600 Lichess范围 基准。红色虚线标记了上截止线。

同样的模式在全范围10+0快速和更长的快速中可见 文物。中心建模决策不变：使用经验 教练主张的稳定范围，避免假装简单的习惯曲线 永远保持同样的信息。

表 5 中的数字也出现了相同的模式。坡度是 报告为每 100 个 Elo 点的指标值变化。易位，小件发展、重复棋子运动和早期皇后运动都显示清晰 运动跨越600–1600。高于 1600 时，这些区域的坡度要小得多 例子，这与评级较高的玩家打破的想法是一致的 更有选择性、更成功的基本原则。

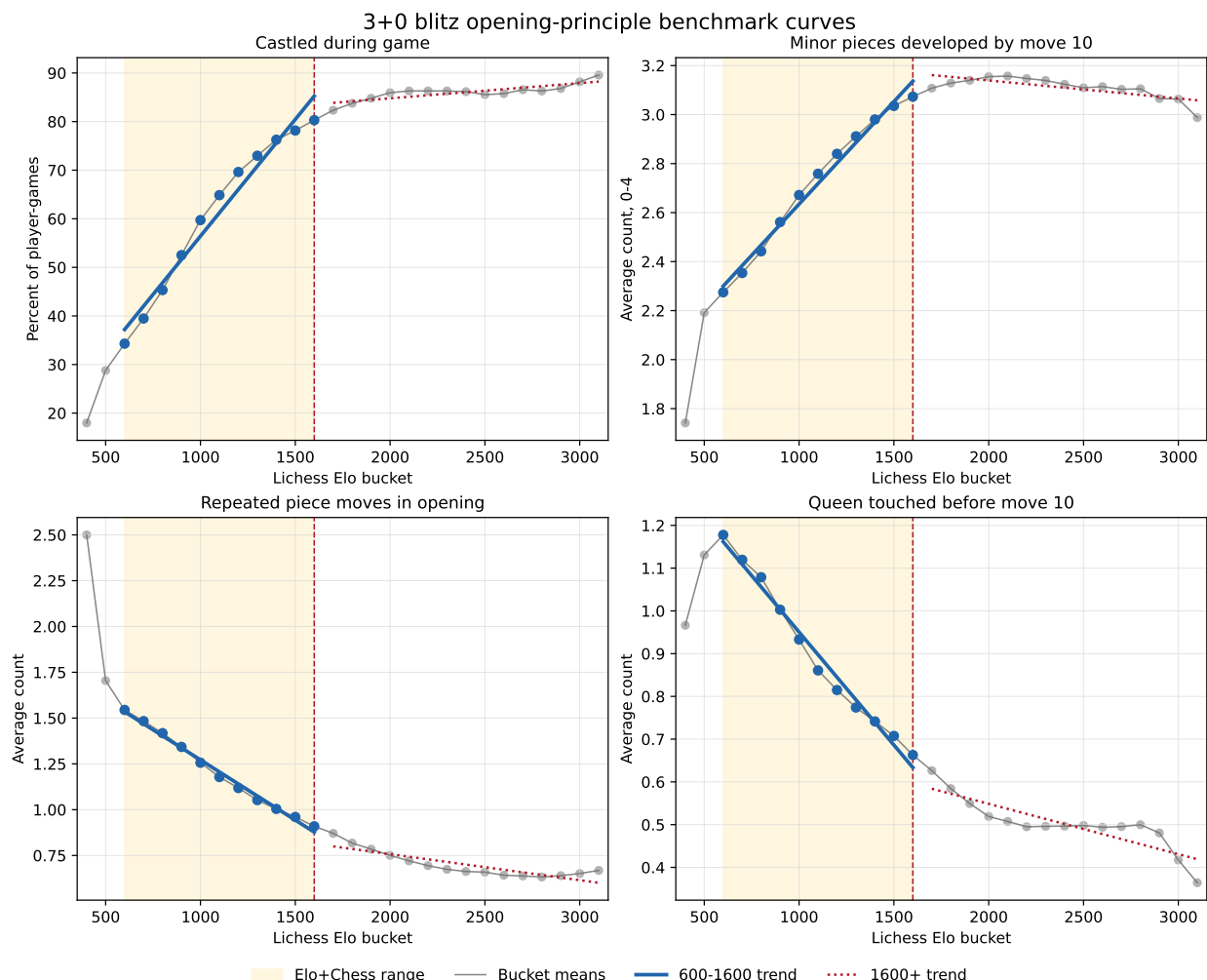


Figure 5: 3+0闪电战精选开放原理基准曲线。的 在 Elo+Chess 教练范围内，关系是最牢固、最有用的。高于上限时，一些关系会大幅趋于平缓。

第二个机制是 饱和度。对于某些习惯，该指标达到了实际上限或甜蜜 现货：强者已经养成习惯了，没理由 曲线继续急剧向上移动。更多并不总是更好 那一点。因此，数据会趋于平缓，因为更强的玩家可以打破 规则在具体的位置上，因为很多基本的习惯已经达到了 他们的有用操作范围。

10 为什么上限很重要

Elo+Chess的目标用户是初级到初级高级玩家。对于这些 玩家的基本习惯与评级密切相关：开发 棋子，避免重复的早期棋子移动，避免不必要的早期皇后移动，城堡 及时连接车、管理材料并避免留下松散的零件或 悬挂。

在较高的收视率下，那些相同的表面习惯变得不那么具有诊断性。一个强大的 由于中心关闭，玩家可能会延迟易位，请尽早移动王后 因为该位置包含具体的战术，或者将同一棋子移动两次 因为它赢得了节奏或迫使结构性让步。原始习惯依然存在 有意义，但是习惯和评级之间的关系变得更加重要 有条件的。这就是为什么 Elo+Chess 使用上限作为主要训练的原因 声明以及为什么高于该范围的用户会被警告说紧密关系 原则习惯和评级之间的关系崩溃了。

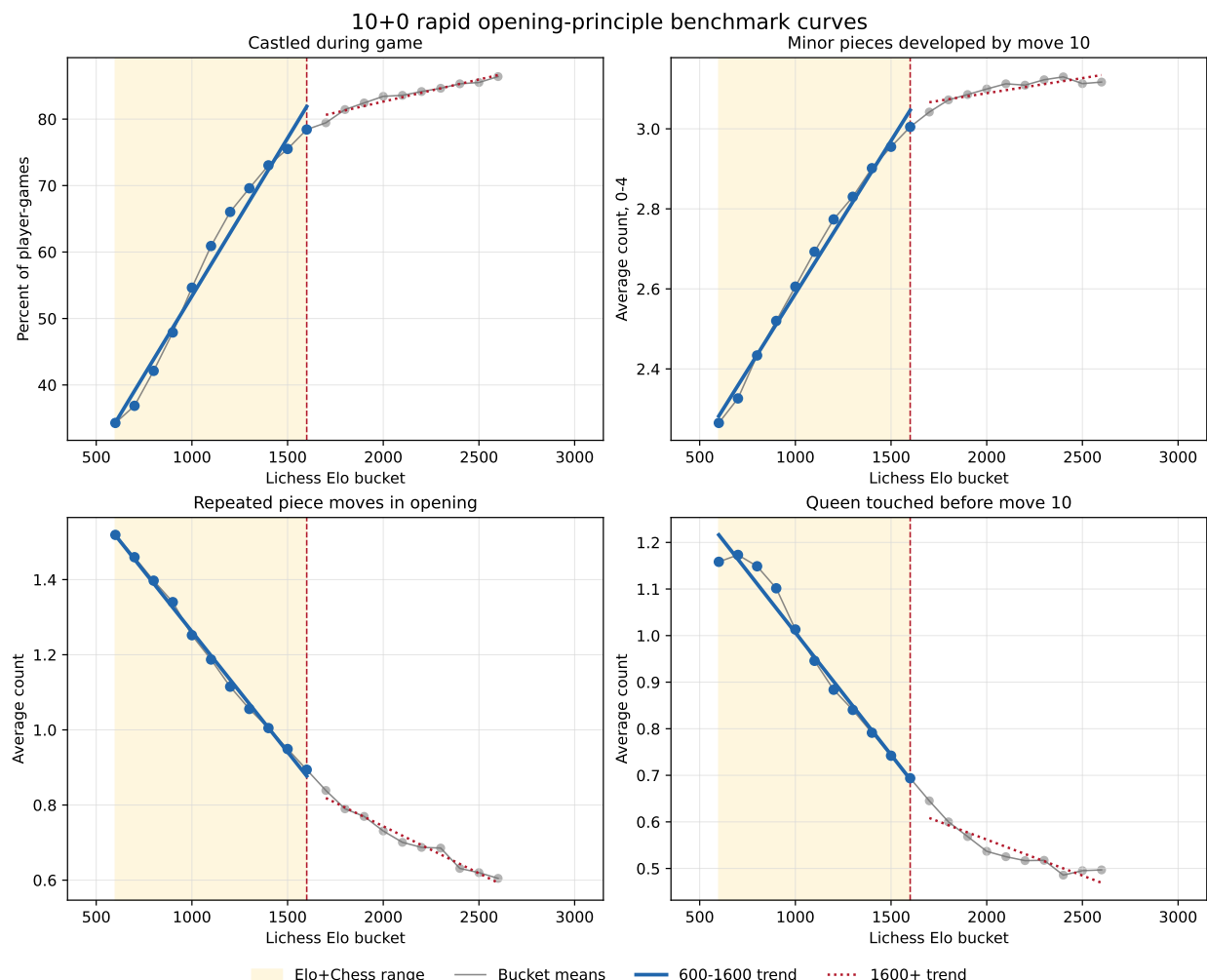


Figure 6: 10+0快速游戏的开局原理基准曲线精选。的 在 Elo+Chess 教练范围内，关系是最牢固、最有用的。高于上限时，一些关系会大幅趋于平缓。

11 双边战术指标和百分位目标

某些指标的形狀与开放原则示例相反。叉子 是一个有用的例子，因为事件本质上是双向的。一位球员 发现并转换比同行更有用的分叉是在做好事，但是 转换后的叉子的数量随着评级的增加而下降，因为更强 对手允许更少的分叉机会并防御战术威胁 更积极地。

图 8展示了这个效果。两个改装货叉均由 报告玩家和对手转换的叉子变得不那么频繁，因为 收视率上升。这并不意味着执行分叉是不好的。这意味着原始事件 比率部分由对手错误率和游戏纹理控制。对于指标 像这样，Elo+Chess 因此在 玩家自己的 Elo 桶，而不是仅依赖于全球桶平均趋势。

为了构建这些百分位数目标，Elo+Chess 按游戏类型对玩家进行采样，Elo 桶，然后使用每个样本球员最近的资格赛。的 当前的目标集使用 11 个桶，从 600 到 1600，每个桶 200 名玩家 桶，以及每个球员最近最多 50 场资格赛 游戏类型。这为每种游戏类型提供了 2,200 个玩家样本，大约 每个游戏类型的桶内百分位有 110,000 个玩家游戏行 分布。

图 9 说明了最终的目标 3+0 闪电战中采样玩家转换分叉率的分布。每个盒子都是 样本玩家的桶内分布，而不是桶内的平均值 个人比赛。在报告 UI 中，将用户的价值与相关值进行比较 用户

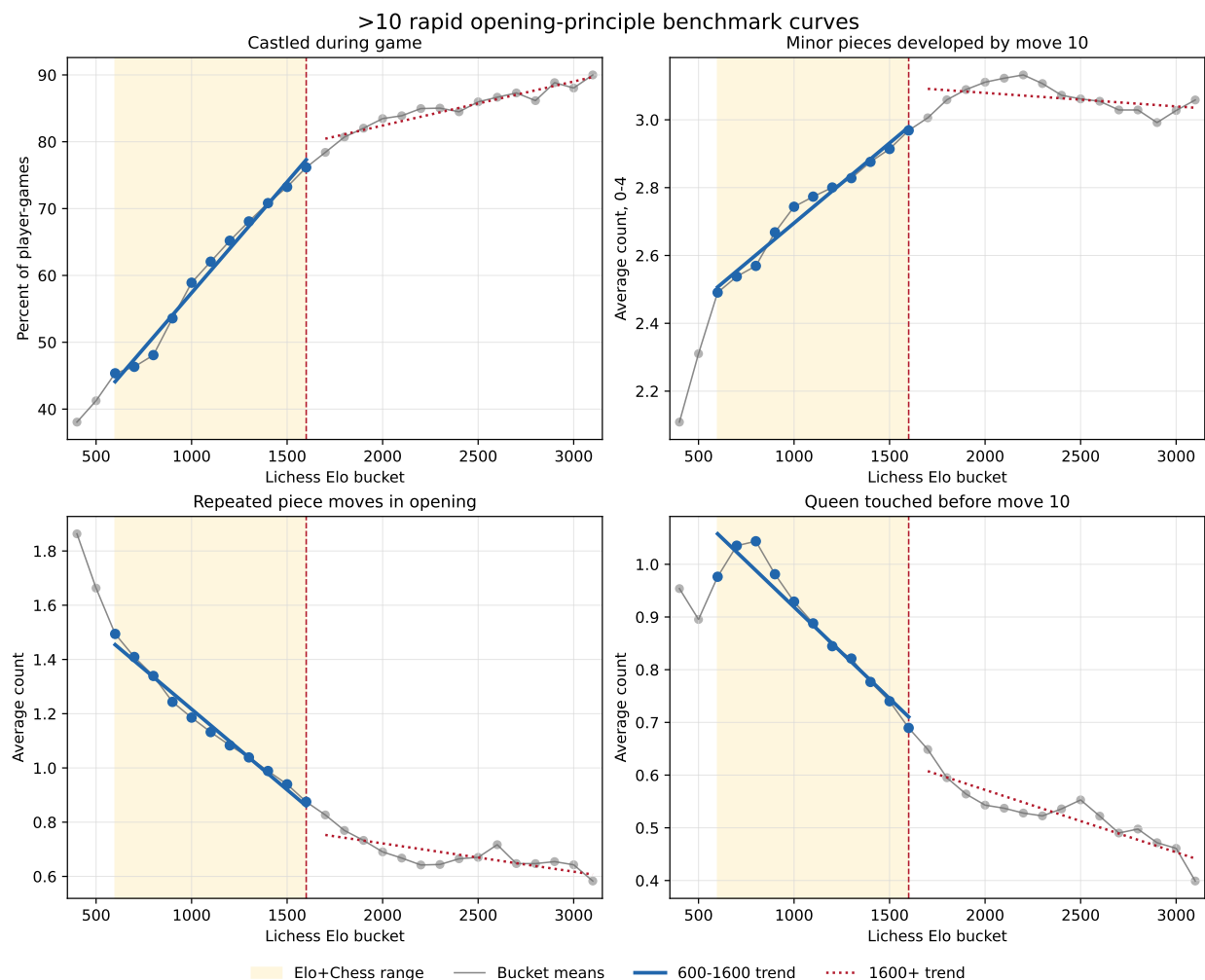


Figure 7: 为较长的快速游戏选择的开局原理基准曲线。

映射的 Elo 存储桶的框。这是诸如此类的陈述的基础 相对于同龄人来说，玩家的分叉转换率是高还是低 相同的水平，即使总体人口事件发生率随评级而下降。

12 复制成本

该方法原则上是可重复的，但不是轻量级的。重建 基准曲线要求：

- 访问大型 Lichess 每月游戏历史文件；
- 按游戏类型、评级范围和资格进行分层过滤 规则；
- 能够扩展的合法棋步解析器和局态评估器 数以千万计的移动状态；
- 开局、材料、典当结构、国王安全、车的特征逻辑 活动、战术、时间管理和比赛结果；
- 从移动状态到玩家游戏特征的聚合；

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: 四个示例开口原理的近似桶曲线斜率 跨制作游戏类型的指标。

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: 当前同行历史百分位数目标集。这些是 用于估计桶内分布的玩家历史样本 百分位式指标。

- 每个报告指标的桶级摘要；
- 在使用每个候选基准曲线之前对其进行手动审查 教练报告。

对于3+0闪电战来说，这意味着近50万场比赛，近100万场比赛 玩家游戏行，以及超过 2800 万个移动状态行。跨越四 基准测试游戏类型，当前基准神器包含超过2.55 万个游戏，超过 510 万个玩家游戏行，以及超过 147 个 万个移动状态行。规模是价值的核心 基准测试：它让报告向用户展示游戏中实际发生的情况 由接近他们水平的棋手来决定，而不是仅仅依靠一般的国际象棋口号。

13 局限性

基准曲线是经验描述，而不是因果证明。一名球员 仅仅通过匹配桶平均值并不能获得评级。指标也是 复杂局面的不完美总结：早期的皇后棋步可能会非常出色 在一种立场上而在另一种立场上鲁莽。曲线的目的是 识别反复出现的习惯并将其与同伴行为进行比较，而不是为了 取代计算或指导判断。

这些曲线也是 Lichess 比例曲线。当用户携带Chess.com游戏时 历史记录，Elo+Chess 将相关评级桶映射到 Lichess 基准 使用单独的跨平台映射模型进行缩放。该映射层是 在 Lichess-to-Chess.com 的配套 Elo+Chess 研究说明中进行了描述 评级转换：[Estimating Cross-Platform Chess Rating Mappings with 模态回归](#)。此映射不应被理解为保留百分位 排名。它保留了两个评级之间的估计评级等效性 系统；百分位数排名仍然是特定于平台的，因为活跃玩家 分布不同。

Two-sided converted-fork rates by Elo bucket

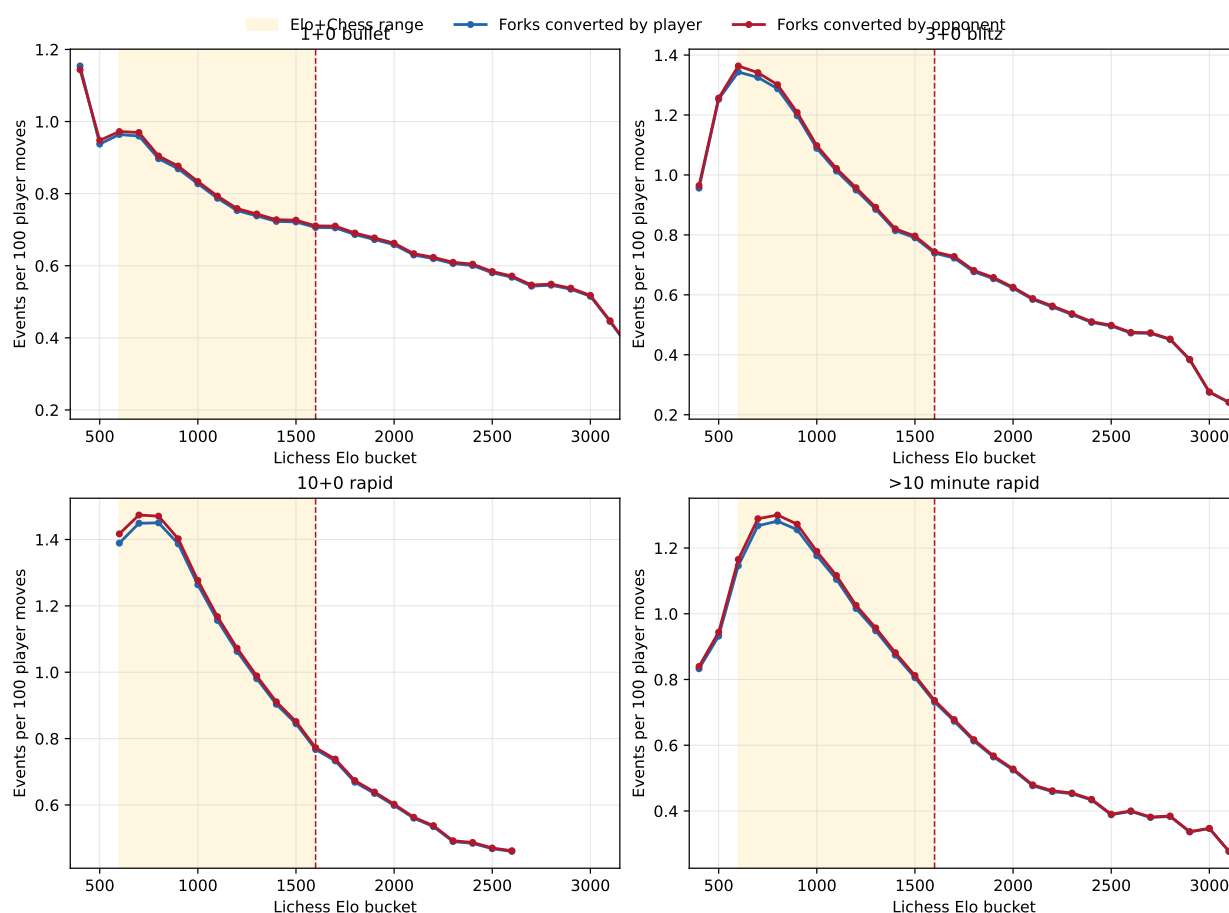


Figure 8: Elo 桶的两侧转换叉率。转换后的分叉事件 由于双方防守更好且干净叉子更少，收视率下降 可以进行转换。

14 生产报告图表

生产报告将基准工件转变为指标级别的面板 无需了解完整的构建管道即可阅读。人物 ~ 10 and 11 显示来自实时 Elo+Chess 报告设计的三个示例。

图 10 开启原理面板 使用普通的基准曲线演示。每张卡都以 度量标准系列和度量标准名称，然后给出简短的解释规则。的 右上角的大等级是故意与原始的分开的 指标值：它回答了用户面临的问题“这个习惯如何评分”与基准相比？”，而下面的行保留测量值。四个摘要行显示“一生”、“过去 10 场”、“胜利”和“失败”。这使得 面板诊断而不仅仅是装饰。如果 Lifetime 和 Last 10 一致，这个习惯可能是稳定的。如果输赢不同，衡量标准可能是 取决于结果或与游戏状态相关，而不是简单的习惯。

图表区域使用内敛的深色背景、金色趋势线和彩色 四个切片的标记。趋势注释明确指出是否 基准关系随 Elo 上升或下降。这很重要，因为 有些指标越高越好，例如小件开发，而 其他则是越低越好，例如早期的皇后运动。因此该卡 不要求用户单独从坡度推断方向。决赛“这意味着什么”和“指导提示”行翻译了基准结果 转化为具体的国际象棋教学，同时保留经验背景。

图 11 显示百分位式 用于原始率不能很好地表示的指标 单一的全球趋势。获胜分叉是一个战术示例：人口率 转换后的叉子数量会随着玩家技能和对手错误率的变化而变化。的 因此，卡

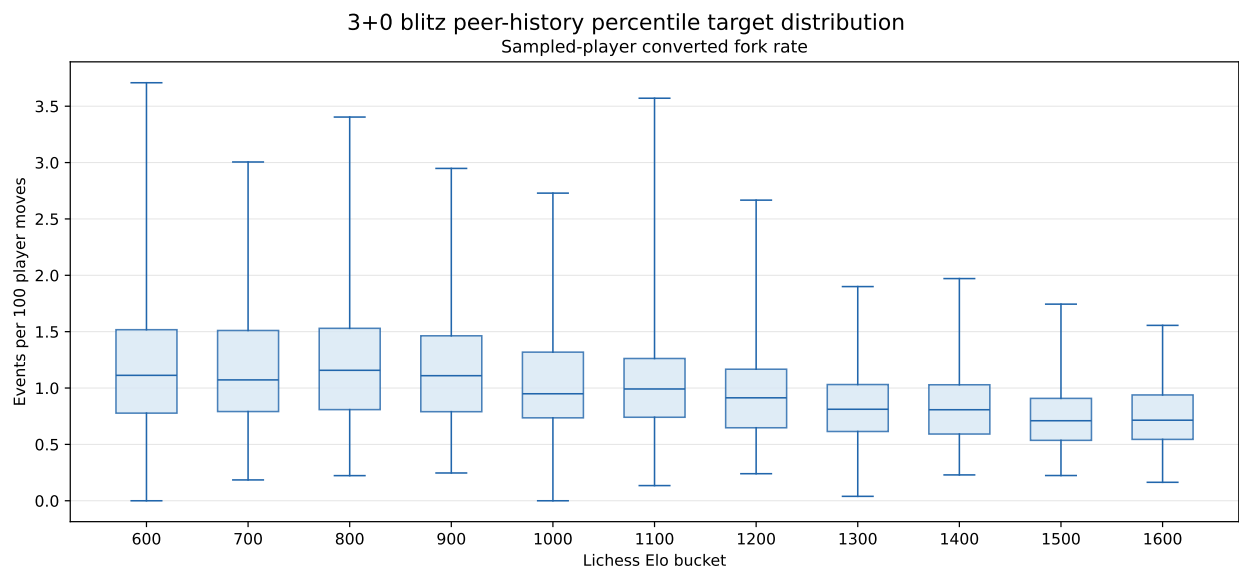


Figure 9: 样本玩家的 3+0 闪电战同行历史目标分布示例 转换后的分叉率。百分位数是在 Elo 桶内计算的，因此报告 将球员与面临大致相似战术机会的同龄人进行比较 费率。

将用户的价值与同行玩家的历史记录进行比较 映射的 Elo 桶。箱线图显示了对等分布，包括 四分位距、中位数和须线。用户的标记绘制在 相同的尺度并标有原始值和百分位等级。相邻 比较桶以静音形式保持可见，因此用户可以看到 对等存储桶是本地的，而不是通用规模。

这种设计使百分位数指标可审计。用户看到原始值， 百分位数、用于比较的对等存储桶以及翻译后的值 Chess.com 到 Lichess 存储桶映射。相同的一生、最后 10 场胜利、以及 保留损失行，因此仍然可以检查战术百分位数分数 近因效应和结果依赖性。

15 结论

Elo+Chess 基准报告是根据大量移动状态计算构建的 分层 Lichess 样本。每个指标都以一组具体的板状态开始 和移动状态变量，成为玩家游戏特征，然后进行平均 通过 Elo 桶形成经验基准曲线。开放原则 例子说明了为什么这种方法有用：简单的国际象棋习惯具有很强的， 与目标范围内的评级有明显的关系，而那些 高于上限的关系通常会趋于平缓或变得更加有条件。这就是 Elo+Chess 将其教练主张集中在评级上的核心原因 数据最清楚地支持它们的范围。

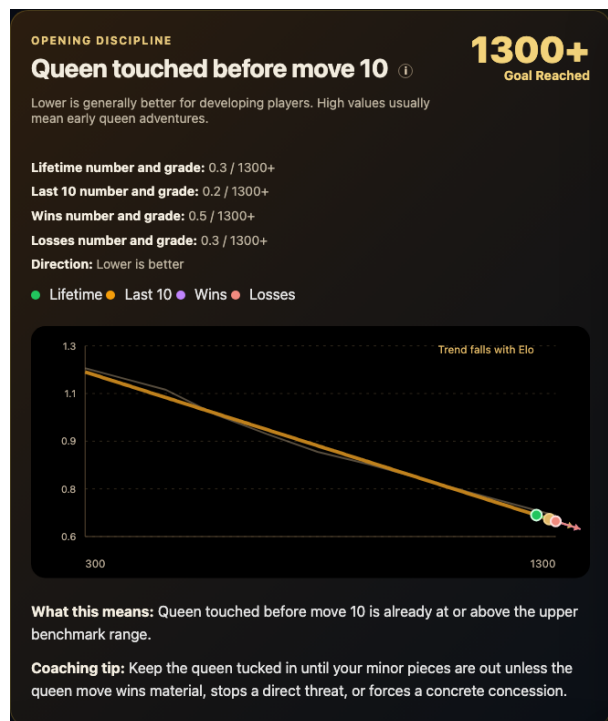
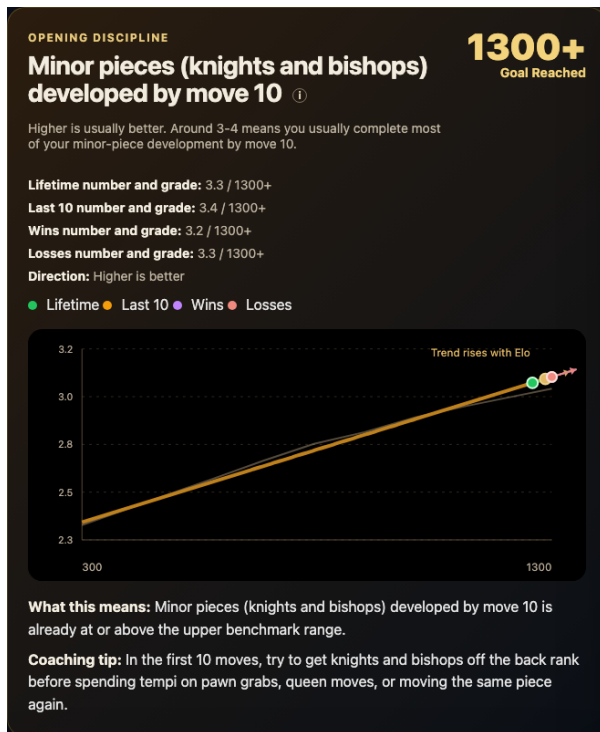


Figure 10: 生产开口原理面板。相同的视觉语法处理 越高越好的指标、小件开发和越低越好的指标，早期的女王运动。两个面板都展示了一生、过去 10 场、胜利和失败 因此报告可以将稳定的习惯与最近的或依赖结果的习惯区分开来 模式。

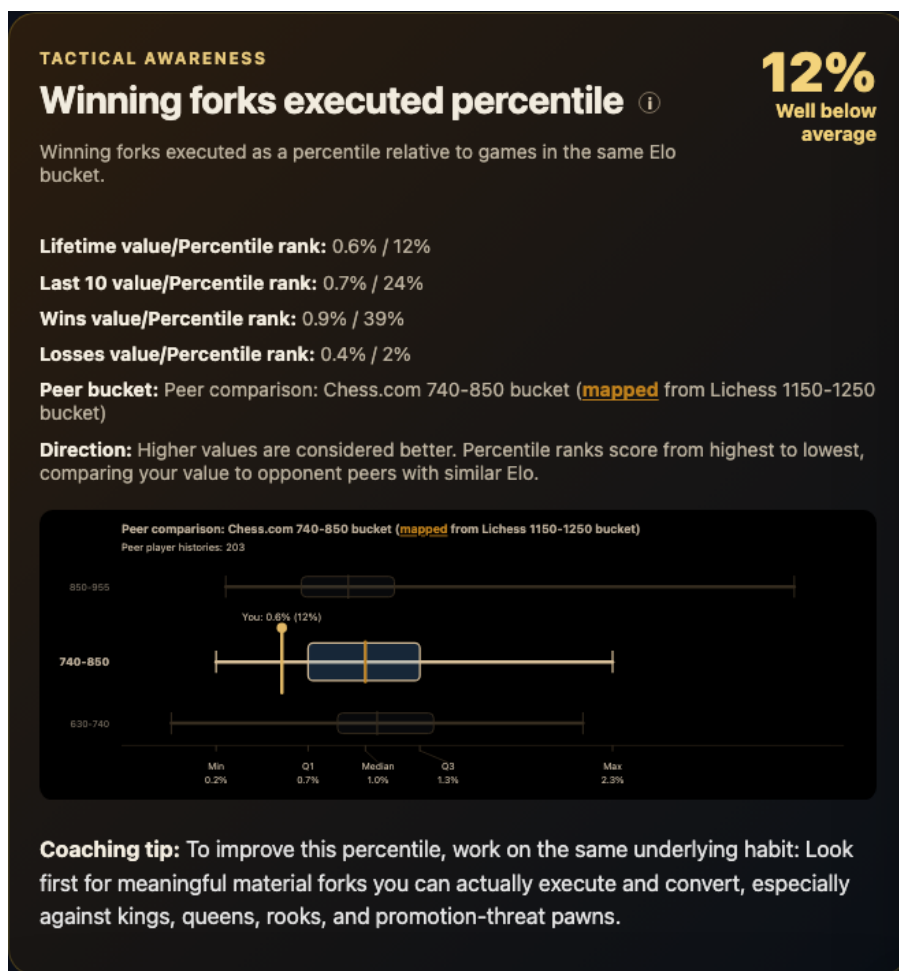


Figure 11: 执行获胜叉子的生产百分位面板。箱线图 显示了报告使用的同桶同行分布，而黄金 标记显示用户的原始值和百分位数排名。