Elo Evolution: Redefining Chess Ratings

Introduction

Chess ratings are used to measure a player's skill, yet many players believe the current FIDE rating system does not accurately reflect true ability. In 1970, FIDE, the International Chess Federation, began using the Elo ratings, a system designed by Arpad Elo. In recent times, commentators have argued that this rating system is flawed due to its struggles with inflation, K-factor, new and fast improvers and player inactivity. Many also claim it oversimplifies what chess strength really means. This project investigates the fairness and accuracy of the FIDE system by analysing surveys from elite Irish chess players, performance accuracy data, and geographical factors such as access to tournaments.

Aim

This project aims to establish the relationship between a player's rating and their skill level, and use the results collected to develop an alternative rating system which reflects an individual's true playing strength.

Method

Examination of Player Attitudes towards the FIDE Rating System

Following examination of typical tournament fields (40–50 players), 45 was identified as a representative population size, and a sample size of 41 calculated for the study. Rated players in the Irish chess community were surveyed on their views and experiences of the Elo rating system used by FIDE.

Estimation of Player's Skill Level using a Puzzle Survey

A sample size of 28 was identified from the population of chess players between the ages of 12 and 21 that competed in the 2025 Malahide Millennium Tournament.

In a survey, basic demographic and chess-background information was collected. 20 positions taken from *Khmelnitsky's Chess Exam and Training Guide* were presented. Each position posed two questions: one on evaluation and one on the superior move, with four options given. Responses to both were scored and converted to a rating estimate.

Stockfish 17.1 was used to calculate the accuracy for each player in their previous three OTB (over-the-board) games.

The relationship between each player's FIDE rating and estimated rating was then compared to accuracy and demographic information including player geographic location and gender.

Creating and Evaluating the Unified Chess Stress Index

Based on the information gathered, a new rating system was devised in which a player's 'score' is calculated using their game accuracy.

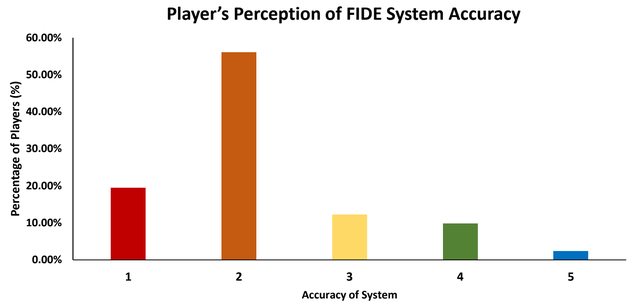
$$R_{old} + K \times (S - E) = R_{new}$$

$$S = \text{Accuracy} \times \textit{Multiplier}$$

It was used in a 5-game chess tournament among amateur unrated players who were all assigned an identical initial rating. Its performance was compared to that of the Elo rating system through analyses of the coefficient of variation and standard deviation.

RESULTS

1. How Accurately Do You Think The Current Rating System Represents Your Skill?

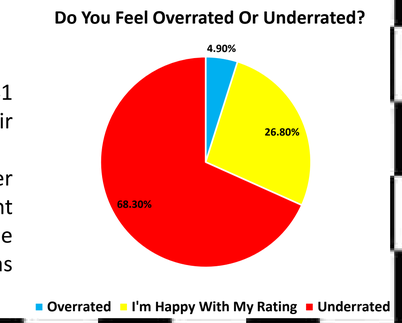


Of the 41 elite Irish chess players surveyed, most are unhappy with the current FIDE rating system. Only 2.4% felt their rating was highly accurate, while 75.6% believed it was moderately or highly inaccurate. This showed that the FIDE system was indeed an error and needed an alternative.

2. Do You Feel Overrated or Underrated?

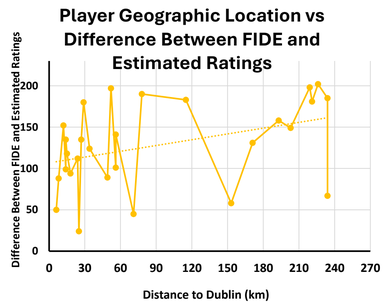
This pie-chart shows the opinion of the 41 players surveyed regarding how accurate their rating is.

Here it is shown that 68.3% of players consider themselves underrated while 26.8% thought the they were happy with their rating. The remaining 4.9% described themselves as overrated.

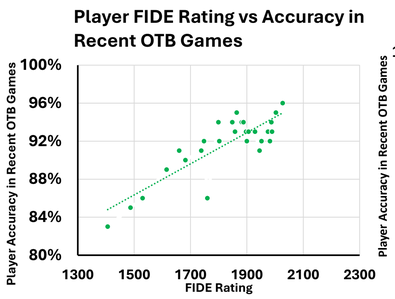


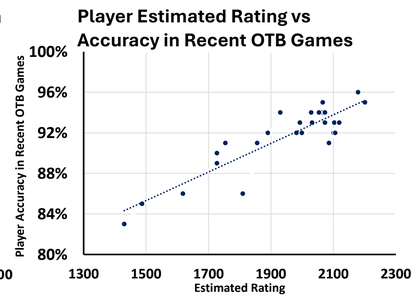
3. Player Geographic Location Between FIDE and Estimated Ratings

A positive Pearson correlation of 0.385736 was observed and a statistically significant p-value of 0.042633 was noted. This tells us that players living further from the capital are more likely to have ratings that do not match their true ability.



4. Comparing FIDE Ratings and Estimated Ratings

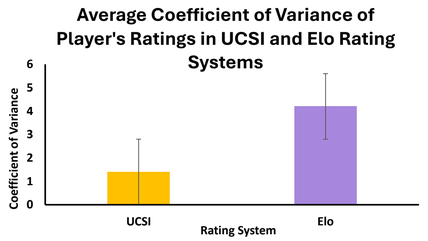




An estimated rating derived from a chess quiz demonstrated a strong and statistically significant Pearson correlation with players' FIDE ratings. However, it showed an even stronger correlation with actual playing accuracy (r = 0.89) than FIDE ratings did (r = 0.84), indicating that it may provide a more accurate reflection of true playing strength.

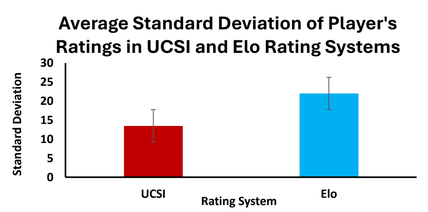
5. The Average Coefficient Of Variance in the UCSI v Elo Rating Systems

The coefficient of variance for Elo ratings was 4.24496189 while the CV for UCSI ratings was 1.23. This tells us that UCSI ratings are more stable than Elo. The lower CV in the UCSI System suggests the system is more stable and consistent.



6. The Average Standard Deviation in the UCSI v Elo Rating Systems

The average standard deviation for Elo ratings was 22.07748 while the SD for UCSI ratings was 13.7188. This tells us that UCSI ratings change less from game to game. They are more stable and reliable.



Conclusion

Examination of Player Attitudes towards the FIDE Rating System

Most elite Irish chess players surveyed are unhappy with the current FIDE rating system. Only 2.4% felt their rating was highly accurate, while 75.6% believed it was moderately or highly inaccurate. Most players (68.3%) felt they were underrated, especially juniors.

Players identified fast-improving young players as a major cause of rating instability, as their low initial ratings lead to large rating swings for opponents. Limited access to tournaments outside Dublin was also seen as a factor in slowing rating progress for rural players.

Estimation of Player's Skill Level using Puzzle Survey

An estimated rating calculated using a chess quiz showed a strong and statistically significant correlation with FIDE ratings. However, the estimated rating had a stronger correlation to actual playing accuracy (r=0.89) than FIDE ratings (r=0.84), suggesting it better reflects true skill.

Distance from Dublin did moderately correlate with the difference between FIDE and estimated ratings. Players living further from the capital were more likely to have ratings that did not match their true ability.

Creating and Evaluating the Unified Chess Stress Index

A new rating system (UCSI), which includes move accuracy, was tested. Compared to Elo, UCSI produced similar trends but lower volatility, meaning ratings changed more smoothly.

On average, UCSI had a significantly lower standard deviation (13.7188 vs 22.07748) and coefficient of variation than Elo (1.23 vs 4.24496189), showing it is more stable and consistent while still tracking performance over time. This suggests UCSI gives a fairer and more reliable measure of a player's true strength.