Handelshøyskolen BI

GRA 19703 Master Thesis

Thesis Master of Science 100% - W

Predefinert informasjon

Startdato:	09-01-2023 09:00 CET	**Termin:**	202310
Sluttdato:	03-07-2023 12:00 CEST	**Vurderingsform:**	Norsk 6-trinns skala (A-F)
Eksamensform:	T		
Flowkode:	202310||11184||IN00||W||T		
Intern sensor:	(Anonymisert)		

Deltaker

Navn:	Noman Bashir

Informasjon fra deltaker

Tittel *:	Efficacy & Fairness in eSports Tournament Design
Navn på veileder *:	Jonas Moss

Inneholder besvarelsen konfidensielt materiale?:	Nei	**Kan besvarelsen offentliggjøres?:**	Ja

Gruppe

Gruppenavn:	(Anonymisert)
Gruppenummer:	58
Andre medlemmer i gruppen:	Deltakeren har innlevert i en enkeltmannsgruppe

FINAL THESIS REPORT

EFFICACY & FAIRNESS IN E-SPORTS TOURNAMENT DESIGN

Author	Noman Bashir
Hand-in Date	03-July-2023
University	BI Norwegian Business School
Supervisor	Jonas Moss
Programme	Master in Business Analytics

Faculty Jonas Moss	Department Data Science and Analytics

Author
Noman Bashir

Title
EFFICACY & FAIRNESS IN E-SPORTS TOURNAMENT DESIGN

Subject: Business Analytics	Level: Masters
Month and year: June 2023	Number of pages: 33

Abstract

Tournaments are mechanisms used to rank alternatives, usually in sporting contests whereby contestants are pitted against in each other in matchups to determine the best contestant. Tournament design is a widely researched issue in economic and operations research literature where it is necessary to determine a design that best approximates the “true” or latent ranking of a contestant. Tournament design has been studied in many different fields of study but literature regarding tournament design when it comes to eSports is negligible. eSports employ unique tournament designs not considered in past studies on the matter and their intensive scheduling requires a slightly different skillset in terms of staying power to win. Not to mention these tournament designs often have multiple matchups in elimination instead of a single matchup. In this study, the eSport StarCraft II and its flagship event – the Global StarCraft II League is considered. Various formats including knockout tournaments, Swiss systems and bespoke eSports tournament designs are compared using a Monte Carlo approach. The study also endeavours to separate the effect of seeding from tournament design and whether seeding has a variable effect on different tournament designs. It is found that Round Robin formats present the best approach to determine the “true” ranking, followed by custom eSport tournament design utilized in the Global StarCraft II League. However, Round Robin has more than twice the number of matches compared to any other design. Triple Knockouts offer an alternative to the existing GSL format when it comes to approximating the True Rankings of the best players but are not the best at minimizing discrepancies in finishing positions for other players. These results can be interesting for emerging and existing eSports where tournament design is neither solidified, changes every year, and varies from tournament to tournament within each eSport.

Keywords Tournament Design, Efficacy and Fairness, Monte Carlo, Starcraft II, GSL

Depository BI Norwegian Business School

Additional information

Contents

Tables and Figures

I. Introduction

Finding the optimal design of sporting tournaments is an important question not only in terms of scientific research but also from an entertainment perspective. With increasing global appeal of sports and esports, making sure that tournament designs allow the top players to come to the fore but also that each game is fair is very important. While the organizers can do nothing about the results of individual match-ups, they can influence the final result of the tournament through the way the tournament is designed – the format (Round Robin, Knock-out, Swiss Style), whether players will be seeded or not and various other considerations. The ideal tournament makes sure that the best players win, the games all have stakes, they are fair and that the contestants maximize their effort within each game. There has been a lot of research on the ideal tournament design in traditional sports and board games such as chess. However, research on eSports has been limited. eSports are particularly interesting to look at from a tournament design standpoint because of the sheer novelty of different tournament designs espoused by the managers of games such as Starcraft II and Counter Strike: Global Offensive.

eSports has become a phenomenon in the past decade with the prize pool of tournaments such as The International for Dota 2 has reached a staggering $40 million. This dwarfs some traditional sports like cricket where the prize pool for the T20 World Cup was $5.6 million. Games such as Starcraft II, Defence of the Ancients 2 (Dota 2), and League of Legends (LoL) have had a number of competitions in the last decade. The organizers have experimented with many different tournament designs in order to make sure that the considerations of an ideal tournament are kept into account.

The importance of developing the ideal tournament design has long been recognized by researchers in economics and sports science, but also by the organizers themselves. The primary goal of effective tournament design is to make sure that the tournament is perceived as fair by the competitors and the spectators. There are other goals as well – such as making sure the best teams/players have the best chances to win (referred hereafter as efficacy), and that each participant maximizes their performance in each match-up during the tournament and the matchups preceding it. Lastly, for spectator sports, a good tournament design keeps and attracts interest. However, numerous contests are still plagued by inefficiencies.

As shown by Haugen and Krumer, even large, long-standing organizations such as FIFA create tournaments in a form where the rules can easily be abused and the standards of efficacy, fairness and effort are not maintained. This problem also exists in eSports tournaments. This research proposes to identify the best tournament design for eSports with respect to efficacy and effort. It will also the effect of seeding in eSports.

The main aim of this research will be to determine which form of tournament is most likely to produce a winner that is considered the “best” amongst all the entrants. However, this research does not take into account other considerations such as a certain length of time in which a tournament should be completed i.e., scheduling in terms of overall tournament length does not feature in this research.

StarCraft 2 is a science fiction real-time strategy video game developed by Blizzard. Essentially, a player takes control as the military commander of various species and endeavours to overwhelm their opponent’s home base. The main tournament being considered in this study is the Global Starcraft II League where the best players for each year come together to determine the yearly champion. The League currently employs a mixture of dual elimination and knockout style tournaments in multiple stages.

II. Literature Review

There is a surprising lack of published research on tournament design in sports journals. Indeed, Haugen and Krumer (2019) mention that since 2014, only two papers on tournament design were published in the top 3 sports management journals. This is interesting because there have been several prominent cases of tournament design working against the intentions of the organizer and how the sport is played. The two most notable cases are “The Disgrace of Gijon” where West Germany and Austria ostensibly colluded in a game of football to settling for a 1-0 score-line to the detriment of the a third team – Algeria. More amusingly, there is another case in the 1994 Carribean Cup Qualification where both teams would benefit from scoring an own-goal in regular time because of a rule that meant that every goal in extra-time counted for two. Barbados scored an own goal to force extra-time and then did their best to prevent Grenada from scoring an own goal. This event was a travesty of a game of football where the intent is to score goals against the other team, not yourself. These events are not restricted to just football, Taylor and Trogdon (2002) examined whether there are incentives to lose in NBA and discovered that teams that have no chance of reaching the play-offs tend to undertake “tanking” where they lose games in order to get better draft picks the next year. In this case, it is tournament design that incentivizes poor performance and lack of effort after a certain period of time. Thus, there is a distinct need to consider the effects of tournament design on the efficacy and efficiency of a tournament in selecting the best winner from amongst the contestants.

However, if we examine the existing research on tournament design, we discover that it is primarily split into two distinct areas:

1. Literature on tournament design which looks at different types of tournaments and their performance with respect to efficacy and getting the “best” winner
2. Economic literature on tournament design which focuses on incentives to lose in specific tournaments / tournament designs

Within the space for assessing tournament design, a review shows that past studies are primarily focused on understanding and modelling sports tournament only. The seminal work in this area has been done by Appleton in 1995 where he analyzed several different formats in football and concluded that the Round Robin format has the highest chance of being efficacious. Meanwhile, several other studies have determined the exact probability pij of player i finishing in place j for a small number of players (David, 1959; Glenn, 1960; Searls, 1963)

Meanwhile, Scarf and Yosuf (2009) have analyzed the different types of tournament formats currently in use in various competitions in the world and how seeding introduces uncertainty into whether one format or the other is more successful in drawing out the best winner. Csato (2021) has analyzed four different hybrid approaches to determine the best tournament design for the IFC Handball Championship.

Similarly, there are a number of other studies about effective tournament design in sporting contests analyzing the situation from several angles – efficacy, fairness, effort maximization and reducing match fixing. However, the landscape for tournament design in eSports is rather sparser. Sziklai et al. (2022) is the most recent study where they discuss that eSport tournaments have yet not solidified into a single coherent design and are still in the experimentation phase. The only other research of eSports tournament design focuses more on tournament theory perspective rather than the design of the tournaments themselves (Coates and Parshakov 2016).

In research where focus is on incentives to lose, Krumer, Megidish, and Sela (2020) analyze why round-robins are not the best format from an effort-maximization standpoint. Depending on the results of the first round, players can have an incentive to lose the next round to maximize their pay-offs. Curiously, this runs contrary to Appleton's work where Round Robin tournaments were considered the best from an efficacy standpoint. The implication of this difference of results is that different tournament designs perform better in different sports. Similarly, Dagaev and Sonin (2018) showcase that tournament systems that rely on knock-outs and round robins are incentive incompatible. The example they use is of the Russian Premier League where a team had an incentive to lose their final game instead of winning in order to have a better chance of qualifying for a European Cup.

Kendall and Lenten (2017) examine rules in various sports such as football, tennis and basketball where the desire to make business-efficient decisions (sports teams are operated as businesses in the modern age) tend to result in unwanted results such as teams attempting to draw or lose deliberately. While these occurrences are rare and limited, they still bring sporting integrity into question. Vong (2017) goes a step beyond and recommends that in tournaments which employ multiple stages following a round-robin group stage, only the top most qualifier should be allowed to qualify the next stage as anything else results in incentive incompatibility and teams/players do not exert effort in certain matches even when effort exterion is costless.

Despite the abundance of literature within these two areas, there is a noticeable gap in research about tournament design in eSports. This is even more relevant as eSports not only grow exponentially year-on-year in terms of prize money, participants and viewership but also because they employ unique formats that have not been seen in traditional sports. The division of the tournament into winner and loser brackets following the first match-ups has been hitherto unexplored in research on tournament design.

Bibliographic coupling is one of the widely used scientific mapping techniques applied to research in order to identify which areas of research are similar and which studies have had a great impact on that particular sphere of study. Bibliographic coupling works on the principle that if two published research studies cite the same sources, they must pertain to the same subject matter. A shared citation becomes a link between two studies. Studies that are cited more often are considered seminal and have larger nodes and thus considered as important to that field of research. Bibliographic coupling also allows the determination of the structure of the field of research i.e., are there any major groupings where research is conducted.

This analysis is replicated for this review. The data was gathered from Web of Science portal where studies under tournament design were gathered from the past 20 years. This data was then clustered using Rs bibliomatrix library.

When clustering, the documents tend to broadly divide into two main clusters – the blue one is operations research while the red one is economic literature. Economic literature usually deals with agents, moral hazard and tournament design as a general concept to be applied in business. Operations research on the other hand focuses more on the subject matter to be considered in this study – for example, it focuses on determining mathematical models, algorithms, and decision-making frameworks that address challenges related to tournament design. These challenges can range from scheduling problems, seeding, ranking, fairness and efficacy among other things. Csato (2021) outlines how operations research can help tournament design.

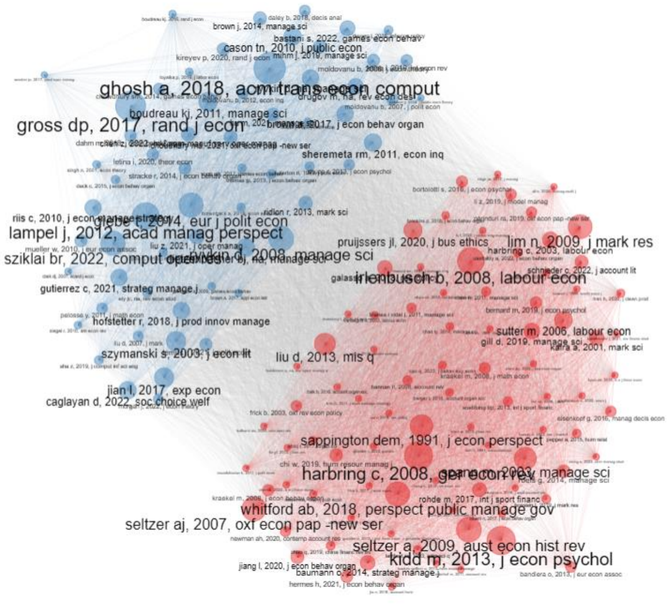

Figure 1 - Clustering of Research Studies Conducted on Tournament Design

Fig 2 examines the kinds of topics that tend to occur together. It can be noted that tournaments, naturally, are the central point in any sort of keyword analysis. However, it should be noted that economic literature dominates this space as can be gleaned from the fact that "incentives" is a frequently occurring keyword in titles alongside prizes and professions. Management and productivity is also something that is considered as many corporate projects are considered as tournaments (such as the bidding process for projects to be undertaken).

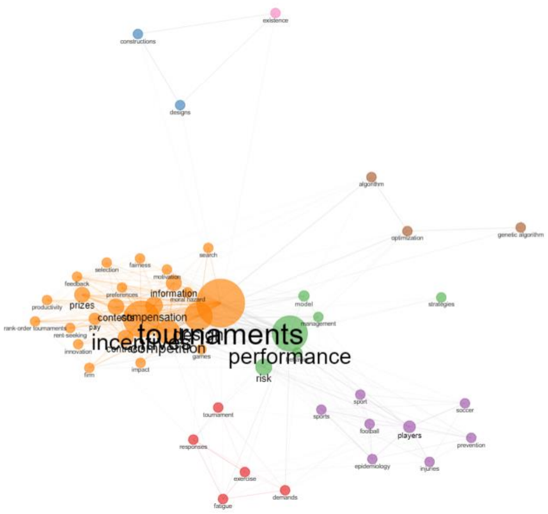

Figure 2- Clustering of Keywords Used in Tournament Design

It can be noted that "designs" occupies a smaller subspace on the outskirts of the main areas where research is conducted. More importantly, eSports is non-existent when doing a title analysis. However, this is not cause for concern as eSports and sports employ very similar structures when it comes to determining outcomes of tournaments and therefore, one's results should be transferrable to the other.

III. Research Methodology

The primary research method for this study will modelling and simulation of different tournaments using the Monte Carlo method. The eSport to be discussed will be StarCraft 2 and its annual competition called Global Starcraft2 League. The reason this eSport was chosen was because of the wealth and availability of data. Tournaments since 2012, their competitors, matchups and rankings have been preserved for StarCraft 2.

The Monte Carlo method is utilized as it is the industry standard for simulating processes that involve uncertain variables. The underlying principle relies on repeated random sampling in order to obtain numerical results. The idea is that by repeatedly sampling from a distribution, the expected value of the final calculation should approximate the "true" value. This approach is extremely useful in circumstances like evaluating the results of a match between two teams and players (and hence tournament design) where it is otherwise impossible or incredibly difficult to gauge results. One could, theoretically, use Machine Learning to train a predictive model but there is an underlying randomness to any matchup, where the better player can simply lose just by chance. The hope is that by using a Monte Carlo simulation and approximating the result of a match (or tournament) in this case, the "true" value can be reasonably ascertained.

3.1 Tournament Formats

It is almost impossible to derive the full-ranking of a tournament because in most tournaments, once a participant is knocked out, there is no way to actually determine where they would finish in the overall standings. Given that this study wishes to approximate "real" world, there will be no additional matches taken between the players. The players that were knocked out at any given point are given the highest ranking they could possibly achieve in that knockout stage. For example, the players knocked out in the quarter-final stage will all be given the ranking "9". There is significant discourse on the best way to handle these scenarios of incomplete information. Sziklai et al. (2022) utilizes complete tournaments i.e. players continue playing matches even when knocked out to determine their overall final ranking. On the other hand, McGarry and Schutz (1997) simply assign players into a final ranking based on when they were knocked out and how many players they had to beat to get to that final stage. It is this latter approach that will be used in this study.

One of the other difficulties in gauging efficacy of tournament designs is the difficulty of evaluating ties or using tie-breaker mechanisms. Luckily, in eSports like StarCraft 2, there can be no ties therefore this concern is one this study does not address. Ties when they occur in final standings in Round Robin or group stages of multi-stage tournaments are resolved by considering the pre-tournament ranking of players. This is a valid assumption to make as the GSL is the culmination of the Starcraft II season and players are promoted/demoted to the GSL based on their ranking. Therefore, their success in the GSL itself is also resolved by their prior ranking.

The GSL always has 32 teams that reach the final tournament through various qualifiers or by their performance in the previous iteration of the competition. The following tournament designs are implemented in this study:

- **GSL Format**: The format in use since the inception of the GSL. It is a multi-stage format consisting of 2 group stages (32 teams divided into 8 groups, followed by 16 teams divided into 4 groups). The matches are all best of 3. In the first group stage, the 4 players of each group are split into two pairs and play each other. The winners face each other in the winner's match where the victor places first and advances to the next round. The losers of the initial match play each other in an eliminator, the loser being eliminated from the tournament and the winner playing the loser of the winner's match. The victor of this final match places second and advances to the next stage while the loser is eliminated. After the first two group stages, the knockouts are structured like regular knockout competitions. However, the matches in the quarterfinals and semifinals are best of 5 while the final is a best of 7.
- **Multi-Stage Tournament with 8 Groups:** This is the tournament format used in most football competitions including the UEFA Champions League and the FIFA World Cup. 32 teams are split into 8 groups with all teams in each group playing each other once. Each win grants 3 points and losses grant 0 points. The top 2 teams in each group advance. The next stages are a regular knockout format.
- **Round-Robin:** Each team plays each other once. The team's finishing order depends on the number of points they accumulated.
- **Swiss Tournament**: This tournament style has several different versions. For the purposes of this study, the Dutch system (which is the most widely used) will be considered. It is a non-eliminating format with a fixed number of rounds where the 32 teams or players are paired randomly (or according to seeding). In each successive round, teams with the same points are paired together. Each team cannot play each other more than once. When several teams have a similar number of points, tiebreakers are resolved by using Opponent Match Win (OMW) Percentage – a metric defined as the percentage of wins all of a team's opponents have had. In the Dutch system, in each round, teams are split up into groups with similar points and OMW

percentages. The winner is the player having the most wins at the end of the tournament.

- **Simple Knockout**: In each round, players are paired to play a match. The loser is eliminated, while the winner proceeds to the next round. The process is repeated until a sole winner remains, which requires n rounds for 2 n players.
- **Triple Knockout**: Same as simple knockout but the teams in each stage play each other 3 times.
- **CSGO Swiss**: This tournament style mirrors a regular Swiss tournament with some minor differences. Teams are paired together in the first round either randomly or by seeding. The winners go into one bracket, the losers into another. In each subsequent round, the teams in each bracket are randomly paired and teams with similar records are bracketed together. In essence, a team needs 3 wins to go to the next round while 3 losses will eliminate them from the tournament. There are six rounds in total, but in practice, a number of teams will be eliminated and progress by the 3rd round. After this initial “group stage”, the rest of the tournament is like a regular knockout tournament. This tournament design is best explained in the image below. Note that CSGO Swiss tournaments are for 16 teams and will be adapted for 32 teams in this study.

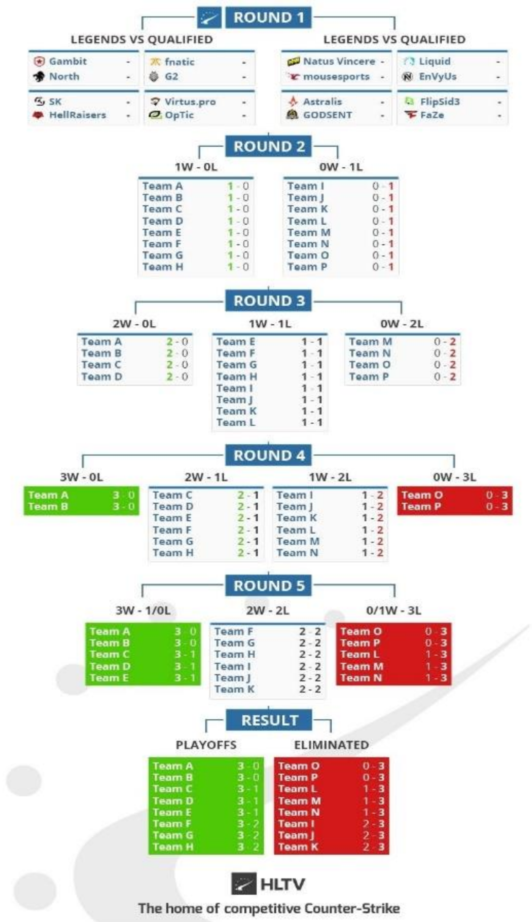

Figure 3 - Counter Strike: Global Offensive Swiss Style Tournament Design

- **Double Elimination**: In contrast to a single elimination tournament where a team is knocked out if it loses one game, a double elimination requires teams to lose twice in order to be eliminated. The most common method of running double elimination tournaments is to have two brackets – a winners bracket and a losers bracket after the first round. The winners proceed to the winners bracket while the losers go to the losers bracket. The winners bracket is conducting in the same manner as the first round, except the losers drop down into the lower bracket. This is best explained in the figure below:

Figure 4– Bracket of a Double Elimination Tournament for 24 Teams

3.2 Seeding

The effects of seeding need to disentangled from the overall tournament design. Therefore, in this study, the baseline is random seeding i.e. all matchups will be randomly selected. However, since the study endeavours to answer the question whether seeding improves tournament efficacy, the simulations will be run once again but with seeding (or in the case of knockouts, brackets which make sure that highest ranked players are on opposite ends of the bracket). It is assumed that the “true” seeding of the player is known beforehand. For eSports such as StarCraft 2, this is not a wild assumption to make as there is a wealth of data to be found and pre-tournament seeding can be determined with ease. In short, the seeding will remain the same in all simulations for a particular tournament player-set.

The effect of seeding in group stages and tournaments has been extensively studied in the past. Most notably, for knockout tournaments, Hwang (1982) proposed the impact of seeding that is used in many bracket managers today. It is this bracketing mechanism that is utilized in this study.

Standard seeding is utilized in all tournaments. For tournaments with group stages, the Top 8 teams are placed into different groups. For example, Seed 1 goes into group 1, seed 2 goes into group 2 and so on. This iteration is repeated for the next

8 teams so seed 9 goes into group 1, seed 10 into group 2 and so on. It is important to note that for knockouts that follow a group stage, there is no seeding done. In fact, the only restriction for knockouts in multi-stage tournaments are that players cannot play the same team in the knockouts that they played in the group stage.

For Swiss style tournaments and the double elimination tournament, the first round of matchups requires seeding. In this case, the top ranked team plays the bottom ranked team. For subsequent rounds, the previous rounds results and OMW% are considered.

For tournaments with brackets, the following bracketing mechanism is used:

Figure 5 – Bracket When Seeding Is Used in Knockouts

3.3 Match Prediction and Simulation

The fixed winning probability $\rho_{AB} = 1 - \rho_{BA}$ determines the likelihood that Player A wins against Player B. There are several different mechanisms utilized to predict the result of a matchup. For example, Appleton (1995) and Raghavachari (2000) use normally distributed ratings. Meanwhile, Scarf et al. (2009) use historic data from the UEFA Champions League to build a predictive model. It has to be noted that these predictive models have only been used for traditional sports as eSports are a relatively newer addition to the space gaining prominence in the past decade. Therefore, these predictive models may not easily apply to StarCraft 2.

However, due to the digital nature of eSports, there is significantly more data available to base models on. In order to avoid reinventing the wheel, this study recreates the predictive model employed by Aligulac (Karpov, 2011) to determine the result of a single matchup. In essence, the predictive model leans heavily on the rating system. The rating system is self-created but relies heavily on the Glicko rating system with a couple of minor changes. The Glicko system itself is the estimation of a player's true strength via a Bayersian approach. The system accumulates previous information and continuously updates the parameters based on match results (much like the ELO system). The model is used most frequently in chess where, in a particular period, a player has a rating and a "volatility" which determines how much the performance of a player can fluctuate (Glickman, 2012).

The rating system used by Karpov has a few additional features: each player has an overall rating and a rating for each matchup. The overall rating is always the mean of the matchup ratings. No approximations for maximising the likelihood function are used, numerical optimisation algorithms are utilized directly. The usefulness of the Glicko system (now updated to Glicko-2) has been assessed several times compared with other rating systems like ELO, most notably in Vecek et al. (2014) where its efficacy is measured and compared with ELO.

In short, using the Aligulac rating for each player, and its associated variance, the probability that Player A defeats Player B depends on the difference of their ratings. This rating difference is inputted into a logistic cumulative distribution function with mean zero, variance 1 and scaled by a factor of 1000).

This model is highly efficient in terms of determining match outcomes as shown in Fig 1.6. The x-axis shows the predicted win-rate of the stronger player and the y-axis shows the actual win-rate. Up to win-rates of 80%, the model is able to predict with a high degree of accuracy. However, win-rates of over 80% should be taken with a grain of salt.

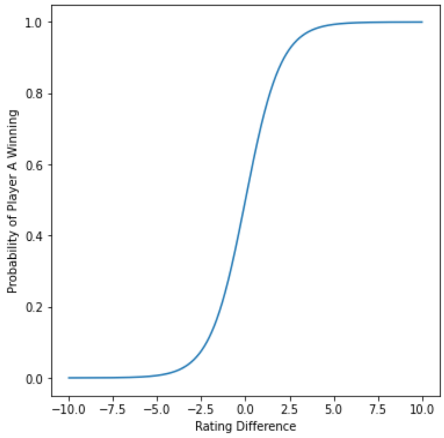

Figure 6 - Probability Function for Win Percentage

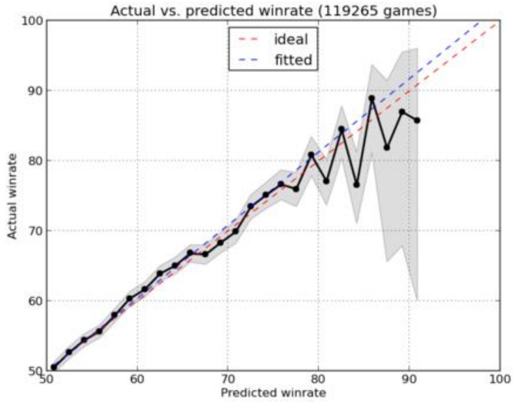

Figure 7- Prediction Results

Just using the rating itself would quite possibly mean there are no upsets. However, there are variances associated with each player's rating in Aligulac's database. For the purposes of this study, the player's rating was sampled from a normal distribution with mean 0 and SD 1. The player's rating would then be finalized as their initial rating +/- variance.

3.4 Tournament Metrics

Usually, efficacy is quantified by looking at the differences between real and observed rankings. The "real" rankings would be the pre-tournament rankings based on the rating system whereas the observed rankings would be the outcomes after the tournament. There has been significant research on which tournament metrics are the most suitable to determine efficacy. One of the most common methods is the average of where the Top *k* players finish. Scarf et al (2009) applied this metric when assessing the efficacy of tournaments in football, particularly the FIFA World Cup. Another common metric is number of inversions (i.e. the number of players who finished above their real ranking).

Meanwhile, Sziklai et al. (2022) applied a unique weighted inversion metric arguing that the spectators are more concerned about where the top players finish rather than where the 32nd seed finishes so number of inversions would not be the most accurate metric. Can (2014) also applies this technique. In short, the reciprocals of the logarithms are summed up from the real ranking to the observed ranking. For example, if a player with a rank 5 finishes in 2nd place, the weighted inversion (for this player) would be $\frac{1}{\ln 5} + \frac{1}{\ln 4} + \frac{1}{\ln 3}$. The weighted inversions for all players who precede their ranking are added and that is the final weighted inversion metric for a tournament.

Table 1 shows an illustrative example of how these measures are calculated:

Table 1
Examples of Calculations of Metrics

Pre-Tournament Ranking	Scenario A	Scenario B	Scenario C
1	2	1	3
2	1	2	5
3	3	5	1
4	4	3	4
5	6	4	2
6	5	6	6

The cells highlighted in grey show players finishing ahead of their positions and are relevant for the aforementioned metrics. In Scenario A, the number of players who finished ahead of their pre-ranking positions are 2 (2^{nd} rank finishes 1^{st} and 6^{th} rank finishes 5^{th}). Hence, the number of inversions are 2. Similarly, in scenario C, two players finish ahead of their rankings and the number of inversions is 2. Whereas the number of inversions in Scenario B is only 1 (only Rank 5 finishes ahead of their rank).

The weighted average calculation is slightly more complication. For scenario 1, Rank 2 finishes in 1^{st} position, so only jumps from 2 to 1. Hence, their weighted inversion score will be 1/ln2. Similarly, Rank 6 finishes in 5^{th} position, so their inversion will be 1/ln 6. In total:

Scenario A: $\omega_A = \frac{1}{ln2} + \frac{1}{ln6}$

In Scenario B, only rank 5 finishes ahead of their "true" position but by 2 positions (5^{th} and 4^{th}). Hence, for this scenario:

Scenario B: $\omega_B = \frac{1}{ln5} + \frac{1}{ln4}$

In Scenario C, multiple players finish ahead of their pre-tournament rank. We can calculate the weighted inversion score of each and add them together. For player 3 it's ln/3 + ln/2 and for player 5 it's 1/ln5 + 1/ln4 +1/ln3. Hence, in totality, the weighted inversion score will be:

Scenario C: $\omega_c = \frac{1}{ln3} + \frac{1}{ln3} + \frac{1}{ln5} + \frac{1}{ln4} + \frac{1}{ln3}$

Lastly, based on previous research, one can imagine that the Round Robin format is the most efficacious in terms of mirroring pre-tournament rankings as it exhaustively pairs all players together at least once. Indeed, Appleton (1995) points towards this conclusion. However, it has to be noted that the Round Robin format requires the most matches by a significantly large margin. Therefore, total number of matches is another important tournament metric to consider. The ideal tournament balances the number of matches with other efficacy metrics.

This study considers the following efficacy metrics when evaluating tournaments:

1. Weighted Inversions
2. Number of Inversions
3. Average Finishing Position of Top 1 Player
4. Average Finishing Position of Top 8 Players

Regarding the average finishing position of the Top 8, it has to be noted that this will vary between tournaments. Round Robin formats provide a conclusive finishing position for each player whereas tournaments like knockouts do not. Based on the system of final classification being used, the lowest average finishing position (or a perfect finishing order) will have an average of 2.8.

All of these metrics will be compared with the number of matches. The expectation, as shown by Appleton (1995) is that the more the matches, the lower these metrics (A lower number in all of these metrics is preferable to a higher number).

3.5 Data Collection

There were two major steps in collection of data:

1. Extracting Tournament Data (players)
2. Extracting Ratings and Deviations

For extracting tournament data, a web scraper was built in Python that parses all GSL tournaments on liquipedia.net and provides the players. The second step is to use these players and ping the Aligulac API to gather information about the players' ratings and deviations for their time period. This dataset is then used in custom-designed tournaments in Python using a Monte Carlo approach to determine efficacious outcomes. All GSL tournaments from 2012 – 2019 were considered.

IV. Results

One thousand simulations were done for each tournament and the results aggregated. While most research papers conduct over 100,000 simulations in order to test the hypothesis, it was discovered that this was not necessary in this study's simulations. In a preliminary analysis, the running mean of weighted inversions was looked at for several tournaments. The average stabilizes around 400-500 iterations and remains constant thereafter. This is shown in the figure below; the simulation is for the tournament style that CSGO tournaments utilize i.e. a modified version of the Swiss Tournament.

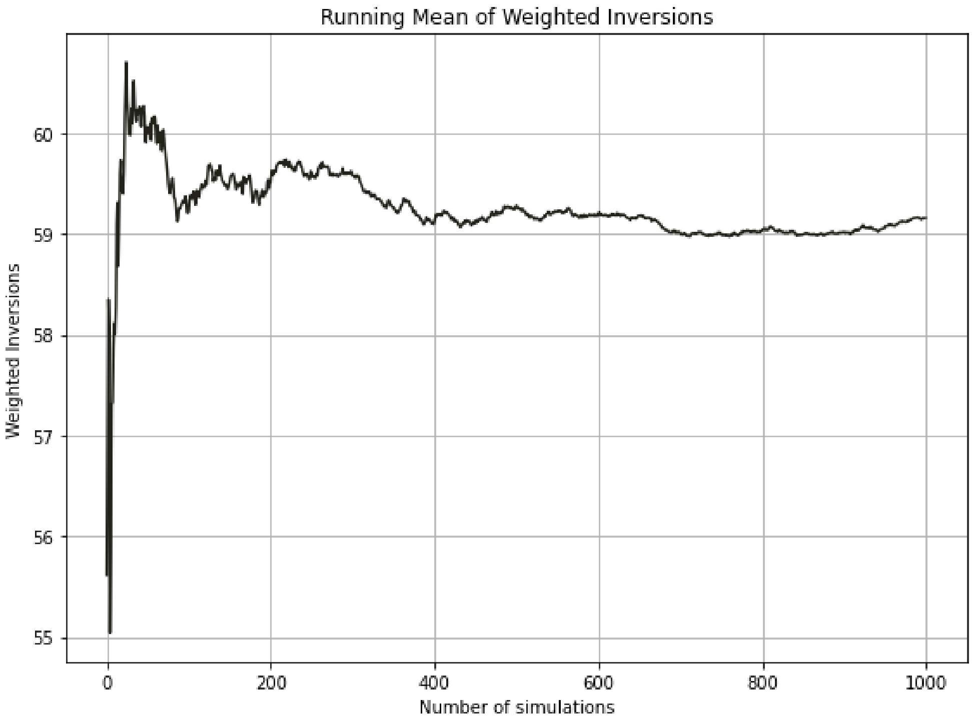

Figure 8 - Average of Metric by Number of Simulations

The table below summarizes the total number of games played:

Table 2
The number of matches in a tournament with 32 players

Tournament	Number of Matches
GSL	217
Multi-Stage 8 Groups	63
Round Robin	496
Swiss Style Tournament	96
CSGO Style Tournament	71
Simple Knockout	31
Triple Knockout	93
Double Elimination	62

Predictably, Round Robin has the highest number of matches whereas the Simple Knockout has the lowest number. The GSL ranks second highest in the number of matches because of its unique tournament structure where there are three matches played in each matchup during the group stages and 5 and 7 in the knockouts and finals.

4.1 Seeded and Unseeded Results

The four major metrics for each of the tournaments conducted without seeding are below:

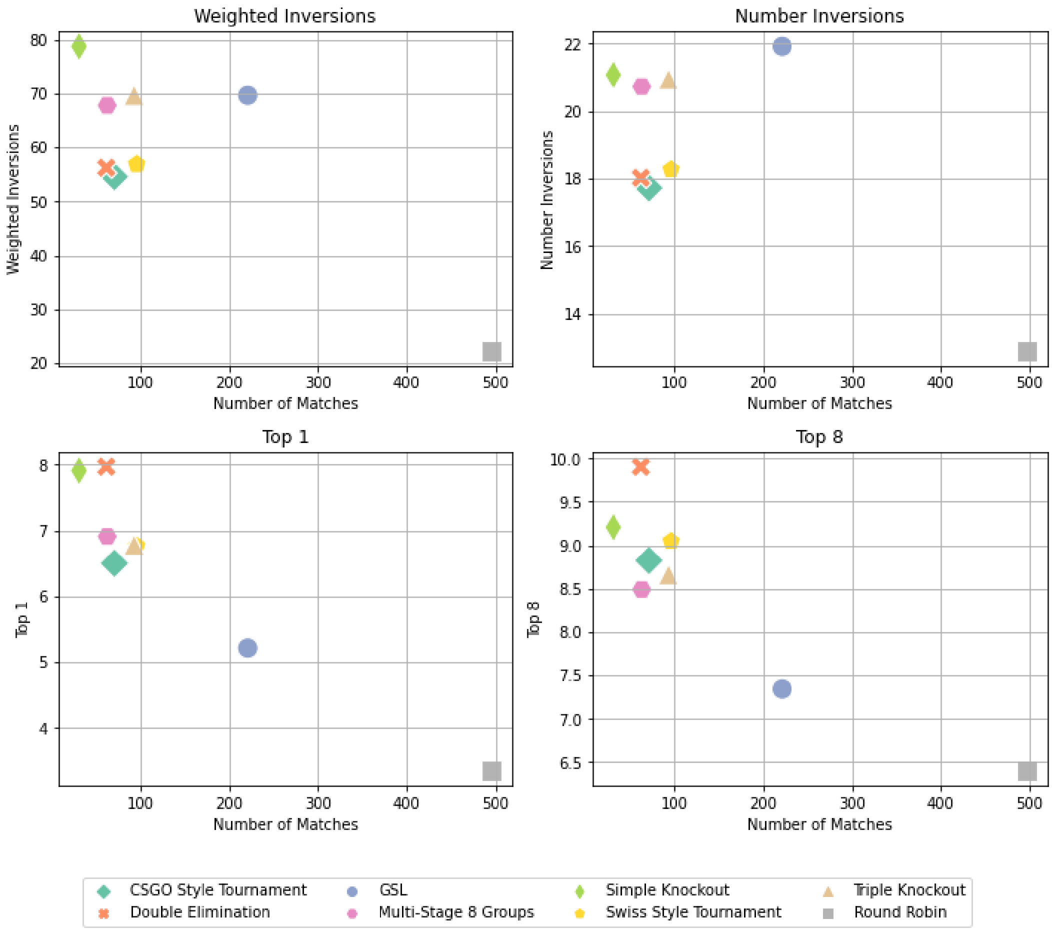

Figure 9- Results for Top Metrics Compared to Number of Matches (Unseeded)

The first thing that stands out is that Round Robin is indeed the best tournament design if the desire is to have the best player win the tournament. Out of the 8 tournaments measured, Round Robin has the lowest average finishing position of both the top player and the top 8 players. However, this comes at a cost as it has more than double the amount of matches than any other. It also scores extremely well on weighted inversions, and number of inversions as predicted. However, given that the GSL and other eSports tournaments are conducted in a small timeframe, it would simply not be feasible to conduct that many matches. The argument could be made that the GSL tournament design itself has over 200 matches, but most of those matches are between the same set of players i.e. in quarterfinals, each matchup has 5 individual matches each but between the same two quarterfinalists. With Round Robins, the scheduling would become incredibly more complex as each player has to play every other player.

The GSL format itself is extremely efficient in terms of sorting through players. After Round Robin, it is the best in terms of Top *K* players classification. However, the GSL format does poorly on weighted inversions and number of inversions. This is understandable as it employs a knockout style format in the later rounds and dual elimination in the earlier rounds. Compared to the Swiss Style tournaments or Round Robin where players are never knocked out and have a chance to recover later in the tournament, players defeated twice in a GSL format are knocked out. An unfortunate pairing would result in the number of inversions being larger.

It is also readily apparent that multi-stage and knockouts are the worst formats when it comes to minimizing inversions. Simple knockouts especially result in a large number of players finishing out of order. This is easily explainable as without seeding, the players end up in positions where strong opponents face each other early on.

The only surprise is Double Elimination tournaments. While they are indeed great at minimizing inversions and making sure most players finish at their pre-rank positions, they do not help in selecting the best winner with their efficacy being comparable to Simple Knockouts when it comes to placement of the Top 1 and Top 8 players. Since the double elimination tournament is a bracket tournament, it is understandable that it suffers from some of the setbacks that simple knockouts also suffer from. However, the result is counter-intuitive in the sense that double eliminations were designed precisely to better select the winner. This result is matched by in Sziklai et al. (2022) where draw and process performed poorly in terms of top players placement.

The results for seeded tournaments are below:

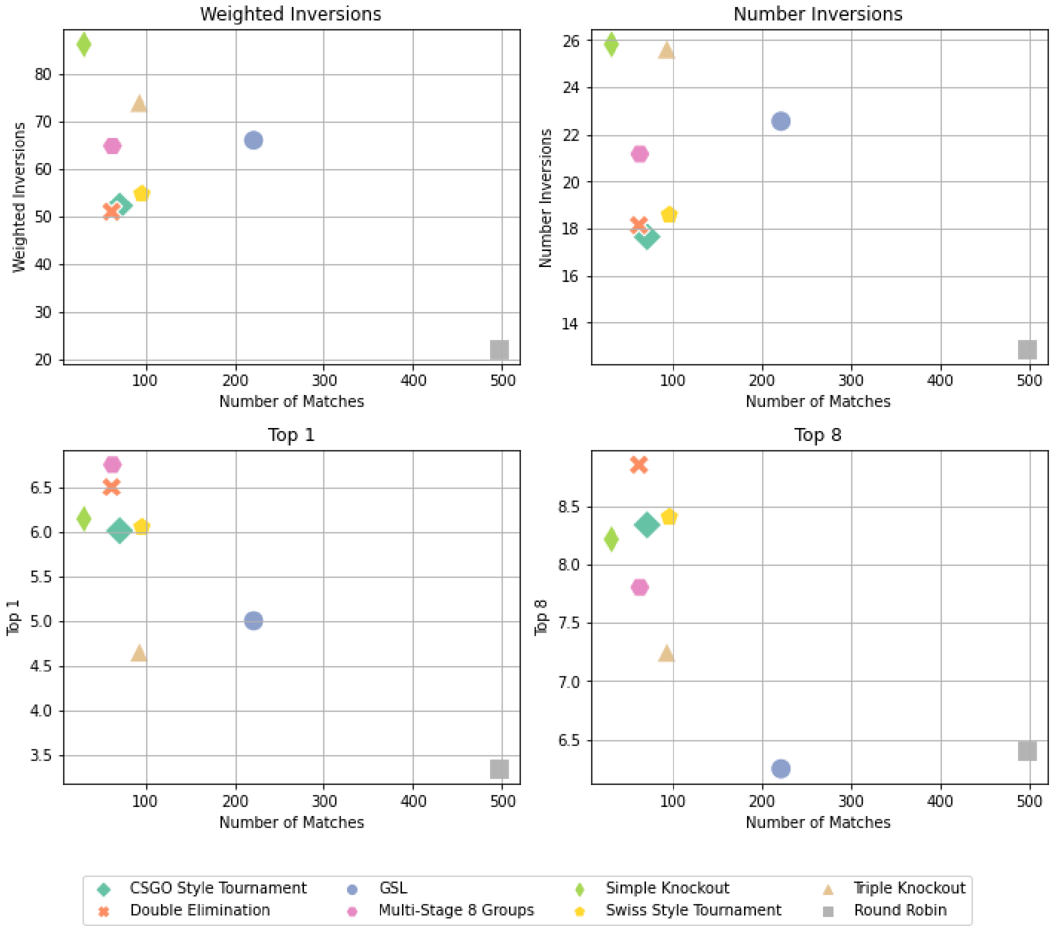

Figure 10- Results for Top Metrics Compared to Number of Matches (Seeded)

The results do not differ too much from the unseeded brackets and groups. Once again, the Round Robin format is the best performing format in most metrics. However, it seems that by seeding, both the GSL and the Triple Knockouts benefit greatly. Indeed, the GSL format outperforms Round Robin when it comes to ranking of the Top 8. However, the biggest surprise is Triple Knockouts which has a drastic performance improvement in terms of placing the Top 8. These results are not altogether unsurprising for the knockout formats. The bracketing in these formats when the tournament is seeded intends for the best players to reach the furthest in the competition. It seems that multiple matchups eliminate the possibility of higher ranked players getting knocked out earlier when playing in a seeded bracket and that is why Triple Knockouts seem highly efficacious.

However, they perform poorly once again when it comes to inversions. The brackets are designed in a way that makes similarly ranked players meet up in the first round of matches. For example, Seed 15 and 16, 17 and 18, 19 and 14 all play each other in the first round. The exits of any of these 3 will add to the number of inversions and the weighted inversion. Comparatively, in unseeded tournaments, it is possible that lower seeds face each other in the first round more frequently. For example, 31 could play 32 and either's exit does not significantly add to the weighted inversion. Therefore, due to the design of the bracket itself, the results are understandable.

The rest of the results mirror the unseeded simulations with double elimination and Swiss style tournaments resulting in similar performance in terms of inversions.

4.2 Impact of Seeding

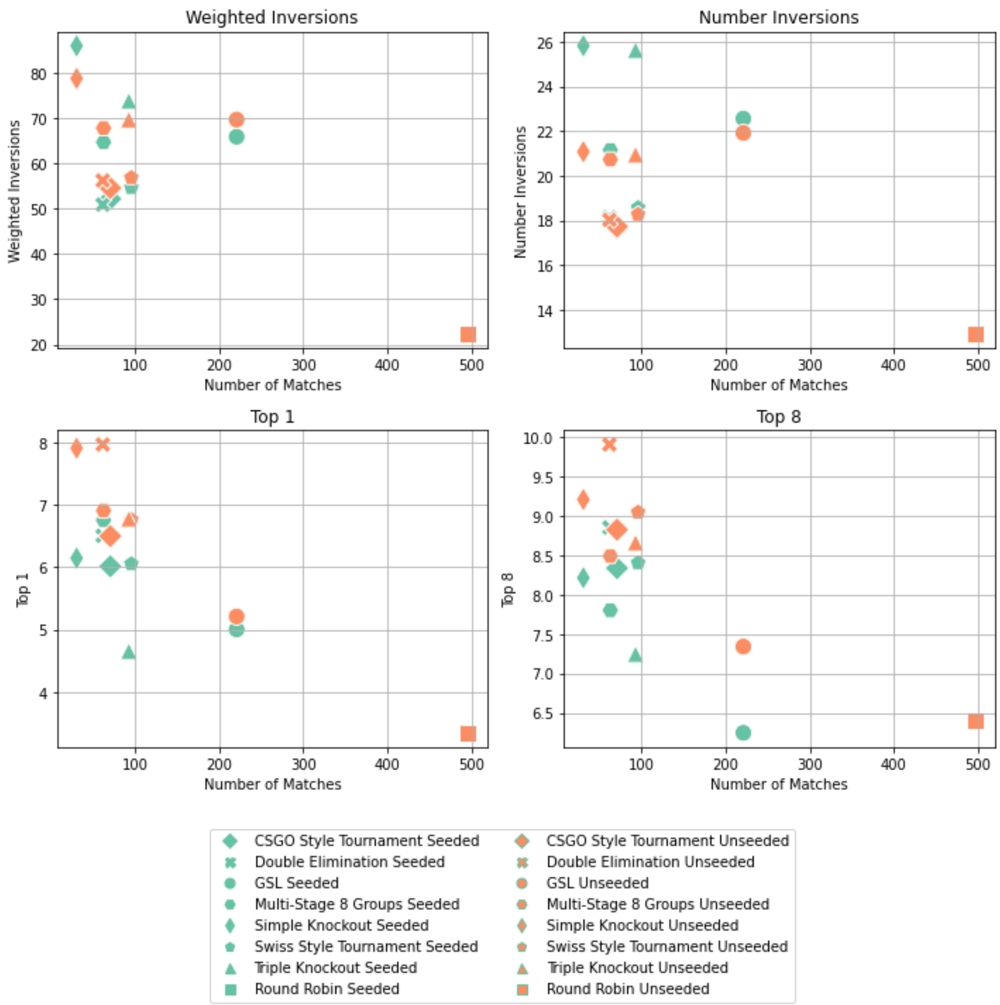

Figure 11- Comparison between Seeded and Unseeded Tournaments

Figure 11 compares the results across tournaments for both seeded and unseeded simulations. It is clear that seeding has a large impact for the finishing positions of Top 8 and Top 1 while there is negligible (or even negative impact) in terms of Weighted Inversions and Number of Inversions. These results are consistent but the magnitude is larger in some tournaments. The knockout formats gain the greatest benefit from seeding whereas the Swiss style tournaments do not. This is understandable as the Swiss style tournament is designed to determine the best players from incomplete information. After the first round, the seeding is discarded in any case and players with similar results are matched up. For example, players who won their first matches are paired and players who lost their first matches are paired. This automatically introduces a seeding-like effect into the tournament and thus these styles of tournaments benefit the least from seeding. Tournaments where randomness reigns supreme like the knockouts (and the GSL to a degree) benefit the greatest from separating stronger players in the earlier rounds.

Therefore, it is clear that the advantage of the Swiss system is when the competition is more unbalanced and there is no significant chance of knowing the true “seeds” of the players.

In terms of number of inversions, there is negligible impact. Seeding is designed to make sure the best players reach the furthest in the tournament and therefore the impact on number of inversions being minimal is unsurprising. There is a slight improvement in weighted inversions seen across all tournaments.

4.3 Distribution of Metrics

The metrics that were calculated earlier were based on an expected value. **Fig 12** shows that the distribution of the weighted inversion and Top 8. Since the distribution can be approximated by the normal curve, averages are a reasonable approximation for the efficacy of the tournaments.

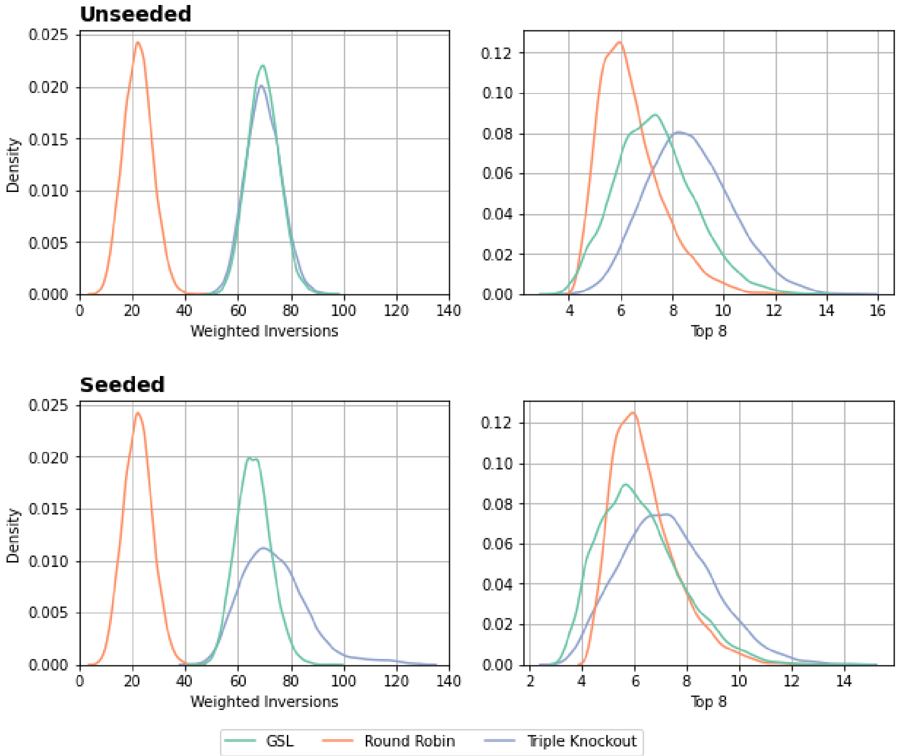

Figure 12 – Distribution for Tournament Results

Moreover, the distribution of tournaments can also help discuss how well certain designs can predict the "True" ranking. In terms of unseeded, Round Robin has a significant advantage but this does not hold true as strongly when it comes to seeded tournaments. In fact, in some cases, the GSL tournament design was able to outperform the Round Robin in terms of Top 8 classification (as we also saw earlier).

4.4 Impact of Ratings Variance

The simulations were conducted on not only one single year or tournament, but across tournaments spanning 8 years. Different players with different ratings took part in these tournaments and in some years, the contestant pool was of a similar quality. The following section assesses whether players being closely matched had any significant impact on the efficacy of different tournament types. To do so, the ratings of each contestant in each tournament were gathered, and then the standard deviation was calculated. The standard deviation was similar across all tournaments ranging from 0.48 to 0.7 at most. While this range is small, there were small impacts in terms of tournament efficacy.

When looked at holistically, there does not seem to be any major changes in terms of metrics i.e. a closely matched group vs a more uneven group had not impact on the efficacy of overall results. Nor could there be any clear distinguishing factor when it came to seeded vs unseeded groups.

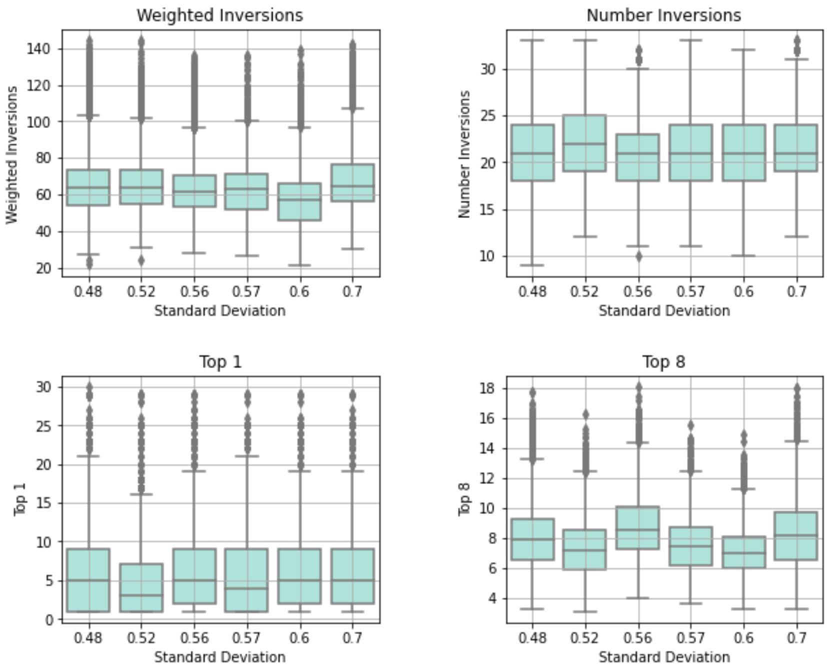

Figure 13- Tournament Metrics Based on Player Balance (Seeded)

This is the unseeded one:

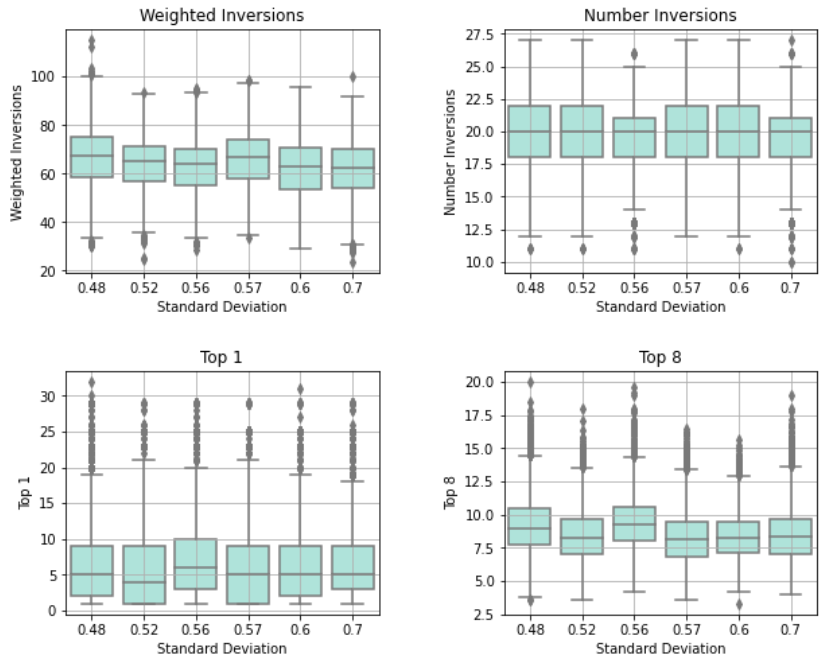

Figure 14- Tournament Metrics Based on Player Balance (Unseeded)

However, significant changes can be seen when it came to specific tournament formats. For example, **Fig 15** shows a boxplot of player standard deviation when compared with metrics for the GSL. It is clear that as the players are more spread out in terms of skill or ability, the GSL format is better able to reduce weighted inversions. However, number of inversions stays constant. This means that although the same amount of players are placed ahead of their positions, the higher ranked players do not place out of position more often i.e. they are better placed. Indeed, the distribution of Top 8 does confirm this hypothesis.

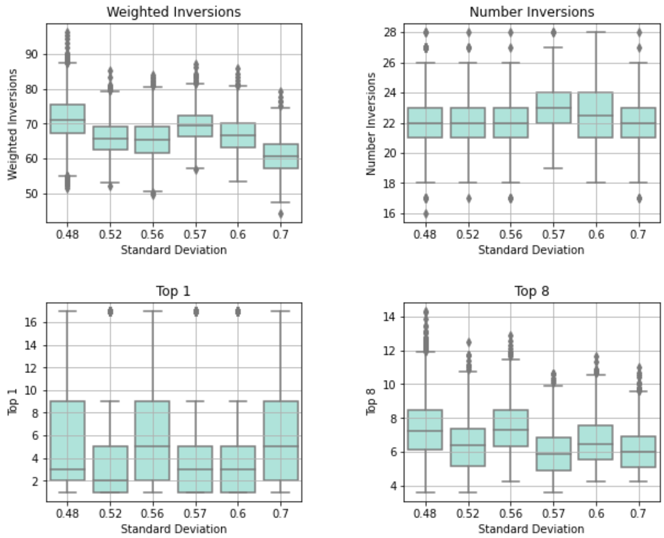

Figure 15 - Tournament Metrics Based on Player Balance (GSL)

The same impact can be seen in Round Robin formats:

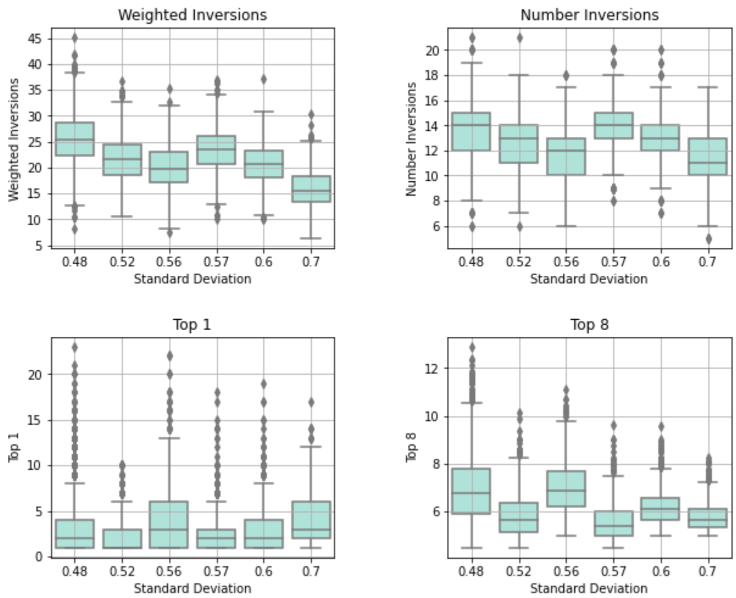

Figure 16 - Tournament Metrics Based on Player Balance (Round Robin)

However, this kind of impact is noticeably absent from all Swiss style tournaments. This is probably due to the fact that the Swiss system is adept at matching up players with similar ability after the first rounds. This result is corresponded by other research on the subject. Most notably, (Sziklai et al., 2022) state that

> “The corresponding Swiss-system is robustly preferred to any group-based tournament (multi-stage with 4 or 8 groups, double group). In the case of real data, the advantage of the Swiss-system is found to be higher if the competition is less balanced, that is, in tennis. The gain from the Swiss-system compared to a simpler design is the lowest in chess, which can be surprising because the Swiss-system is applied in this particular sport, showing the strength of traditions. The tournament measures of the Swiss-system converge to the corresponding measures of round-robin since the latter is equivalent to a Swiss-system where the number of rounds is the number of players minus one.”

Table 3 contrasts the results across all metrics for each one of the tournament designs. The results are compared for both seeded and unseeded tournaments. Some of the interesting observations are that Double Elimination style tournaments do not gain much from having a larger standard deviation when it comes to the contestants’ ratings while the GSL does.

Table 3

Results by Tournament Year and Seeded Status

	Seeded				Unseeded			
	WI	NI	Top 1	Top 8	WI	NI	Top 1	Top 8
2012 | SD = 7.2								
CSGO Swiss	51.8	17.5	5.0	8.1	59.2	18.1	5.8	8.8
Double Elimination	50.2	18.0	5.5	8.6	55.2	17.9	7.2	9.5
GSL	69.9	22.7	4.3	5.9	73.3	22.1	4.5	7.2
Multi-Stage	63.3	21.0	5.9	7.5	66.3	20.6	6.0	8.3
Simple Knockout	84.7	26.0	5.9	8.2	82.6	21.3	7.5	9.1
Swiss	53.8	18.4	5.2	8.2	61.9	18.6	6.0	9.0
Triple Knockout	72.0	25.7	4.1	6.9	73.6	21.2	6.1	8.6
Round Robin	26.5	14.3	2.6	6.1	26.5	14.3	2.6	6.1
2013 | SD = 7.8								
CSGO Swiss	54.1	18.2	5.2	7.7	54.3	17.5	5.4	8.5
Double Elimination	52.9	18.7	5.6	8.2	58.4	18.5	6.9	9.3
GSL	64.0	22.5	3.8	5.7	67.9	21.8	4.1	7.0
Multi-Stage	66.8	21.7	5.5	7.2	70.0	21.2	6.3	7.9
Simple Knockout	89.6	27.0	5.7	7.9	77.0	21.1	7.3	9.0
Swiss	56.8	19.0	5.3	7.9	55.5	17.9	5.6	8.8
Triple Knockout	76.9	26.9	4.0	6.8	67.7	20.8	5.9	8.3
Round Robin	21.8	12.8	2.2	5.8	21.8	12.8	2.2	5.8
2014 | SD = 9.0								

CSGO Swiss	44.8	16.9	6.4	7.4	53.2	17.8	6.1	8.4
Double Elimination	42.9	17.6	6.3	7.9	49.2	17.5	7.9	8.9
GSL	64.4	23.0	4.7	6.0	69.1	22.1	4.7	7.1
Multi-Stage	57.7	20.7	7.0	7.2	61.5	20.3	7.3	7.9
Simple Knockout	78.9	25.3	6.1	7.5	78.7	21.3	7.5	9.0
Swiss	47.7	18.1	6.4	7.6	55.9	18.1	6.6	8.7
Triple Knockout	65.6	25.2	4.7	6.6	69.2	21.2	6.4	8.4
Round Robin	20.7	12.9	3.2	6.2	20.7	12.9	3.2	6.2
2015 | SD = 8.6								
CSGO Swiss	50.7	17.7	5.7	8.0	56.8	18.2	5.6	8.2
Double Elimination	49.3	18.2	6.2	8.4	55.3	18.0	7.5	9.6
GSL	67.7	23.3	4.3	5.3	71.4	22.4	4.2	6.6
Multi-Stage	64.4	21.3	6.6	7.7	67.6	20.9	6.7	8.3
Simple Knockout	84.1	25.8	6.1	8.0	81.2	21.5	7.0	8.8
Swiss	53.2	18.4	5.9	8.0	59.2	18.7	5.8	8.4
Triple Knockout	72.1	25.8	4.5	7.1	71.8	21.3	6.1	8.1
Round Robin	23.6	13.9	2.6	5.6	23.6	13.9	2.6	5.6
2016 | SD = 10.5								
CSGO Swiss	57.6	18.5	6.6	9.0	47.3	17.1	7.0	7.9
Double Elimination	56.4	18.9	7.4	9.5	61.7	18.6	8.7	10.7
GSL	58.4	21.9	5.5	5.6	63.2	21.3	5.9	6.7
Multi-Stage	70.4	21.9	7.3	8.2	73.3	21.3	7.3	9.0
Simple Knockout	92.9	26.5	6.8	8.7	73.0	20.8	8.2	8.7
Swiss	59.9	19.4	6.6	9.1	49.4	17.7	7.0	8.1
Triple Knockout	80.7	26.3	5.2	7.7	63.1	20.5	6.7	8.1
Round Robin	15.9	11.1	4.1	5.8	15.9	11.1	4.1	5.8
2017 | SD = 8.4								
CSGO Swiss	52.0	17.2	6.9	9.3	51.9	17.3	7.4	9.4
Double Elimination	51.1	17.7	7.4	9.9	55.6	17.7	9.1	10.9
GSL	63.8	22.2	5.8	7.0	67.4	21.7	6.3	7.8
Multi-Stage	64.0	20.5	7.9	8.6	67.1	20.1	7.7	9.3
Simple Knockout	84.4	24.4	6.5	8.9	76.3	20.7	8.5	9.6
Swiss	54.2	18.1	7.0	9.4	53.9	18.0	7.7	9.6
Triple Knockout	73.8	24.4	5.4	8.1	67.4	20.6	7.7	9.1
Round Robin	20.0	11.8	4.3	7.0	20.0	11.8	4.3	7.0
2018 | SD = 7.2								
CSGO Swiss	57.0	18.1	5.4	8.5	60.0	18.5	6.9	9.6
Double Elimination	56.0	18.5	6.5	9.0	60.4	18.3	7.2	9.9
GSL	73.1	23.1	5.4	6.9	76.0	22.4	5.8	8.0
Multi-Stage	69.3	21.4	6.6	7.8	71.9	21.1	6.6	8.4
Simple Knockout	91.8	26.6	5.8	8.4	84.2	21.5	8.6	9.7
Swiss	60.3	19.0	5.7	8.5	62.9	19.2	7.5	9.7
Triple Knockout	80.0	26.3	4.3	7.3	76.0	21.4	7.6	9.3
Round Robin	26.8	14.4	3.7	7.1	26.8	14.4	3.7	7.1
2019 | SD = 7.2								
CSGO Swiss	49.5	17.1	6.9	8.7	54.2	17.3	7.8	9.9
Double Elimination	48.6	17.6	7.2	9.3	54.2	17.7	9.2	10.4
GSL	66.1	21.8	6.3	7.5	69.2	21.4	6.3	8.3
Multi-Stage	62.0	20.8	7.2	8.2	65.0	20.4	7.4	8.8
Simple Knockout	81.9	24.8	6.3	8.3	77.5	20.5	8.7	9.8
Swiss	51.6	18.1	6.4	8.7	56.2	18.0	8.0	10.0
Triple Knockout	70.0	24.5	5.1	7.4	69.0	20.4	7.8	9.4
Round Robin	23.1	12.1	4.0	7.7	23.1	12.1	4.0	7.7

4.5 Comparison with Earlier Research and Summary

Lastly, it is worth comparing the results of this study to research conducted on this subject earlier. While there is scarcely any research on eSports in particular, the results of studies on tournament design should yield similar results. Appleton (1995) showcases that the Round Robin is the best format when it comes to determining where the best player finishes after the tournament. Other notables results are that the Swiss style tournament does not perform well in this regard. Sziklai et al. (2022) contradict this second statement where they find that the Swiss system is not dominated by the Draw and Process and is even better in certain scenarios. The results of this study back up Appleton's results as the Round Robin tournament does vastly outperform every other tournament design not only in terms of where the Top 8 and top player places but also in terms of weighted inversions and number of inversions. Only after taking into account the effect of seeding is Round Robin dislodged by the GSL format (something not considered by any other study) and that too only in Top 8 placement and only marginally. McGarry and Shutz (1997) note that the simple knockout format is the weakest format in terms of any sort of efficacy and that conclusion is reinforced by this study. The simple knockout format was one of the weakest performers in all metrics regardless of whether the tournament was seeded or not.

Contrary to Sziklai et al. (2022), this study finds that the Swiss style tournament is not a significant improvement over other tournament designs, most noticeably when the seeding closely approximates the true rank of the players prior to the tournament. However, that study compares Swiss tournaments with rounds between 5-10 whereas this study only compares one with 6 rounds in order to match up more closely with the CS:GO Swiss tournament design. Sziklai et al (2022) also discovered that increasing the number of matches while keeping the tournament design the same is not optimal. Therefore, going from a simple knockout to a triple knockout is not efficacious. However, this study finds that triple knockouts are extremely efficacious when it comes to ranking the top player in the correct place when seeding is done properly. There are questions as to whether perfect seeding improves the quality of the triple knockout as the best player is less likely to go out due to a fluke loss compared to all other styles of tournament where a fluke loss would put the best player on the back foot. Indeed, in single elimination tournaments, the player would be immediately knocked out. In double elimination, they would have to take part in the loser bracket and play many more matches with potential for another upset. In Swiss style tournaments, the best player would suddenly be pit against worse players but would be out of the running of the championship as a single loss can mean someone else wins the championship with a win-score of 6-0. It is only the GSL format and the triple knockout format where a single loss does not have disastrous results for the top-ranked player. Therefore, both GSL and triple knockouts perform well in this regard.

The results of this study can be summarized as follows:

1. The Round Robin system is more efficacious than any other formats. However, this efficacy is a trade-off with the high number of matches played.
2. The existing GSL format provides the best balance between high metrics and number of matches, especially when seeded.
3. Seeding improves the performance of knockout tournaments dramatically in terms of placement of the Top 8 and the topmost player. However, in terms of inversions, these tournaments are not the best design.
4. Draw and Process, Swiss and CSGO Swiss tournament designs are not the ideal tournaments for determining the top-ranked player regardless of seeding. However, these tournaments do tend to minimize inversions. Therefore, if you want lower ranked players to stay in the lower ranks, these tournaments work the best. Swiss style tournaments are not impacted by either seeding or the spread of player ability whereas tournaments like GSL and Round Robin perform better on established metrics if the difference in player ability is larger.
5. Triple knockouts provide a useful alternative to the GSL format if seeding is done accurately.
6. New tournament designs introduced by eSports (such as the GSL and CSGO Swiss) that have not been considered in prior research prove to be highly efficacious on all metrics.

V. Conclusion

It is useful to remind ourselves that the purpose of a tournament is manifold. Some of the considerations which tournament hosts must take into account are:

1. Making sure the best players win
2. Schedule the entire tournament with minimum resources used (matches played)
3. Make sure the matches are exciting for spectators to witness.

These goals can be considered as trade-offs. This study (and previous research) has shown that the Round Robin format is the best format when it comes to minimizing player inversions, making sure the best players finish the highest. However, it comes with several drawbacks: it has more than double the number of matches required and a lot of the matches later on in the tournament are dead rubbers where the lowest ranked team has nothing to play for. This problem can be witnessed in the English Premier League where teams in mid-table have neither relegation nor European qualification to compete for and spend the last few months of the season in limbo.

Conversely, simple knockout tournaments utilize the smallest number of matches and every match has high stakes, however they result in an inefficient approximation of “true” rank. The GSL combats this by adopting a double elimination format in the first two rounds and a knockout format in subsequent rounds. Doing so keeps matches to a reasonable amount but also removes any stakeless games where the outcome does not matter, and also approximates the true rank of a player very well, especially if seeding is considered. The triple knockout system presents itself as a useful alternative to further reduce matches while retaining the accurate placement of the GSL and Round Robin formats. Consequently, the triple knockout format might be worth considering in some eSports matches especially where time considerations are a serious concern.

There are several avenues for further research and several limitations of this study that could be explored further to solidify results. This research was conducted using data only from StarCraft 2 where there is an abundance of data on players and an existing system that is highly capable of presenting “true” ratings and rankings of players and approximating a win-rate. It would be useful to know if these conclusions hold true in eSports where such data is not available or ratings are more murky and upsets are more common such as CS:GO or Defence of the Ancients. Secondly, this study does not solve the conundrum of a player’s final finishing position when they are knocked out of the tournament. Each player does not play the same amount of matches in any and all tournaments. Perhaps a certain amount of matches could be introduced in a “loser” bracket of sorts to determine final finishing position that does not increase the number of matches drastically. Moreover, there are other eSports tournament designs that were not considered which might outperform the GSL format, for example. Lastly, it is clear that certain tournament designs are geared towards performance in a certain metric. The GSL foregoes number of inversions for improving Top 8 and Top 1 rankings while the double elimination tournament (counterintuitively) does otherwise. Further research could look towards creating a composite mix of these metrics to better gauge tournament or developing a normalized metric that cannot be improved upon by simply seeding.

Codebase

The code and results for this study can be found here.

References

Appleton, D. R. (1995). May the best man win? *Journal of the Royal Statistical Society: Series D (The Statistician), 44*(4), pp. 529-538.

Can, B. (2014, March). Weighted distances between preferences. *Journal of Mathematical Economics, 51*, pp. 109-115. https://doi.org/10.1016/j.jmateco.2014.01.002

Coates, D., & Parshakov, P. (2016). Team vs. Individual Tournaments: Evidence from Prize Structure in eSports. *Higher School of Economics Research Paper No. WP BRP 138/EC/2016*, 138. https://doi.org/10.2139/ssrn.2787819

Csató, L. (2019, June 14). A simulation comparison of tournament designs for the World Men's Handball Championships. *International Transactions in Operational Research, 28*(5), pp. 2377-2401. https://doi.org/10.1111/itor.12691

Csató, L. (2021). *Tournament Design: How Operations Research Can Improve Sports Rules.* Springer Nature.

Dagaev, D., & Sonin, K. (2017, May 16). Winning by Losing: Incentive Incompatibility in Multiple Qualifiers. *Journal of Sports Economics, 19*(8), pp. 1122-1146. https://doi.org/10.1177/1527002517704022

David, H. A. (1959). Tournaments and Paired Comparisons. *Biometrika, 46*(1/2), 139-149. https://doi.org/10.2307/2332816

Glenn, W. A. (1960). A Comparison of the Effectiveness of Tournaments. *Biometrika, 47*(3/4), 253-262. https://doi.org/10.2307/2333297

Glickman, M. (2012). Example of the Glicko-2 System. *Boston University*, 1-6.

Haugen, K. K., & Krumer, A. (2021). On importance of tournament design in sports management: Evidence from the UEFA Euro 2020 qualification, manuscript. In V. Ratten, *Innovation and Entrepreneurship in Sport Management* (pp. 22-35). Edward Elgar Publishing. https://doi.org/10.4337/9781783473960

Hwang, F. K. (2018, April 11). New concepts in seeding knockout tournaments. *The American Mathematical Monthly*, pp. 235-239. https://doi.org/10.1080/00029890.1982.11995420

Karpov, I. (2011, November). *Aligulac*. Retrieved from Aligulac: http://aligulac.com/about/api

Kendall, G. (2017). When sports rules go awry. *European Journal of Operational Research*, *257*(2), 377-394. https://doi.org/10.1016/j.ejor.2016.06.050

McGarry, T., & Schutz, R. W. (1997). Efficacy of traditional sport tournament structures. *Journal of the Operational Research Society, 48*(1), 65-74. https://doi.org/10.1057/palgrave.jors.2600330

Mendonça, D., & Raghavachari, M. (2000). Comparing the efficacy of ranking methods for multiple round-robin tournaments. *European Journal of Operational Research, 123*(3), 593-605. https://doi.org/10.1016/S0377-2217(99)00110-1

Scarf, P., Yusof, M. M., & Bilbao, M. (2009). A numerical study of designs for sporting contests. *European Journal of Operational Research, 198*(1), 190-198. https://doi.org/10.1016/j.ejor.2008.07.029

Searls, D. T. (1963). On the Probability of Winning with Different Tournament Procedures. *Journal of the American Statistical Association , 58*(304), 1064-1081.

Sziklai, B. R., P. B., & Csató, L. (2022). The efficacy of tournament designs. *Computers & Operations Research, 144*, 105821. https://doi.org/10.1016/j.cor.2022.105821

Taylor, B. A., & Trogdon, J. G. (2002). Losing to Win: Tournament Incentives in the National Basketball Association. *Journal of Labour Economics, 20*(1), 23-41. https://doi.org/10.1086/323930

Veček, N., Črepinšek, M., Mernik, M., & Hrnčič, D. (2014). A comparison between different chess rating systems for ranking evolutionary algorithms. *Federated Conference on Computer Science and Information Systems.* Warsaw: IEEE. https://doi.org/10.15439/2014F33

Vong, A. I. (2017). Strategic manipulation in tournament games. *Games and Economic Behavior, 102*, 562-567. https://doi.org/10.1016/j.geb.2017.02.011