

Elo: The New Point System For USTSA

By: **Mury Johnson**
USTSA Points Specialist

As most of you are aware the point system that USTSA uses to rate and rank foosball players has changed starting this year. Hopefully by this time a good portion of you understand the basics behind an Elo rating system, either from previous experience, from reading threads on the EZ Board, or talking with fellow foosers. I could write a short book on the history, all the ins and outs, and Elo's application for our own foosball points system, which I don't have the time to do, but what follows is a somewhat lengthy summary of what foosers need and want to know.

Elo was first used to rate chess players and was invented by a physics professor and master chess player named Arpad Elo, hence the name. Chess still uses his points system to this day, as well as many other sports and Elo has even made its way into the online gaming world.

Why are we switching to Elo for foosball? Historically foosball players have been ranked in what can best be described as an upward march with the most points being awarded to those who play the most often, the longest, and attend tournaments in which those points are turned in. Without the aid of computers the old point system was not a bad solution, but it had its limitations in being able to dynamically and accurately rate players based on actual performance with a strong bias toward the most recent results.

At the highest level the old ratings reflect and reward those who attend the most tournaments. Therefore it was possible that the best player(s) in the world would not be rated in the top five. For seeding purposes that was a problem.

At the lowest ranking, there were rookies (now called amateurs) that are levels in skill better than their competition. It can be quite discouraging for a new player attending their first tournament to have to face someone who has been playing foosball 15 years and can run circles around them. Elo goes a long way to fix those problems as well as many others.

Elo gives and takes points to and away from the players based on the probability of winning or losing based on past performance. It is not a true indicator of talent and skill, but it tries to approximate that as closely as possible based on previous wins and losses. Chess players praise Elo for its accuracy in predicting an outcome of a match. There is a higher degree of luck and "good bounces" in foosball, but from our research and back testing we believe that the ratings reflected by Elo will be seen as an accurate reflection of how each member of the foosball player base stacks up against the rest of the field.

Please note that Elo is a rating system, and not a ranking system. Elo starts at the point where a match result is recorded and ends at the point where the calculations of all the match results available have been done and a point list is generated. There are no ranks as far as Elo is concerned. Rankings come into play when the powers that be draw somewhat arbitrary lines somewhere in the points list. In recent history one of those lines has been drawn at 1200 points to distinguish the difference between an amateur and a semi-pro. Even though those lines are subjective they serve a very useful purpose in providing different level of events to increase the interest and competitiveness of the whole player base.

Elo is more dynamic than the old points system and it will also lead to a different points distribution for the player base as a whole. Elo will provide a relative comparison for every player based against the whole. For example, if the player base is comprised of a lot of less skilled players and a lot of extremely talented players with not so many medium skilled players, you will see a result of lots of players grouping on the two ends. On the flip side if the player base is comprised of mostly good players with a few extremely talented players and a few beginners you will see most of the players clumped together in the middle. This is a good thing. It might seem weird to see a bunch of players all ranked around the same point level, but if that happens it will be an accurate reflection of what the player base looks like and will provide more accurate input for formulating a ranking system that provides for the best competition throughout the field of players.

I bring this up because I want you to be aware that those lines where we separate different ranks will most likely change over time. As an example, it may be desirable to have X number of pro-masters. Let's say it is decided that there should be 100 active pro-masters. If that's the case, it would be unlikely that the 100th player will have exactly 2400 points, which is where the line is currently drawn. Another alternative would be to say X percentage of the player base should be pro-masters.

Please note that while the point level of where those distinctions are drawn may change it will be done with a lot of care to encourage the highest level of competition, fun, and excitement for players of all level. In addition, any changes that need to be made will be done with how we all historically picture the skills of the existing rankings.

The Elo formula will move a player toward their correct rating over time. It is important to note that the further a player is away from their correct rating the faster they will move toward it. As they zero in on their correct rating the point movements will become smaller and smaller.

This is done by awarding a variable of points based on the probability of success. If a 1400 point player plays another 1400 point player the winner will gain the base amount of points and the loser will lose the base amount of points. Let's assume that the base amount of points is 16. After the match the winner will have 1416 and the loser will have 1384. Please note the points in this example are strictly examples. There are many other factors that determine how many points are awarded and taken away, which I will get to later.

Now Let's say a player with 2400 points plays someone with 1400 points. We are going to assume that the 2400 point player is better and is expected to win. Let's assume that historically that the odds of a 2400 point player beating a 1400 point player is 7:1. Let's also assume that both of the players are currently rated correctly. That means that if those players play eight matches the 2400 point player should win seven matches and the sum of those points gained will equal closely what that player will lose when they lose the one match. If the 2400 player wins, in this example they would gain 4 points and the 1400 point player would lose 4 points. However, if there is an upset, we assume that either the 1400 point player is underrated or the 2400 point player is over-rated, so Elo makes an adjustment for that. So, as an example, the 2400 player would lose 28 points and the 1400 point player would gain 28 points. As I mentioned before if they were both rated correctly the outcome of eight matches will look like:

7 matches times 4 points equals a 28 point swing = 1 match times 28 points equals a 28 point swing in the other direction.

So after 8 matches both players would end up approximately where they started.

Arguments are made that upsets will happen and there are certain matchups between certain players that don't reflect the norm. This is true. There will be some bouncing around, which is unavoidable, but in the long run over dozens or hundreds of matches players will trend toward their correct rating.

Some people say, "Well what if I have a real bad tournament? I'm going to lose a lot of points." It is true that if someone has a real bad tournament they will end up rated a little lower following that tournament, but since they are now unnaturally rated lower the next time they play they will be in a position to gain more points at a faster pace. Remember, the farther you are from where you should be rated the faster you will end up back where you belong after subsequent matches.

As you may have noticed I've been talking about matches. Unlike the old system where points are awarded based on where you place in an event, Elo uses every match for data. This is very useful. Let's take a very good amateur in open singles. Even though that amateur might not finish in the top 16, they may have had some very tough matchups early on that they won. Maybe they had to beat 4 semi-pros before facing some pro-masters. Elo will adjust their points upward for all those early upsets even though they didn't finish in the money.

Likewise, one pro-master's tour to the finals may have been much more difficult than another pro-master's path. It is possible that the loser of the finals could end up with more points gained throughout the event than the actual winner.

The goal of Elo is to provide an accurate view of player's probability of winning versus the rest of the whole. So in the example above, even though the apparent better player did not win the finals going forward it will show the predictive probability that they will do better than the other player in the future. It's important to remember that Elo is not attempting to take the place of a title. A title is still a title, and should be what everyone is striving for and giving credit for. Elo provides an accurate predictor for who might win an event going forward with all things being equal. Of course, a foosball event has many variables such as matchups, how every player is performing that day, and a degree of luck.

What else is new with Elo? For the first time as far as I'm aware there will be separate singles and doubles ratings. A lot of players feel that they are either a better singles player or a doubles player. Singles and doubles are completely different games, and a person's mix of skills greatly affects their success in one versus the other. Think about that rock solid goalie that everyone wants tending the nets for them. Some of those goalies don't have the fine tuned five bar to have the same sort of success in singles. Think about that dangerous singles player that is all energy with incredibly fast reflexes that can't put a complete game together in doubles.

As I said going forward players will have a doubles and a singles rating. Because of that it will be possible that some players will be ranked one thing in doubles and another in singles. For example, there will be people that have a semi-pro doubles ranking and an amateur singles ranking. It is extremely important to note that it will be up to the individual promoter on how they deal with that. Some promoters for scheduling purposes may say that you can only play in your highest ranking and above, instead of letting you play down in your lower ranking for that type of event. As a step in the right direction however, at least the data will be available for a promoter to be more flexible.

Elo will also fix the disparity in points between men and women. This is not a knock on the talent of women. We all know that women can be incredible foosball players. The problem lies in the fact that historically women were gaining points at a faster rate than men due to their extra events. As a result many women were rated past their skill level and had a tough time competing in their rankings. Certainly not all women had that problem, but it wasn't uncommon either, as evidenced by many women asking to be able to play down a rank.

What events will count? First of all, I can say with complete certainty that any event where you pick your own partner (DYP) or any funsy event (rollerball) will not count toward your rating. We are only going to use events where we know that players are usually going to play their best and offer a true reflection of their probability of winning in the future.

Is it possible for players to sandbag? I'm not even sure I want to bring it up, but I've been asked by a few people, so let me answer it by saying the following. Yes, it is possible for player's to try to sandbag. However, not only are we asking players and promoters to notify us of forfeits or other attempts to lose points, but we have written triggers into the software to catch weird patterns. In addition, players purposefully trying to lose points will not automatically move down a rank, which brings up the following.

What happens when a player loses enough points to move down a rank? Let me state that our primary objective is to maintain an accurate rating system. How that rating system affects the ranking system can be done in many different ways. The current plan is laid out so that you must fall below a certain threshold and remain there for a period of time before USTSA will consider lowering your ranking. For example, if you gain enough points to hit 1800 you will become a pro, but you will need to fall below 1600 points and remain there for an extended period of time before USTSA will consider adjusting your ranking.

We also hope to implement a field that will record a player's highest attained rating. The usefulness of that field is both for prestige and for potentially maintaining a player's ranking at their highest rating attained.

Switching to the Elo system is going to take patience, work, and a lot of change. The Elo system is sound, but its implementation for foosball is going to be a bit of an ongoing process. Please be patient with us as we go through that process and make adjustments along the way.

In these last few paragraphs I'm going to explain a few of the foosball specific adjustments to the Elo formula that we have implemented.

Another problem that is being addressed is what I'll call the coattail points. While its common for players to seek out players of their same caliber in doubles, there are many times that a very strong forward is teamed up with a goalie of weaker ability. Consider one of the very top players in the world playing with a 1000 point amateur against two 1700 point semi-pros. Both teams will have approximately 3400 points. Who do you think will win? Chances are that the top player as a forward can control the game enough in the middle of the table to come out a winner more than 50% of the time. Because of that we are weighting the points in doubles. The higher rated player's points will count more toward the team rating than the lower rated player. In addition, when points are awarded or taken away the higher rated player has more to gain or lose. How many more points are awarded to the higher rated player will be based on the point differential between himself/herself and their partner. For example, when a 1300 point player plays with a 1500 point player they will gain/lose very close to the same amount of points. When a 2400 point player plays with a 1600 point player the spread of the points awarded or taken away will be a little bit higher. its nothing dramatic. It will not be like the 2400 point player gains 30 points and the 1600 point player gains 5. It will be closer to 21 and 14 respectively.

Weighting of matches will also be done. For example, the outcome of a 3 out of 5 will be much more accurate than the outcome of a race to 7. When points are being calculated a 3/5 match will count as 100%. A 2/3 match will be weighted somewhere around 75%. That means that if you would have gained 12 points for a victory if it was 3/5 you will instead gain 8 points if it was a 2/3. A race to 7 will be weighted around 50%.

While the weighting will not be extreme for majors vs regionals we are going to factor that in as well. It will be based on bracket size and the amount of money at stake. The reason for this is different players treat different tournaments differently. There are a lot of pro-masters that will prepare very hard for a major, but might use a regional as a warmup event and might not even play with their normal partner. On the flip side a regional is a pretty big deal for new players and players that don't tour very much. As I stated the weighting is not going to be extreme, but there will be a slight discount on points exchanged at smaller tournaments.

I hope that this provides you with a good overall picture of how Elo works and how things will change moving forward. Elo has been a tremendous rating tool used in other sports, events, and even in other foosball communities. We've ran tournament results back from 2000 forward using Elo to test the formula and the results look very promising. Changing rating systems is going to take some work and time for everything to settle in. Please be patient through those changes. It will be worth it!

Mury Johnson