Review of the NZASP Ratings System

Final Report

Prepared by: Steven Brown (Rating System Manager), on behalf of the Ratings Review Panel

Panel: Dylan Early (convenor), Steven Brown, Patrick Carter, John Foster, Alastair Richards

15 January 2019

(v2 - minor typos fixed 23 January 2019)

REVIEW OF THE NZASP RATINGS SYSTEM

Terms of the Review

Terms

The NZ rating system, as far as we still have records, was established by Roy Vanini (the 1988 system). This was modified by Nigel Richards (the 1999 system), and again by Steven Brown (the 2017 system).

The Association has asked us to consider:

- how fair and effective is the 2017 system?
- if it could be more fair and/or effective, what further modifications should be made?
- when should any modifications take effect?

with getting a fair, workable, and widely-accepted system being our primary objective.

Method

To achieve this, we need to balance:

- fairness (both absolute and perceived, and to all kinds of players)
- simplicity (of understanding, implementing, and administering the system)
- continuity (i.e. minimising disruption to the ongoing operation of the system and associated workloads)
- consistency with other rating systems.

Timing

End of year

We consider that any changes (and any rebasing of an existing system) should only take place at the end of a calendar year. This is for four main reasons:

- that all players in a year's worth of tournaments should be treated equally, regardless of where they live and which tournaments they can get to at different times of the year;
- that this discourages over-frequent changes to the system;
- that it is much easier (and less error-prone) to implement and administer the system when all changes can be associated with whole years;
- and that future analysis of the system in future reviews will be easier and more meaningful if data is available in larger and more consistent chunks.

Method

Implement obvious and easy-to-implement changes for the beginning of this year (2019), and save anything more complex, uncertain, or time-consuming for further analysis, and implementation at the beginning of a subsequent year following a subsequent review.

The Ratings Manager will rate the upcoming Wellington tournament under the new system, and generate new rating lists from these results to be used for grading the upcoming Pakuranga tournament. Should the management meeting at the Pakuranga tournament reject the proposed changes, the Wellington tournament will be re-rated under the current (2017) system, and these ratings used to rate the Pakuranga tournament.

Analysis

This report will examine the major aspects of the rating system, looking at: what was in place in the 1999 system for each aspect, and any problems with this that the 2017 changes were intended to address; the changes introduced for 2017, and how they appear to have worked in the following two years; and our recommendation for further change, if any.

Detailed analysis referred to in this report is available in the #ratings-review channel at nzscrabble.slack.com

Ratings curve

1999 system

The 1999 system used the following ratings curve, using Excel spreadsheet functions:

NormSDist(AbsoluteValue(PlayerRating - OpponentRating) / (Sqrt(2)*200))

The inverse of this function was used in the process of calculating provisional ratings:

AverageOpponentRating + NormSInv(WinPercent) * (Sqrt(2)*200)

The main complaint about this ratings curve from players over the years was that expectancies were too high for players at the top of each grade, and too low for the players at the bottom of each grade. The use of Excel-specific functions also made it hard to implement the ratings system on other platforms.

In 2015-16, the Ratings Manager analysed results from the first 17 years of the 1999 system, and compared how the results matched up to: the existing curve; the old Australian curve; the current Australian curve; the current WESPA curve (almost exactly the same as the current NASPA curve); and the runner-up for being chosen as the current WESPA curve. The current WESPA curve was clearly the best fit, followed by the WESPA runner-up, followed a little way back by the current Australian curve, with the 1999 NZ curve and the old Australian curve being less fair by almost an order of magnitude.

2017 system

The 2017 system uses the logistic curve used in the WESPA system, where EXP(*x*) = *e* to the power of *x*:

1/(1 + EXP((PlayerRating - OpponentRating)/-313))

The inverse (scaled to avoid infinities) is used for provisional ratings, where LN(*e* to the power of *x*) = *x*:

LN(1/winRatio - 1) * -313

Recommendation for 2019

Keep the 2017 system curve. Further, more rigorous, analysis of the 18 years of the 1999 system, and of the 2 years using the new curve, show the same results as the analysis undertaken in 2016 - that it is the best fit of any of the curves that have been used or considered for use by major rating systems. Some other curves, not used in or considered by other ratings system might be an even better fit, but this is outweighed by the benefits of continuing to use the same curve as WESPA (and virtually the same as NASPA).

Multiplier or k-factor

1999 system

The k-factor, also known as the multiplier, is the value by which the difference between an established player's win ratio and their expectancy is multiplied to give their ratings change. The higher the value, the more affect the current tournament has on a player's rating (up or down), compared to the affect of all their previous tournaments in reaching that rating. The formula used in the 1999 system was: **(3000-startRating)/50**

This can be restated as: **((3000-startRating)/100) * 2** (for ease of comparison with alternatives).

The unfairness of the 1999 ratings curve meant that the ratings of players at the top or bottom of grades changed more than it otherwise should. The value of the k-factor was also somewhat higher than in comparable systems, but probably needed to be to compensate for the problems with the ratings curve. The 3000 in the formula established this as a theoretical maximum for ratings in the system - as players' ratings increase, so do their expectancies, but the k-factor steadily decreases. In practice, the unfairness of the ratings curve in the 1999 system meant that players rarely got much over 2000.

2017 system

The 2017 system uses: **((3000-startRating)/100) * (gamesInTourney/10)**

This means that a player on exactly 2000 points would have a k-factor equalling the number of games they played in the tournament, while a player on exactly 1000 points would have a k-factor of twice this value.

Adopting this formula was seen as slightly reducing the k-factor on average, in light of the fairer ratings curve, while giving more weight to results gained at longer tournaments. We now consider that while it was a move in the right direction for minimising volatility in the system, the implementation was over-complicated from a player's point of view.

Recommendation for 2019

Change the k-factor formula to: **((3000-rating)/100) * 1.5**

This is equivalent to the k-factor for a 15-game tournament under the 2017 system, and thus provides the best continuity between the 2017 system and what we are recommending.

Participation points

1999 system

The 1999 system did not include this feature. Under the 1999 system, players would sometimes avoid tournaments if they were likely to be seeded at the top of a grade so as to avoid losing too many rating points. The level of the top-rated players in the system had also been steadily falling since around 2008, which appeared to be a symptom of ratings deflation. It was thought that participation points would help by giving more weight to the ratings of players who played more often, and thereby encourage more players to play more often, and to help replace points lost to the system as higher-rated players stopped playing.

2017 system

Participation points were introduced in the 2017 system, at a rate of 1 point per 3 games played by each player. They appear to have been too confusing to players to be worth retaining for any other benefits.

Recommendation for 2019

Remove entirely, but investigate in a future review the possibility of pairing their reintroduction with some other method of adjusting the overall level of ratings.

Treatment of New, Provisionally-rated, and Historically-rated players

1999 system

Under the 1999 system, a player received a provisional rating at a tournament if they had played fewer than 35 tournament games in NZ before that tournament. Players were only allowed to be placed in a grade other than indicated by their rating, provisional or otherwise, if they had previously played no games in NZ.

As noted above, the formula used to help calculate provisional ratings is the inverse of the formula used to calculate expectancies, whatever the ratings curve. So, given that the 1999 ratings curve gave an unfairly high expectancy to players rated higher than their opponents, its inverse gave an unfairly low provisional rating to players who achieved a win percentage of less than 50%; and where the 1999 ratings curve gave an unfairly low expectancy to players rated lower than their opponents, its inverse gave an unfairly high provisional rating to players who achieved a win percentage of more than 50%.

Also under the 1999 system, provisionally-rated players received a new rating after each new tournament, based only on their results at that tournament (compared to the ratings of their opposition at that tournament), with no reference to any previous provisional rating, except as used to place them in a grade. So, a talented club or overseas player could be placed in a high grade for their first tournament, and do okay, getting a first rating around the average of the grade they'd been placed in. There would then be no further discretion as to which grade to place them in. They could then have an off day at a one-day tournament, going down almost 300 points for winning only one game, whereas a player with a similar, but established, rating in the same grade would only go down 50-odd points on the same result. The provisional player would then have difficulty getting those points back at all quickly, both while still provisionally-rated and once their rating became an established one.

While the 1999 system was still in use, discretion for the placement in higher or lower grades than their rating would indicate (by the tournament director, with the agreement of the player concerned) of all provisional players and of players with historic ratings (i.e. those who had played no tournaments for at least 2 full years) was introduced. This introduced a risk that the opponents of a historically-rated player placed in a higher grade would be unfairly disadvantaged if that player performed better than their rating would indicate.

2017 system

The 2017 system reduced the number of games a player would remain provisional for from 35 to 30, and initially established a system where a new provisional rating would be averaged with the previous provisional rating, weighted according to the number of games being added. This made things fairer for players being provisionally rated, and slightly reduced the duration of having to estimate pre-tournament expectancies for grade with provisional players, at the expense of otherwise complicating the system.

When the World Seniors in 2017 was rated under the NZ system, a major flaw in the provisional rating procedures became apparent. A similar flaw would also have become apparent in the 1999 system, if it had still been in place. The flaw was fixed by calculating provisional ratings iteratively rather than as a one-off calculation, and ratings were re-calculated back to the beginning of 2017, as if the new system had worked that way from the start.

In early 2018, a further minor flaw in the provisional rating procedures was realised, but too late to be worth re-rating from the beginning of the year, or from the beginning of the previous year again. Keeping the 2017 system as it is would require implementing a fix for this problem.

Recommendation for 2019

We recommend changing the system for provisional ratings to a system where a player is still classed as provisional until they have played at least 30 NZ tournament games, but where the iterative performance rating mentioned above is calculated only for their very first tournament, whereafter they are treated by the rating system as having an established rating. This would both greatly simplify this aspect of the ratings system, and would mean that all grades could have exact pre-game expectancies easily calculated, except those that include brand-new players. It would also remove the need to implement the second post-World Seniors fix.

The new system would then run the risk, however, of a new player being placed in an inappropriately low grade, and not then having the opportunity of further provisionally-rated games to improve their initial rating. We recommend that this be addressed by keeping acceleration points (and feedback points for opponents) for such players, and for players with a historic rating (but with a higher threshold, as discussed below).

Acceleration and Feedback points

1999 system

The 1999 system did not include this feature. There was a small but recognised risk in the 1999 system that any player whose playing strength had increased markedly compared to their rating, such as some players with historic ratings among others, could be disadvantaged by the time it would take them to reach their new 'true' rating, and that their opponents would suffer from any sudden ratings gain made by the player.

2017 system

Acceleration and feedback points were introduced in 2017 to attempt to cater for players in the situation above in a manner consistent with the NASPA system (https://scrabbleplayers.org/w/Rating_system_overview) (as the details of the WESPA system are not available). The NASPA formula did not work well with our k-factor formula, so was changed so that if a player won more than 3 1/3 games more than their expectancy, the points above this threshold were doubled. Each opponent received 0.05 of this amount per game against the player as feedback points. In practice, this threshold meant that acceleration points were achieved more often than had been intended.

Recommendation for 2019

Remove Acceleration and Feedback points (except as recommended for simplified provisional ratings). For provisional and historic players, increase the threshold to 3.5 games, making them slightly harder to earn.

Other matters

World Seniors 2017

We believe that this tournament should not have been rated in the NZ system, but consider the question of which tournaments should be rated is outside the scope of both this review and the responsibilities of the Ratings Manager, and should be better defined in the Constitution.

Range of ratings in the system

There is sufficient disagreement about the causes, severity, and appropriateness of recent rises in the ratings of top-rated players that we consider that this matter should be left for further consideration by a subsequent review, when the system has had time to adjust to the new (2017) ratings curve and there is a further year or two of data to analyse.