Ratings System Review

Implementation Report

Prepared by: Steven Brown

September 2017

RATINGS SYSTEMS REVIEW

Background

The management meeting at the 2014 Nationals asked Glenda Foster to convene a panel to review the ratings system. Glenda advised the management meeting at the 2015 Nationals that she had nothing to report, and felt that I would be in a better position to review the system. Over the following year I was reviewing the system, posting material to the #ratings-review channel at nzscrabble.slack.com, and seeking feedback from others interested in the review. I presented an interim report to the management meeting at the 2016 Nationals, and received useful feedback. I undertook to produce a final report for consideration at the October 2016 management meeting (at the Mt Albert tournament). This final report should be read in conjunction with the interim report. It will primarily consist of a list of amendments to the interim report's recommended course of action; followed by a detailed description of the current system; the proposed changes to the current system required by the recommended course of action; and any transitional measures needed.

Interim Report

The interim report covered the following topics: Background, Recommended Goals of Review, Approach to Review, Rating System Theory and History, Analysis of ratings database, Issues with Current System, Recommended Course of Action, and Feedback Sought on Review.

The only material from the interim report that I will reproduce here is the recommended goals of the overall review; and the interim report's recommended course of action (with recommended amendments added).

Final Report

~~The interim report covered the following topics: Background, Recommended Goals of Review, Approach to Review, Rating System Theory and History, Analysis of ratings database, Issues with Current System, Recommended Course of Action, and Feedback Sought on Review.~~

~~The only material from the interim report that I will reproduce here is the recommended goals of the overall review; and the interim report's recommended course of action (with recommended amendments added).~~

Recommendations of Final Report

Adopt NASPA/WESPA curve (close enough is good enough, ease of development). [no change]

*Modify K factor from (3000-currentRating)/50 to (3000-currentRating)/1000*ratedGames.* [no change]

Allow back-to-back one-day tournaments to be rated as if they are one two-day tournament.

It may be that this becomes a requirement (for both ease of administration of the ratings system, and its accuracy), i.e. that overnight re-rating between tournaments run by the same club would be disallowed. Players would still be rated according to the number of games actually played. I.e. someone who only played in the first day of two one-day (7-game) tournaments would be rated on the 7 games they played, while someone who played both days would have an expectancy out all 14 games they (the second player) played.

Derive initial rating from first tournament only (in line with NASPA). [new recommendation below]
Reduce provisional games to 30 (in line with WESPA). [no change]
Allow all New, Provisional, Historic players to be re-graded (with agreement). [no change]
Rate all New, Provisional, Historic players on the same basis as established players. [new rec. below]
Will now revert to provisional ratings for each tournament until the number of provisional games has been exceeded, but will base each provisional rating on all games so far, not just the current tournament (as in the current system). Also the provisional rating will be an average of per-opponent provisional ratings (vs the player's win percentage for the tournament) rather than the current provisional rating being based on the average opponent rating. (I.e. change the provisional rating methodology to be more in line with the current methodology for calculating expectancies for those with established ratings.

Add 'participation points' to each new rating (one point per game rated).
Change to one point per three games rated - this will bring the number of points inserted into the system closer to the number lost when experienced players exit the system.

Start new system from first tournament next year. [no change]

Add 'accelerator' and 'feedback' points to the current system.
As per the NASPA system (and others), accelerator and feedback points are designed to account for players who are rapidly rising in the ratings, and may have been placed in a higher grade than their rating would indicate. Any player who does particularly well in comparison to their expectancy is assumed to have been severely underrated going in to the tournament, and receives extra points to help speed them up to a higher rating. Players who played that player receive feedback points to compensate for points they would not have lost if the player had been higher rated in the first place.

Current System

Provisional ratings - ratings curve.
Under the current system, ratings are calculated in two passes. In the first pass, provisional ratings are calculated for each player who has played fewer than 35 tournament games in the NZ system before the current tournament (including players new to the NZ rating system). These are calculated using:

$$\text{AverageOpponentRating} + \text{NormSInv}(\text{WinPercent}) * 282.842712474619$$

where NormSInv(p) is the function for looking up the inverse of the Cumulative Normal Standard Distribution, which is built in the Windows Excel, and 282.842712474619 is 200 times the square root of 2. If the application of this formula results in a negative rating, the new rating is rounded up to zero. The average opponent rating includes the ratings of players starting the tournament with either established or provisional ratings, but does not include any players new to the NZ system, although results against such players are included in the win percentage. If all opponents are new to the NZ system, the player is assigned a provisional rating of 500, regardless of whether the player themself is new to the NZ system or (as happened at Kāpiti Coast some years ago) they already have a provisional rating from previous tournaments. Nigel comments in the source code that "this needs some more work, although I would doubt that this situation will ever occur".

Note that any provisional rating from previous tournaments has no input into working out the player's new provisional rating (besides having determined their grading and thus the average rating of their opponents). This means that a player's first established rating could be based on anything from the last 23 games out of their first 35 to the last 7 games out of their first 42, depending on the number of games they have played before the tournament that gives them over 35 games, and the number of games in that tournament.

Established ratings - ratings curve.

In the second pass, the expected win percentage for each game for players with an established rating is calculated using:

NormSDist(AbsoluteValue(PlayerRating - OpponentRating) / 282.842712474619)

where NormSDist(x) is the Cumulative Normal Standard Distribution function built in to Microsoft Excel. These values are summed to give the player's expectancy for the tournament. The OpponentRating for new and provisional players is their new provisional rating following the first pass.

Established ratings - k-factor.

For players who started the tournament with an established rating, the difference between expected wins and actual wins are then multiplied by a k-factor to give the number of points the player should gain or lose, using this formula:

(3000-currentRating)/50

A player who starts a tournament with a rating of 1000 thus has a k-factor of 40, while a player who starts on 2000 has a k-factor of 20, etc. If subtracting (wins below expectancy) * k-factor from a player's initial rating gives a negative rating, it is rounded up to zero.

Historic ratings.

If a player has played no tournament games for at least 2 years before the current tournament, the player may be placed in any grade agreed on by the player and the tournament director (as is the case with players new to the NZ rating system). If agreement cannot be reached, the player must be placed in the grade indicated by their "historic" rating (or in the bottom grade in the case of "new" players). In such cases, if the player with a historic rating is placed in a higher or lower grade than their rating would indicate, then once the new ratings have initially been calculated, the player's new rating is taken to be their new initial rating, and the ratings are re-calculated on this basis. This has a similar effect to the role of 'accelerator' and 'feedback' points in other systems.

Proposed Changes

Provisional ratings - ratings curve.

I recommend that the number of games where the provisional rating methods are applied for tournaments where the player starts with less than that number of games be reduced to 30, as in the WESPA system. I also recommend that the provisional rating achieved after the tournament before the tournament in which a player passes 30 games be saved, along with the number of games in all prior tournaments, and be used to calculate the new provisional rating for each new tournament. The provisional rating from the tournament in which the player exceeds 30 tournament games then becomes their first established rating for entering their next tournament. The number of games included in the final provisional rating calculation will therefore vary

between 30 (if the player starts an n-game tournament having played 30-n games beforehand) and 44 (if the player starts a 15-game tournament having played 29 games before).

I recommend changing the curve from the current Normal Standard Distribution curve to a logistic curve with the same scaling as used by both NASPA and WESPA. Provisional ratings would use this formula for each game being included in the provisional rating (including games against opponents new to the NZ system):

OpponentRating - Log(1 / WinRate - 1) * 313

where the OpponentRating is what the opponent in question's rating was at the time of the game, and the WinRate is for the tournament the game was part of. The resulting provisional ratings would be summed along with the total per game provisional ratings from previous tournaments and then divided by the total number of games. Log(p) is a widely available function giving a natural (i.e. base-e) logarithm. 313 is the empirically derived scaling factor arrived at independently by NASPA, WESPA, and my analysis of the NZ ratings database. The actual figure used by NASPA is close to 1/313, with multiplication and division swapped as appropriate in the formulas.

Established ratings - ratings curve.

For players with established ratings, I recommend calculating the expected win percentage for each game using:

1 / (1 + Exp(AbsoluteValue(PlayerRating - OpponentRating) / -313))

where Exp(x) is a widely available function which gives the value of e raised to power of x (i.e. the inverse of the natural logarithm). The value 313 is as explained above. The calculation of expectancies for players with established ratings would otherwise be as in the current system.

Established ratings - k-factor.

I recommend amending the formula for calculating the k-factor to:

(3000-currentRating)/1000*ratedGames

which is equivalent to half of the current k-factor, multiplied by 1/10th of the number of games that the new established rating is being calculated for. I.e. playing 20 rated games in a tournament would give the same k-factor as the current system, playing 10 games would give half the current k-factor (up or down).

Accelerator and feedback points.

I recommend that points gained in excess of 5 points per game played be doubled, and opponents of anyone in receipt of accelerator points receive 0.05 points per accelerator points, per game played against the player who received them. Very roughly, if someone in a 14-game double round robin exceeds their expectancy by around 4 games, they would receive around 25 accelerator points, and each of their opponents would get 2 or 3 feedback points (based on having played the player twice). I recommend that these rates of accelerator and feedback points be reviewed after the first two years of operation of the new system.

Accelerator points would not apply to the rating achieved by a new player in their first NZ tournament, as there is no previous rating to compare it to.

Participation points.

I recommend that 'participation' points be awarded, partly to combat ratings deflation (as higher-rated players leave the system and new players join the system), and partly to recognise that a new rating is more reliably comparable to other ratings than the rating of someone who may not have played for a long time.

I recommend trialling a rate of 1 participation point for each 3 games played in a tournament (i.e. 6, 7, or 8 games played would get 2 points added after the new rating is calculated, 12, 13, or 14 games played would get 4 points, etc.).

New, provisional, and historic ratings.

I recommend that any player who is provisionally-rated at the beginning of a tournament may be placed in a different grade to that indicated by their rating (if the player and tournament organiser agree on a suitable grade), in the same way that new players and players with historic ratings may already be re-graded.

I recommend that the current method of re-rating players with historic ratings if they are played in a different grade be discontinued, in light of the adoption of accelerator, feedback, and participation points.

Transitional Measures

Provisional ratings.

For players listed as having provisional ratings in the end-of-year ratings (i.e. who have played fewer than 35 games in the current system), if they have played at least 30 games, their current provisional rating will become an established rating for the start of the new system. If they have played fewer than 30 games, their provisional ratings after each of their prior tournaments will be used in the new system as if it was a partial provisional rating under the new system.

For example, take a player who has played a 7-game tournament, and a 13-game tournament under the current system, and then plays a 7-game tournament under the new system. The overall calculation will be to add 7 times the provisional rating after the first tournament to 13 times the provisional rating after the second tournament, and then add the 7 per-game provisional ratings from the third tournament separately. Then divide the whole lot by 27. Subsequent tournaments can then revert to the new system - i.e. add 27 times the initial provisional rating to the new per-game ratings, then divide by the new total games.

Major Recommendations

i) That the above proposed changes and transitional measures be adopted for tournaments starting with the first tournament next year (2017).

ii) That I produce a cross-platform implementation of the new system, that can be used to easily administer it.

iii) That the new system be reviewed after two years of operation.

(*Please note that some formulas above omit the rounding required to give whole-number ratings. This is solely for the sake of clarity - any rounding that is part of the current system will carry over to the new system, as appropriate.*)

I would particularly like to thank Chris Handley, Ed Okulicz, Glenda, Olivia, and those who attended the June management meeting for their input into various stages of this review.