

Critique of the USAPA Tournament Player Rating (UTPR) System.....by Jim Kloss

"EVERYTHING (WELL, ALMOST EVERYTHING) ABOUT THE UTPR SYSTEM EXPLAINED. AND HOW TO FIX THE BROKEN SYSTEM.

There have been and continue to be a lot of posts about the UTPR system and its problems. Along with two others (Carl Tietze and Keith Sturdivant), I spent a lot of time studying the system, trying to talk to USAPA, trying to make suggestions about improving the system, etc. I'll try to provide you all here a summary of what I learned. This is all an expression of my opinion on the system — Carl and Keith can give their own opinions. ? I'll provide what I found out about how it was developed, how it works, the flaws I see, and, as I always do, my suggestions on how to fix a broken system.

1. How was it developed?

USAPA went to pt.com and pickleballcentral.com, asking for help. As I learned, the main work was that of pickleball central. It appears USAPA designated the k factor (a key component explained further below), but did little to nothing else to develop or oversee the product.

2. What is the UTPR and what is it supposed to be?

As I understand it, and as appears to be the intent from the USAPA site, the UTPRs are to be used as a rating and ranking system. Usable by leagues, clubs, rec groups, tournaments, etc. There's no guideline from USAPA of which I'm aware saying the UTPR is limited in its usability.

However, pickleballcentral states that its essentially only a tournament snapshot. Not a grade, not a "batting average", and NOT to be used by leagues, clubs rec groups, etc. It's to be used primarily if not exclusively as a tournament seeding system.

Thus, we have the initial problem — there appears to be a great disconnect between what USAPA wanted, and what it got. I've seen no clarification from USAPA or pickleballcentral advising that the UTPRs should not be used as a grade, as a batting average, or as a method for leagues and clubs to place people.

This history is key, as perhaps the #1 complaint I hear from players is how clubs and leagues are using or misusing the UTPRs. This is troubling when it appears the system as created was never intended for that use. But, if not so intended, why have we, the USAPA members, not been told this?

3. The collection of initial data

The UTPR system began, as you will recall, with collection of data. Mainly from the pt.com archives. But, that archived data was never kept with the intent to use it for a rating system. TDs and players were never told that the accuracy of the data was important. If the right team got the gold medal, did anyone care if the system recorded the winning team as losing in the final? Sure enough, when players were asked to review their tournament data, USAPA reported an "error rate" of 3%. But, that "error rate" was only a statement that 3% of all players reported that they looked at their data told USAPA about errors! How many people looked at their data? How many didn't see errors that did exist? How many saw errors but didn't bother to report them to USAPA? Obviously, LOTS. If we assume even 10% of players who played in at least one tournament looked at their data and all who had an error saw it and reported it, that would mean an actual "error rate" (as USAPA used the term) of 30%! And, do you think even 10% of players looked at their data? I doubt it. So, we don't know the error rate, but it's certainly not insignificant.

I would like to clearly say right here that I in no way put even the tiniest of blame

on pt.com for any initial errors or other problems with the UTPRs. Tournament results prior to the UTPRs were not intended for the UTPR system and there was no reason to ensure 100% accuracy of collected data. As it was not important, there is no reason we should expect pt.com to have ensured the data was accurate, even if such a system was possible at the time.

4. How is the UTPR system supposed to work?

According to the USAPA website, the UTPR system is based on the ELO system, which comes from chess. It was later modified for other sports, such as tennis. There are also other sports with rating/ranking systems, such as badminton. So, let's look at those other sports and their systems.

In chess, which has a good system, you have a rating algorithm. That algorithm calculates your chance of winning v your opponent. The greater the rating disparity, the greater the better player is favored. The more they are favored by, the less the favorite can move up by winning, the more the underdog can move up by winning — this is called the "k factor". The k factor is extremely important. Set it high, and your rating can move a lot. Set it low, and there's less variability.

Chess also has in their algorithm an "Rd factor". The Rd factor takes into account how many tournament results you have. The more data you have in the system, the lower your standard deviation, and thus the more stable your rating will be. For example, if you have 100 games in the system, your rating is twice as stable as someone with 50 games. Makes complete sense, right? How would anyone design a system that lacked the RD factor, as it's plainly a good idea. (For more on the chess system, google "Glicko system")

Tennis — as I understand it, tennis factors in the score of the match, not just the result. Based on your rating and that of your opponent, the system projects a score. Say, something like 6-4, 6-4. If the favorite still wins, but by 7-5, 7-5, the winner will lose rating points, even though they won. Makes sense, doesn't it? If you play Ben Johns and Kyle Yates, and lose 11-9, 9-11, 11-9, it makes no sense that you would LOSE rating points, right? How would anyone design a system that doesn't take into account the score?

Badminton — last example. In badminton, your ranking is based on 52 weeks of results. So, your tournament performance is only 1/52d of your ranking. Of course, if you play fewer tournaments, then each one matters more, but that just makes sense, same as chess.

5. How does the UTPR system actually work?

Does it use tournament history, like badminton? NO. If your rating is 5.250 based on 200 tournaments, it's treated the same as if you are 5.250 based on 5 tournaments.

Does it have an Rd factor, like chess? NO

Does it take into account your score, like tennis? NO

So, even though the USAPA site talks about chess, ELO, other sports, our UTPR is different (I would say wildly different) from those other sports. But why? Unfortunately, I have no answers for you. In my opinion, we should copy chess. The chess ELO appears to be considered as the best system - - why not copy it?

Also, remember the k factor? The UTPR k factor is too high, meaning each tournament result makes your rating move too much. All of you who have complained about big rating swings, this is one reason why. The UTPR computes a team rating, based on the rating of the two partners. That's compared to the opponents' team rating. Say 2 5.200s play 2 5.000s. The system then based on the k factor, predicts who will win and what percentage of time. In my example, the UTPR k factor will predict the 5.2s will win 77.1% of the time. And, the system is set up so that you can win/lose up to .1 rating points per match. So, if the 5.0s win, they

will go up to about 5.0771, while the 5.2s will go down to about 5.1229. I could be off a bit on that math, but you get the idea.

There's no adjustment for the score, or how many events you have played, or any of the other factors other sports use. If you play sand baggers, or new players who self-rate and self-rate too low, or the many other things that can come up, you can lose .2 or more on your rating in a single tournament. If you play in a big round robin and have a bad day, you can literally go from a 5.000 to a 4.250 in one day, even if you have played in 50 events that year and won 50 gold medals.

6. What are the problems?

There's a disconnect between what system USAPA apparently wanted and what we have. But, no one is being told that, so the system is being used for things that pickleballcentral says they didn't intend.

The historical data initially used was not what we needed, in terms of accuracy. Those errors were baked into the system.

There was a failure to use an Rd factor.

There was a failure to have lengthy tournament history factor in.

The k factor is not set correctly, leading to wild swings in the data.

Players self-rating leads to bad data. For example, you could have a great tennis player self-rate at 3.0, beat some 5.0s, and significantly lower the 5.0s ratings.

No one appears to be available to consistently and promptly answer player questions on their ratings.

Score of the match is irrelevant.

7. How do we fix it?

Broadly, i would say the USAPA board should be responsive to member concerns, but as we have seen with many prior issues, the board has not done that. There are several new board members though, and as I'm ever the optimist, I'm hoping one or more of them will see this, ask some questions, and that it will lead to a task force being set up to fix all the issues. I would hope that task force is not made up of "yes men" only, and that it be 100% transparent — reporting to members thoroughly and consistently. But, perhaps i should just focus on hoping my Cleveland Browns win a Super Bowl — that appears to be more likely.

More specifically:

A. Put in an Rd factor, like chess. This is a layup, no brainer. How anyone could object to this addition is truly beyond me.

B. For new players, with no tournament history, have their results not count, in any way, until they have played a certain minimum number of events. I suggest 5, but I'm open to another number. These new players would compete normally, consistent with their self-rating, but their results would not effect their opponents. Once they play 5 events, the computer would generate them a rating and they play normally. Again, a very easy fix.

C. Change the algorithm to account for score. Losing 11-0, 11-0 is simply not the same as losing in 3 games.

D. Change the algorithm to lower the k factor. If 5.2 plays 5.0, i think the 5.2 should only be favored by like 55-45, or 60-40. For it to be about 78-22 is just too high. This is leading to wild rating swings. Smooth it out.

E. Put someone in charge of responding to member rating inquiries. Set a standard that all

reasonable inquiries be responded to with 5 days. Publicize the availability of this service.

F. Figure out what the UTPR is to be used for and what its not to be used for. If its not for leagues and clubs, lets say so and publicize the info. Why the secrecy on this stuff?

That's some history and my opinions on the issues. As always, my interest is in improving the game, improving USAPA, and getting max info to all players. I welcome all positive responses, designed for the same goals, whether you agree or disagree with my ideas.

"Let's disagree with each other most enthusiastically . . . But lets always stand together, no matter the differences" — Abhijit Naskar. I dont think he was talking about pickleball, but he could have been — we are all bound together by a love of pickleball, lets not forget that."