Chess Style Ranking Proposal for Run5 Ladder Participants
Version 3.2

This proposal is based upon a modification of US Chess Federation methods for calculating ratings of chess players. It is a probability based method that attempts to quantify the probability that a ranked player will beat another ranked player in a game played between the two of them. The current version represents my good faith effort of estimating where consensus lies, based upon discussions at Run5 discussion board. I've included some commentary in blue to explain some of the rationales for the choices.

Definitions:

R_A = player A's current ranking
R_B = player B's current ranking

W_A = expectation of player A in a game against player B

$$W_A = \frac{1}{\left(1+10^{(R_B-R_A)/1000}\right)}$$

Ideally W_A is the probability that A will beat B. We don't know the actual probability distribution for the players, so this must be estimated from actual game play data points. So W_A is actually an estimate of that probability. The longer the system is in effect and the more games A and B have played, the more accurate the estimate.

The use of 1000 in the above formula means that a player who is ranked a thousand points higher than another has an estimated 90% probability of winning a game played between the two of them. Conversely 10% probability of winning if he is a thousand points below his opponent. The US Chess federation (according to my Googled sources) uses a value of 400. I think we want a larger spread in rating numbers for a variety of reason. We have a larger variety of results than in Chess so a larger spread would be helpful. Also, 1000 is a nice round metric-like number.

I) Initial Ranking at the start

Players registered on the ladder at the start of the scheme will have their rank determined in one of two ways. If they have not played any reported games their initial rank will be 2000.

If they have played in reported games their initial rank shall be based on an algorithm based on their current Won/Loss record as of the moment the scheme starts. These rankings are said to be "established".

The initial ranking was determined in the following manner:

The current database of games played had who won what game against whom and when it occurred. It does not contain information about the level of victory. However, the player stats are included which show the number of games played, the total wins, losses and points earned.

Originally some proposals were put forth on the forums of ways to run through the games and assign a victory level to the winner so that the players could get an initial rank based on the games they had already played. The most popular method seemed to be one using a formula using the winner's # of games won (*ww*), the winner's # of games played (*wg*) and similarly for the loser's # of games won and played. The formula below shows how *V*, the "victory percentage", is calculated from those numbers.

$$V = \frac{\left(ww/wg\right)^3}{\left(\left(ww/wg\right)^3 + \left(\ell w/\ell g\right)^3\right)}$$

V is an estimate of the level of victory for the game based upon this information. For example, *V* represents approximately a decisive victory when a very good player plays beats an average player:

$$V = \frac{\left(10/11\right)^3}{\left(\left(10/11\right)^3 + \left(16/27\right)^3\right)} = 0.78$$

An overwhelming victory for a win over a below average player:

$$V = \frac{\left(10/11\right)^3}{\left(\left(10/11\right)^3 + \left(1/11\right)^3\right)} = 0.99$$

and Draws vs. equally good players:

$$V = \frac{\left(10/11\right)^3}{\left(\left(10/11\right)^3 + \left(10/11\right)^3\right)} = 0.5$$

Through a sequence of detective work Pterrok was able to figure out to very good accuracy what the victory levels for most games for most people were and the initial rankings were based on this deductive work. The process of figuring this out drew on the above formulae as aids; but the actual process is described more fully in the appendix at the end of the proposal.

II) Initial Rank of a newly registered ladder participant

A newly registered player will enter with a ranking of 2000 points.

There has been discussion of developing a provisional ranking system for newly registered players; but none of the schemes discussed to date seem to work properly. Consequently this aspect of the scheme may be revisited in the future.

III) Ranking Adjustment after a played game

S = equals the score a player obtains from a game against another ranked player.

S = +1 for an overwhelming win
S = +0.85 for a decisive win
S = +0.65 for a marginal win
S = +0.5 for a draw
S = +.35 for a marginal loss
S = +.15 for a decisive loss
S = + 0 for a overwhelming loss

Chess only has wins, draws, and losses for results so in the chess version S = 1 or 0.5 or 0 respectively. The draw value needs to have a value of 0.5 so that two players of equal ranking who play a drawn game will not have their rankings adjusted. (See below) In determining the level of victory it is considered legitimate to include "handicap points" your opponent has granted you (or has bid).

Player A plays a game against player B: Player A's ranking will be adjusted after the game according to:

$$\boxed{R_A = R_A + e \times k \times (S - W_A)},$$

and similarly for player B.

k is determined by the player's own current ranking.

k = 40 points if the player's ranking is 4000 points or higher
k = 60 points if the player's ranking is 2500-3999 points
k = 80 points if the player's ranking is below 2499 points

The k value is basically the number of ranking points that change hands after a game, adjusted by various factors, such as the probability of a win. For example, if you had a zero expectation and won, you'd get 80 points if you were a low ranked player.

The brackets for the *k* values aren't necessary. One could simply use a value of 80 for all rankings. The chess world uses the above bracketing system (scaled accordingly) in order to combat ranking "inflation" and "deflation". This is a phenomenon where high ranked players could artificially raise their ranking relative to low ranked players by only playing amongst themselves. It also encourages play between higher and lower ranked players.

The e-value for each scenario will be determined by the following formula:

$$e = \frac{\mathrm{Min}(N,24)/24 + \sqrt{\mathrm{Min}(U,500)/500}}{2}.$$

Where N = the number of turns in the scenario and U = the number of units in the scenario.

The "golden" values of 24 and 500 used in the above formula were chosen to insure that the standard long scenarios in all the DBWII releases to date would have a value of one.

The chess world has what they call half *k* and quarter *k* events. (Things like speed chess events, or blind chess, etc). Presumably they feel victories in those types of events aren't as reflective of playing skill as regular tournament games. We have a similar situation in DBWII, where the shorter scenarios are viewed as being less reflective of playing skill. More complex scenarios are more reflective of a player's skill. Discussion has shown that scenario length and number of units are rough measures of this complexity. Consequently such long, high unit number scenarios warrant a higher exchange of ranking points than the smaller ones. Of course, at some point this increase in number doesn't really lead to additional complexity (a version of the "law of diminishing returns"). Part of the statistical basis for the scheme depends on the (points exchanged)/1000 ratio not being too large. This implies that *e* should be no larger than one.

IV) Odds and Ends

a) There has been some debate has to whether or not there should be inactivity penalties established. A consensus has been formed to not have such a penalty at this time. It is understood that in the future if such a penalty is deemed desirable it will be possible to establish without disrupting or changing any of the above.

b) Similarly consensus has been formed to **not** have a floor to the ranking points. Again it is understood that a floor can be established in the future if deemed desirable without disrupting or changing any of the above.

c) Several people from SSG (Tempest, Troutie and Roger) will be given an initial rank of 90% of the top rated player at the start of the scheme. If the formula rates them higher than that, then naturally they'll get the higher rating.

IV) References

My information on chess style rating schemes comes from

http://www.sizes.com/sports/chess_ratings and

http://www.tutorgig.com/encyclopedia/getdefn.jsp?keywords=Elo_rating

V) Acknowledgements

I would like to acknowledge all who participated in the discussions at the Run5 site regarding this ranking scheme. In particular, Pterrok, Twinkle and Fantassin, who contributed valuable insight from their experience with the DBWII game system and who have contributed time in checking calculations, collecting scenario data and otherwise improving the original scheme.

Special thanks go to Joss, who has done much of the website coding for the ladders.

Joel Rauber (Joel137)

VI) Proofs

$$W_A = \frac{1}{\left(1+10^{(R_B-R_A)/1000}\right)} \text{ and } W_B = \frac{1}{\left(1+10^{(R_A-R_B)/1000}\right)}$$

Therefore,

$$W_A + W_B = \frac{1}{\left(1+10^{(R_B-R_A)/1000}\right)} + \frac{1}{\left(1+10^{(R_A-R_B)/1000}\right)}$$

$$= \frac{\left(1+10^{(R_A-R_B)/1000}\right)+\left(1+10^{(R_B-R_A)/1000}\right)}{\left(1+10^{(R_A-R_B)/1000}\right)\left(1+10^{(R_B-R_A)/1000}\right)}$$

$$= \frac{2+10^{(R_A-R_B)/1000}+10^{(R_B-R_A)/1000}}{1+10^{(R_A-R_B)/1000}+10^{(R_B-R_A)/1000}+10^{(R_A-R_B)/1000}10^{(R_B-R_A)/1000}}$$

$$= \frac{2+10^{(R_A-R_B)/1000}+10^{(R_B-R_A)/1000}}{1+10^{(R_A-R_B)/1000}+10^{(R_B-R_A)/1000}+10^{\left[(R_A-R_B)+(R_B-R_A)\right]/1000}}$$

$$= \frac{2+10^{(R_A-R_B)/1000}+10^{(R_B-R_A)/1000}}{2+10^{(R_A-R_B)/1000}+10^{(R_B-R_A)/1000}} = 1$$

VII) Appendix (Pterrok's Commentary on initial rankings)

One method that did NOT seem to garner much support was just assigning an Overwhelming victory to the winner, even though Twinkle's experience was that he got all Overwhelming wins and a few Draws, which was also MY personal experience, although I have a lot of overwhelming losses to counter the overwhelming wins!

So what I did was make a dump of all the played games sorted by players from those with least amount of played games and increasing all the way down to Fairfax who had an astounding 58 games played!

I then looked at the first player's played game and since I knew what the point's multiplier for that scenario was, it was easy to say with certainty that this game HAD to have been THIS victory level since there is only one way to get that amount of points in one game.

I then went into the games played list and added the victory level to it and recompiled the list and now it could give out points for the victory level. The first player showed up with a SET!!! flag and if I scrolled down to his opponent in the list, I would see that he had x of y points needed to achieve his first year ladder points total.

After that it was a simple matter of inspection--it was quite like those Sum-it Squares puzzles in crossword books!--as the players with low number of games HAD to have certain combos. Entering them all in and re-running the list yielded more SET!!! players and more with x of y points totals.

A major problem, however, is that not all the game data is present and/or correct. For example, it might say a player played 5 games and yet there would only be 3 games with his name in the list, or much more common was that it said he had played 4 games and yet there were 5 games in the list!

This came about because the tables were edited by-hand when a game result was entered by mistake and/or when the database was recovered after it crashed. I decided to always use the number of games played and the total points as the arbiter and so I was actually able to deduce some games that should have been removed from the played list. But as the data was not perfect, perfect results could not be achieved by ANY method.

After getting through all the games that could be determined absolutely I was left with a mess of games where the players were all interconnected. Surprisingly, however, by either assuming the player got the MINIMUM possible number of points in all his games if he was 'bad' or the MAXIMUM possible points in each game if he was 'good' got totals for those players which was very close to their season total.

So I tweaked a game or two to get those players into the proper points total. When all was said and done I was left with Robjess and Fairfax who should get a medal for playing against themselves more than most people have played games! However, I got Fairfax to come out correctly, but Robjess was off. In fact, I only got him to be at 252 of 265 season points when all was said and done. This is a difference of 13, or 4.9% of his total.

My first thought was to just up his Initial Rank by 4.9% after I determined what it should be, but then I got tricky...I changed one of the victory levels of one of his games and noted how many points his points total went up. Then I ran that game list through the ranking code and saw how much his Rank increased! Empirically it turned out that whatever amount his points changed by, his Rank would change by:

(difference from previous total) / (season total) * previous Rank / 2

It came out close to that empirical formula if I gave him MORE points by upping a victory level or LESS points by lowering a victory level! It happened for one other player as well, and so I ran with that as the adjustment factor.

Out of 105 players with a recorded game, 90 came out with the proper points totals, 7 with too many points and 8 with too few, so those 15 had an adjustment applied to their rank when the game list with victory levels was run through the ranking code.

I can guarantee that all the game results are NOT correct, simply because not all the data was in there properly. Beyond that, it's also because in some cases I had to make a choice and arbitrarily assigned a level and then balanced things out from there. However, I feel that even with this proviso it is in aggregate closer than any method presented on the forum. (And BTW, I personally came out the worst on this method--but I know my game results and entered them properly and that's the breaks!)

The final ranking code gave each player 2000 points and played each game according to the sorted game list and assigned victory level and Ranks got adjusted from there.

I had tried a run where each player was given 2000 PROVISIONAL points and needed 8 game submissions to determine a true rank; however, ALL games had to count and so when a newbie played another newbie, the current provisional rank was used in lieu of an established rank.

The results were odd and this may be due to a bug in the implementation. However, the undefeated Fantassin fell down the list quite a-ways. On one of the forum discussions I had thought that using 8 game submissions was OK, but from Joel's original provisional ranking formula, the e value IS part of the Newbie formula and Fantassin has an e-game season total of 9.99--just barely out of newbie range since he played so many Wiking Whiteout tournament games.

So I ran the code again, this time keeping players as newbies based on e-games instead of games submitted. Fantassin fell even farther down the list. The problem may very well be in counting games of newbies vs. newbies which the formula says DON'T do, but after having gone through all the data so many times I can tell you we DON'T have enough active players with enough e-games themselves to be out of newbie range to be able to afford a newbie coming to the site and playing a game and being told it doesn't count since it was against another newbie instead of an established player.

Consequently, I suggested to Joel and others to NOT worry about any sort of provisional ranking newbie code this year--just give them 2000 points and let them play anyone and everyone. Next year we may have to revisit this if the top players get very far ahead and/or if we grow the number of active, established players.