# What Can 100,000 Chess Players Tell Us About Improvement?

Andrew Ashman[1], Soo Park[#] and Daoud El Dibani[#]

[1]Advanced Technologies Academy, USA
[#]Advisor

## ABSTRACT

The role of memorized chess openings in achieving mastery has long been debated. This study attempts to prove a relation between exposure to different phases of a chess game and the change in skill over the course of a year. Using the Lichess open database, the researcher analyzed game lengths and changes in Elo over a one year period for accounts ranging from beginner to advanced. The method isolates length of chess games from the Elo measurement to allow for tracking of persistent variables over the period of the year. This allows for the examination of a player's tendency towards longer or shorter games, which was compared to their change in skill over the year. Contrary to my hypothesis, the results suggest a near zero correlation between length tendency and change in Elo relative to peers.

## Introduction

"Why do you hate Chess? Being the be- probably, possibly, the best Chess player ever?" an interviewer asked Bobby Fischer who was nearing the end of his life. "Because I know what Chess is all about! It's all about memorization. It's all about pre-arrangement" (Bobby Fischer on Paul Morphy and How Opening Theory Destroyed Chess, ca. 2005). Though Fischer was largely considered a recluse and unstable at the time, his assessment on openings certainly has some merit in modern day; the current highest ranked chess player, Magnus Carlsen, confided that studying openings takes up 80% of a player's time (Magnus Carlsen, 2011). The opening stage of chess generally consists of theory, or book moves that have already been studied. The best moves during this stage of the game are often memorized as far as possible to gain an advantage, sometimes pushing 30 moves (Lars Bo Hansen, 2008). I hypothesize that this problem has only gotten worse, as modern resources have key differences. Since World Chess Champion Kasparov lost to the computer Deep Blue in 1997 chess knowledge has been democratized (Kasparov, 2010). The average person can access a computer capable of running a chess engine that is superhuman in performance. This rids the necessity of opening analysis through masters, and instead places it in the hands of chess engines. This has many implications from learning outcomes to equity, but most notably for this research, it means that a modern chess player can memorize the best moves as far as they see fit. I hypothesize that this new ability to spend time memorizing openings with a computer predicts poorer learning outcomes. Not only does the chess engine do much of the analysis formerly left to the player, but players direct their focus away from deeper, more involved learning techniques more heavily supported by the body of literature, such as meaningful and active learning. This study seeks to prove this correlation between the opening and poor learning through the use of novel datasets and analytic techniques.

## Literature Review

The acquisition of chess knowledge has been of great interest in the study of problem-solving methods in the brain (Van Der Maas & Wagenmakers, 2005). Many psychologists have studied chess masters to gain insight into thinking itself (Morales, 1996; Grabner et al., 2007). Chess fits this role well due to its inherent easily measured properties (Van Der Maas & Wagenmakers, 2005). Part of what makes modeling chess useful is how long the learning process is. It takes years of effortful practice to become a grandmaster in the game and longer to become world-class, much in the same way as expert musicians or mathematicians (Ross, 2006). This allows for longitudinal studies to become very effective for testing learning methods, their key weakness being the cost of data collection and the time investment. Likely because of this, there are few reputable longitudinal studies data.

A longitudinal analysis by Howard (2011) even found that self-reported logged hours of studying had a very low correlation with chess skill, concluding that the number of games played made a far better prediction. Likewise, a ten year study by Gaschler et al. (2014) on German players found similar though less extreme results, with some improvement that wasn't modeled by practice. This opposed prior literature which found significant correlation with length of study (Gaschler et al., 2014). Chowdhary et al. (2023), however, took a novel approach. They employed the Lichess open database to analyze a year's worth of games to determine the predictive value of winning streaks. This database contains statistics about every game played on the website since its creation in 2013 (Duplessis, 2024). Chowdhary et al. (2023) further explored this structure by analyzing the diversity and specialization of opening sequences among players of different skill levels. They observed that expert players tend to specialize in a narrower range of openings compared to beginners, suggesting that expertise involves developing a deeper understanding of specific opening variations and their strategic implications. They also found that beginners tend to go on longer winning streaks than advanced players. This method addresses both difficulties with the longitudinal approach: cost and time investment. Though, while innovative, it fails to bridge the gap of longitudinally studying improvement over time. Further chess literature appears rife with studies employing the Lichess open database that study chess more so than the players.

For example, Holdaway, C., & Vul, E. (2021) were able to create a clever operational definition to model risk taking behavior. It was as simple as measuring the number of possible responses from the opponent, with the more responses, the riskier the move. The logic is that as the breadth of necessary knowledge increases it limits the possible depth. With lower depth means less overall knowledge about the state of the game, hence the risk. Although the Lichess database reports limited statistics, through clever operational definitions, more complex properties can be modeled.

Curiously, none of them have attempted to track player progress over time with the massive collection of data despite the poor understanding and disagreement prevalent with prior longitudinal studies. My study will aim to address this gap with the application of the prominent psychological models used in current chess improvement studies to Lichess datamining. The goal is to merge the two fields so that the Lichess database can be used as a source of data on long term learning. There are two notable theories that chess could provide insight to. Active learning is when a person is conscious of and in control of their learning experience (Armstrong, 2010). Similarly, deep or meaningful learning occurs when one relates knowledge to what one already knows. The goal of education, according to Jonassen & Strobel, 2006 is to achieve meaningful learning. These learning structures are the goals of education due to the high correlation with success and enjoyment. Contrary to this notion, rote learning is when one learns with no effort to integrate concepts with already existing knowledge, often by simply repeating phrases in an attempt to keep them in memory. This form of learning is far more easily forgotten. King and Russell (1966) demonstrated this with an experiment in which participants attempted to memorize a text. One group was instructed to attempt to memorize it word for word, and the other was instructed to just remember main ideas. The latter group was able to recall more information overall.

These definitions can be expanded with the learning model of effortful practice. Ross, 2006 reinforces that it takes years of effortful practice to become a grandmaster in chess and longer to become world-class, much in the same way as musicians or mathematicians. He touts the ten year rule of effortful practice as the mark of a master. He also brings up the fact that people don't improve in driving despite practicing it every day. That's an example of non-effortful practice. Similar findings have been independently reached by Ericsson, K. A. (2008). They found that any improvement in performance comes from the integration of complex intersensory performance only achieved through deliberate practice. Ericsson applies this to chess with a simple learning strategy: replay a chess master's game and predict their moves. If your move matches theirs, then there's nothing to change. If it doesn't, then you must reevaluate your thought process. This method of learning can be thought of as similar to a progressive overload in weight lifting, where you train heavier and heavier weights. In this model, the reason people don't get any better at driving over time is because there is no new challenge.

## Methodology

The design of the study will feature a correlational analysis between the length of chess games for an account, and their improvement over a year. This study will require two operational definitions. The first is the measure of chess skill that will be tracked over time. The Glicko-2 rating system is included in the Lichess database and is generally considered the most reliable measure of chess skill, albeit with a more computationally complex algorithm. This is ideal for my study as the rating was precomputed. It is so frequently used because of its validity as a measurement. Veček et al. (2014) showed that the Glicko-2 system was the most accurate way to rank players. The only downside noted was the complicated formula, but the Lichess database contains the measurement precalculated for all data points. This measurement is also frequently referred to as Elo.
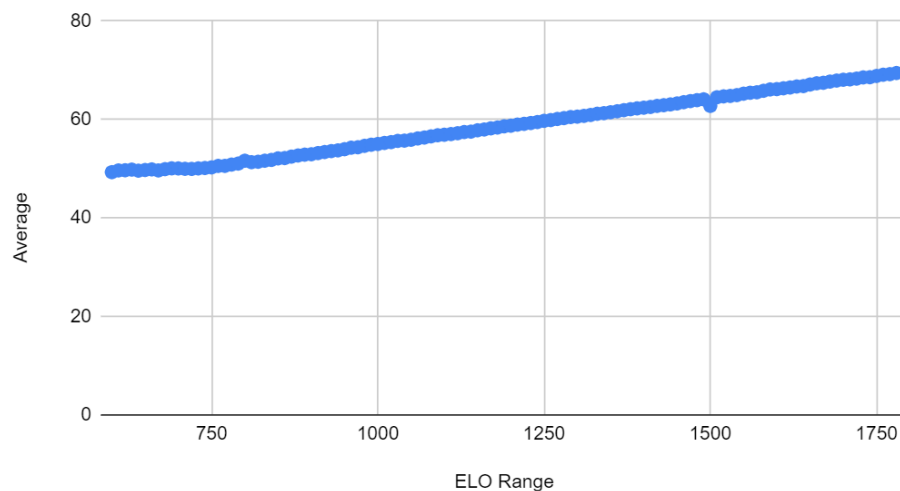


**Figure 1.** The correlation between the Elo of a game and the length

One challenge involves defining a metric for "length of chess games" that is independent of a player's skill level (ELO). A single number that describes their percentile compared to other players independent of skill is desired. A naive approach would be to take every game the account has played over the year and average the length of each game. This fails because average move length is dependent on Elo (Fig 1). This would skew the results because an account with a higher average game length will simply reflect their higher skill, not their

exposure to different game phases. To account for this effect, players must be compared to their peers in a similar Elo.

A better method is to calculate the average length of game for various Elo ranges. Then the length of each game can be compared to how long a player of their skill level would normally be playing. These z scores will then be independent of skill and can then be averaged for each account to understand how above or below average they were in game length across the year compared to their skill.

## Data Source

Like many other studies using the psychometric approach on chess, the Lichess open database will be used. The database contains about five billion games at the time of writing, making it the largest open chess database. This great amount of data lends itself well to a longitudinal study. This is because each account will have games stored across several years. For the purpose of this study, that will allow for analysis of the variables correlating with growth.

## Statistical Analysis

Skill improvement will be calculated with two variables, first observed Elo, and last observed Elo. Players were then ranked by how much their Elo exceeded (or fell short of) the average within their skill group. Skill groups encompass groups of 50. Groups of much lower size would have been ideal, however the sample sizes would become less statistically significant. Groups of 50 strike a balance between sample size per group and the continuity of data. The data for the first column is interpreted as the following: the average player, starting with an Elo between 700 and 750 exclusive, in January increased about 123 points over the year, with a standard deviation of about 187.

| ELO Range | Average | Standard Deviation | Sample Size |
|---|---|---|---|
| (700, 750) | 123.3421053 | 186.8796676 | 456 |
| (750, 800) | 115.5167224 | 173.2057351 | 598 |
| (800, 850) | 118.682448 | 181.4876854 | 866 |
| (850, 900) | 100.2618357 | 159.275196 | 1035 |
| (900, 950) | 98.21144279 | 159.8485151 | 1206 |
| (950, 1000) | 95.28535032 | 167.9427012 | 1570 |
| (1000, 1050) | 76.55684455 | 151.5582774 | 1724 |
| (1050, 1100) | 78.85074627 | 157.7181354 | 1876 |
| (1100, 1150) | 77.0757257 | 160.4739933 | 4754 |
| (1150, 1200) | 73.3255814 | 162.7601274 | 5246 |
| (1200, 1250) | 67.46834911 | 158.9811773 | 4755 |
| (1250, 1300) | 62.81563394 | 156.3760824 | 5245 |
| (1300, 1350) | 59.05852417 | 155.2969391 | 5109 |
| (1350, 1400) | 54.62604784 | 143.0733685 | 4891 |
| (1400, 1450) | 49.67707106 | 127.1489661 | 5094 |
| (1450, 1500) | 44.28781084 | 118.9276274 | 4906 |
| (1500, 1550) | 90.46483971 | 256.997622 | 7736 |
| (1550, 1600) | 38.17579505 | 112.4758059 | 2264 |
| (1600, 1650) | 33.45595654 | 106.5786109 | 5154 |
| (1650, 1700) | 29.18510111 | 104.783095 | 4846 |
| (1700, 1750) | 25.43037975 | 105.0113657 | 5293 |
| (1750, 1800) | 25.30868919 | 99.80455239 | 4707 |
| (1800, 1850) | 24.02098575 | 94.52505326 | 5194 |
| (1850, 1900) | 18.55867665 | 96.07459484 | 4806 |
| (1900, 1950) | 14.91426487 | 89.01083217 | 5482 |
| (1950, 2000) | 12.94754316 | 87.1789776 | 4518 |
| (2000, 2050) | 11.71952516 | 85.80898449 | 5644 |
| (2050, 2100) | 7.119146006 | 84.92952324 | 4356 |
| (2100, 2150) | 8.396664374 | 79.7030019 | 5816 |
| (2150, 2200) | 5.592734226 | 81.99038739 | 4184 |
| (2200, 2250) | 5.899248785 | 84.77102087 | 4526 |
| (2250, 2300) | 3.709385113 | 77.31997863 | 3090 |
| (2300, 2350) | 3.711672474 | 77.24349957 | 2296 |
| (2350, 2400) | 1.029452055 | 77.82099164 | 1460 |
| (2400, 2450) | 2.314393939 | 74.37441861 | 1056 |
| (2450, 2500) | -0.5888324873 | 74.94537091 | 591 |

**Figure 2.** The average increase and standard deviation computed for change in Elo over the year

Similarly each skill group was calculated for average length of chess game, along with the standard deviation.

| ELO Range | Average | Standard Deviation | Sample Size |
|---|---|---|---|
| 600-610 | 49.30869165 | 38.86981828 | 55421 |
| 610-620 | 49.66343007 | 38.06154215 | 46757 |
| 620-630 | 49.69459265 | 37.9325422 | 52854 |
| 630-640 | 49.82752601 | 37.63458244 | 59006 |
| 640-650 | 49.61523246 | 37.45017999 | 66227 |
| 650-660 | 49.74599147 | 37.07854502 | 74092 |
| 660-670 | 49.86963965 | 36.87054609 | 82947 |
| 670-680 | 49.65966966 | 36.56252947 | 92087 |
| 680-690 | 49.92173572 | 36.46609392 | 102090 |
| 690-700 | 50.09011597 | 36.4201957 | 114164 |
| 700-710 | 50.07228347 | 36.19840661 | 127083 |
| 710-720 | 49.99068564 | 35.82929695 | 139140 |
| 720-730 | 49.97324964 | 35.65682904 | 153568 |
| 730-740 | 50.07696369 | 35.60795365 | 166689 |
| 740-750 | 50.15613877 | 35.38989921 | 183465 |
| 750-760 | 50.28194023 | 35.24064995 | 201110 |
| 760-770 | 50.58064879 | 35.13423925 | 213289 |
| 770-780 | 50.58545493 | 34.87410104 | 229191 |
| 780-790 | 50.85108213 | 34.96180661 | 247660 |
| 790-800 | 51.05347578 | 34.75028658 | 269019 |
| 800-810 | 51.63724597 | 35.07486663 | 301885 |
| 810-820 | 51.33991121 | 34.66556721 | 306786 |
| 820-830 | 51.42541445 | 34.58113198 | 326940 |

**Figure 3.** Sample of the average length of a game for each Elo category and its standard deviation. Shortened for brevity.
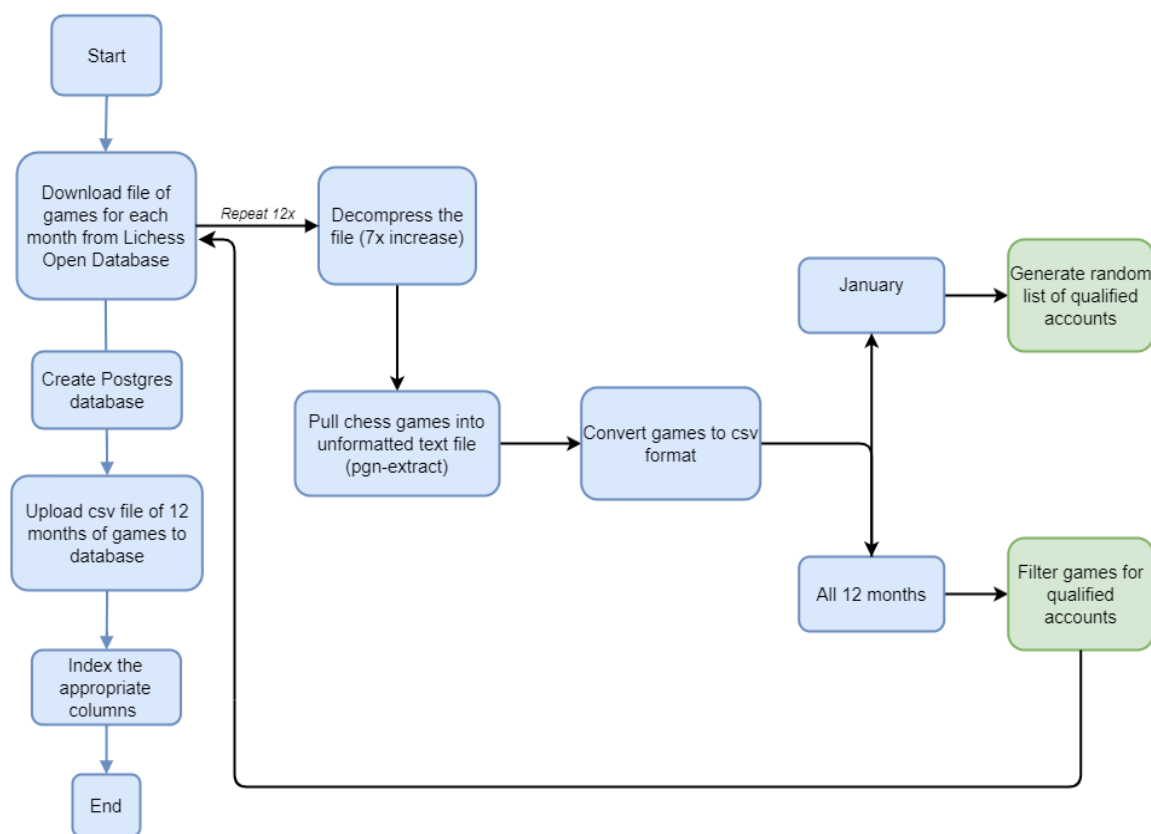
## Data Pipeline



**Figure 4.** Data Pipeline

In order for these calculations to be performed, a data pipeline was constructed for the main purpose of data accessibility and modularity. The first step in the data pipeline was to download the twelve months of PGN files, which stored collections of about one hundred million games. Each month is about 30 GB.

| | | | |
|---|---|---|---|
| 2023 - December | 31.7 GB | 96,909,211 | .pgn.zst / .torrent |
| 2023 - November | 30 GB | 92,389,636 | .pgn.zst / .torrent |
| 2023 - October | 30.8 GB | 94,922,297 | .pgn.zst / .torrent |
| 2023 - September | 30.3 GB | 93,218,629 | .pgn.zst / .torrent |
| 2023 - August | 31.2 GB | 96,118,124 | .pgn.zst / .torrent |
| 2023 - July | 30.8 GB | 95,300,285 | .pgn.zst / .torrent |
| 2023 - June | 31.2 GB | 96,641,906 | .pgn.zst / .torrent |
| 2023 - May | 33.7 GB | 104,193,153 | .pgn.zst / .torrent |
| 2023 - April | 32.9 GB | 101,706,224 | .pgn.zst / .torrent |
| 2023 - March | 34.9 GB | 108,201,825 | .pgn.zst / .torrent |
| 2023 - February | 31.8 GB | 98,471,537 | .pgn.zst / .torrent |
| 2023 - January | 33.5 GB | 103,178,407 | .pgn.zst / .torrent |

**Figure 5.** The list of all chess games downloaded including size and number of games.

The second step was to decompress each file, which increased each file sevenfold to about 220 GB per file leading to 2.5 TB total. One common option which was employed by Holdaway and Vul (2021) at this point is to decompress the files and analyze the chess games while they are still in memory. This would completely eliminate the need to store 2.5TB of data. I rejected this, because the difficulty of programming required would take more time than it would save. Further, the potential speed increase wasn't necessary as in the next step, each file would only take a few hours to process. Parsing was done with pgn-extract, which is a program written in C that can parse large PGN files.

## Conversion from PGN Format

The PGN format inside of the Lichess database was unsuitable for this project. The data size was far too large, including information that wasn't necessary to the project. Moreover, sequential searching for all of the games played by a user took over an hour, which wouldn't be feasible with a sample size of over one hundred thousand users.

```
[Event "Rated Bullet tournament https://lichess.org/tournament/yc1WW2Ox"]
[Site "https://lichess.org/PpwPOZMq"]
[Date "2017.04.01"]
[Round "-"]
[White "Abbot"]
[Black "Costello"]
[Result "0-1"]
[UTCDate "2017.04.01"]
[UTCTime "11:32:01"]
[WhiteElo "2100"]
[BlackElo "2000"]
[WhiteRatingDiff "-4"]
[BlackRatingDiff "+1"]
[WhiteTitle "FM"]
[ECO "B30"]
[Opening "Sicilian Defense: Old Sicilian"]
[TimeControl "300+0"]
[Termination "Time forfeit"]

1. e4 { [%eval 0.17] [%clk 0:00:30] } 1... c5 { [%eval 0.19] [%clk 0:00:30] }
2. Nf3 { [%eval 0.25] [%clk 0:00:29] } 2... Nc6 { [%eval 0.33] [%clk 0:00:30]
3. Bc4 { [%eval -0.13] [%clk 0:00:28] } 3... e6 { [%eval -0.04] [%clk 0:00:30
4. c3 { [%eval -0.4] [%clk 0:00:27] } 4... b5? { [%eval 1.18] [%clk 0:00:30]
5. Bb3?! { [%eval 0.21] [%clk 0:00:26] } 5... c4 { [%eval 0.32] [%clk 0:00:29
6. Bc2 { [%eval 0.2] [%clk 0:00:25] } 6... a5 { [%eval 0.6] [%clk 0:00:29] }
7. d4 { [%eval 0.29] [%clk 0:00:23] } 7... cxd3 { [%eval 0.6] [%clk 0:00:27] }
8. Qxd3 { [%eval 0.12] [%clk 0:00:22] } 8... Nf6 { [%eval 0.52] [%clk 0:00:26
9. e5 { [%eval 0.39] [%clk 0:00:21] } 9... Nd5 { [%eval 0.45] [%clk 0:00:25] }
10. Bg5?! { [%eval -0.44] [%clk 0:00:18] } 10... Qc7 { [%eval -0.12] [%clk 0:
11. Nbd2?? { [%eval -3.15] [%clk 0:00:14] } 11... h6 { [%eval -2.99] [%clk 0:
12. Bh4 { [%eval -3.0] [%clk 0:00:11] } 12... Ba6? { [%eval -0.12] [%clk 0:00
13. b3?? { [%eval -4.14] [%clk 0:00:02] } 13... Nf4? { [%eval -2.73] [%clk 0:
```

**Figure 6.** Sample of a full chess game in PGN format

Therefore, I used pgn-extract to extract the players, their Elos, the result, the time control and the length of the game. Each of these values is necessary to filter for the correct games, and to construct the length of the game and change in Elo variables. The WhiteElo and BlackElo refer to the Glicko-2 rating system. The first widely used rating system in chess was created by Arpad Elo. For historical and convenience reasons, many in the chess community refer to any rating system as "Elo" resulting in the naming convention on the PGN format (Veček et al., 2014).

Rapid games in chess have a time control closer to ten minutes, compared to blitz or bullets five or two minutes respectively. I wanted longer time per player so that the game length wouldn't be as heavily influenced by time. I decided against longer classical games because there weren't as many of them. After extracting these games from the month of January, I created an account list. The account list contained every person who had played a game longer than 10 minutes in the month of january. This yielded about 1 million accounts. The account list was then reduced to about 130,000 through random stratified Elo sampling. This would allow players of all skill levels to have equal representation within the correlational analysis. For this, I chose ten thousand accounts from each Elo range 800-2500. Finally the games from all of those accounts were uploaded to a Postgres database using the following columns (fig. 7). Relational databases are prevalent among research leveraging the Lichess open database, however the exact type chosen didn't appear to have a significant impact.

## Databasing

public.games: 18,372,452 rows total, limited to 1,000

| # | game_id | winner_elo | winner | ply_count | winner_account_id |
|---|---|---|---|---|---|
| 1 | 1 | 1,821 | ottomanamongus | 121 | 49,408 |
| 2 | 2 | 1,811 | TORKVE333 | 49 | 69,871 |
| 3 | 3 | 2,126 | Kalvo_Toni | 94 | 107,032 |
| 4 | 4 | 1,271 | Georges5400 | 10 | 22,111 |
| 5 | 5 | 1,648 | ne_ya | 47 | 72,083 |
| 6 | 6 | 1,770 | Pestr1 | 87 | 73,234 |
| 7 | 7 | 1,457 | Braunsfelde | 13 | 52,322 |
| 8 | 8 | 1,973 | Zakaryabouid | 87 | 90,576 |
| 9 | 9 | 1,552 | Hristo1956 | 101 | 76,098 |
| 10 | 10 | 2,098 | OPC_RapideOuedzem | 75 | 103,129 |

**Figure 7.** Resulting database structure for chess games.

Prior to the database, there were twelve .csv files containing games. These files were not organized by name and they were difficult to search entirely due to being separate. The database solved both of these issues. The Postgres database stored all of the games and players to be analyzed. The games tab contained about 18,000,000 games, each with the winner's name, Elo, and how many moves it took them to win. This allows for the searching of all the games of an account across the year within a few milliseconds. The structure of the database lends itself to my project.

**Figure 8.** Resulting database structure for chess accounts.

The accounts table contains the accounts list of about 100 thousand accounts from earlier, as well as two measurements: first Elo and last Elo. These values are how I compare skill increases or decreases over the year. The database also had a games table with the necessary information. An extra column of account_id was added to the games table. By adding this column and indexing it, the search for all of the games matching an account went down from two minutes to several milliseconds.

## Statistical Analysis

After the database was set up to be queried by username, the application of statistical techniques could begin. A factor that measured how much a player grew compared to his peers was necessary. For this, I took several ranges of values for the first Elo measurement and computed the difference (Fig. 2). This was necessary to accommodate for the learning curve. Improvement gets much slower over time, so growth shouldn't be measured in absolute form, but rather in reference to peers. For example, an exceptional increase in Elo from a highly rated 2000 Elo player would be 50. An exceptional increase in an average 1500 rated player would be 400.

Next, to calculate game standard deviations for a player, each player had their corresponding games pulled from the database. First, the python script would check what their Elo was while they were playing that game. Next, it would search the table to figure out what range they were in. This is necessary to figure out how long their game should be on average. It would then compute a z score for that game, based on how above or below average the length of their game was. After creating and averaging all of the z-scores for an account, the result is a single number that represents how above or below average they were in game length over the year. After performing this operation for every account, they can be ranked in terms of their games' lengths over the year. That value can then be compared to the rank of their growth compared to their peers. This is known as a Spearman correlation. It is used here as the density of the length of chess games is not a normal distribution. It has a right skew due to the lack of restrictions on total game length.
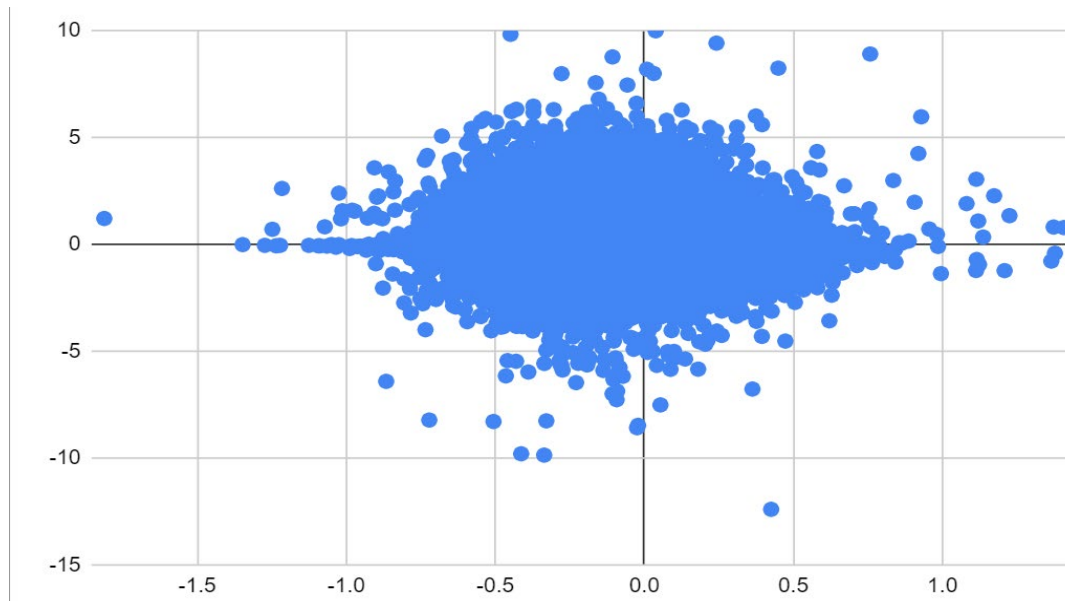
**Figure 9.** Ply is on the x axis and change in Elo is on the y axis. Change in Elo is measured in standard deviations; ply is unitless

## Discussion

The data yielded a correlation coefficient of 0.0632. This is a very low correlation coefficient suggesting near zero correlation between increase of skill and length of games. The percentile length of chess games cannot predict to any extent improvement over the course of a year. The null hypothesis, however, cannot be proven due to the limitations of the method. In this particular method, only statistically significant positive correlations can disprove the null hypothesis. This is because the imperfection of the operational definition of skill could be at play. Elo is an approximation for skill and, although highly accurate, contains slight error. It attempts to approximate the true skill of the player.

Consider a player whose Elo is slightly inflated compared to their true skill. Due to this overestimation, the player is more likely to be matched against opponents who are stronger players than them. These stronger opponents would likely take longer to beat, leading to longer games for the player in question. However, since they're facing tougher opponents, they might lose more often. This could lead to their Elo decreasing over time, reflecting their true skill level. As their Elo adjusts downward to reflect their true skill, their game length might normalize. This creates a pattern where longer games are initially associated with a decrease in Elo, which is the opposite of what you'd expect for skill improvement.

The same is true for a slightly underestimated player. Due to this underestimation, they might be matched against opponents who are actually weaker. These weaker opponents might be defeated more quickly, resulting in shorter games for the player. Again this is the opposite, where we'd expect to see lower increases in skill from short games.

This effect is exacerbated in two ways. First, through the pairings in Lichess. Cheng and Camargo (2023) found that the vast majority of games occurred within a 100 rating point differential. This means a very narrow normal distribution with pairings that are highly sensitive to slight changes in rating. This contributes to the opponent's skill being more heavily affected by an under or over approximation. The greater the change in skill, the greater the aforementioned effect becomes. Secondly, the effect is amplified by winning and losing streaks. Chowdhary et al. (2023) found that most players tended to exhibit streaks while they were playing.

Because the real skill of a player lies somewhere in between the winning and the losing streaks, their Elo likely diverges from their underlying skill more often.

Both overestimations and underestimations by the Elo system can introduce "noise" into the data. This noise makes it hard to distinguish between the true effect of game length on skill development and the fluctuations caused by the imperfections of the Elo rating.

Therefore the null hypothesis cannot explicitly be proven leading to an inconclusive result. Although this uninterpretable result is limited to studies that yield negative to zero correlation, it suggests the need for new statistical methods capable of handling these inconsistencies.

The spearman rank also might not have been able to overcome the skew of chess games. They were heavily right skewed, leading to a nonlinear relationship in probability. This is because a chess game cannot be less than zero moves, but it can be virtually infinitely long. Techniques that can normalize that skew may have revealed a stronger relationship.

The method for sampling Lichess accounts was to take a file for all games played in a month, scrape the names of the players from each of the games, and choose randomly among them. This means I was only looking at accounts that played in the month of January. In order to remove bias, I removed duplicate names so that every account would have the same chance of being chosen. There is still a slight sample selection bias here however. An account that plays more games is more likely to have played sometime in the month of January than a less active player. This means I am slightly selecting accounts that play more chess more often. Another possible resulting sample selection bias is the month specific demographic changes, such as hemisphere and age of the player. The effect size of these sample selection biases is likely negligible and they likely have nothing to do with the dependent variable of my study. Even so, the account selection could be more robust.

## Conclusions and Future Directions

This study investigated the relationship between the length of games an account plays, and their improvement in Elo. The correlational choice was motivated by the existence of big datasets including the Lichess.org open database. Contrary to the hypothesis, the analysis revealed a near zero correlation, suggesting that the length of games played doesn't predict Elo increase over time. However, this conclusion cannot be fully justified. The reliance on the Elo measurement, which is prone to swings could confound the analysis.

Future operational definitions should be reviewed via survey to test whether they properly measure memorization. The design should have participants submit their Lichess accounts username and rank 1-10 how much they feel they study openings compared with other methods of study. If the correlation between their survey number and the number assigned to them through statistical analysis of their account are highly correlated, it would lend this study more credibility. Further, this type of survey could be used to design operational definitions designed to be correlated with a specific factor. Therefore a followup study should inductively apply a more fitted technique, like a survey, instead of making assumptions.

Future research could also rely on the novel methods of evaluating skill. While Elo is the most common, others include accuracy, which is generally a measure of how well your moves align with the engine.

For the definition of a memorized opening, Cheng and Camargo (2023) developed a technique for detecting chess accounts that were able to harness theoretically less successful openings to great success. This could be a potential next correlation analysis as that would appear to match the description of a player who studies the opening. Future research could see this field of data mining chess games become a rich source of human decisions appropriate for testing behavior models.

## Acknowledgments

## References

Armstrong, J. S. (2010). Natural Learning in Higher Education. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1928831

*Bobby Fischer on Paul Morphy and how opening theory destroyed chess*. (n.d.). Www.youtube.com. https://www.youtube.com/watch?v=P349BdHUxlc

Cheng, I., & Camargo, C. (2023). *Machine learning to study patterns in chess games*. https://doi.org/10.13140/RG.2.2.30894.52807

Chowdhary, S., Iacopini, I., & Battiston, F. (2023). Quantifying human performance in chess. *Scientific Reports*, *13*(1), 2113. https://doi.org/10.1038/s41598-023-27735-9

De Marzo, G., & Servedio, V. D. P. (2023). Quantifying the complexity and similarity of chess openings using online chess community data. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-31658-w

Duplessis, T. (2024). *The best free, adless Chess server*. Lichess.org. https://lichess.org

Gaschler, R., Progscha, J., Smallbone, K., Ram, N., & BilaliÄ‡, M. (2014). Playing off the curve - testing quantitative predictions of skill acquisition theories in development of chess performance. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00923

Grabner, R. H., Stern, E., & Neubauer, A. C. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica*, *124*(3), 398–420. https://doi.org/10.1016/j.actpsy.2006.07.008

Holdaway, C., & Vul, E. (2021). Risk-taking in adversarial games: What can 1 billion online chess games tell us? *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43. Retrieved from https://escholarship.org/uc/item/403764rd

Holdaway, C., & Vul, E. (2021). Risk-taking in adversarial games: What can 1 billion online chess games tell us? *PsyArXiv (OSF Preprints)*, *43*. https://doi.org/10.31234/osf.io/vgpdj

Howard, R. W. (2011). Longitudinal Effects of Different Types of Practice on the Development of Chess Expertise. *Applied Cognitive Psychology*, *26*(3), 359–369. https://doi.org/10.1002/acp.1834

Jonassen, D. H., & Strobel, J. (2006). Modeling for Meaningful Learning. *Engaged Learning with Emerging Technologies*, 1–27. https://doi.org/10.1007/1-4020-3669-8_1

Kasparov, G. (2010). *The Chess Master and the Computer*. https://web.mit.edu/6.034/wwwbob/kasparov-article.pdf

King, D. J., & Russell, G. W. (1966). A comparison of rote and meaningful learning of connected meaningful material. *Journal of Verbal Learning and Verbal Behavior*, *5*(5), 478–483. https://doi.org/10.1016/s0022-5371(66)80064-6

*Magnus Carlsen – "I don't quite fit into the usual schemes."* (2011, December 22). Chess News. https://en.chessbase.com/post/magnus-carlsen-i-don-t-quite-fit-into-the-usual-schemes-

Morales, E. M. (1996). LEARNING PLAYING STRATEGIES IN CHESS. *Computational Intelligence*, *12*(1), 65–87. https://doi.org/10.1111/j.1467-8640.1996.tb00253.x

Ross, P. E. (2006). The Expert Mind. *Scientific American*, *295*(2), 64–71. https://www.jstor.org/stable/26068925

The Editors of Encyclopedia Britannica. (2019). Bobby Fischer | Biography & Facts. In *Encyclopædia Britannica*. https://www.britannica.com/biography/Bobby-Fischer

Van Der Maas, H. L. J., & Wagenmakers, E.-J. (2005). A Psychometric Analysis of Chess Expertise. *The American Journal of Psychology*, *118*(1), 29–60. https://doi.org/10.2307/30039042

Veček, N., Črepinšek, M., Mernik, M., & Hrnčič, D. (2014). *A comparison between different chess rating systems for ranking evolutionary algorithms*. 511–518. https://doi.org/10.15439/2014F33