On Improving Ranking in FTC

Author: Chuck Spohr
5119 Baryons Mentor
chuckspohr@gmail.com

Executive Summary

In this paper, I explore how different systems for calculating TBP and RP affect tournament rankings after qualifying rounds. I will also examine the effects of Swiss-system scheduling, along with a modification to the Swiss system for FTC specifically.

Definitions

FTC First Tech Challenge, a program under the FIRST umbrella, founded to promote STEM education through robotics mentorships and competitions at the middle and high school levels.

RP Ranking Points. 2 points earned for a win, 1 for a tie, 0 for a loss.

TBP Tie Breaker Points. Points earned at the end of a match equal to the pre-penalty score of the losing alliance.

Rank The “place” a team is in. Determined by sorting teams by RP, then TBP. A rank of 1 means the team is in first place with the most RP and the highest TBP of those teams of equal RP.

Methodology

A number of modifications suggested by the FTC community are tested by means of a simulator capable of re-running a tournament under different rules and schedules. The simulator reports outcomes and statistics which are then tabulated in Google Sheets, charted, and compared.

Results

The best option for changing TBP alone is to use the alliance’s own score (as opposed to the losing alliance’s score), but improvements are marginal at best, at about 15%.

Changing RP to the alliance’s own score is much more effective, and it doesn’t hurt much to add a win bonus. Improvements measured at 31% - 41%, depending on the implementation.

Further improvements come with Swiss-system scheduling. Combined with TBP improvements, we can achieve a noticeable improvement to rankings, at 45% - 55%, depending on the implementation.

Introduction

Having been a mentor of a successful FTC team for the last seven seasons, I've experienced the highs and lows of the FTC experience right along with the kids on the team. I've found no other pursuit as exciting and gratifying as organizing a group of high-schoolers to the common goal of building an awesome robot to do things they couldn't imagine, and competing with friends or complete strangers working on the same goals.

I've become a huge supporter of FTC and extended my volunteerism to include judging, refereeing, inspecting, and training. Each one of those is a unique experience, and I treasure the opportunity to continue for years to come.

With this passion comes the desire to improve the competition experience for the teams involved. Over the years, I've been a part of many tournaments where the qualification rankings seemed terribly out of touch with actual team performance. This sometimes helps, and often hurts teams trying to advance to the next level.

Online there have been many complaints about rankings, specifically around tie-breaker points (TBP) not being a fair way to break the ties especially at the highest rank. The alliance captains are often not the best performing teams. In regions where there are precious few advancement slots to get to the World Championships, this can be disastrous for a team who would otherwise deserve to advance.

In the paper that follows, I set out to study, define, and propose improvements to the FTC ranking system, scheduling, and rules around TBP and RP. I will identify many proposed fixes, test them, and examine the results to identify some changes FTC could implement to make the competition experience better for everyone.

Part I: Problem Definition

Let's define the problem we're trying to solve.

Problem Statement: Team rank does not correlate well to team performance

Why is that a problem? If you've attended a handful of FTC tournaments, you may have witnessed some of the effects:

- The best teams may not be the alliance captains
    - In local and regional tournaments, this can be disastrous for well-performing teams who do not attain a high enough rank to advance.
    - When advancement slots are few, there may only be enough to advance one captain.
- Teams are overlooked for selection
    - A team unprepared to be an alliance captain may not be aware of teams who deserve a spot in elimination rounds.
- Teams' performance not captured in rank value
    - Participants and spectators look at rank to measure how well their team did that day. You may have seen a team putting up more points than most, but have fallen to a surprisingly low ranking.

How can we know how well a team performed when their scores are combined with their partners'? Fortunately, in both FTC and FRC, this has been solved in a very clever way: OPR. This is a mathematical estimation of a team's average contribution to their alliance scores, taking the strengths and weaknesses of a team's partners into account.

It's one of the few tools available to measure an individual team's performance short of direct observation. Scouting is important, of course. But this offers a way to calculate the individual team's scoring averages based on the alliance scores publicly posted on numerous websites.

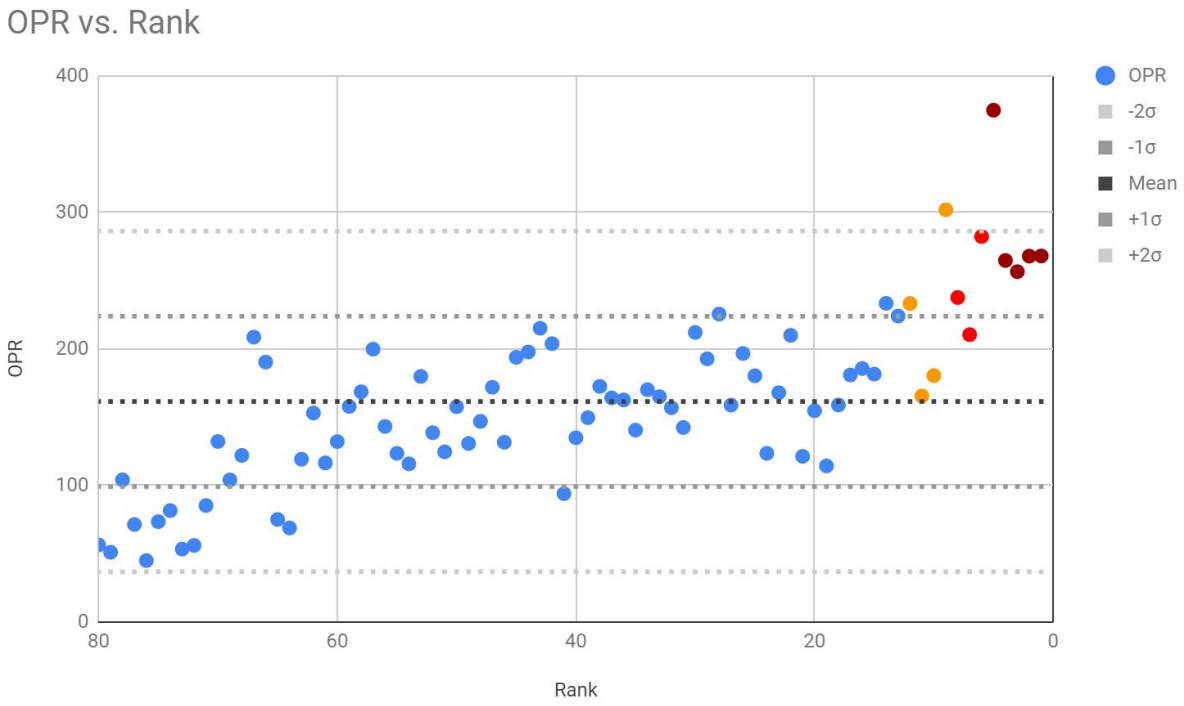
Unfortunately, this is not available to spectators at a tournament. They only have RP, TBP, and rank to look at on the scoreboard. Some regions do post live results to websites such as theorangealliance.com and ftcscores.com, which do show OPRs, but many teams and spectators are unfamiliar with them.

Experienced teams do this, though, to quickly gauge a team's scoring potential. The wiser of the alliance captains know that rank is not accurate, and will use a combination of OPR and their own scouting data to choose the best partners in alliance selection.

If you are not familiar with OPR, I recommend The Math Behind OPR — An Introduction as a starting point for understanding what it means and how it is calculated.

For teams to leverage OPR during competition, they must know how to calculate it and have written software do so, or use one of a handful of existing apps to do that for them.
For the purposes of this study, I will be using OPR for simulating match scores in theoretical matchups and comparing OPR values and team ranks to see how well a given ranking system correlates to team performance.

Let's start by looking at a couple of actual tournaments. The following charts show OPR vs. Rank. First up is Detroit Edison 2019.



The dot colors indicate RP. Dark Red is 18, Bright Red is 16, Yellow is 14, then everything else is blue. The dotted lines are standard deviation bars showing the first and second standard deviations above and below the mean.

You can see from this chart that the team with the highest OPR actually ranked 5th. The top two ranked teams were about equal in OPR, but were 4th and 5th in OPR values.

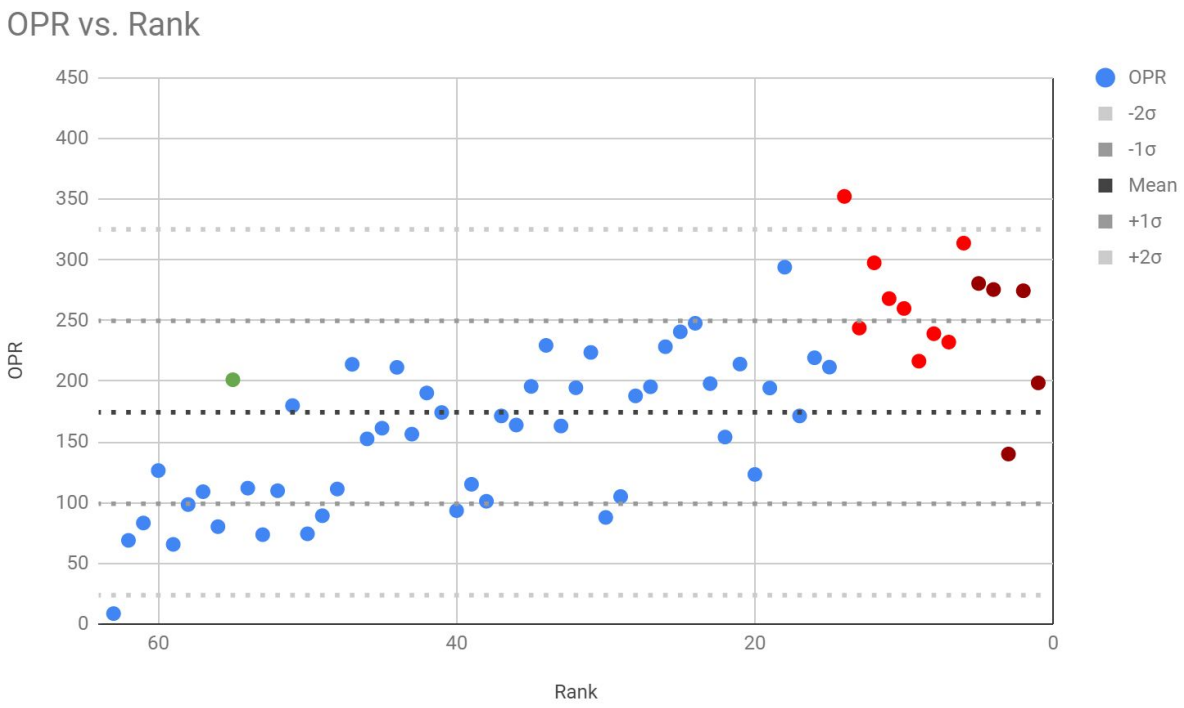
When you get into the RP=16 group, the 7th ranked team actually had a lower OPR than a large number of teams behind them.

Furthermore, when you get into the middle 40 or 50 ranks, there is no apparent correlation between rank and OPR. It just seems completely randomly arranged. With a better system, we

would hope to see the dots arranged in a tight, ascending configuration, especially at the top where so much is at stake in identifying alliance captains.

I did not personally attend the World Championships in Detroit in 2019. I was, however, at the Houston event in 2018 with my team, the Baryons. There we experienced one of the awful effects of the current ranking system.

Let's look at the OPR vs Rank chart for Houston Jemison 2018.



In this case, there were no undefeated teams, so dark red is RP=14 and bright red is RP=12. The green dot is my own team, 5119 Baryons at 55th rank. Now on to some observations...

All of the top 4 teams, the alliance captains, had many teams below them with a higher OPR. In fact, the Baryons at 55th rank had a slightly higher OPR than the #1 team. The #3 team's OPR was below average for the entire field.

Within the top two RP groups (14 and 12), there appears to be a negative correlation between rank and OPR. In other words, as rank goes up, OPR goes down. It's as if the ranking order reversed within an RP group.

Throughout this entire chart, the dots are scattered widely and rank order is only very loosely correlated to OPR.

From my own experience and by talking to others, reasons that rankings get so far off include:

- Varying difficulty of schedules
- Top teams facing off or not
- Ties broken by TBP alone
- Teams at the top are almost always tied in RP
- TBP is a result of schedule, not performance

Goals

So now that we've defined the problem, what do we want to achieve? In this study, by testing many proposed solutions and measuring outcomes, I will identify a small number of targeted improvements to accomplish the following:

- Significant improvement to the the quality of rankings, real and perceived
- Reduce incentives to game the system
- Incremental changes to the current system
- Easy-to-understand for spectators and participants
- Maintain flow, fun, and excitement of current system

With regard to improving the quality of the rankings, we seek to measure and improve specifically:

- The degree of scattering on the OPR vs. Rank chart
- The number of "places-off" a teams rank is vs. where it would be if ordered by OPR Rank. A value of 0 would indicated that teams are ranked by their OPR values.
- The number of Top 7 teams in the Top 7 ranks

Part II: Methodology

Test Cases

When evaluating new scoring, scheduling, and ranking models, we need to define what we're testing and how to measure outcomes.

In this study, I will alter, test, and measure outcomes for the following:

- TBP Calculation
    - Losing Score (no change, for comparison)
    - Own Score
    - Total Score
    - Own Score + Losing Score
    - Top x Scores (teams best 3 out of 5, or 5 out of 9 matches)
    - Highest Score
- RP Calculation
    - Win * 2 + Tie * 1 (no change, for comparison)
    - Own Score
    - Own Score + Win * 100 + Tie * 50
    - Win * 2 + Tie * 1 + Achievement * 1
    - Win * 2 + Tie * 1 + (OwnScore >= 200) * 1 + (OwnScore >= 400) * 1
- Scheduling System
    - Random Scheduling (no change, for comparison)
    - Swiss-System Scheduling
    - FTC-Modified Swiss System

Own Score is the alliance's own pre-penalty point total from each of their matches. Win or lose, I count the alliance's own value.

Total Score is the sum of both alliances' pre-penalty point totals.

Top x and Highest Score is the alliances own top pre-penalty scores or single best score.

These test cases will be run through a purpose-built event simulator, capable of rerunning an actual event under different RP/TBP rules and alternate scheduling systems. More on this in the next section.

Since not all tournaments are equal in FTC, I will test events of various sizes. The TOA Event Keys used for these simulations are:

Small:	12-team tournament	1819-AZ-FTCAZNMD
Small:	16-team regional tournament	1819-NJ-LTNJ6
Medium:	24-team league championship	1819-FL-SCLT1
Medium:	32-team regional tournament	1819-CO-FTCCOCQ
Large:	64-team world championship	1718-CMP-HOU2
Large:	80-team world championship	1819-CMP-DET1

Measurements

We will measure outcomes with a few different metrics:

- Spearman Correlation - A statistical correlation coefficient for measuring the noisiness of ranked values. 0 is no correlation, 1 is perfect correlation.
- OPRRankDif - The average difference between a teams OPR rank and their FTC Rank.
- Top7InTopRank - The number of Top 7 OPR teams in the Top 7 FTC Rank.

The last metric, Top7InTopRank, focuses on the Top 7 because if each of the first three Alliance Captains pick another Alliance Captain in the selection ceremony, the 7th place team will become the last Alliance Captain. All Top 7 teams will be either captain or 1st pick. In most regional tournaments, it will be from this pool that advancements are made.

For model comparison purposes, I will calculate a percent improvement, which is the delta (measured - baseline) over the best possible change. For example, if the baseline correlation measured 0.79, the best possible is 1.0, and the test case is 0.86, then the improvement is (0.86 - 0.79) / (1.0 - 0.79) = 0.07 / 0.21 = 0.33

Finally, to rank models at the end, I will average the three measurements' percent improvements for an overall improvement.

The Simulator: FTC EventSim

FTC EventSim is a command-line application which takes a user-specified options file and executes a tournament according to the rules specified. The output is written to the console, which can easily be directed to a file. What is output is also configured in the options file. Sample options files are available in the GitHub repository, documented in the README.md, and a number of additional articles are available in the Wiki.

To reproduce to results discussed here, download the software and options files from GitHub, and execute the commands provided here.

Demonstration: Actual Results

First, let's look at how to run an actual tournament as it happened so we can take some measurements. We'll recreate Detroit Edison 2019 and Houston Jemison 2018. First up, Detrioit:

```
EventSim StudyOptionFiles\DetEdActualRankings.xml > DetEdAct.txt
```

The software will download actual results from theorangealliance.com, tally up match results and report actual rankings and various statistics calculated along the way. An excerpt from the rankings output:

```
Rank  Number  PPM    RP  TBP   OPR    OPRRank  OPRDif  PPMRank PPMDif Dfclty
1     9971    268.0  18  2928  268.0  4        3       4       3      1476
2     8680    267.9  18  2892  267.9  5        3       5       3      1235
3     12231   256.5  18  2877  256.5  7        4       7       4      1284
4     10435   264.7  18  2641  264.7  6        2       6       2      1314
5     11115   374.7  18  2437  374.7  1        -4      1       -4     1279
6     8479    282.2  16  2772  282.2  3        -3      3       -3     1447
7     10918   210.4  16  2513  210.4  15       8       15      8      1108
```

We also get some metrics about the tournament as a whole:

```
Teams Matches High Low Avg    OPRDif TopX TopXDif OPRTopX PPMTopX OPRErr TOPRErr OPRCor
TOPRCor
80    180     642  10  482.29 11.83  7    3.86    6       6       29.5   4.2     0.74   0.11
```

The key metrics we're interested in for this study are:

Detroit Edison 2019 Actual

OPRCor = 0.74	Spearman Correlation OPR vs. Rank
OPRDiff = 11.83	Average abs value of the difference between OPRRank and Rank
OPRTopX = 6	Number of Top 7 OPR teams in the Top 7 Rank

Now run the same analysis for Houston Jemison 2018:

```
EventSim StudyOptionFiles\HouJemActualRankings.xml > HouJemAct.txt
```

Houston Jemison 2018 Actual

OPRCor = 0.76	Spearman Correlation OPR vs. Rank
OPRDiff = 9.83	Average abs value of the difference between OPRRank and Rank
OPRTopX = 4	Number of Top 7 OPR teams in the Top 7 Rank

Demonstration: Simulated Results from Actual Schedule

FTC EventSim can rerun an actual tournament with the same teams and strengths (OPR) by creating a new random schedule with theoretical matchups and determine outcomes. It will generate the same ranking report and summary data.

Match scores can be generated with the straight OPR value, or by optionally adding random noise to the scores. The noise generation algorithm attempts to simulate actual score variances, following a similar distribution and spread from actual tournament data. The details behind this are in the Wiki article, Score randomness and distribution analysis.

For the purposes of this study, score randomness settings are set to simulate the distribution and spread seen at the Detroit Edison 2019 event.

Since the schedule and scores will be different each time, it is useful to run several iterations of these simulations to see how results change from run to run. I will demonstrate by running Detroit Edison 2019 three times to see what happens:

```
EventSim StudyOptionFiles\DetEdRandomRankings.xml > DetEdRnd.txt
```

Run 1

Rank	Number	PPM	RP	TBP	OPR	OPRRank	OPRDif	PPMRank	PPMDif	Dfclty
1	11115	374.7	18	3011	412.7	1	0	1	0	1609
2	6931	185.6	18	2500	239.4	13	11	25	23	909
3	8479	282.2	16	3004	286.0	4	1	3	0	1734
4	12231	256.5	16	2830	268.3	8	4	7	3	1504
5	11316	215.1	16	2591	180.5	31	26	13	8	1047
6	10918	210.4	16	2497	211.3	19	13	15	9	1534
7	14270	301.8	16	2145	274.6	6	-1	2	-5	1071

Detroit Edison 2019 Random Schedule
OPRCor = 0.73 Spearman Correlation OPR vs. Rank
OPRDiff = 13.25 Average abs value of the difference between OPRRank and Rank
OPRTopX = 3 Number of Top 7 OPR teams in the Top 7 Rank

Run 2

Rank	Number	PPM	RP	TBP	OPR	OPRRank	OPRDif	PPMRank	PPMDif	Dfclty
1	12231	256.5	18	2721	316.5	2	1	7	6	1269
2	10918	210.4	18	2614	251.9	7	5	15	13	1402
3	11316	215.1	16	2610	222.6	16	13	13	10	580
4	11115	374.7	16	2604	359.7	1	-3	1	-3	1290
5	14270	301.8	16	2559	291.9	5	0	2	-3	965
6	8479	282.2	16	2542	286.7	6	0	3	-3	1211
7	12589	233.2	16	2532	245.3	9	2	10	3	1002

Detroit Edison 2019 Random Schedule
OPRCor = 0.81 Spearman Correlation OPR vs. Rank
OPRDiff = 10.98 Average abs value of the difference between OPRRank and Rank
OPRTopX = 5 Number of Top 7 OPR teams in the Top 7 Rank

Run 3

Rank	Number	PPM	RP	TBP	OPR	OPRRank	OPRDif	PPMRank	PPMDif	Dfclty
1	9971	268.0	18	2601	264.9	4	3	4	3	1251
2	10030	167.9	18	1903	181.7	28	26	35	33	1067
3	11115	374.7	16	3511	430.5	1	-2	1	-2	2026
4	12538	208.6	16	2806	241.0	9	5	17	13	833
5	8479	282.2	16	2727	248.7	7	2	3	-2	745
6	12589	233.2	16	2613	259.8	6	0	10	4	950
7	12231	256.5	16	2465	234.4	10	3	7	0	1239

Detroit Edison 2019 Random Schedule

OPRCor = 0.79 Spearman Correlation OPR vs. Rank

OPRDiff = 11.20 Average abs value of the difference between OPRRank and Rank

OPRTopX = 4 Number of Top 7 OPR teams in the Top 7 Rank

For comparison, the Actual metrics from earlier:

Detroit Edison 2019 Actual

OPRCor = 0.74 Spearman Correlation OPR vs. Rank

OPRDiff = 11.83 Average abs value of the difference between OPRRank and Rank

OPRTopX = 6 Number of Top 7 OPR teams in the Top 7 Rank

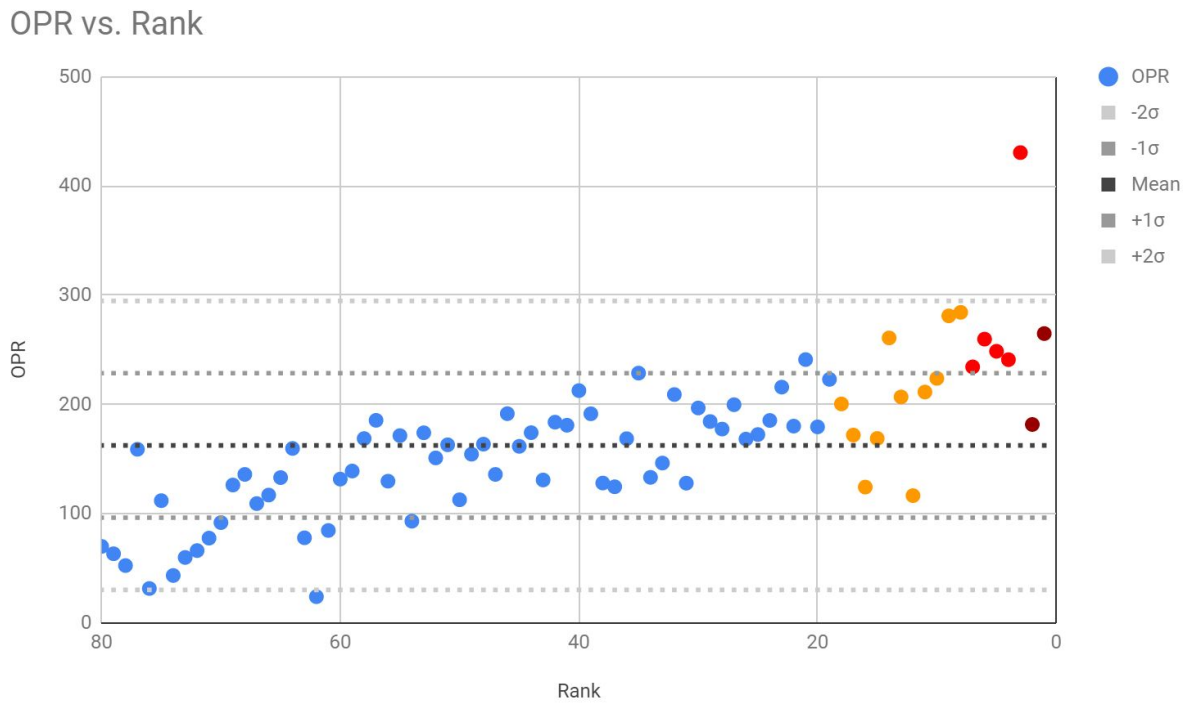
You can see by studying the team numbers in the top 7 ranks that the lineup changes dramatically for each run.

The metrics change as well, but actual event metrics generally fall within the range of values from the random events.

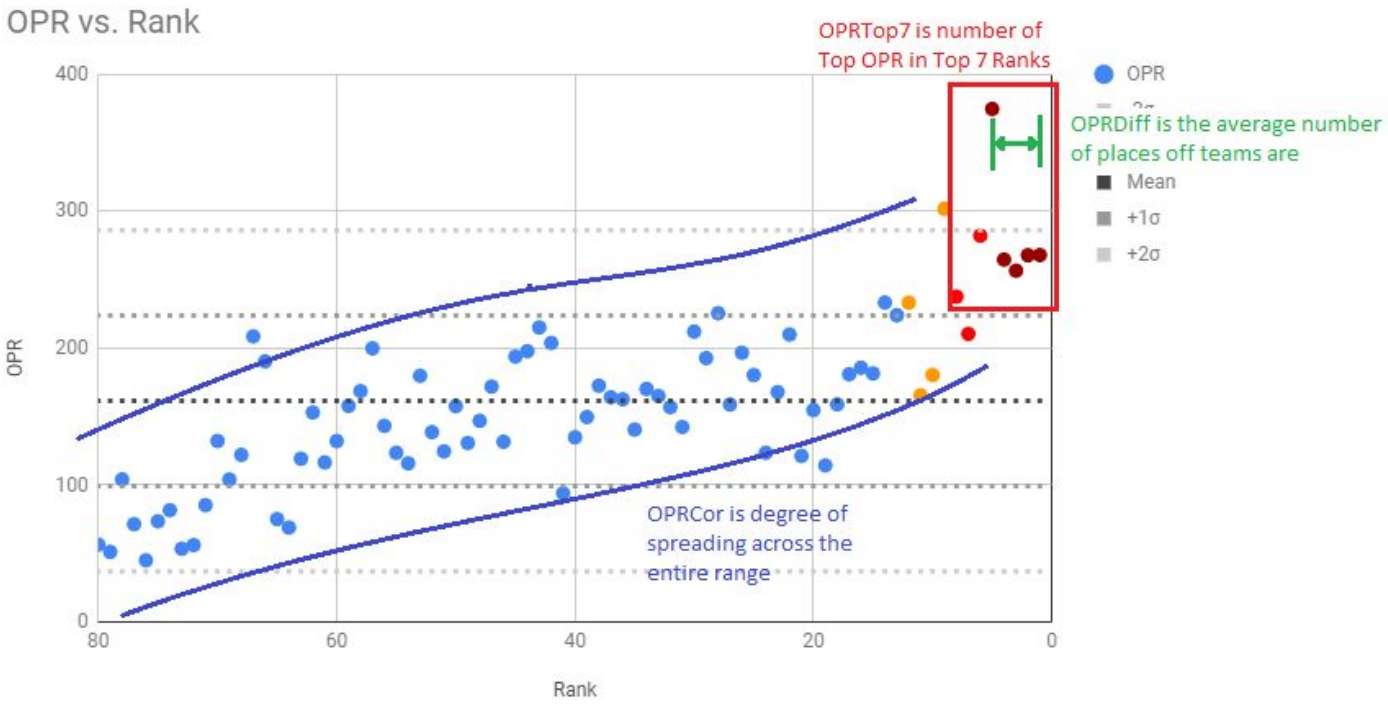
Later in this study, test results will be based on 1000 runs with the metrics averaged over the 1000 run set.

On the next page, let's visually compare that last event to the actual results with the actual results. Annotations on the last graph show what we're measuring.

Detroit Edison 2019 Random Run 3



Detroit Edison 2019 Actual



Part III: TBP Improvements

Rationale for Changing TBP

In this section we will be looking only at changing the way TBP is calculated. Reasons to change the TBP calculation include:

- Teams “game” TBP by scoring for opposite alliance to boost losing score
- If not gamed, TBP is a value outside a team’s control and is affected more by random schedule
- If gamed, a team’s OPR is affected negatively. Potential alliance partners monitor OPR.
- FTC publicly stated in April 2019 that they would be reevaluating the TBP rule and is looking for viable alternatives
- Spectators find the current Losing Score system is confusing.

Drawbacks from Changing TBP

What can go wrong by changing TBP? Reasons to leave it alone include:

- Match difficulty no longer measured directly
- May be confusing for spectators if new rule is complicated
- May be gamed in unanticipated ways
- May still be dependent on match schedule
- Only affects rankings within an RP-tied group. Does not affect RP.

Proposed TBP Fixes

In Part I, I listed test cases for the proposed TBP changes. In review, they are:

- Losing Score (no change, for comparison)
- Own Alliance Score
- Total (Both Alliance) Score
- Own Alliance Score + Losing Score
- Top x Own Alliance Scores
- Highest Own Alliance Score

Testing TBP Fixes By Simulating Events

To test these I will run 1000 iterations of each test case. A test case is the TBP rule applied to each test tournament. For example, to test “Own Alliance Score”, I will run it 1000 times for

each of the tournament sizes listed earlier in Part I. That's 6 tournaments, for a total of 6000 simulated events per rule. With 6 rules to test, we'll be simulating 36,000 tournaments.

Schedules are random and different in each simulation, and scores are calculated along a typical random distribution and spread with OPR set as the mean for each team, as discussed in Part II.

Results are averaged over each of the 1000-trial test case and compared.

We use a Google sheet, Ranking Study - TBP Test Cases, to lay out the test cases and format the commands for the simulator. These commands are copied into a script to run all at once.

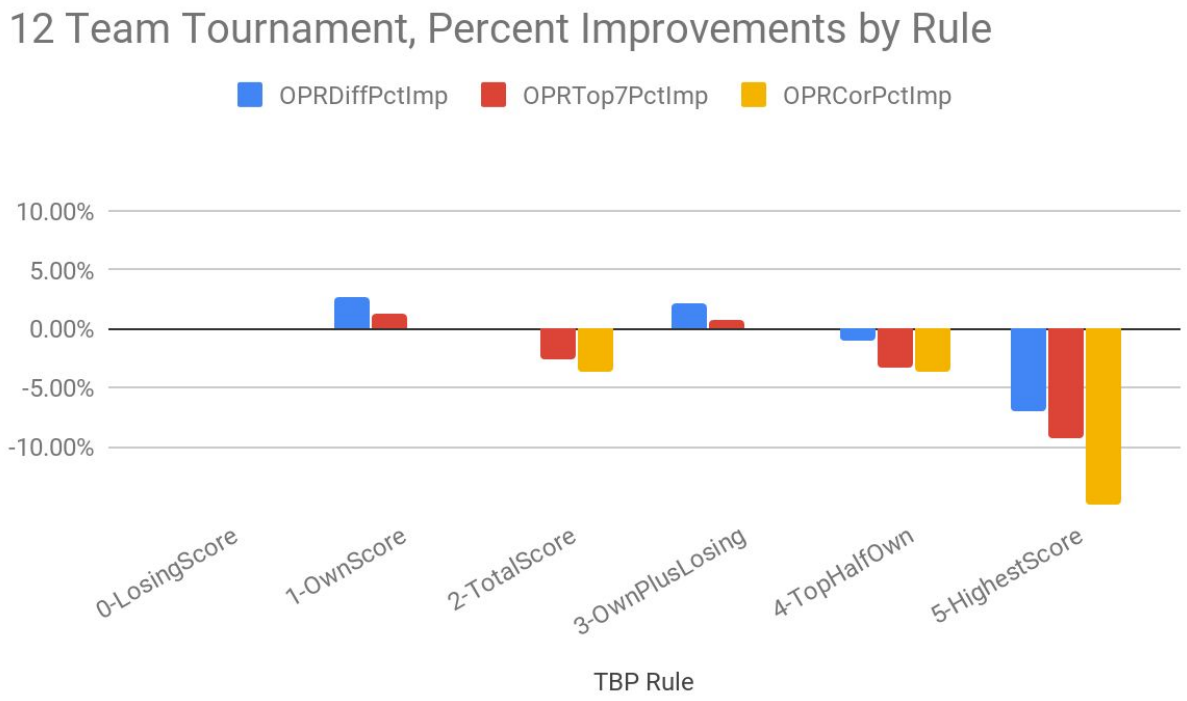
The results are then copied into a new tab in the sheet, formatted, and filtered on the next page.

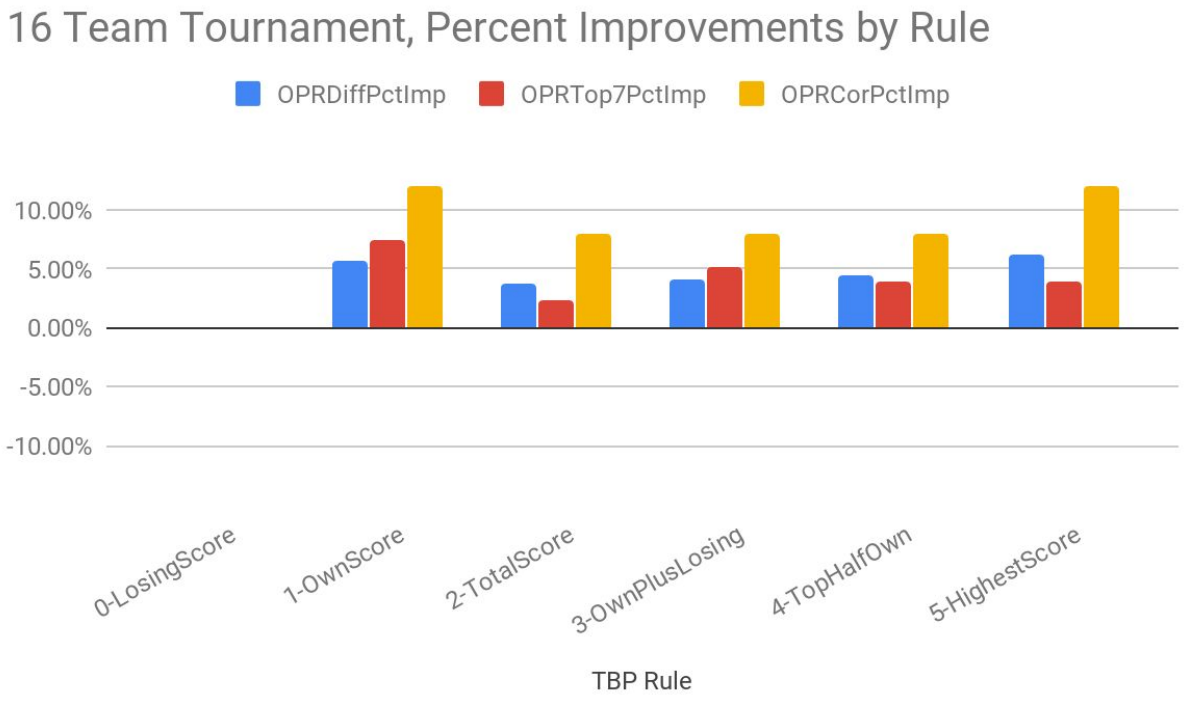
TBP Results

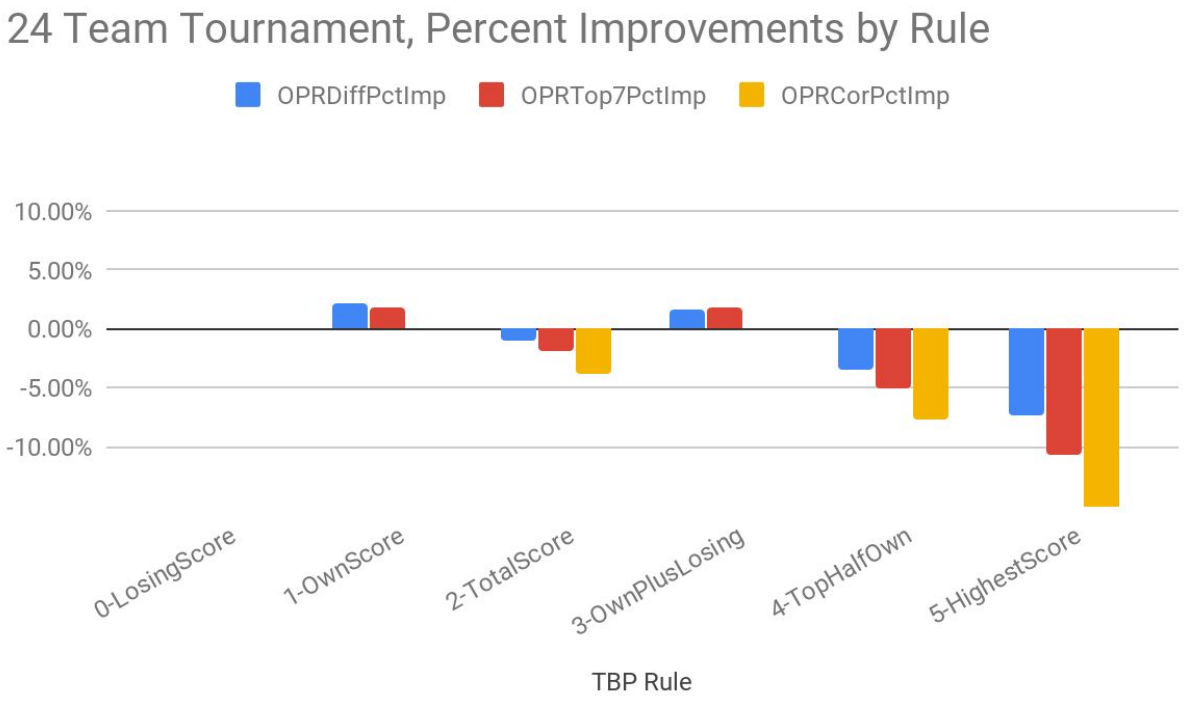
Title	Teams	OPRTopX	OPRTop7PctImp	OPRDif	OPRDiffPctImp	OPRCor	OPRCorPctImp
0-LosingScore	12	5.5	0.00%	1.85	0.00%	0.73	0.00%
1-OwnScore	12	5.52	1.33%	1.8	2.70%	0.73	0.00%
2-TotalScore	12	5.46	-2.67%	1.85	0.00%	0.72	-3.70%
3-OwnPlusLosing	12	5.51	0.67%	1.81	2.16%	0.73	0.00%
4-TopHalfOwn	12	5.45	-3.33%	1.87	-1.08%	0.72	-3.70%
5-HighestScore	12	5.36	-9.33%	1.98	-7.03%	0.69	-14.81%
0-LosingScore	16	5.24	0.00%	2.45	0.00%	0.75	0.00%
1-OwnScore	16	5.37	7.39%	2.31	5.71%	0.78	12.00%
2-TotalScore	16	5.28	2.27%	2.36	3.67%	0.77	8.00%
3-OwnPlusLosing	16	5.33	5.11%	2.35	4.08%	0.77	8.00%
4-TopHalfOwn	16	5.31	3.98%	2.34	4.49%	0.77	8.00%
5-HighestScore	16	5.31	3.98%	2.3	6.12%	0.78	12.00%
0-LosingScore	24	5.4	0.00%	3.78	0.00%	0.74	0.00%
1-OwnScore	24	5.43	1.87%	3.7	2.12%	0.74	0.00%
2-TotalScore	24	5.37	-1.88%	3.82	-1.06%	0.73	-3.85%
3-OwnPlusLosing	24	5.43	1.87%	3.72	1.59%	0.74	0.00%
4-TopHalfOwn	24	5.32	-5.00%	3.91	-3.44%	0.72	-7.69%
5-HighestScore	24	5.23	-10.63%	4.06	-7.41%	0.7	-15.38%
0-LosingScore	32	4.61	0.00%	4.77	0.00%	0.77	0.00%
1-OwnScore	32	4.81	8.37%	4.51	5.45%	0.79	8.70%
2-TotalScore	32	4.81	8.37%	4.7	1.47%	0.77	0.00%
3-OwnPlusLosing	32	4.81	8.37%	4.6	3.56%	0.78	4.35%
4-TopHalfOwn	32	4.72	4.60%	4.73	0.84%	0.77	0.00%
5-HighestScore	32	4.6	-0.42%	4.92	-3.14%	0.76	-4.35%
0-LosingScore	64	4.01	0.00%	8.08	0.00%	0.83	0.00%
1-OwnScore	64	4.24	7.69%	7.8	3.47%	0.84	5.88%
2-TotalScore	64	4.2	6.35%	7.9	2.23%	0.84	5.88%

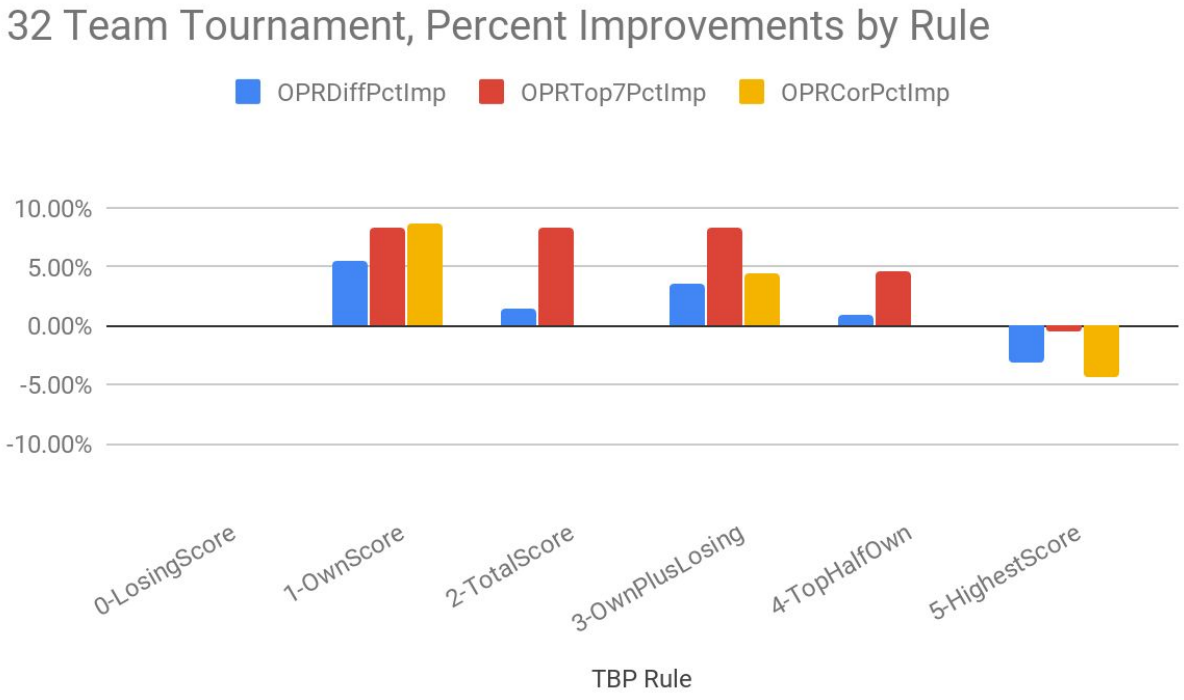
3-OwnPlusLosing	64	4.17	5.35%	7.84	2.97%	0.84	5.88%
4-TopHalfOwn	64	4.15	4.68%	7.83	3.09%	0.84	5.88%
5-HighestScore	64	4.16	5.02%	8.09	-0.12%	0.83	0.00%
0-LosingScore	80	4.48	0.00%	10.15	0.00%	0.83	0.00%
1-OwnScore	80	4.76	11.11%	9.85	2.96%	0.84	5.88%
2-TotalScore	80	4.68	7.94%	10.01	1.38%	0.83	0.00%
3-OwnPlusLosing	80	4.72	9.52%	9.93	2.17%	0.83	0.00%
4-TopHalfOwn	80	4.7	8.73%	10.05	0.99%	0.83	0.00%
5-HighestScore	80	4.47	-0.40%	10.67	-5.12%	0.81	-11.76%

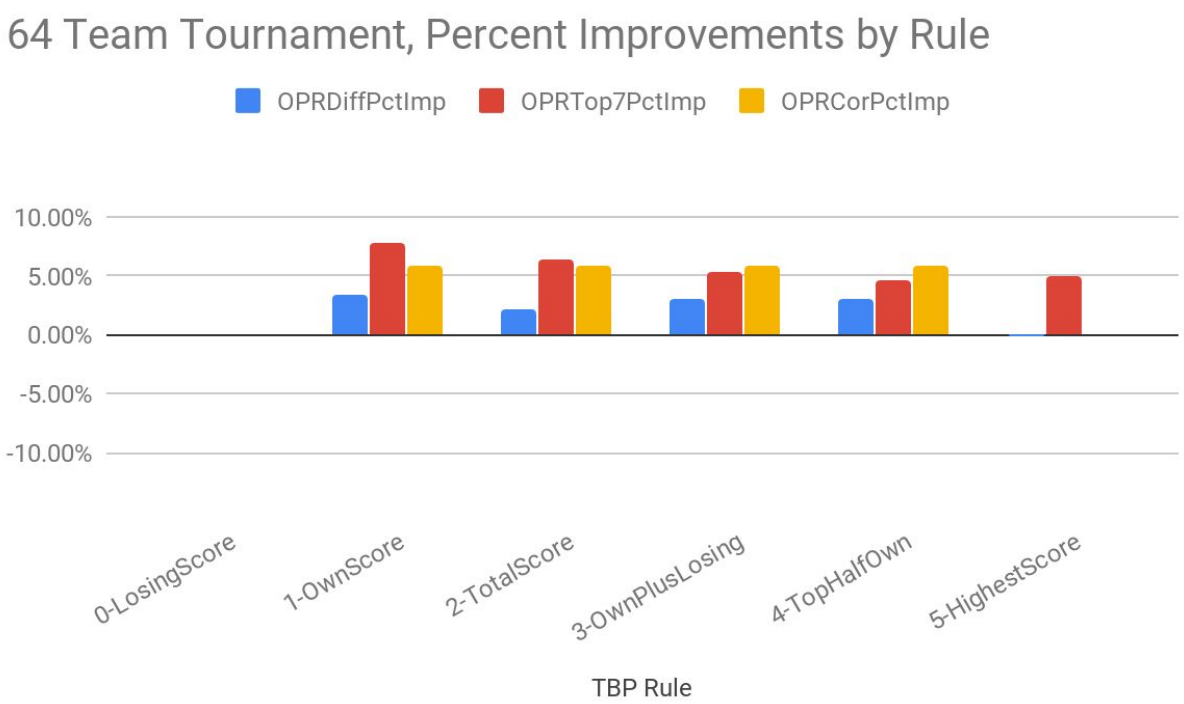
Furthermore, we can visually compare Percent Improvement figures with a series char, one for each event size:

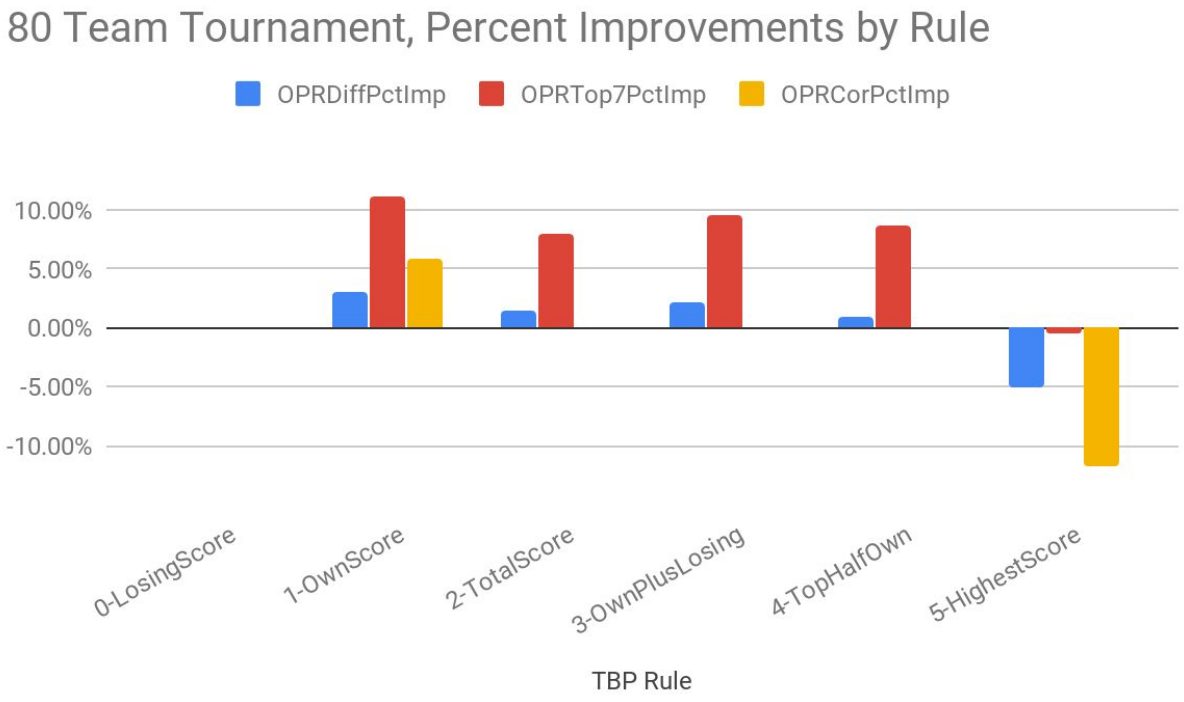


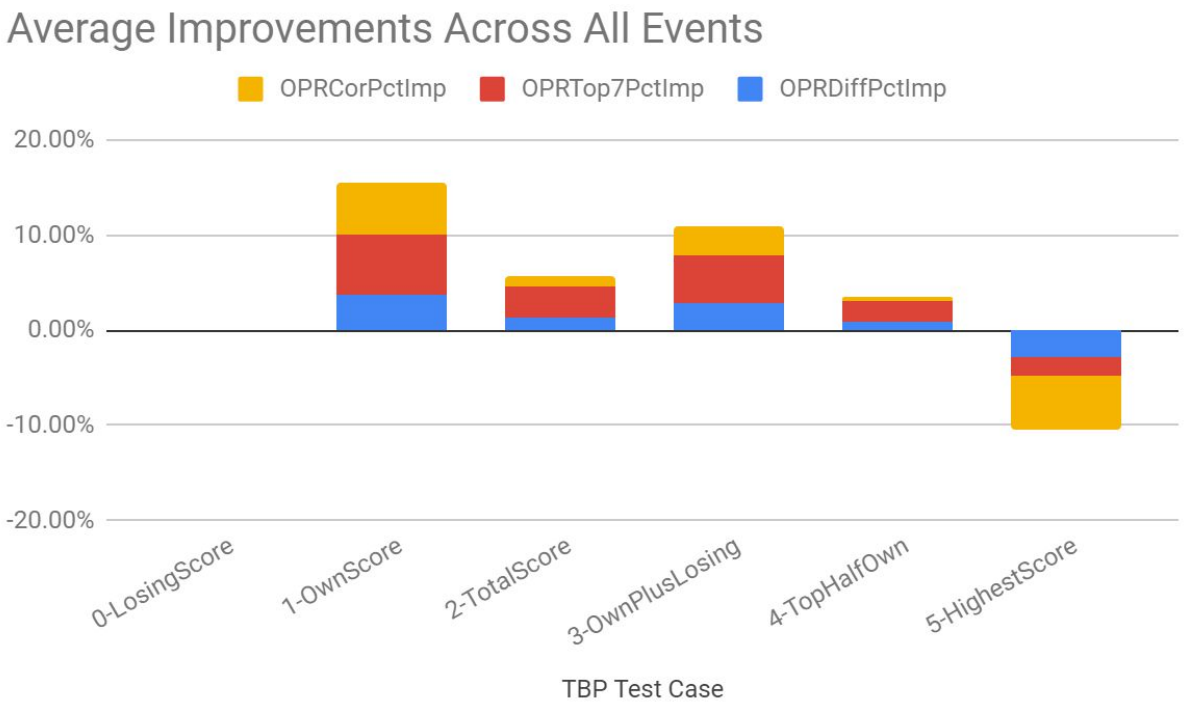

16 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
10.00%
5.00%
0.00%
-5.00%
-10.00%
0-LosingScore
1-OwnScore
2-TotalScore
3-OwnPlusLosing
4-TopHalfOwn
5-HighestScore
TBP Rule


24 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
10.00%
5.00%
0.00%
-5.00%
-10.00%
0-LosingScore
1-OwnScore
2-TotalScore
3-OwnPlusLosing
4-TopHalfOwn
5-HighestScore
TBP Rule


32 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
10.00%
5.00%
0.00%
-5.00%
-10.00%
0-LosingScore
1-OwnScore
2-TotalScore
3-OwnPlusLosing
4-TopHalfOwn
5-HighestScore
TBP Rule


64 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
10.00%
5.00%
0.00%
-5.00%
-10.00%
0-LosingScore
1-OwnScore
2-TotalScore
3-OwnPlusLosing
4-TopHalfOwn
5-HighestScore
TBP Rule


80 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
10.00%
5.00%
0.00%
-5.00%
-10.00%
0-LosingScore
1-OwnScore
2-TotalScore
3-OwnPlusLosing
4-TopHalfOwn
5-HighestScore
TBP Rule


Average Improvements Across All Events
OPRCorPctImp
OPRTop7PctImp
OPRDiffPctImp
20.00%
10.00%
0.00%
-10.00%
-20.00%
0-LosingScore
1-OwnScore
2-TotalScore
3-OwnPlusLosing
4-TopHalfOwn
5-HighestScore
TBP Test Case

Observations

The measurements for OPRCor, or degree of scatter on the OPR vs Rank charts, shows the least improvement in most event sizes by any TBP rule change. We can conclude that OPR/Rank correlation is not helped much by the TBP calculation.

Similarly, the OPRTop7 measurements are affected by only small percentages. The best case improvement we get is +10%. 10% of 7 is 0.7, so that would mean on average, we would get one more of the top 7 teams in the top 7 ranks in about two thirds of the tournaments. We can conclude that TBP changes will only improve top rank placement a little, and not all the time.

OPRRankDiff shows the best improvement by TBP changes, with a max of almost 15% in tournaments of 32 and above. Unfortunately, with an average OPRDiff of about 4.7 in 32-team events, we're only improving team ranks by less than 1 on average.

OwnScore shows the most improved OPRRankDiff across most event sizes. But at a max of less than 15%, it's not a significant improvement.

TopHalfOwn and OwnPlusLosing are next in most improvement, TopHalfOwn performs best in large events where OwnPlusLosing performs best in smaller events.

HighestScore is inconsistent across event sizes, sometimes performing best, sometimes worse than LosingScore.

Part IV: RP Improvements

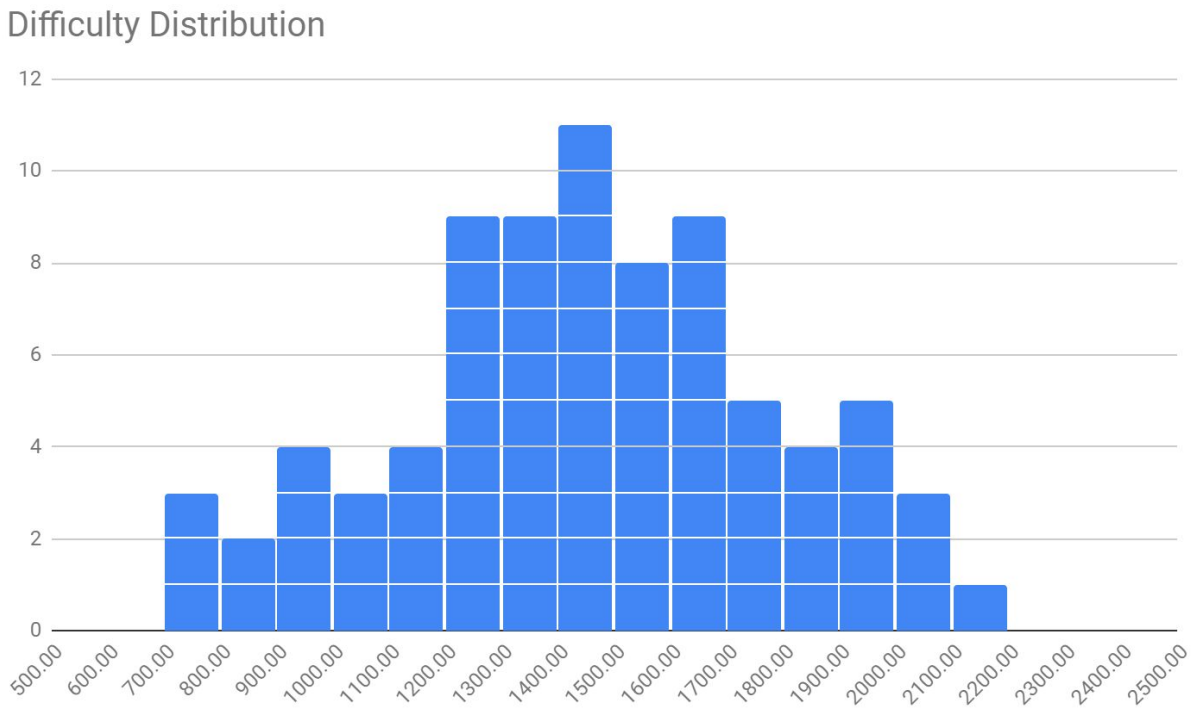
Rationale for Changing RP

It's apparent from the TBP tests above that something more is needed to make noticeable improvements. Since TBP can only affect rank within an RP-tied group, we need to look at how RP can be modified such that higher performing teams' RP reflects their true rank. Some reasons cited for RP being an inaccurate measurement of performance include:

- Random scheduling affects RP
- If the 2nd best team loses to the best team, that team falls well out of 2nd position, ranked below the teams who did not face the best team
- RP is highly dependent on match difficulty, which can vary up to a factor of 3 when measured by Difficulty = Opponent1OPR + Opponenet2OPR - Partner1OPR

Let's take a look two charts to visualize how schedule difficulty comes into play. First, take a look at Detroit Edison's schedule difficulty distribution.

Detroit Edison 2019 Actual

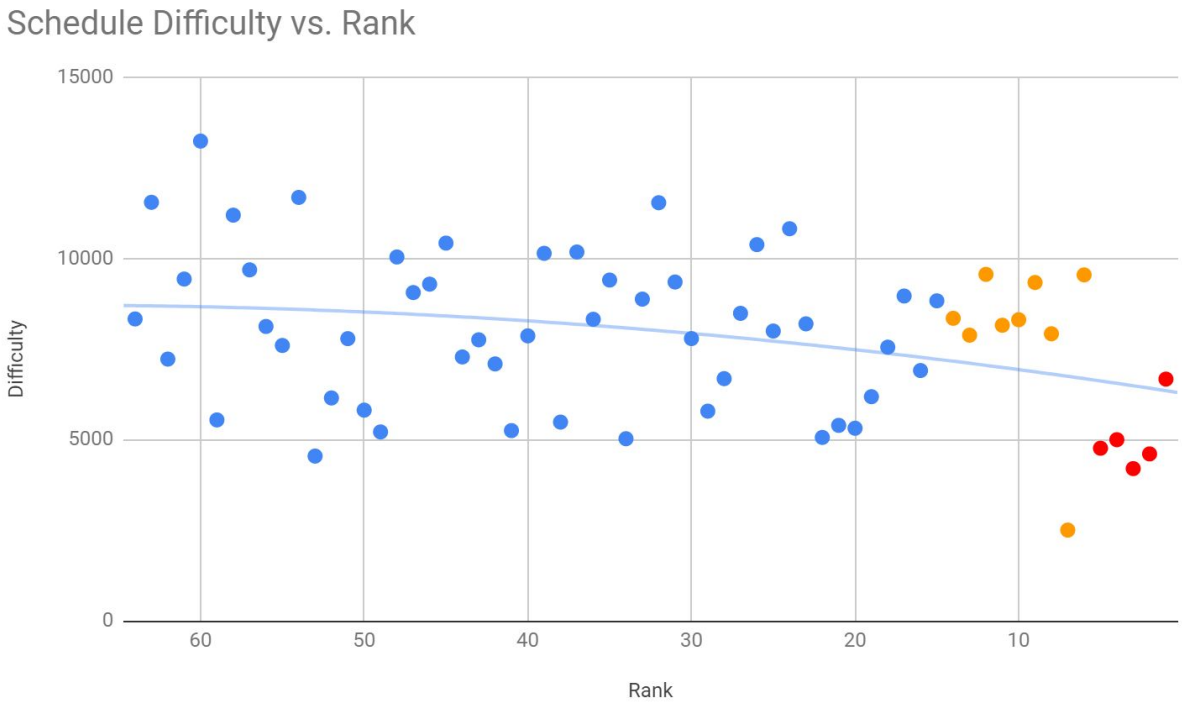


Schedule Difficulty is reported by the simulator by totalling each team's opponents' OPR's, and subtracting their partners' OPR's. A team's own OPR is not a factor in this equation, so any team facing the same schedule will face the same difficulty level.

You can see from this distribution that difficulty ranges widely, from about 700 up to about 2200. That's about a factor of 3. In other words, teams on the right face a challenge 3 time greater than teams on the left to win their matches and gain RP.

Now let's see how Difficulty and Rank are related, this time in Houston.

Houston Jemison 2018 Actual



As expected, Difficulty falls as you move to the tops of the ranks. In fact, 6 of the top 7 ranked teams had schedules of below-average difficulty. In 5 of those, they were some of the least difficult schedules of the entire division. That's not the teams' fault. They were given their random schedules like everyone else.

Drawbacks from Changing RP

We've all come to understand RP's relationship with win/loss record. If we change RP, what does that mean? We could award RP to game achievements similar to FRC, or we could just turn match points into RP like we do TBP. But any change to RP will move it away from indicating number of wins. There are some drawbacks:

- Match win/loss results lose importance
- May be confusing for spectators
- Most sports and gaming leagues rank primarily on W/L record
- Achievement-based RP loses effectiveness at high ranks since we can assume all high-performing teams are attaining the achievement

Proposed RP Fixes

The simulator can calculate RP based on Wins, Losses, Ties, various point values, and achievements. For the purposes of the simulator's expression evaluator, the value for Win is 1 if the alliance wins the match, and the value for Tie is 1 if it's a tie. Both alliances get that value in a tie. The expressions within parenthesis are true/false evaluations, returning 1 or 0.

As mentioned in Part I, we'll be testing the following RP formulas:

- Win * 2 + Tie * 1 (no change, for comparison)
- Own Score
- Own Score + Win * 100 + Tie * 50
- Win * 2 + Tie * 1 + Achievements * 1
- Win * 2 + Tie * 1 + (OwnScore >= 200) * 1 + (OwnScore >= 400) * 1

Testing RP Fixes By Simulating Events

Again, I will use the simulator to run 1000 trials of each event size, using a random schedule and random scores.

I will use the existing rules for TBP (Losing Score) so the only effect observed will be due to the RP rule changes. I will measure and compare the same factors as I did earlier, and look at the resulting charts for visual comparison.

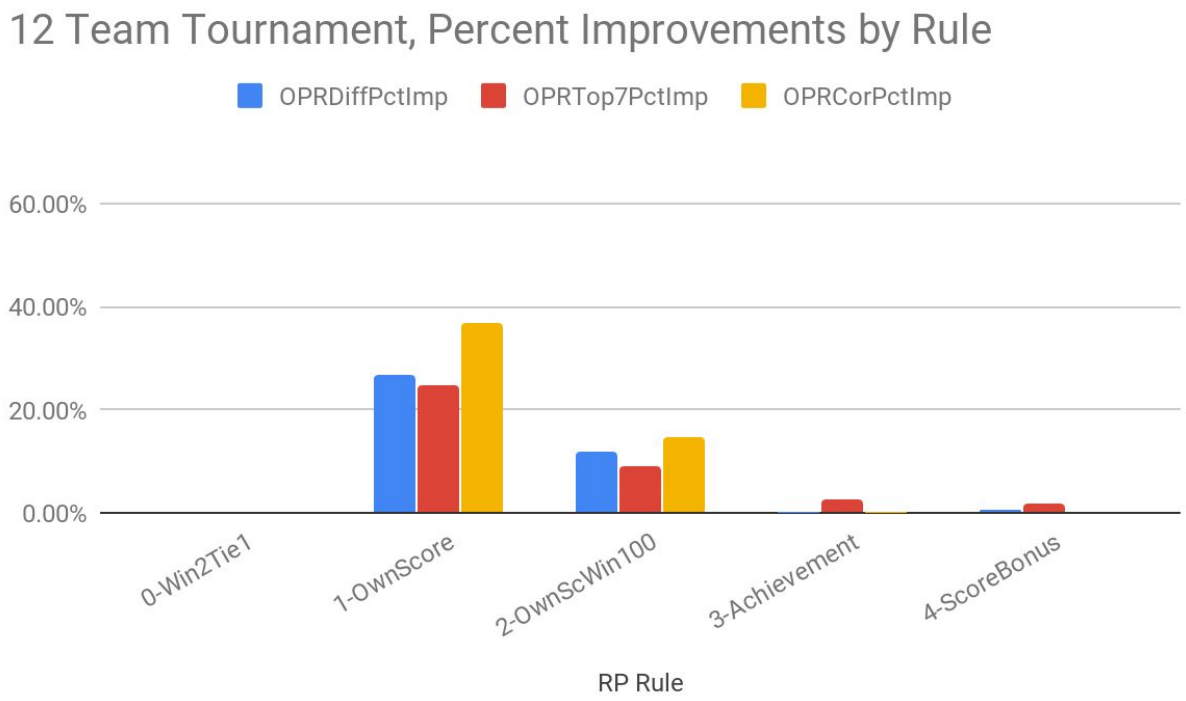
Test cases, results, and charts can be found on the Google sheet, Ranking Study - RP Test Cases.

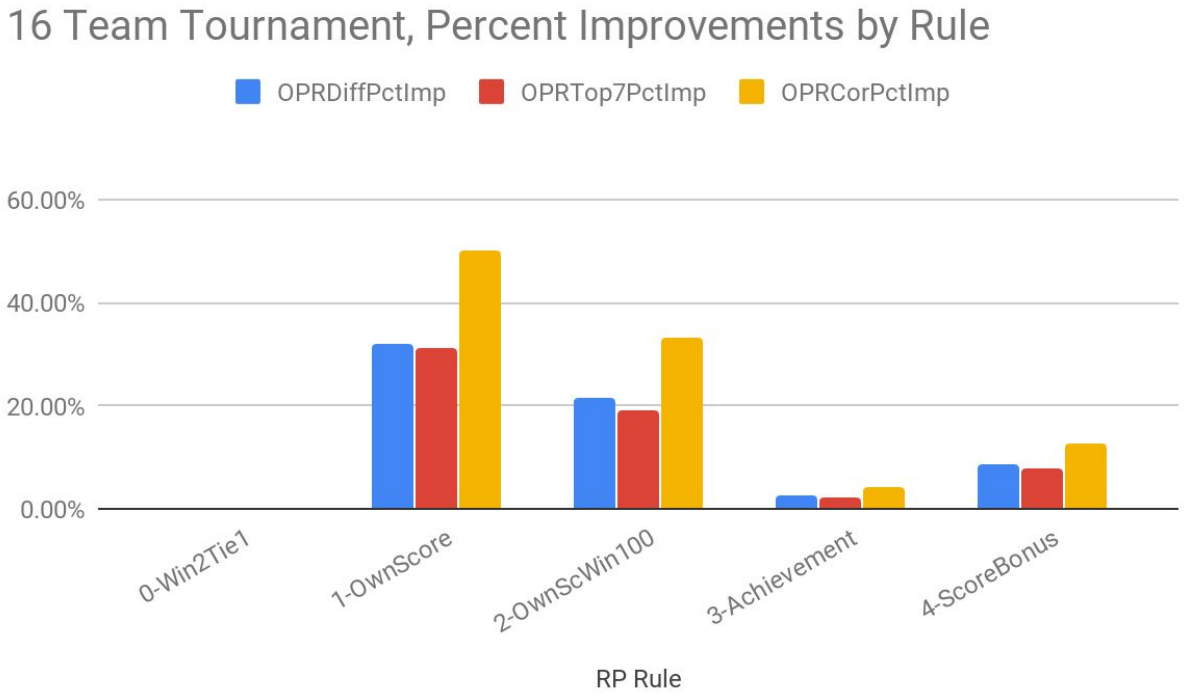
RP Results

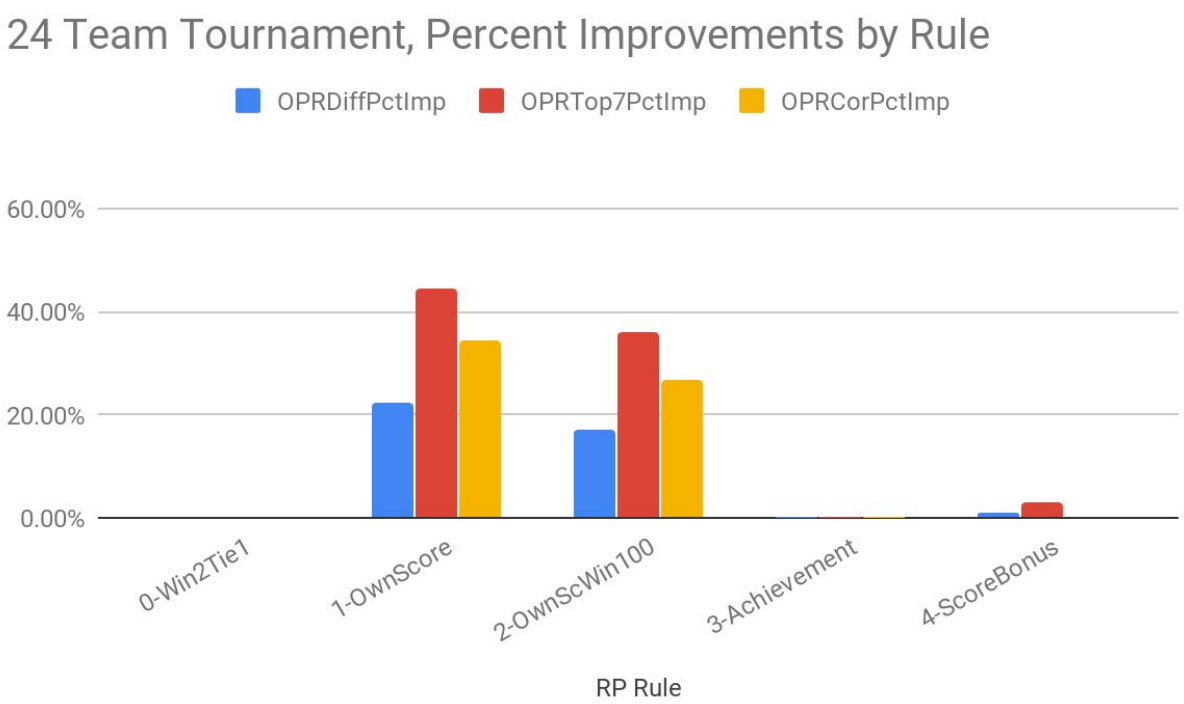
Title	Teams	OPRTopX	OPRTop7 PctImp	OPRDif	OPRDiff PctImp	OPRCor	OPRCor PctImp
0-Win2Tie1	12	5.47	0.00%	1.86	0.00%	0.73	0.00%
1-OwnScore	12	5.85	24.84%	1.36	26.88%	0.83	37.04%
2-OwnScWin100	12	5.61	9.15%	1.64	11.83%	0.77	14.81%
3-Achievement	12	5.51	2.61%	1.87	-0.54%	0.72	-3.70%
4-ScoreBonus	12	5.5	1.96%	1.85	0.54%	0.73	0.00%
0-Win2Tie1	16	5.23	0.00%	2.43	0.00%	0.76	0.00%
1-OwnScore	16	5.78	31.07%	1.65	32.10%	0.88	50.00%
2-OwnScWin100	16	5.57	19.21%	1.91	21.40%	0.84	33.33%
3-Achievement	16	5.27	2.26%	2.37	2.47%	0.77	4.17%
4-ScoreBonus	16	5.37	7.91%	2.22	8.64%	0.79	12.50%
0-Win2Tie1	24	5.36	0.00%	3.78	0.00%	0.74	0.00%
1-OwnScore	24	6.09	44.51%	2.94	22.22%	0.83	34.62%
2-OwnScWin100	24	5.95	35.98%	3.13	17.20%	0.81	26.92%
3-Achievement	24	5.35	-0.61%	3.84	-1.59%	0.73	-3.85%
4-ScoreBonus	24	5.41	3.05%	3.74	1.06%	0.74	0.00%
0-Win2Tie1	32	4.59	0.00%	4.75	0.00%	0.77	0.00%
1-OwnScore	32	5.55	39.83%	3.11	34.53%	0.9	56.52%
2-OwnScWin100	32	5.27	28.22%	3.76	20.84%	0.85	34.78%
3-Achievement	32	4.66	2.90%	4.74	0.21%	0.77	0.00%
4-ScoreBonus	32	4.81	9.13%	4.63	2.53%	0.78	4.35%
0-Win2Tie1	64	3.96	0.00%	8.08	0.00%	0.83	0.00%
1-OwnScore	64	5.43	48.36%	4.79	40.72%	0.94	64.71%
2-OwnScWin100	64	5.24	42.11%	5.19	35.77%	0.93	58.82%
3-Achievement	64	4.32	11.84%	7.21	10.77%	0.87	23.53%
4-ScoreBonus	64	4.5	17.76%	6.43	20.42%	0.89	35.29%
0-Win2Tie1	80	4.5	0.00%	10.18	0.00%	0.83	0.00%
1-OwnScore	80	5.9	56.00%	6.2	39.10%	0.93	58.82%
2-OwnScWin100	80	5.82	52.80%	6.79	33.30%	0.92	52.94%

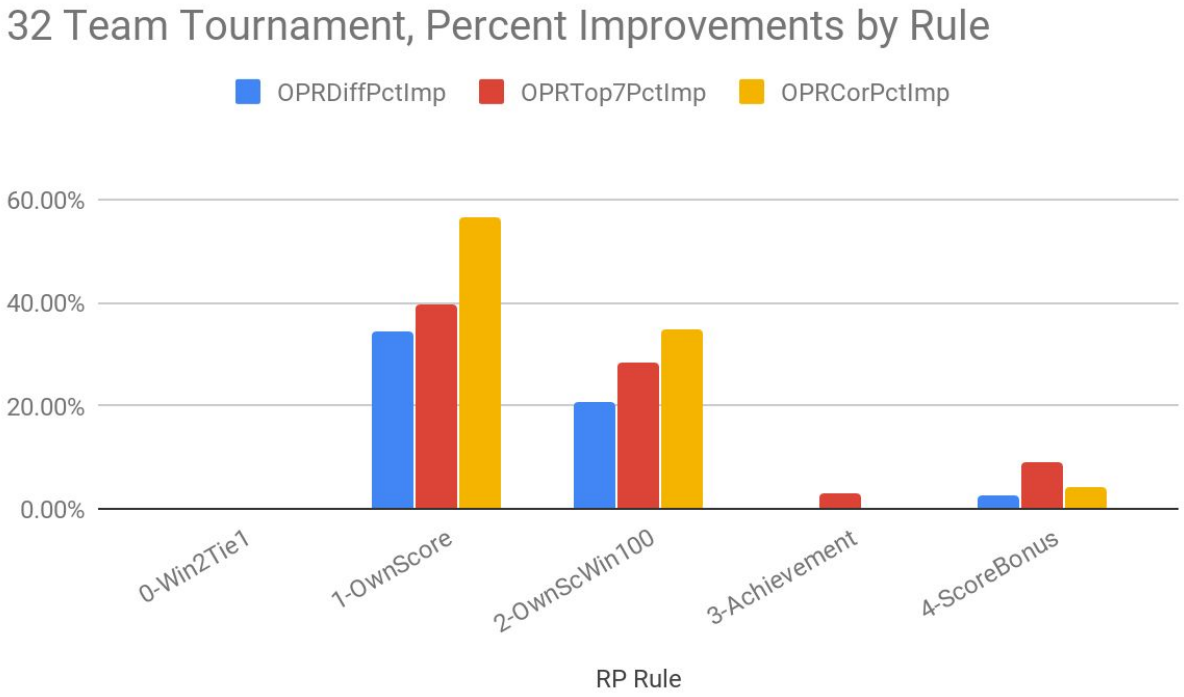
3-Achievement	80	4.96	18.40%	9.28	8.84%	0.86	17.65%
4-ScoreBonus	80	5.46	38.40%	8.72	14.34%	0.87	23.53%

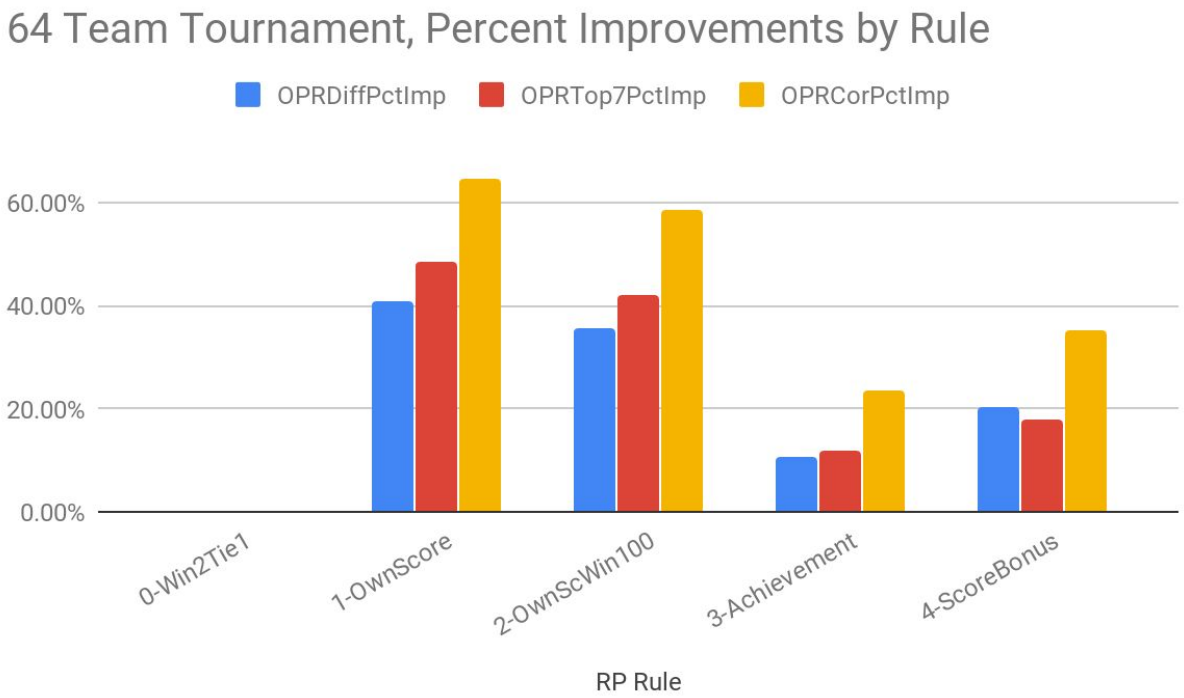
Now we can visually compare Percent Improvement figures with a series char, one for each event size:

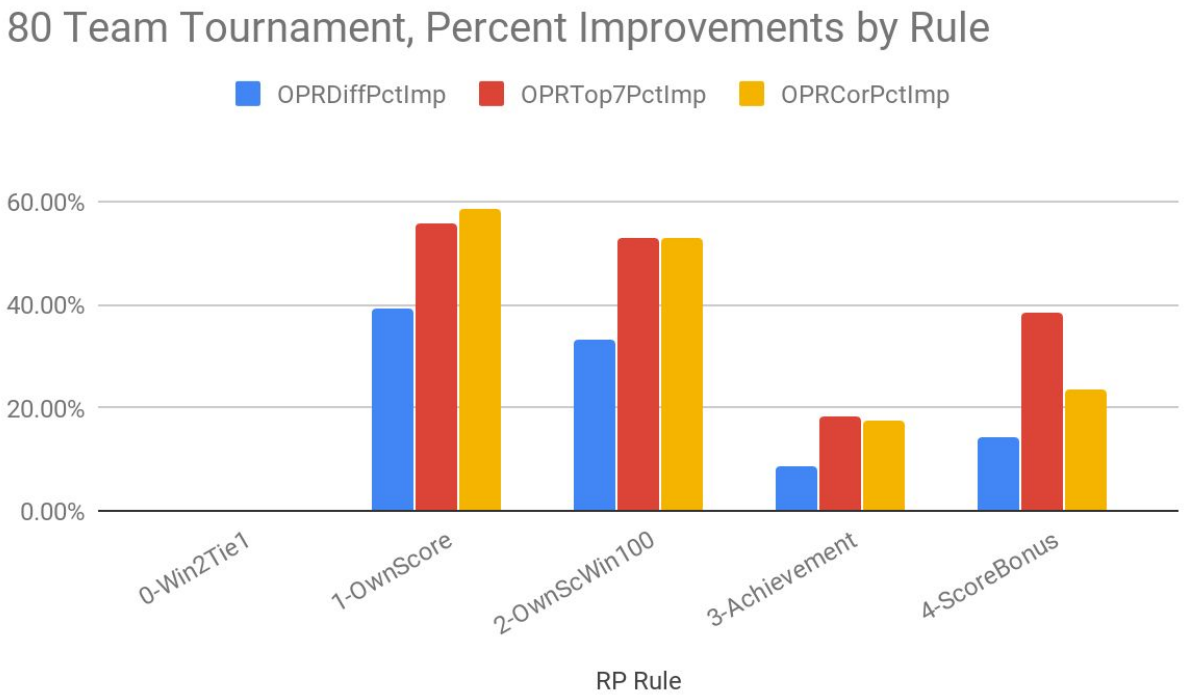


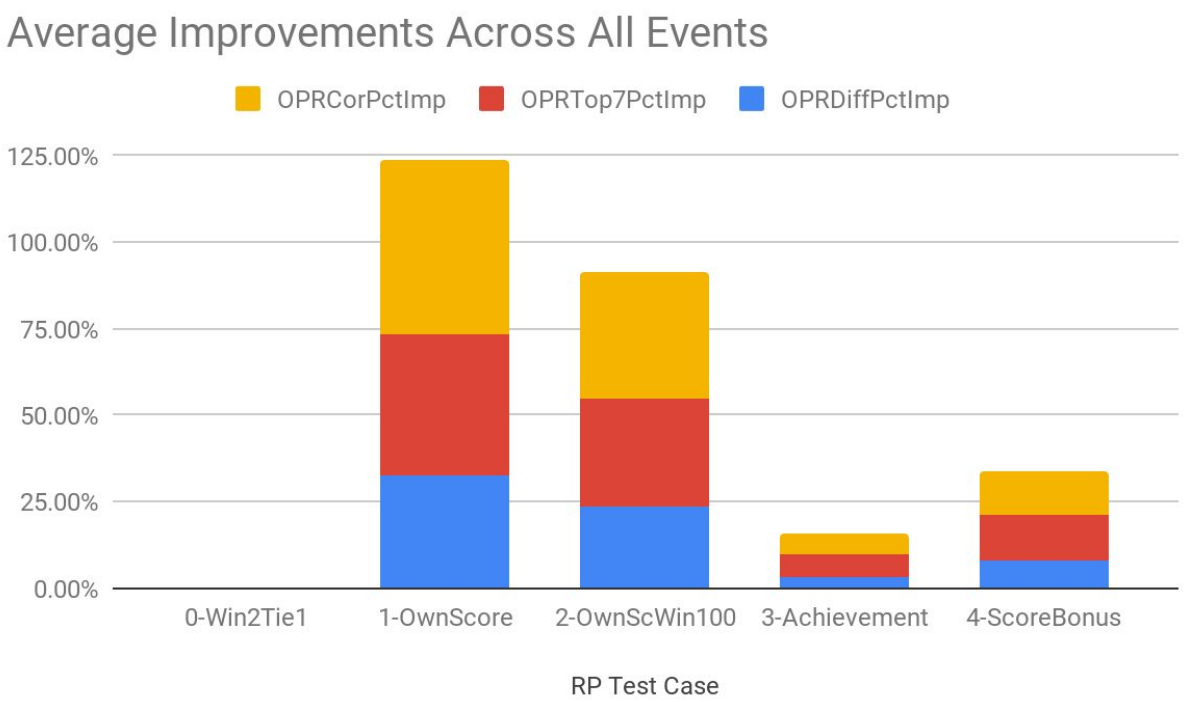

16 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
60.00%
40.00%
20.00%
0.00%
0-Win2Tie1
1-OwnScore
2-OwnScWin100
3-Achievement
4-ScoreBonus
RP Rule


24 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
60.00%
40.00%
20.00%
0.00%
0-Win2Tie1
1-OwnScore
2-OwnScWin100
3-Achievement
4-ScoreBonus
RP Rule


32 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
60.00%
40.00%
20.00%
0.00%
0-Win2Tie1
1-OwnScore
2-OwnScWin100
3-Achievement
4-ScoreBonus
RP Rule


64 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
60.00%
40.00%
20.00%
0.00%
0-Win2Tie1
1-OwnScore
2-OwnScWin100
3-Achievement
4-ScoreBonus
RP Rule


80 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
60.00%
40.00%
20.00%
0.00%
0-Win2Tie1
1-OwnScore
2-OwnScWin100
3-Achievement
4-ScoreBonus
RP Rule


Average Improvements Across All Events
OPRCorPctImp
OPRTop7PctImp
OPRDiffPctImp
125.00%
100.00%
75.00%
50.00%
25.00%
0.00%
0-Win2Tie1
1-OwnScore
2-OwnScWin100
3-Achievement
4-ScoreBonus
RP Test Case

Observations

The two RP rules based on score, OwnScore and OwnScore + Win Bonus, showed dramatic improvements in all measurements.

It’s interesting that all RP rule measurements improve generally as the tournament size grows. This may be because there are fewer ties due to the wider range of RP points won by teams. Teams spread out more across RP.

Part V: Scheduling Improvements

Rationale for Changing Scheduling

When RP is based on wins, a team's schedule has as much to do with their final ranking as the team's performance. When someone asks me if I think we'll advance from this tournament, my reply is always, "We'll have to be lucky AND good." I'm wishing for luck on the random draw, in that we draw many winnable matches with high scores on both sides, plus good performance to win those matches.

If we draw many winnable matches against much weaker teams, our TBP will suffer. We end up with high RP, but last in the RP group, where we might miss an alliance captainship.

If we draw matches against the top teams, partnered with much weaker teams, RP suffers and fall further down the ranks.

In other sports and gaming tournaments where the number of matches is small relative to the number of players, many have chosen Swiss-System Scheduling as a means to accurately rank players without having to resort to an elimination or round-robin format.

We'll get into the details below, but benefits include:

- Breaks ties with RP
- Maintains win/loss/tie meaning of RP
- Reduces number of undefeated teams
- Teams of similar rank will face off as opponents
- May be familiar to those involved in other Swiss-based leagues

Drawbacks from Changing Scheduling

This is a pretty dramatic change to the way FTC currently schedules tournaments, where the matches are set early in the day, and not changed once set.

Swiss schedules are dynamic. A team will only know their next match partner and opponents because each round is scheduled based on results of the previous round. When Round 1 is complete, Round 2's schedule is built based on RP and TBP.

This could introduce noticeable delays in the flow. There would be a pause in play while the new schedule is generated and teams prepare with their partners.

The delay can be mitigated by scheduling further ahead, and I will propose details on an FTC-specific modification to Swiss scheduling in the pages to follow. We'll see how this system performs against straight Swiss scheduling.

Finally, Swiss scheduling is best done by computer because it's fairly complicated. A new scheduling program would have to be developed and integrated with the scoring software. Fortunately, FTC EventSim has the algorithms built-in, and is open source. This code is freely available for integration into whatever FTC might need.

Proposed Scheduling Fixes

The Swiss System

I highly recommend starting with Wikipedia's article on Swiss-System Scheduling. There are different variations, some of which are supported by FTC EventSim. Almost all variations include ranking players by wins, then by tie-breakers, which fits nicely with RP and TBP we're familiar with. In every case, though, Swiss scheduling is designed for 1v1 play, so the 2v2 format presents an extra challenge, but not insurmountable.

The basic idea is to schedule teams as opponents who have the same RP values. Play is divided into rounds, where each team competes once, and after each round, the field is divided into groups by RP and new matchups are created between those teams.

To get to the 2v2 format, first we find each team their one ideal opponent. From that list of 1v1 pairs, we match them up with another 1v1 pair to end up with 2v2. Each team in the match is facing their ideal opponent.

Folding

Starting with 1v1 matchups, we have a couple of ways of finding a team's opponent with their RP group. One technique is to "fold" the RP group. Say, for example, you have a group of 4 teams who are undefeated, and by tiebreakers are ranked 1-4.

1
2
3
4

To fold this group, it's as if you've written these on a piece of paper and fold it in the middle. 1 will meet 4 and 2 will meet 3.

1v4
2v3

Sliding

The other way to approach the 1v1 matchups is to “slide” the teams. Instead of folding the paper, you cut it in half and slide the bottom to the top, so…

1
2
3
4

...goes to…

1v3
2v4

This can be done with every group, any size, as long as there is an even number of teams in that group. If the number is odd, just take the top team from the next group and make it the last team of the group you’re folding or sliding.

Things Get Complicated

We’d probably like to avoid having teams oppose each other twice in the same tournament. Sometimes a fold or slide gives us that situation. To remedy that, we match a team up with either the next neighbor up or down the list, and find a similar suitable match for the the team left behind. Algorithmically, this gets very complicated, and sometimes a group is so small you must reach into the next group to find a match.

Fortunately, there is a straightforward way to solve this using graph theory. You populate a graph where each team is a node, or “vertex” on the graph, and each potential matchup is an “edge”. You assign “costs” to the edges based on your preference for that matchup, and send the graph to a “minimum cost perfect match” graph solver. I decided to implement this approach after reading Swiss Tournament Scheduling: Leaguevine's New Algorithm.

It works very efficiently, returning a list of matchups in the most “ideal” or optimal way such that the fold or slide is preserved with as few deviations as possible to maintain our other constraints.

The code I’ve implemented is all open-source, and free to use however FTC sees fit.

2v2

So now that we have 1v1 matchups optimally paired, we can treat each of these pairs as vertices on a new graph and match them up with other 1v1 pairs. Through much

experimentation, I've found that matches are more balanced if we simply divide the whole list of 1v1's in half and slide the bottom half to the top to get 2v2.

In a 12 team example after one round we'd have 6 teams of RP=2, and 6 teams of RP=0.

Rank	RP
1	2
2	2
3	2
4	2
5	2
6	2
7	0
8	0
9	0
10	0
11	0
12	0

We would fold each of these 6 team groups to get:

Red	Blue
1	6
2	5
3	4

Red	Blue
7	12
8	11
9	10

We would then divide this list of 6 pairs in half, and slide the bottom group to the top to get 2v2 matchups like this:

Red	Blue
1 & 7	6 & 12
2 & 8	5 & 11
3 & 9	4 & 10

Scheduling The First Round

To be effective, Swiss scheduling depends on ordered lists. If at all possible, the first round should be scheduled from a seed-ordered list of teams. I would propose using TBP and RP values from feeder tournaments in the case of Regional and World Championships, and perhaps average or most-recent records of teams participating in leagues and local tournaments.

If that's not possible, the first round can be scheduled randomly, which will then provide the ordering for Round 2. In this case, you lose the advantage of the sorting power of Swiss for the first round and counting on the following rounds to make it up. This effect can be studied with the simulator, but is out of scope for this paper. For our simulations, I will be seeding the first rounds based on team OPR, since that is readily available in our source data.

The first round is a special case for another reason: Teams are all tied at 0 RP and TBP. Through experimentation, and also to keep first-round matches somewhat balanced, I decided to "slide" the first round opponents. This way the top-seeded team will oppose a middle-seeded team in the first round.

These are all configurable options in EventSim, but to limit the number of test cases, We'll be seeding and sliding teams in the first round.

Multiples of 4

In a 2v2 format, for each team to play once per round, there has to be a multiple of 4 teams. In real life, this isn't always the case, so we need a way to work with a numbers like 21, which is the worst case where we're three shy of the next multiple of 4.

FTC currently uses the surrogate system to fill in where odd numbers leave gaps. We could do this as well for Swiss. In the case of 21 teams, however, we'd need three teams to fill in the gaps *every round*. In a 5 round event, that would be 3 * 5 = 15 surrogates needed to complete all rounds. As long as no team serves as surrogate twice, that could be workable, but if there were only 13 teams at an event, it wouldn't be possible.

Other Swiss tournaments commonly use byes. Teams are given 2 RP if they didn't compete in that round. If this were adopted, I would propose choosing the top team in a given group for the bye, and a team can only receive one bye in a tournament. For a meet of 11 teams, this wouldn't be possible since you'd have three byes per round for a total of 15 byes for 11 teams.

A hybrid approach could implement byes if the number of teams is one more than a multiple of 4 and surrogates if one less, thus minimizing the number of byes and surrogates. It could go either way for two more than a multiple of 4.

Another possibility would be to have the last match of the round treated as a sort of “remainder” match, where the gaps are filled by team who have already played in that round, but whose scores will count as their next-round score and would not compete in the next round. To fill the very last remainder match, you would use surrogates whose results would not count.

As of this writing, the simulator does not support any of these approaches, and will not be simulated for the purposes of this paper. I merely point these approaches out for discussion should FTC choose to adopt this system.

FTC-Modified Swiss Scheduling

To avoid delays between rounds, I propose scheduling ahead. Schedule the first two rounds from the seed-ordered roster (or randomly if required). Then, as soon as Round 1 completes, begin Round 2, and at the same time, create the schedule for Round 3. Announce that the Round 3 schedule is available on the big screen.

Likewise, as soon as Round 2 completes, start Round 3 and schedule Round 4.

For multi-day tournaments, we have additional opportunities to schedule ahead. For example, say the schedule goes:

Day 1 - Rounds 1 & 2
Day 2 - Rounds 3 - 7
Day 3 - Rounds 8 & 9

Rounds 1 & 2 scheduled before the start.
Rounds 3 & 4 scheduled overnight based on results of Day 1.
Round 5 scheduled while Round 3 is played, Round 6 during Round 4, etc.
Round 8 & 9 scheduled from Day 2 results.

The simulator supports both simple round-ahead and multi-day scheduling and will be included in the test cases.

Testing Scheduling Fixes By Simulating Events

I will test both straight and FTC-Modified Swiss scheduling. For each type, I will vary the opponent matching methods (fold or slide) for rounds after round 1. Round 1 will always be slide, and based on a starting seed value from the team’s established OPR value.

With these combinations I will also vary the TBP rule with Losing Score (existing system) Own Score, and Own Score + Losing Score.

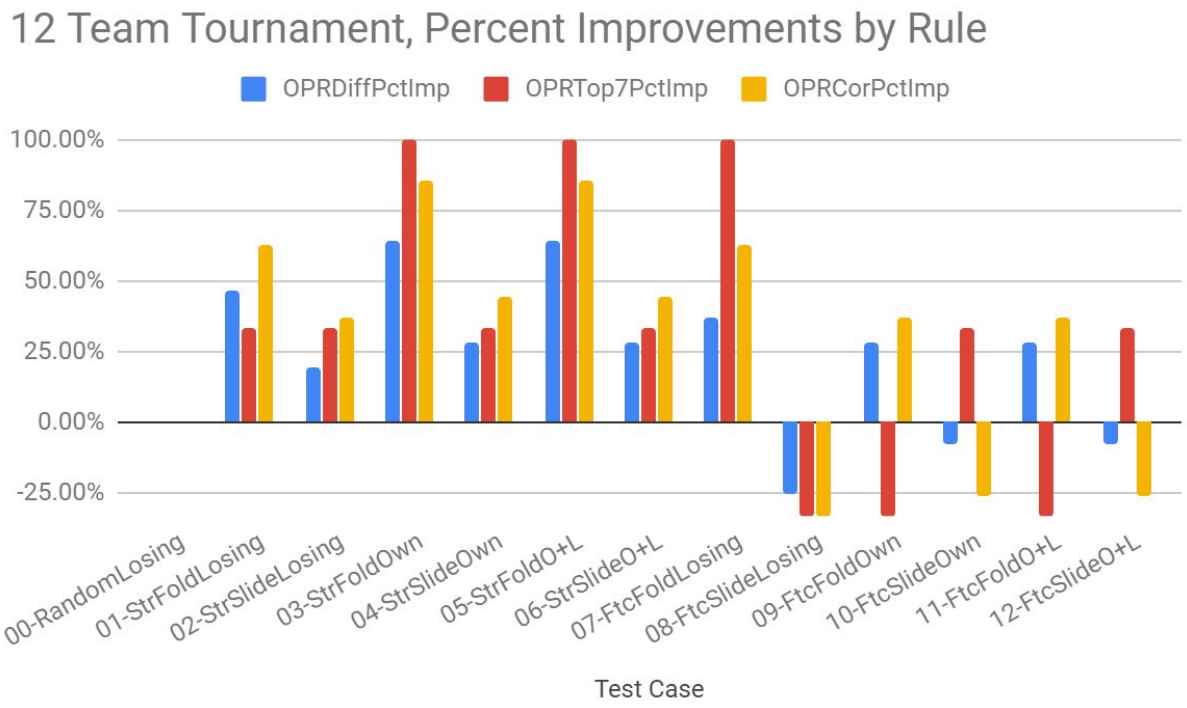
I will run all simulations for all previously-tested tournament sizes.

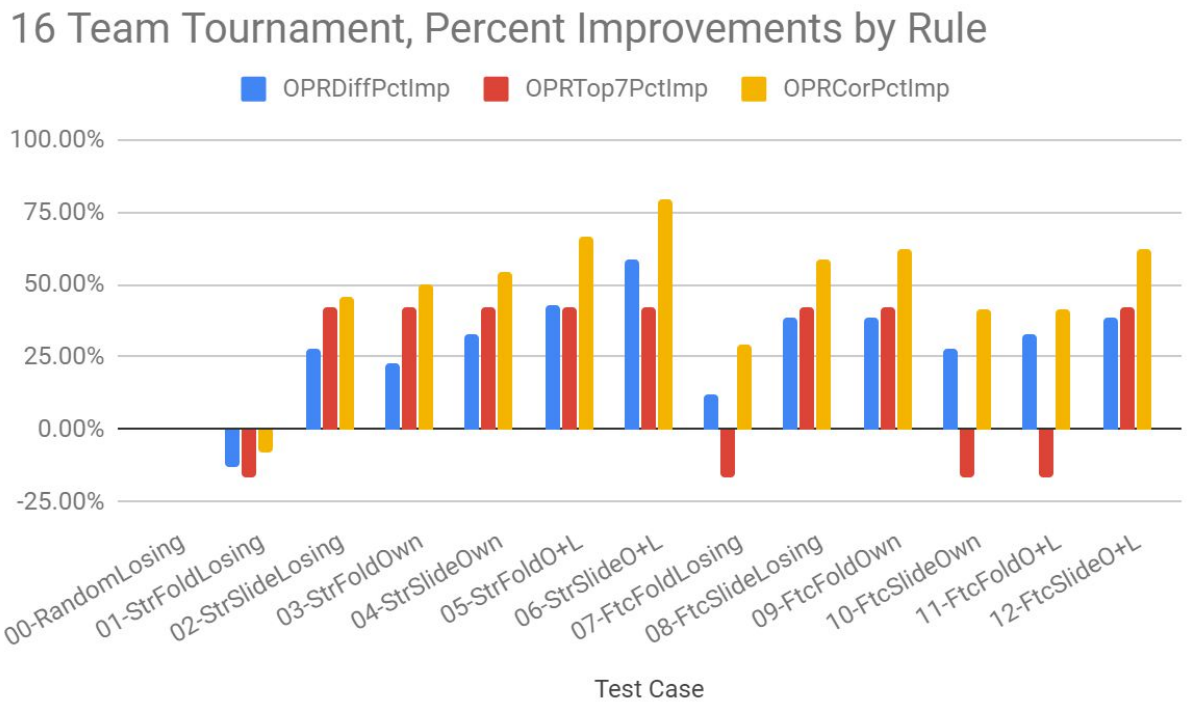
That totals 72 combinations: 1 Randomly scheduled control, 2 Swiss systems (straight and FTC-modified), 2 matching methods, 3 TBP rules, and 6 tournament sizes. At 1000 trials per test case, that's 78,000 simulations!

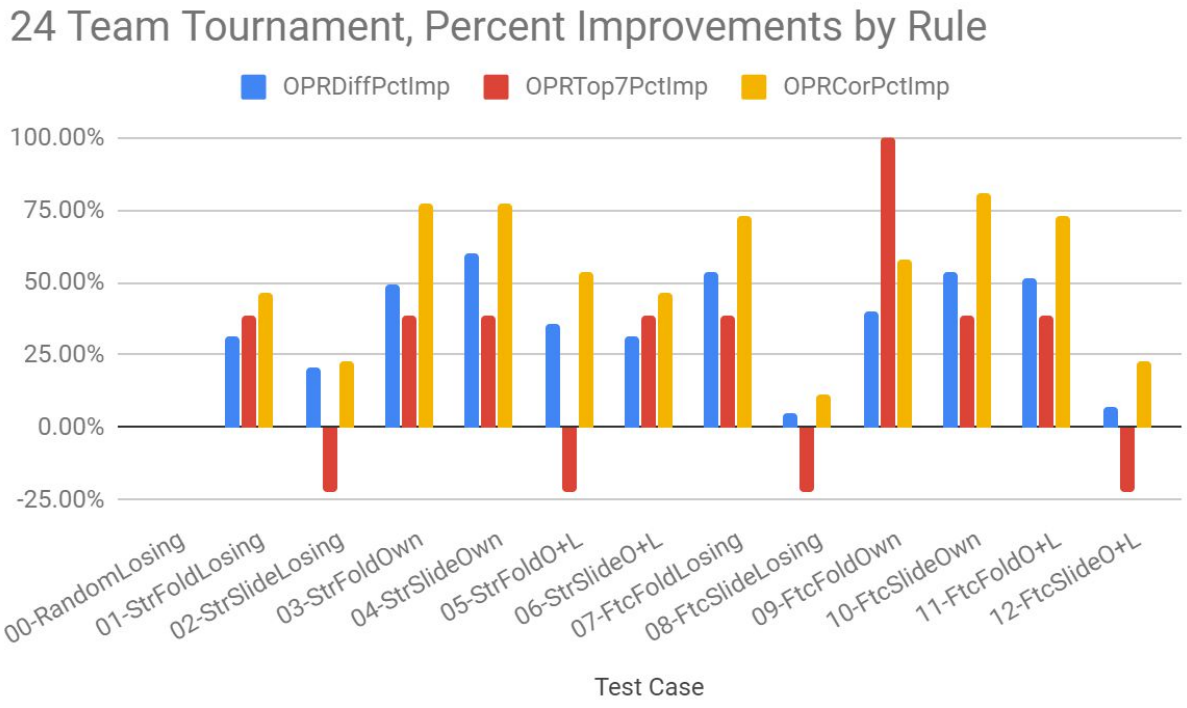
We'll be measuring the same values as the previous test, and charting them the same way.

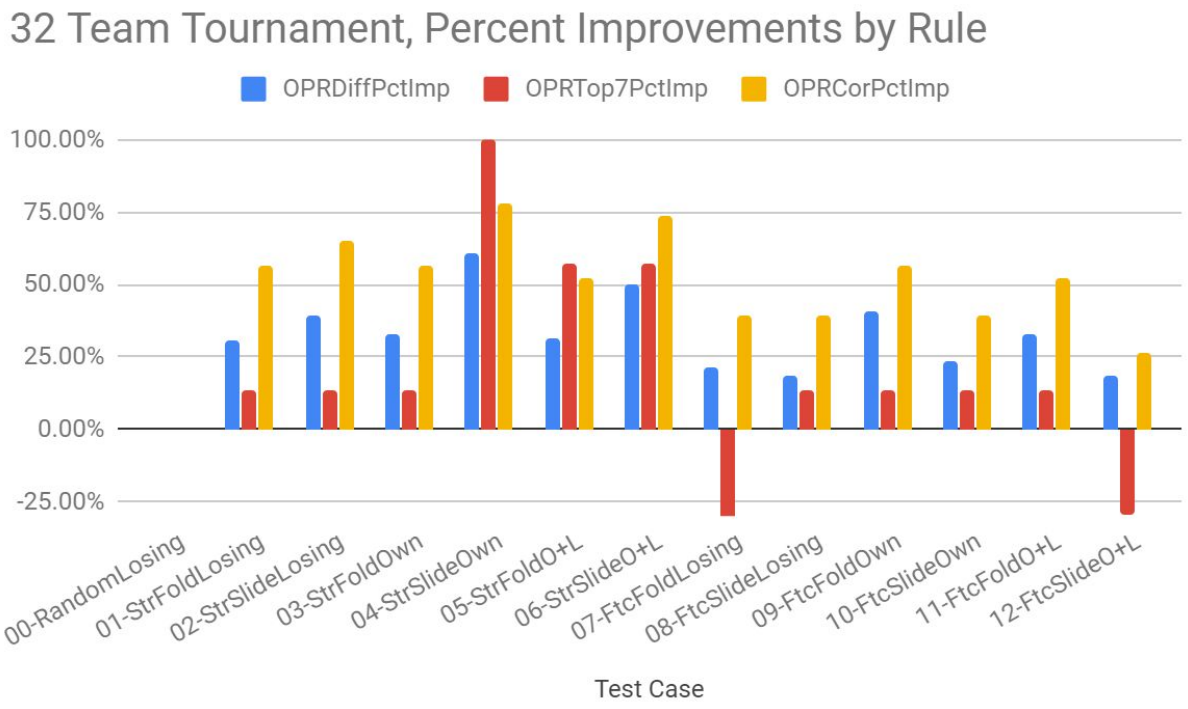
Test cases, results, and charts can be found on the Google sheet, [Ranking Study - Swiss Test Cases](#).

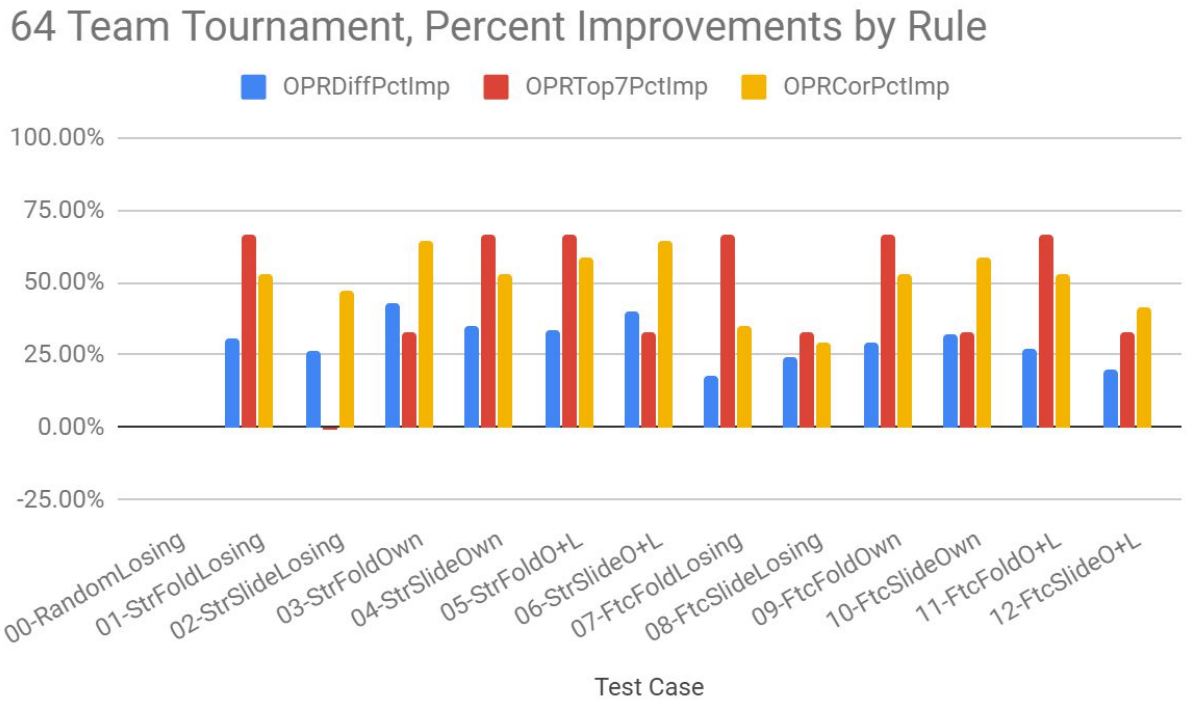
Swiss-system Results

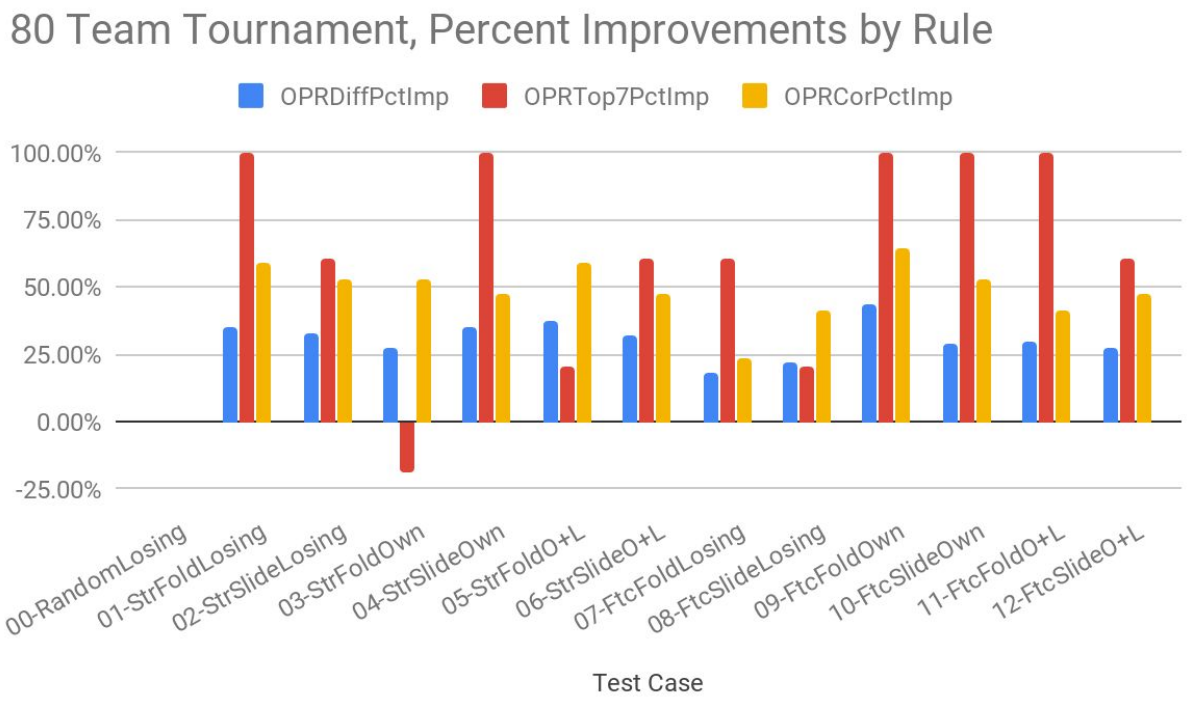


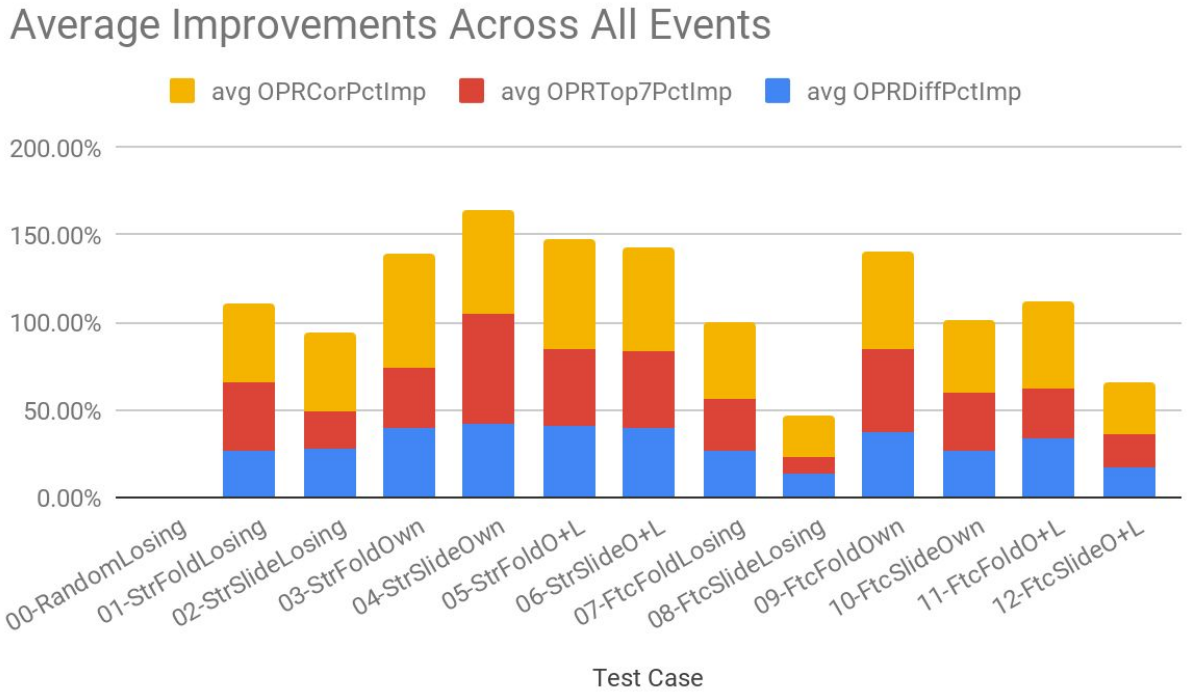

16 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
100.00%
75.00%
50.00%
25.00%
0.00%
-25.00%
00-RandomLosing
01-StrFoldLosing
02-StrSlideLosing
03-StrFoldOwn
04-StrSlideOwn
05-StrFoldO+L
06-StrSlideO+L
07-FtcFoldLosing
08-FtcSlideLosing
09-FtcFoldOwn
10-FtcSlideOwn
11-FtcFoldO+L
12-FtcSlideO+L
Test Case


24 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
100.00%
75.00%
50.00%
25.00%
0.00%
-25.00%
00-RandomLosing
01-StrFoldLosing
02-StrSlideLosing
03-StrFoldOwn
04-StrSlideOwn
05-StrFoldO+L
06-StrSlideO+L
07-FtcFoldLosing
08-FtcSlideLosing
09-FtcFoldOwn
10-FtcSlideOwn
11-FtcFoldO+L
12-FtcSlideO+L
Test Case


32 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
100.00%
75.00%
50.00%
25.00%
0.00%
-25.00%
00-RandomLosing
01-StrFoldLosing
02-StrSlideLosing
03-StrFoldOwn
04-StrSlideOwn
05-StrFoldO+L
06-StrSlideO+L
07-FtcFoldLosing
08-FtcSlideLosing
09-FtcFoldOwn
10-FtcSlideOwn
11-FtcFoldO+L
12-FtcSlideO+L
Test Case


64 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
100.00%
75.00%
50.00%
25.00%
0.00%
-25.00%
00-RandomLosing
01-StrFoldLosing
02-StrSlideLosing
03-StrFoldOwn
04-StrSlideOwn
05-StrFoldO+L
06-StrSlideO+L
07-FtcFoldLosing
08-FtcSlideLosing
09-FtcFoldOwn
10-FtcSlideOwn
11-FtcFoldO+L
12-FtcSlideO+L
Test Case


80 Team Tournament, Percent Improvements by Rule
OPRDiffPctImp
OPRTop7PctImp
OPRCorPctImp
100.00%
75.00%
50.00%
25.00%
0.00%
-25.00%
00-RandomLosing
01-StrFoldLosing
02-StrSlideLosing
03-StrFoldOwn
04-StrSlideOwn
05-StrFoldO+L
06-StrSlideO+L
07-FtcFoldLosing
08-FtcSlideLosing
09-FtcFoldOwn
10-FtcSlideOwn
11-FtcFoldO+L
12-FtcSlideO+L
Test Case


Average Improvements Across All Events
avg OPRCorPctImp
avg OPRTop7PctImp
avg OPRDiffPctImp
200.00%
150.00%
100.00%
50.00%
0.00%
00-RandomLosing
01-StrFoldLosing
02-StrSlideLosing
03-StrFoldOwn
04-StrSlideOwn
05-StrFoldO+L
06-StrSlideO+L
07-FtcFoldLosing
08-FtcSlideLosing
09-FtcFoldOwn
10-FtcSlideOwn
11-FtcFoldO+L
12-FtcSlideO+L
Test Case

Observations

Both straight and modified Swiss systems offer dramatic improvements in all metrics. Both perform nearly equally in the largest tournaments. Straight Swiss has the advantage in the smaller tournaments, where both see quite a variance of results in The OPRTop7 measurement. In smaller tournaments, measurements for modified Swiss are hit and miss.

Note that in 12 and 16 team events, we are most likely talking about league meets where Top7 Ranking doesn't matter as much as overall ranking since there is no alliance selection.

Overall, for straight Swiss, sliding opponents and using a teams own score for TBP edges out the others, but there are close contenders. For modified Swiss, folding for opponents with OwnScore TBP seems to work best, but sliding works two for Own+Losing TBP schemes.

Part VI: Conclusion

Models Ranked

From the data we've collected and shown, we can rank the proposals within each test group by the average overall improvement. Here, I've averaged the three metrics' Pct Improvement values for ranking and comparison.

Changing TBP only, ranked by overall improvement

1	5.15%	1-OwnScore
2	3.65%	3-OwnPlusLosing
3	1.91%	2-TotalScore
4	1.17%	4-TopHalfOwn
5	0.00%	0-LosingScore
6	-3.49%	5-HighestScore

Changing RP only, ranked by overall improvement

1	41.21%	1-OwnScore
2	30.52%	2-OwnScWin100
3	11.19%	4-ScoreBonus
4	5.30%	3-Achievement
5	0.00%	0-Win2Tie1

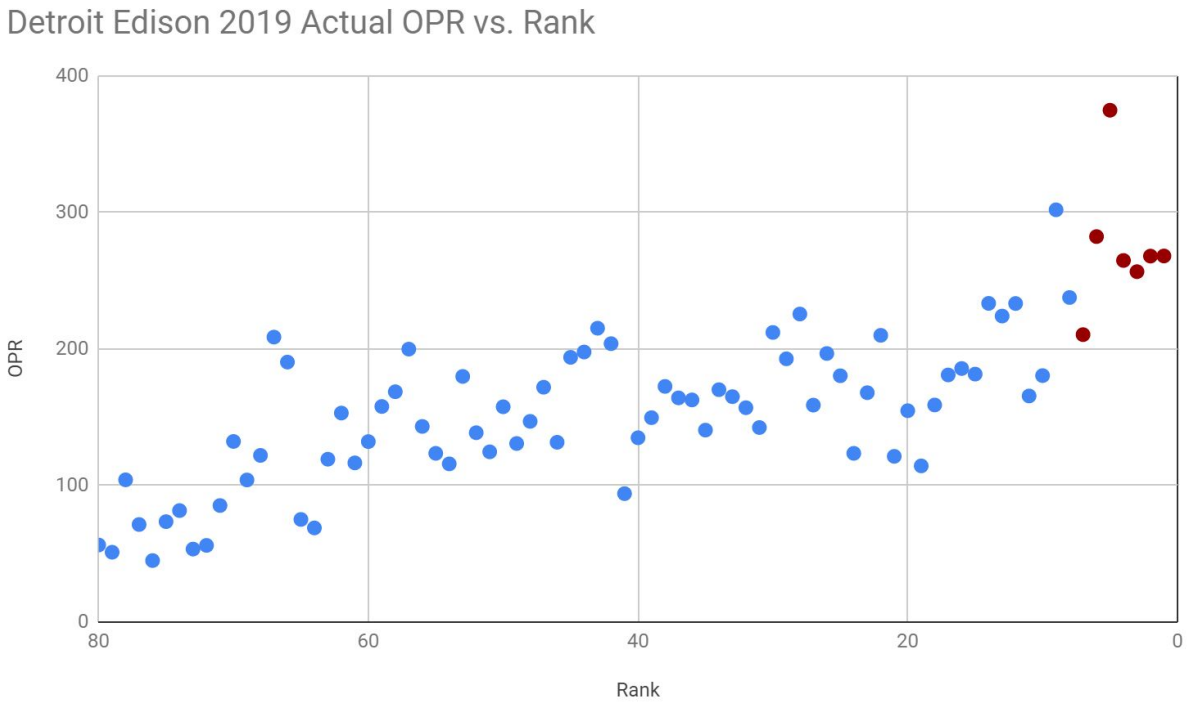
Changing Scheduling & RP, ranked by overall improvement

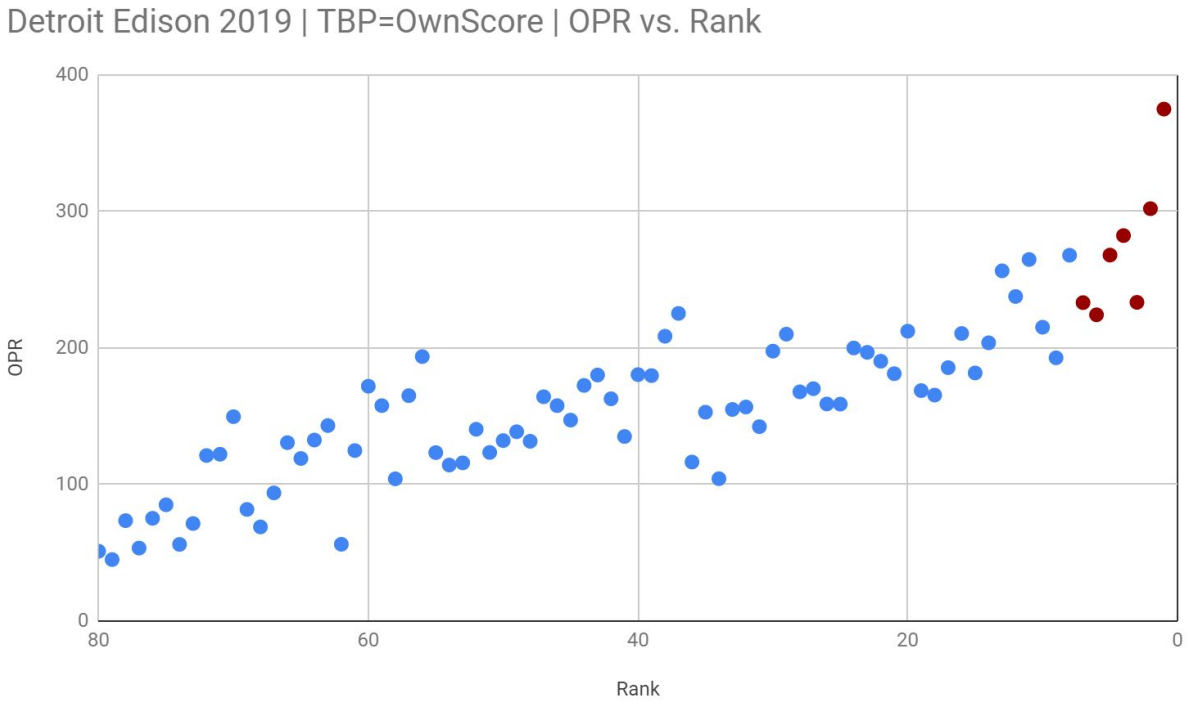
1	54.81%	04-StrSlideOwn
2	49.13%	05-StrFoldO+L
3	47.78%	06-StrSlideO+L
4	46.75%	09-FtcFoldOwn
5	46.32%	03-StrFoldOwn
6	37.19%	11-FtcFoldO+L
7	36.97%	01-StrFoldLosing
8	33.80%	10-FtcSlideOwn
9	33.30%	07-FtcFoldLosing

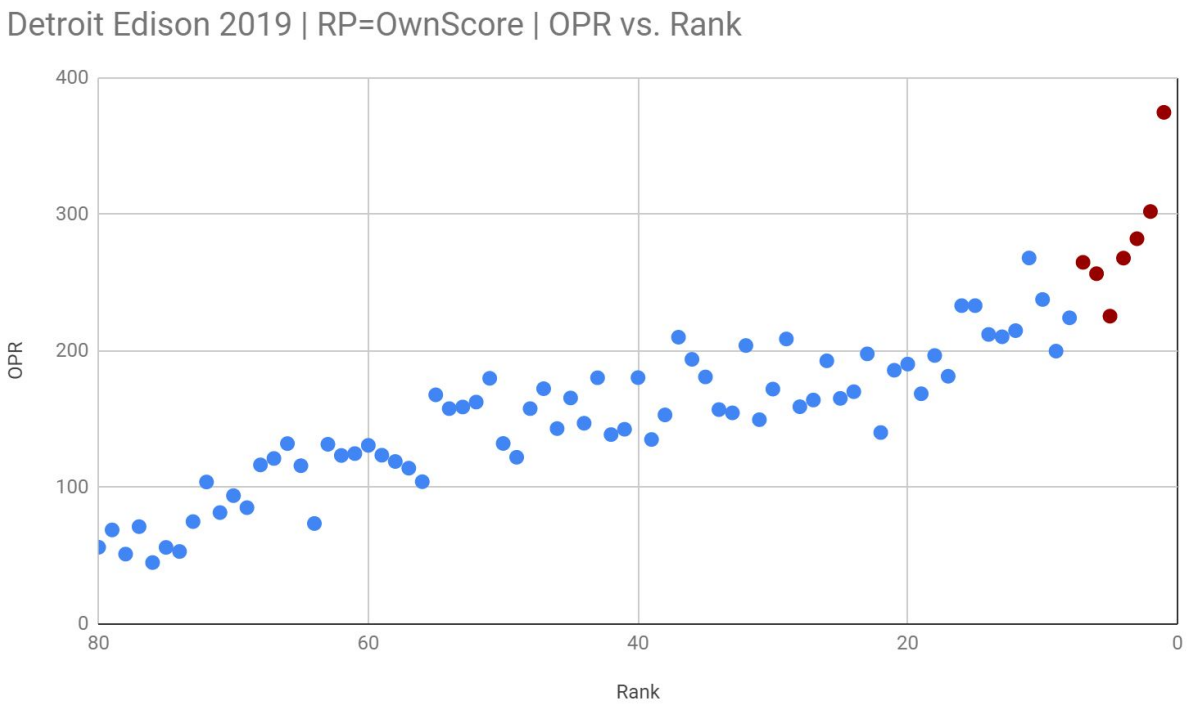
10	31.28%	02-StrSlideLosing
11	21.86%	12-FtcSlideO+L
12	15.66%	08-FtcSlideLosing
13	0.00%	00-RandomLosing

Best Models Compared To Actual

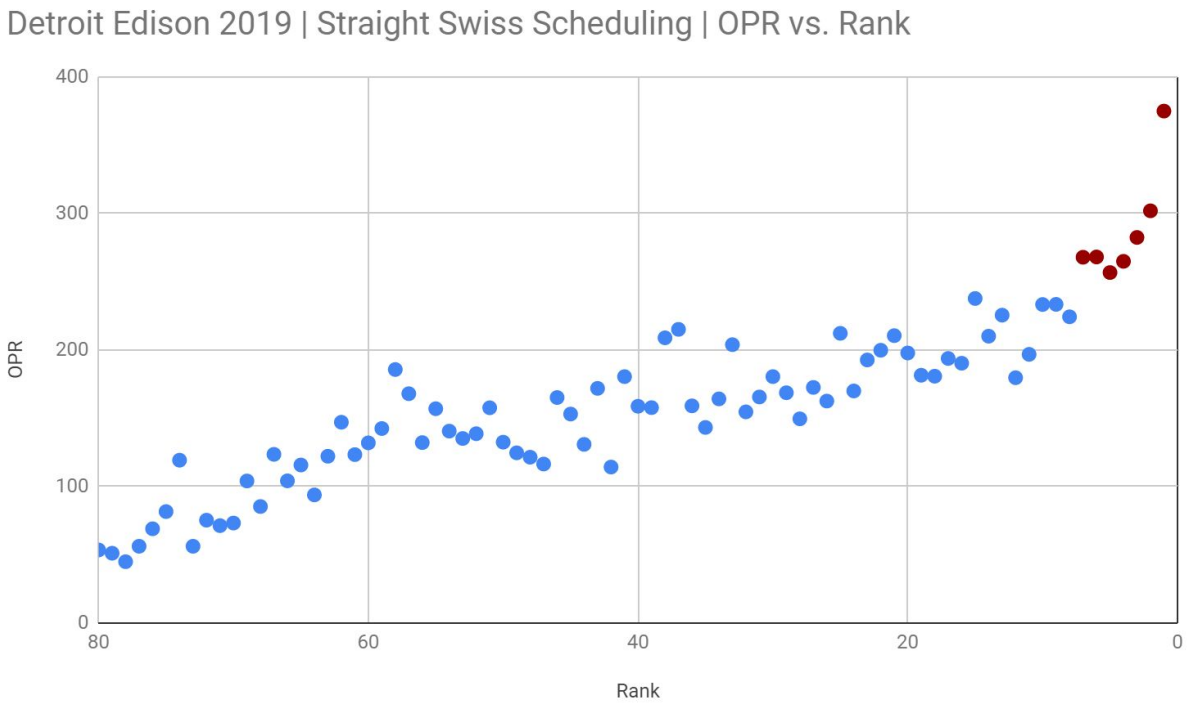
Finally, let's look at OPR vs Rank charts for Detroit Edison 2019, comparing actual results to typical a sample from the best candidate from each system. For Swiss scheduling, I've included the best FTC-modified as well. Sample runs, results and charts can be found on the Google Sheet, Best Model Comparison. The Top 7 ranked teams are highlighted to better discern their placement relative to performance.

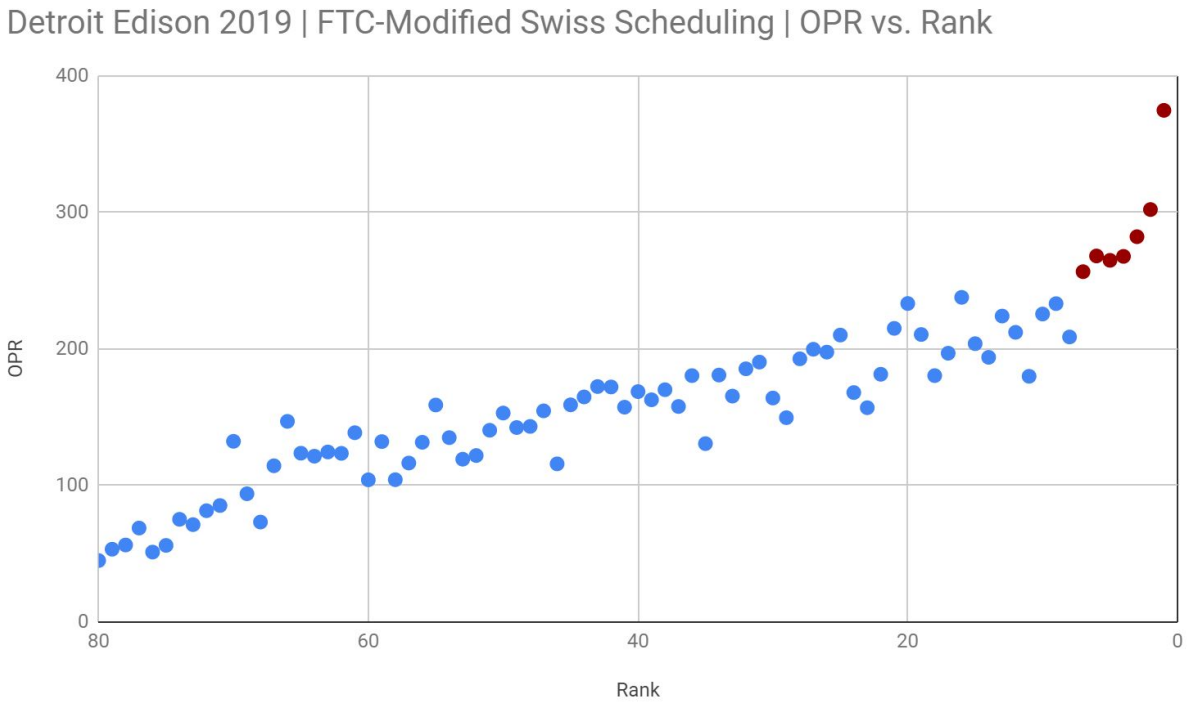



Detroit Edison 2019 | TBP=OwnScore | OPR vs. Rank
400
300
200
100
0
OPR
80
60
40
20
0
Rank


Detroit Edison 2019 | RP=OwnScore | OPR vs. Rank
400
300
200
100
0
OPR
80
60
40
20
0
Rank

.


Detroit Edison 2019 | Straight Swiss Scheduling | OPR vs. Rank
400
300
200
100
0
OPR
80
60
40
20
0
Rank


Detroit Edison 2019 | FTC-Modified Swiss Scheduling | OPR vs. Rank
400
300
200
100
0
OPR
80
60
40
20
0
Rank

Observations

Even the best TBP-only rule change does not help the amount of overall spreading (OPR Correlation) in these charts, but it does fix problems in the top 2.

Changing RP to OwnScore does improve things significantly, straightening out the top 4 nicely, but the rest of the chart is fairly well spread out with some improvement to overall OPR Correlation.

Both Swiss scheduling models tighten things up nicely across the board, especially at the top ranks, and surprisingly, the FTC-Modified system does a very nice job with this tournament lining teams up from end-to-end with the smallest amount of spreading.

Proposed Short-Term Changes

As discussed earlier, Swiss scheduling presents a significant challenge to FTC in its implementation. Scheduling would be done as the tournament progresses and would require scheduling software to be integrated with the scoring system. I doubt there is time left in 2019 to implement these changes before the 2019-2020 season.

Changing the RP or TBP rules in the scoring system, however, might be manageable. Of these rule changes, changing to RP=OwnScore yields the best results overall, but this has disadvantage of taking wins and losses out of the ranking equation. Adding a 100 point RP bonus for wins would help with this, and performs almost as well. This may come down to a philosophical discussion around the importance of wins and losses in FTC ranking.

The next best option would be to change TBP only, and change it to the Alliances own pre-penalty score. There is some improvement across the board, most noticeable in the very top ranks. I would recommend this option if the other options above are not palatable for philosophical or logistical reasons.

Proposed Long-Term Changes

If FTC is willing to move to a dynamic scheduling model, I have presented an FTC-specific modification to Swiss-system scheduling that has proven to have superior results. I propose FTC begin to explore the Swiss system. I am personally offering my technical assistance in doing so. I have not tackled the non-multiple-of-four challenges yet, and there may be issues I have not considered that are outside the scope of this paper. I hope FTC will engage the community to work toward meaningful changes to the scheduling and ranking system.

Acknowledgments

I wish to extend my heartfelt thanks to William Gardner, aka CHEER4FTC, from Charlottesville, VA for his ideas, advice, proofreading, code reviews, and general wisdom.

I also wish to thank the enthusiasts on reddit.com/r/ftc for the many ideas brought forth.

And finally, I wish to thank FIRST and FTC for running an outstanding program capable of igniting the passion in the thousands of us involved as participants.