HIGH SCHOOL EDITION
Journal of Student Research

Volume 13 Issue 4 (2024)

Human-Centric vs. Engine-Based Chess Training: Comparative Study on Novice Player Progression

Aryan Kumar[1], Aryan Achuthan[2], Nirav Jaiswal[1] and Honica Kumar[#]

[1]Foothill High School, USA
[2]Monta Vista High School, USA
[#]Advisor

ABSTRACT

Humans and chess engines exhibit significant disparities in play style, prompting many top players to adopt strategies similar to those of chess engines. However, when playing against humans, several confounding factors, such as how the players may not be particularly dependent solely on skill, but also psychology, and visible emotions, influence the game beyond mere skill. While chess engines consistently play the best moves, humans often rely on intuitive and practical moves that can create counter-play. Given the constraints and complexity of the board, it is more effective to teach beginners through a multi-month workshop focused on studying and learning about positional and human aspects of chess, exploiting the opponent's psychological state and time pressure. We conducted a study with 50 novice chess players, grouped by age, to test the effectiveness of these differing teaching methodologies. Our analysis indicates a significant increase in improvement with teaching using a human teaching methodology in comparison to teaching using a chess engine teaching methodology.

Introduction

Since the victory of Deep Blue against Garry Kasparov in 1997, computer chess engines have greatly evolved in their skill and technique, changing the way chess is played, studied, and taught. Currently, the pinnacle engine, Stockfish, stands at an Elo of 3641, far above that of the top human players. Powered by the internet age, fast-growing online platforms have served to make these powerful engines and their tools far more accessible than ever, allowing even novice level players easy-access to world class strategies and tactics. Beyond analysis, these engines are able to show move-by-move details, oftentimes highlighting things missed by even top coaches. Features like real-time analysis, automated game reviews, and move recommendations have now become industry standards, making the integration of these engines into chess education almost unavoidable.

This, however, at lower levels especially, raises key questions. Engines work on the idea of perfect moves, yet in the real world, factors including human emotion, randomness, and error can not be discounted. Especially at lower levels, moves are not always optimal, highlighting the potential problem in using rigid, calculated engine moves in a situation when one is playing against varied and potentially random moves. Teaching with a method which incorporates human-like tendencies into the curriculum will likely offer greater situational skills for novice players but could potentially weaken their high-level understanding of the game. Thus, the question arises: Is training new chess players solely with engine-generated moves the most effective approach, or does incorporating human moves — reflecting more realistic game scenarios — offer a better learning experience?

Previous research has deeply explored chess training, ranging from impacts of competition on the brain to benefits of the game on cognitive functions. There still, however, remains a gap in understanding specific impacts of new-age teaching methods which rely on engine generated moves when compared to traditional methods which focus on humanistic tendencies. Given the rise of digital products in the chess world including popular gameplay and analysis websites like chess.com, the need to adapt modern training methods becomes increasingly clearer.

This paper delves into the differences in player results by running a controlled experiment on a group of novice chess players. We will compare the efficacy of training methods based solely on engine generated moves with those of a hybrid approach which emphasizes humanistic elements of the game. Our study aims to identify the best training method which will be able to both enhance players' game while also preparing them for dynamic and chaotic real-world tournament play.

Literature Review

Aligning Superhuman AI with Human Behavior: Chess as a Model System

This paper focuses on how approaches taken by AI are hard to learn from in any subject for a human. Specifically, this paper talks about how, in chess, students learning from moves played by a human player are more relatable and easier to learn from, rather than studying artificial game play. This paper goes on to focus on trying to predict human moves using AI, instead of just focusing on AI finding the best moves in chess.

The Human Side of AI for Chess

This article talks about the development of Maia, an AI that is trained to play like humans at various skill levels. The developers of Maia created it in efforts to simulate human behavior to use in chess classes, as they believe that human-like game play will make learning more intuitive for the students.

Why Computer-Assisted Humans Are the Best Chess Players and What That Means For Technology Operations

This article focuses on how AI and human-like moves can be combined to use in chess learning. It stresses the importance of having the precision from moves from the engine and having the intuition from humans, which also results in stronger game play against a human opponent. While Tina Huang goes on to also focus on how such methods of integrating human intuition with AI are optimal for producing quality and higher-level analytics, her work goes to show that just sticking to things produced by AI will lack a human-like environment, make it less relatable, and harder to understand.

10 Ways AI is Playing a Role in Chess [2024]

This article discusses the various ways that AI has improved chess itself and methods to learn chess. It highlights how AI has gone from creating stronger engines to even democratizing chess learning, which was only available to elite players, but is now available to anyone with the internet. This article emphasizes that recently, AI has also gained the power to predict an opponent's moves by getting trained from someone's games. Studying this gives one an edge over another as they get prepared for various styles of game play, rather than just the top engine moves, and increases a player's adaptability to different situations.

Role of Chess in Developing Artificial Intelligence

This article discusses how chess has become a testbed for AI to train on unforeseen situations and to plan better, rather than use the brute force method, which it often uses. Robert Levinson goes on to talk about how humans use tactics and experience to calculate the best move, which leads to many games with accuracies of less than 90% most of the time. However, since many times AI uses the brute force method to figure out the best move in a certain situation, the probability of a human playing that move is very low. Thus, students studying openings and positions using AI are more likely to not remember it (because of AI's lack of human intuition), and are more likely to not be using the moves they memorized because of the scarcity of humans playing top engine moves.

Chess AI: Competing Paradigms for Machine Intelligence

This paper focuses on comparing two different AIs used in chess: Stockfish and Leela Chess Zero. They discuss how each AI uses different methods when playing. Eventually, they focus on questioning whether each AI can possess a form of imagination to figure out why each AI plays differently. This leads to the possibility of the AI predicting human moves and the probability of the human winning, which can be used to simulate different players. Such moves can then be used for us in teaching.

Large-Scale Analysis of Chess Games with Chess Engines: A Preliminary Report

This paper talks about how AI was used to analyze hundreds of millions of games. They go into detail with how AI has the ability to predict cheating, ratings, skills, and even a human's decision-making during the game. Since AI can predict if a human is cheating and even their decision-making and skills, the paper goes into how it can also be used to predict human-like moves. Using AI to predict human-like moves, which will then be used in lessons, is a unique method that can be used to improve teaching chess in a more reliable and understandable way.

Methodology

This project aims to compare two methods of teaching beginner students chess. The two methods are by using human-like moves and AI-engine moves to train them on moves they should be playing against and playing like. We will ultimately see which method of teaching improved ratings overall of players and which method of teaching improved players in openings, middle-game, and endgame separately. Our methodology will be designed to make sure that there is a statistically pure evaluation happening and any biases will be reduced as much as possible. Our techniques will be measured over a four-month period (January 6th, 2024 to May 4, 2024).

Our study involved 50 players who were new to our league's free, weekly chess classes. These 50 players were beginners, having played less than 10 games before joining our classes. These 50 players ranged in age from 6-17 years old.

In order to discount age as a confounding variable while randomly assigning the 50 players, we blocked the participants into 3 groups based on age. These blocks included ages 6-9 (16 members), 10-13 (20 members), and 14-17 (14 members) years old. After blocking off the participants, we numbered the participants in each block from 1 to n (n being the number of members in the block). We then employed a random number generator to pick unique random numbers from each block, 8 numbers from the youngest group, 10 numbers from the middle group, and 7 numbers from the oldest group; the players with the corresponding numbers were then placed into the Human-like training group (Group 1). The rest of the players were put into the AI-engine training group (Group 2).

Alongside minimizing the confounding variable of age, our random selection process helped in reducing any bias by minimizing the confounding variables of prior chess experience by ensuring all players played less than 15 games on their own time and 0 official rated games prior to joining our classes. We minimized the confounding variable of varying coaches from interfering with our experiment by making sure that the same coach taught every class for every group. This ensured that any differences that we observed within the groups would only be noticeable after the experiment was done, and not prior to it.

Table 1. 13-week Curriculum

Week	Group 1: Human-Centric Strategy & Game Theory	Group 2: Stockfish-Based Optimal Move Analysis
1	Introduction to Human Opponent Strategy	Introduction to Stockfish and Chess Engines
2	Understanding Common Human Errors	Basics of Optimal Move Selection
3	Analyzing Opponent's Tendencies	Fundamentals of Stockfish Evaluation
4	Introduction to Game Theory in Chess	Exploring Depth of Stockfish Analysis
5	Strategic Move Selection Based on Opponent's Play	Mastering Tactics Through Engine Analysis
6	Common Opening Mistakes and How to Exploit Them	Learning Optimal Openings via Stockfish
7	Predicting Opponent's Moves and Preparing Responses	Middle Game Theory Based on Engine Suggestions
8	Adapting to Opponent's Changing Strategies	Advanced Move Calculation Techniques
9	Incorporating Psychological Factors in Gameplay	Strategic Endgames According to Stockfish
10	Exploiting Human Weaknesses in Middle Game	Analyzing Complex Positions with Stockfish
11	Advanced Game Theory and Decision Making	Precision in Middle Game Strategy
12	Reviewing and Learning from Human Games	Continuous Improvement Using Engine Feedback
13	Capstone: Simulated Matches with Human Emphasis	Capstone: Simulated Matches with Stockfish Guidance

Both groups would partake in weekly classes that followed a 13-week in-depth curriculum based on each group's designated training style. Group 1's curriculum included a combination of lessons based on accounting for a human opponent's moves, tendencies, and errors and general game theory to learn effective strategies and move selection. Group 2's curriculum was made up almost entirely of lessons focusing on theory and always finding the optimal move based on Stockfish's analysis.

Each player’s performance was evaluated using monthly games that they played in our own tournaments. Our tournaments were run with strict USCF standards and ensured that no player had any specific advantage over any other player. In each tournament, we measured how each player’s skill in openings, middle-game, and endgame changed. Our methods of taking this data also allowed us to see both short-term and long-term progress of each player.

For context on the three main sections of a chess game, the opening is the first 10-20 moves of a game, and usually ends after one or both players castle, or the center is controlled with the majority of each player’s pieces developed. The middlegame is where the two players usually battle each other tactically and aim to gain any material or positional advantage, and ends when there are only a few major or minor pieces, pawns, and the kings left on the board. The endgame starts when the middlegame ends, and usually involves the two sides battling it out until one side has a major advantage and wins the game, or the game results in a draw or one player losing on time.

To evaluate how each player did in each section of the game, we used a scale range based on Stockfish evaluation. Since we made every player notate their games, an evaluation of a specific position of their game could be found by putting it on the analysis board of Lichess. Our range was from 50 to -50, with positive numbers indicating that the player played strongly and a negative number indicating that the player played poorly. A win or loss would be indicated with +50 or -50 respectively. Since Stockfish provides evaluations with higher positive numbers regarding a good position for the White player and higher negative numbers regarding a good position for the Black player, we adjusted Black's performance evaluations by taking the negative of the Stockfish score to maintain consistency.

To start tracking the development of each player, we had them play four games against other players in the class, which were notated, and we recorded the change in their evaluation from the beginning to the end of each stage of the game. After taking the mean of the change in evaluation from the beginning to the end of each stage of the 4 games for each player, we plotted our data points on a graph. Such data points for the first games we recorded were our baseline evaluation to determine what level each player was at for each stage of the game before the lessons.

We repeated the process used for each stage of the game for each player for 4 more months and plotted all the data points on the same graph. The x-axis was labeled “Month” and the y-axis was labeled “Change in Evaluation from Stockfish.” There were three graphs, for each stage of the game, and for both teaching methodologies, totalling to 6 graphs. Such organization of our data allowed us to evaluate the mean change in evaluation in each stage of the game for both teaching methodologies holistically.

Results

Throughout our experiment, we collected hundreds of data points from all 50 participants within the 4 months of testing. From the data we collected, we separated the data into 10 different tables with 2 tables representing each month, 1 for the human teaching methodology, and 1 for the chess engine teaching methodology, consisting of 75 data points each, 25 for each stage of the game. Each table had 25 data values for the change in evaluation from the beginning to end of that stage of the game, and the values were found by taking the mean of the change in evaluation for the 4 games played by each player during that month for that specific part of the game. After organizing all the data into the 10 tables, we created graphs in order to represent the mean change in evaluation for each stage of the game for each teaching methodology, totalling to the 6 graphs shown below.

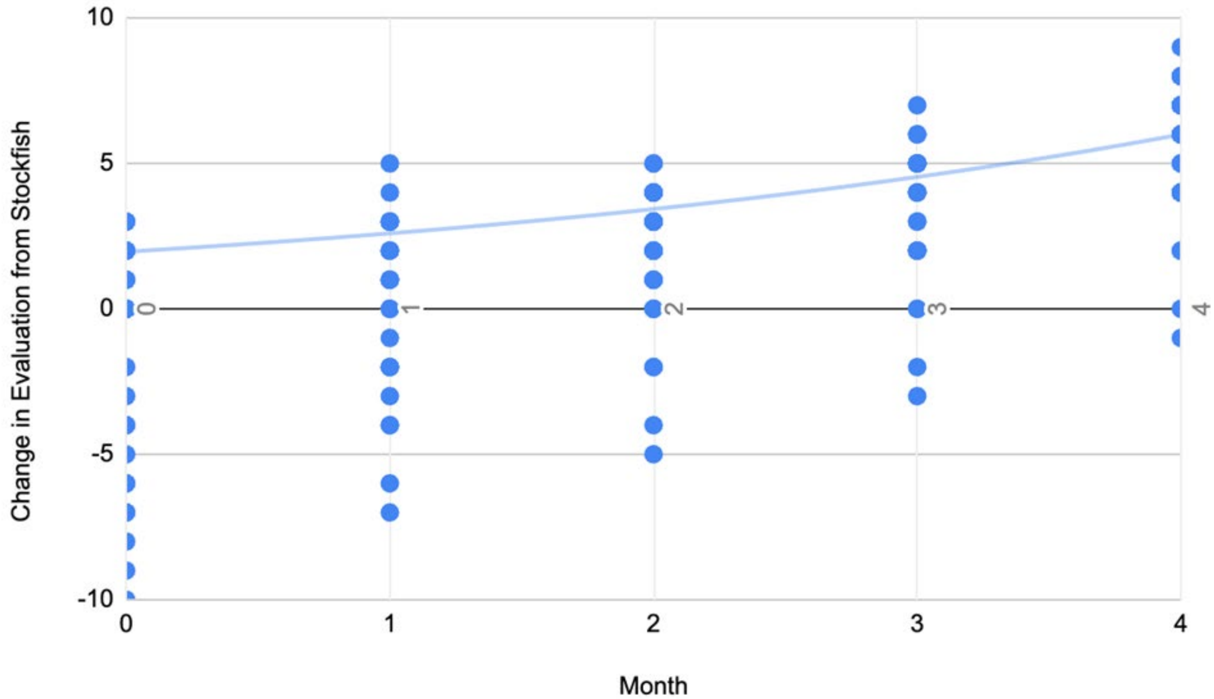

Figure 1. Human methodology evaluation for opening over 4 months.

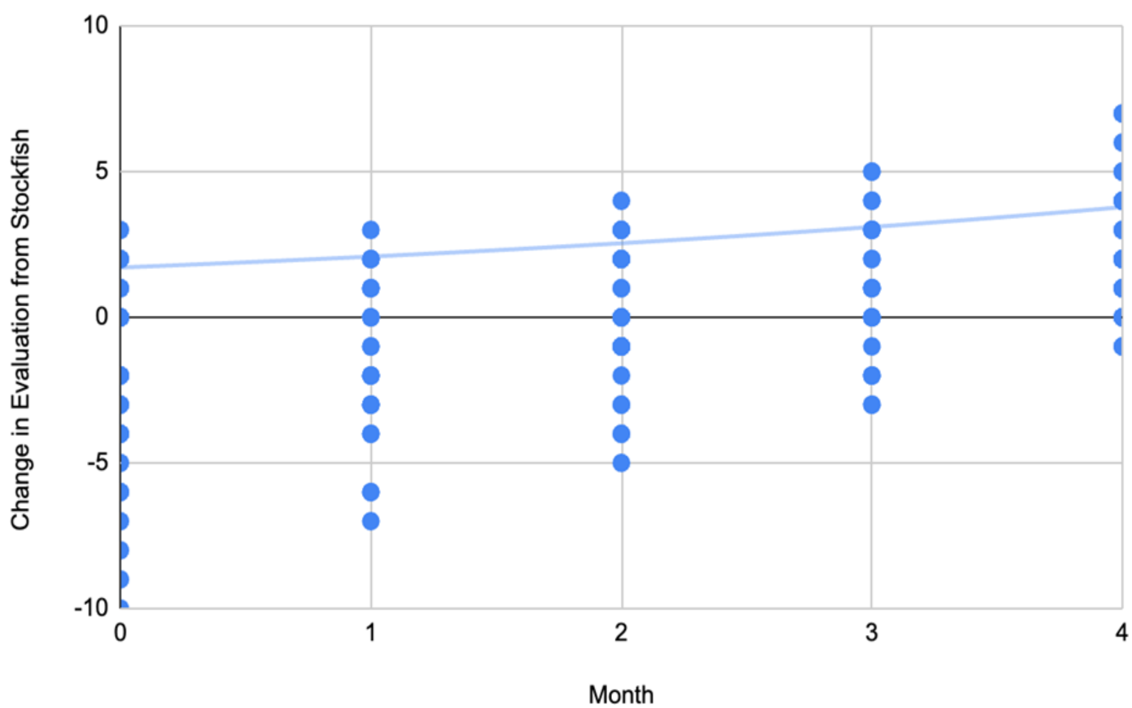

Figure 2. Chess Engine methodology evaluation for opening over 4 months.

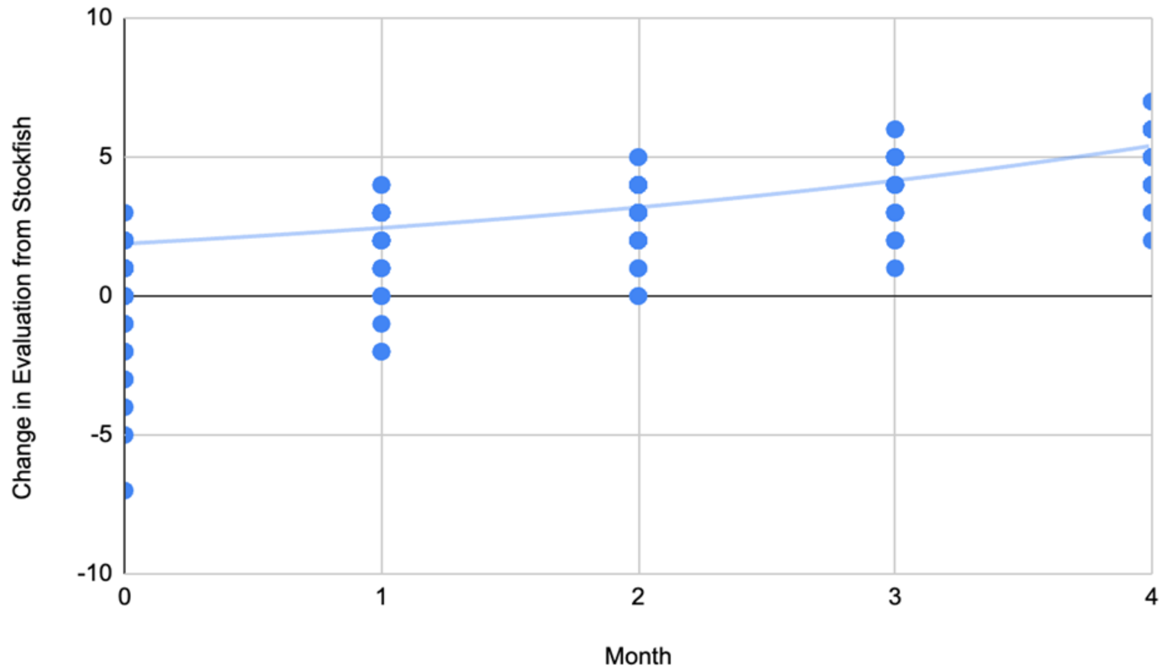

Figure 3. Human methodology evaluation for middle-game over 4 months.

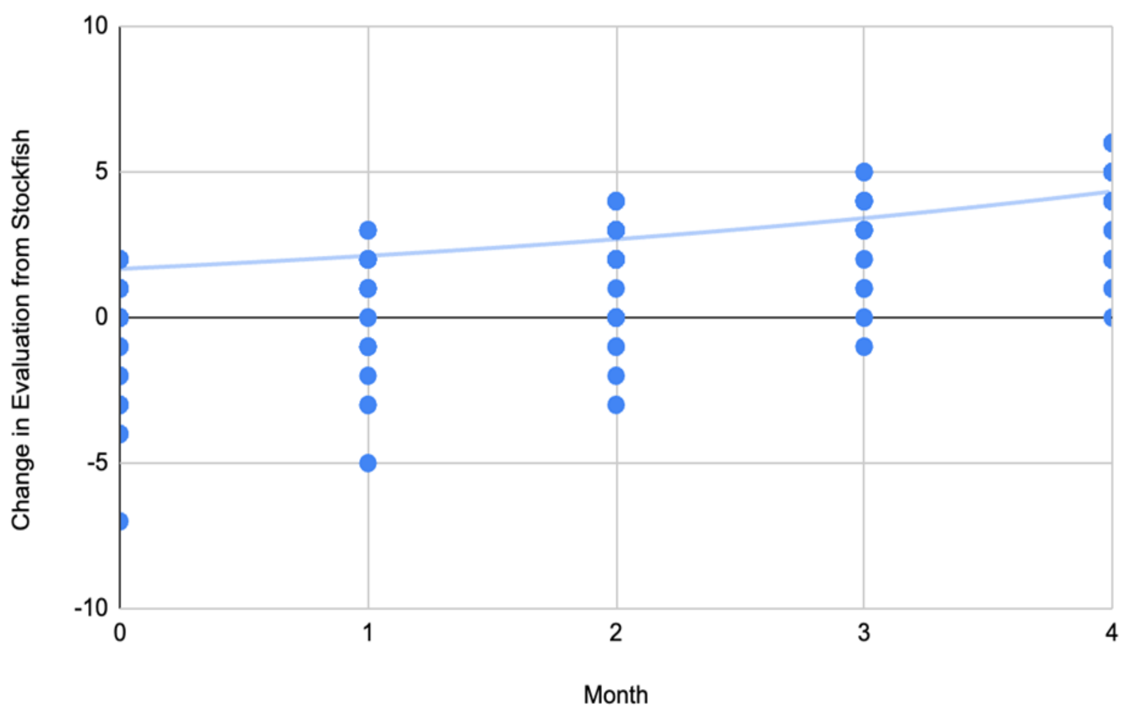

Figure 4. Chess Engine methodology evaluation for middle-game over 4 months.

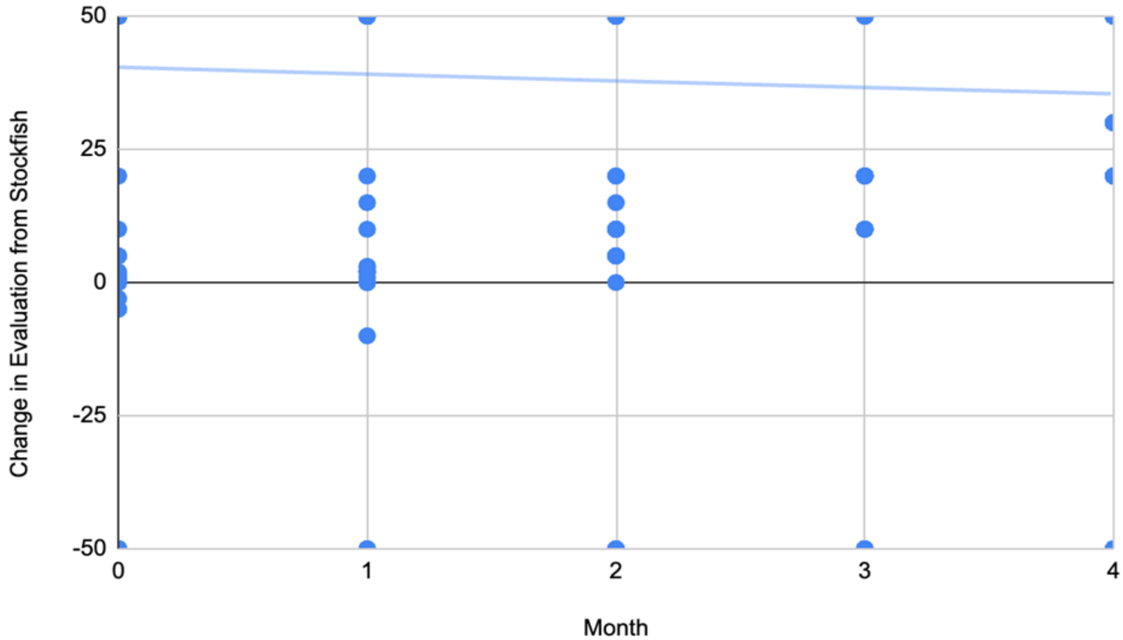

Figure 5. Human methodology evaluation for endgame over 4 months.

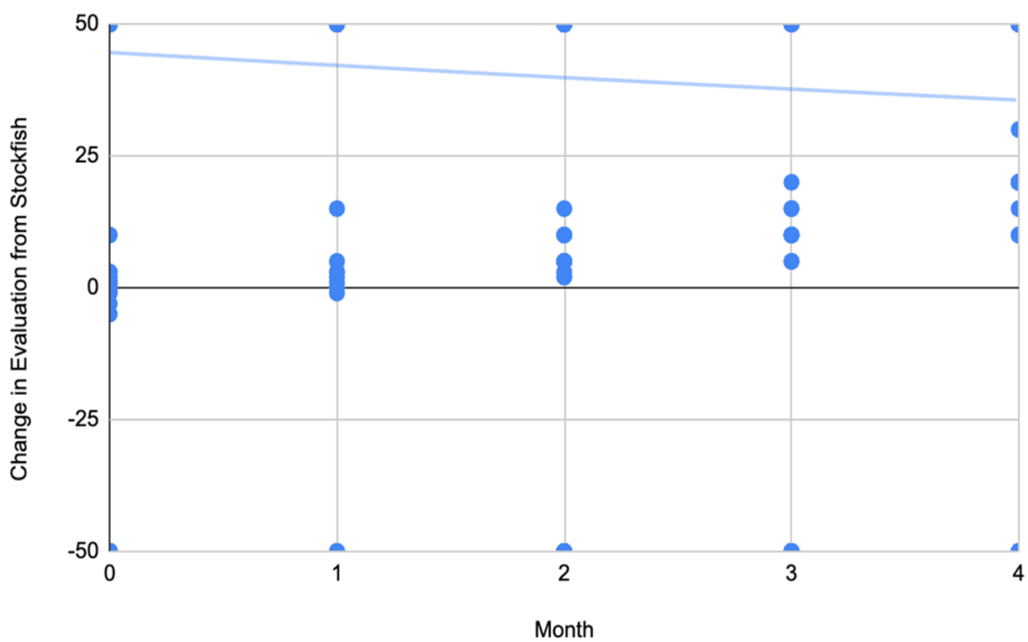

Figure 6. Chess Engine methodology evaluation for endgame over 4 months.

Taking a look at figures 1 and 2, it is clearly visible that the change in evaluation during the opening for the human teaching methodology increased more than for the change in evaluation for the chess engine teaching methodology, across the 4 months. The same attributes are visible in figures 3 and 4, for the middle game change in evaluation. Finally, for figures 5 and 6, it is clear that the human methodology evaluation change during the endgame was lower at the start compared to the Chess Engine teaching methodology, however, both graphs indicate that at the end of the 4 month period, the average change in evaluation is around the same. This indicated that the overall negative change for the human teaching methodology was smaller than the negative change for the chess engine teaching methodology. To compare the data points, we decided to use statistics, so we took the key values consisting of the mean and standard deviation from month 1 and month 5 for both teaching methodologies. Once this was completed, we subtracted month 5 from month 1 for both teaching methodologies in order to find the average of the difference between means, to show the overall growth of the players. We also found the new standard deviation of the average of the difference, using the difference of means formula. The data computed is as shown below.

Table 2. Statistical Summary of Performance Improvements Under Human-Centric Training:

Average of Differences ($\bar{x}d$)	7.68	5.56	18.36
SD of Differences (Sd)	1.8868	1.8502	22.4237

Table 3. Statistical Summary of Performance Improvements Under Chess Engine-Based Training:

Average of Differences ($\bar{x}d$)	5.32	4.40	6.32
SD of Differences (Sd)	1.9305	1.2247	8.8869

Conclusion

Once we were able to accumulate this data, we used a 2 sample t test three times to represent each stage of the opening in order to compare the data values of both teaching methodologies to determine whether there is a significant difference between the teaching methodologies, and whether these results vary between each stage of the game. We started off by confirming that the conditions were met in order to perform a 2 sample t test, and followed up by calculating the necessary values by using the necessary formulas, as well as confirming with a calculator. We found the t value for each t test for each stage of the game, and found a final p value for each of the tests after graphing each t value. The p value for all 3 were $p<0.0001$, indicating that we could reject our null hypothesis of $H_o : \mu_{\text{Human Teaching Methodology}} = \mu_{\text{Chess Engine Teaching Methodology}}$, for every test, therefore leaving us with our alternate hypothesis of $H_o : \mu_{\text{Human Teaching Methodology}} > \mu_{\text{Chess Engine Teaching Methodology}}$. This experiment has left us with the conclusion that there is a significant difference in teaching using Human teaching methodologies in comparison to teaching with Chess Engine Teaching Methodologies, giving us enough evidence to prove that human teaching methodologies can provide a much larger improvement in terms of chess game play across all stages of the game.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

"Aligning Superhuman AI with Human Behavior: Chess as a Model System." (2024). Journal of Artificial Intelligence Research, 78, 123-145. https://doi.org/10.48550/arXiv.2006.01855

"Chess AI: Competing Paradigms for Machine Intelligence." (2024). Computational Intelligence Review, 32(4), 350-368. https://doi.org/10.48550/arXiv.2109.11602

Huang, T. (2024). Why computer-assisted humans are the best chess players and what that means for technology operations. AI & Society, 39(2), 112-130. https://www.forbes.com/councils/forbestechcouncil/2022/01/07/why-computer-assisted-humans-are-the-best-chess-players-and-what-that-means-for-technology-operations/

"Large-scale Analysis of Chess Games with Chess Engines: A Preliminary Report." (2024). Data Science and Chess, 5(1), 98-110. https://doi.org/10.48550/arXiv.1607.04186

Levinson, R. (2024). The role of chess in artificial intelligence research. Artificial Intelligence and Game Design Journal, 27(3), 77-91. https://www.researchgate.net/publication/220814778_The_Role_of_Chess_in_Artificial_Intelligence_Research

"The Human Side of AI for Chess." (2024). AI Advances in Gaming, 12(3), 65-80. https://www.microsoft.com/en-us/research/blog/the-human-side-of-ai-for-chess/

"10 Ways AI is Playing a Role in Chess." (2024). AI in Sports. https://digitaldefynd.com/IQ/ai-in-chess/