



# ELO MECHANICS EXPLAINED & FAQ

By Mario Kovacevic, RankR

## How accurate is the Elo Ranking System?

This is a bit of a philosophical question. In order for us to have a sense of how accurate Elo is, we need to have some sort of PERFECT set of rankings to compare Elo to. If we did have that PERFECT RANKING SYSTEM, we could very easily say that Elo was correct for 87% of the teams, or that Elo displayed a VERY GOOD correlation to our PERFECT ranking system. The problem is that a PERFECT ranking system does not exist – and probably never will.

Each person interpreting the accuracy of the current Elo rankings will have their own **subjective and biased** opinion on the strength of each U SPORTS team – no matter how objective or unbiased they try to be. It would take a super computer to factor in every bounce of the ball, pass of the puck, goalie save, Libero dig and wide-receiver reception to truly account for the complexities involved in comparing two varsity teams and then to sort out those two teams relative to the other 37 teams in their league. This super computer does not exist, so the best that I think that we can be left with is an approximation – or rather an estimate of the probability that any one opponent will defeat any another opponent in the league. This is what Elo does.

The Elo system will not be perfect – there will always be upsets in the every league that only our fictitious super computer will be able to account for. In its most basic sense, Elo works using Bayesian probability (bear with me!) to take into account all of Team 1's previous games (and their margin of victories as well as the strength of their opponents), the same for Team 2 and come up with a probability of Team 1 beating Team 2: this is the PREDICTED RESULT before the game even happens.

That's basically what Elo points are: a way to quantify this pre-game probability. A large difference in Elo points between two teams means that the higher ranked team (the team with more Elo points) has a HIGH probably to win against a lower ranked opponent – there is even a formula to exactly calculate this probability exactly (for example, a difference of 200 points means that the team with 1700 points has a 68.98% chance of beating the team with 1500 points). When the game between Team 1 and Team 2 is finished, the Elo system takes the ACTUAL RESULT that occurred, compares it to the PREDICTED RESULT generated before the game, and then adjusts the probabilities: giving the victor a slightly improved chance to win if the teams met again, while the loser's probability of winning in a rematch would drop slightly. That is, the victor gains Elo points, while the loser loses Elo points.





# ELO MECHANICS EXPLAINED & FAQ

By repeating the calculation described above for every single game that has occurred in a given U SPORTS-sanctioned sport for thousands of games over many seasons, the algorithm effectively makes a probability distribution for all the teams. With that probability distribution, we can look at the number of Elo points that any team has, and make a **reasonable estimate** of the probability that any one team will defeat any other team if they played. By sorting out these probabilities, we have our rankings. The teams with the Top 10 chances of winning make it to the Tuesday Top 10.

If a 45th ranked Toronto team just squeaks into the play-offs, makes a miracle play-off run and wins the National Championship – Elo will not predict this AT THE END OF THE REGULAR SEASON. What Elo will do, is as the play-off run is occurring, Toronto's odds of winning against the next opponent will increase higher and higher with each successive victory. By the time Toronto reaches the national final, they would have beaten so many higher ranked opponents, and their probability would have been adjusted upwards so many times that they would now become one of the "higher-ranked opponents". If they lose in that national final, Elo would simply shift their probability of winning the next game slightly downward – they would once again lose Elo points.

So how accurate is Elo? Not sure. But it does provide a reasonable (and always-evolving) estimate of the relative strengths of teams within a contained league or system. The probability adjustments (AKA changes in Elo points) that are made after each game are built on three, hard to argue against, general principles which are described below. Looking at any individual game (Toronto beats York 2-1) and the change in Elo points (Toronto gained 26 Elo points, York lost 26 Elo points) that occur after the game, it is also hard to argue with that adjustment if you keep the general principles in mind:

- Winning against a better and better opponent should gain you more and more Elo points than winning against a lesser opponent.
- Winning by a larger margin of victory should gain you more Elo points than winning by a smaller margin of victory or tying the game.
- Winning more important games, like at the national championships, are more valuable than winning less important games, like regular season games.





# ELO MECHANICS EXPLAINED & FAQ

Looking at the overall rankings Elo produces for your sport, you will notice that the **overall landscape of teams** makes sense – the teams who had better seasons are generally near the top, and teams with poor runs of games this past year have drifted towards the bottom. Even with these top teams, they are more often than not sorted in such a way that most people would agree: yes three-time national champs York should be first, and perennial-contenders but play-off underperformers Toronto are lower down at 10th position. That's fine and good. But are there going to be star-studded teams that most of the U SPORTS coaches and media would have in their Top 10 but are number 11 in Elo's? Yes. If Elo could speak, she would say that that star-studded team simply did not win by large enough margins of victory, against good enough opponents especially in important games to deserve a Top 10 ranking compared to the other 10 teams who did. Will Elo ever be able to tell you why Guelph is ranked 12th, but Acadia is 11th, despite Guelph just beating Western by 40, and last week Acadia losing to Western by 10? No it won't, only a perfect ranking system running on supercomputer will do that (or the fans/coaches/players from Guelph, Acadia and Western).

## What are the general principles of the Elo Ranking System?

For each sport at the start of the 2012 season, each team was assigned 1500 Elo points. Each game is NET ZERO game. This means that no Elo points are created. The amount of Elo points that the victor gains, is exactly equal to the Elo points that the opponent loses. In the case of a tie game, the lower-ranked team gains the Elo points (albeit less than if they would have won!) while the higher-ranked team loses that exact amount. If you take an average of the Elo points of every university in the entirety of U SPORTS at any point in time, the result will always be 1500. Why? Because teams are simply stealing points from each other – no points are being “created” or “lost”. Over time, the best teams would have stolen the most points for themselves while the lower-ranked teams would have lost the most points. Within this kind of environment, the Elo points are won or lost based on three basic principles:

- Winning against a better opponent should gain you more Elo points than winning against a lesser ranked opponent.
- Winning by a larger margin of victory should gain you more Elo points than winning by a smaller margin of victory or tying the game.
- Winning more important games, like at the national championships, are more valuable than winning less important games, like regular season games.

## Does the Elo Rating System account for difference in team strength?

Yes, a very highly-ranked team will gain minimal points for even blowing out a low-ranked opponent. Conversely, a lowly ranked opponent make a massive jump in points by upsetting a much higher ranked opponent. In the case of a tie, the lower-ranked team would gain points, while the higher-ranked team would lose points.





# ELO MECHANICS EXPLAINED & FAQ

## Does the Elo Rating System account for difference in score line?

Yes, a blowout win is much more valuable than a marginal win – but this is with diminishing returns. In soccer for example, if a 1-0 win yields 20 Elo points, a 2-0 win would yield 30 Elo points, while a 3-0 win would yield 35 Elo points. A 6-0 win might yield 42.5 Elo points while a 7-0 win would yield Elo 43 points. You can see that each successive goal gains you less and less additional Elo points than the previous goal.

This basic principle is applied to other sports while adjusting the underlying formula to account for the differences in the average score line. In basketball unlike soccer, a 1-point win is effectively the same as a 3-point win, where as in soccer, a 3-goal margin of victory is basically a blowout. The basketball formula is adjusted so that a 15 point margin of victory is needed in a basketball game, to yield you the same amount of points as a 3-goal margin of victory in soccer. For volleyball, the basic Elo formula is applied, assuming that you can only win by 3 sets, by 2 sets or by 1 set – this works just fine.





# ELO MECHANICS EXPLAINED & FAQ

## **Does the Elo Rating System account for the level of competition?**

Yes. In Volleyball for example, a 3-2 win at the National Championship will gain you more Elo points than a 3-2 win in the playoffs, which will gain you more Elo points than a 3-2 win in the regular season.

## **Does the Elo Rating System account for exhibition games?**

Yes, it does. In fact, the more data Elo sees, the more accurate the rankings become. Cross-conference games are especially valuable to Elo. Any preseason, mid-season or post-season “exhibition” games can easily be factored into the Elo system. The inclusion of exhibition games in the calculation would vary depending on the sport and gender. While men’s soccer might not value exhibition games, women’s hockey might.

## **Does the Elo Rating System account for Home/Away results?**

It can. Like exhibition games, it depends on the preferences of the individual sport. In men’s soccer where regular season crowds rarely exceed 100 people, home advantage might not be as important compared to women’s volleyball where a packed gym can make all the difference. Inclusion of Home/Away results would vary from sport to sport depending on the expertise of the Top 10 committees.

## **How does Elo account for the difference in parity between conferences? Will a top team in a very poor conference be ranked higher than a mid-table team in the strongest conference?**

Before reading the example consider the basic principle which says that in any single game, the amount of points that the victorious team GAINS is always exactly equal to the amount of points that their opponent LOSES. Also, that the rankings for any sport currently include scores from at least September 2012.





# ELO MECHANICS EXPLAINED & FAQ

Say every team in U SPORTS was assigned a score of 1500 at the very start of that season. This means that in **Sept 2012** the average score in the OUA (let's assume this is our strongest conference) was 1500, and the average score in the RSEQ (our weakest conference) was also 1500. You are right to say that at the end of that very first season in 2012, a top team in RSEQ, would be ranked higher than a mid-table OUA team (who is probably better than the RSEQ team). Over the next 5-6 seasons, as teams from the best conference (the OUA) constantly start beating on teams in the RSEQ in cross-conference play (nationals, exhibitions, tournaments), the OUA will start hoarding all of the Elo points. If we were to take the average points across the OUA in 2018, the average across all OUA teams would no longer be 1500, but higher than 1500. In the RSEQ, which has lost points to the OUA over the last 5 seasons, the AVERAGE Elo points across all RSEQ teams would be less than 1500. This does a very good job of accounting for disparity in conferences: "a tie against a top team in a very tough OUA league becomes more valuable than a win in a weak team in the RSEQ league".

## Won't Elo simply encourage running up of the score line?

The below example is for basketball, but the logic applies for all sports.

The Elo formula does account for goal differential but with this is with diminishing returns. If I am the coach of Windsor and we are ranked No.1, and playing against a team that we know we are going to demolish because they are ranked 40. Let's say its my alma mater Toronto. Because there is such a massive difference in Windsor and Toronto's Elo ratings, if we were to beat Toronto by 10 we might increase by 1.5 Elo points. A 20-point win would get us about 1.75 Elo points while a 40 point win would give us 2.0 Elo points. Clearly, there is no incentive to beat on a team who you are favored to beat when even a close 2-point win against an evenly ranked would give you somewhere in the area of 10 Elo points.

Compare this to a Windsor win against an evenly-ranked opponent like McMaster, a 10-point win would give you 13 Elo points. A 20-point win would get you 15 Elo points. A 30-point win would get you 16 Elo points. A 40-point win would get you 16.5 Elo points. You can again see that you earn the bulk of your points in the first 10 points, and that beyond that you would be running up the score for very little in return.

## How does Elo deal with the transition between seasons? Do all the Elo points get carried over or do all the teams restart at 1500 points?

The answer lies somewhere in between. One extreme would start the national champion Laval team with the **exact** 1875 Elo points that they had at the end of last season. This of course does not make sense, since the Laval team that walks off the field/court/ice in March, is not that same team that walks back on next September. The other extreme is to restart everyone at 1500 points. Three things would happen in this case. First, we would lose all the information that Elo was





# ELO MECHANICS EXPLAINED & FAQ

tracking with regards to the relative parity of the conferences as described earlier in this FAQ. Secondly, it is fair to say that a Laval team that has made it to the national championship final the last five years will probably do well in the coming season. Assigning them an arbitrary and **average** score of 1500 at the beginning of the season would be taking a lot of valuable information from the Elo system which would otherwise “remember” that Laval has been a strong team the last 5 years. Thirdly, games at the beginning of the year, especially those against weaker opponents would be very lucrative for Elo points. Picture a perennially strong Laval team matched up against a perennially weak Toronto team at the very first game of the season. Subjectively, we would expect a Laval blowout after an average looking off-season. If we were to assign both teams 1500 points, Laval would gain a massive 40 Elo point boost after that game because they just blew out an “equally-ranked opponent”. Conversely, if the Elo points remained from the previous season, Laval might only gain a 5 Elo point boost, because the system “knows” that U of T is probably still going to be weak this year.

## **Some conferences play more REGULAR SEASON games than other conferences. How does this affect the rankings?**

Playing more regular season games DOES NOT give a team an advantage than a team who plays less regular season games. Remember, each game is NET ZERO. Meaning that the team that wins gains as many Elo points as their opponent loses. Winning teams are effectively stealing points from losing teams at the end of each game. After a tie, the lower-ranked team would steal points from the higher ranked team. Each game is just as much of an opportunity to gain Elo points as it is an opportunity to lose Elo points.

Therefore, playing more games will certainly make the rankings within that conference more “accurate,” since the algorithm has had more samples of information to factor in. However, this does not mean that playing more games is **more advantageous** – in fact, it is exactly equally advantageous as it is disadvantageous.

