JOURNAL OF AMATEUR SPORT

JAS

HTTP://WWW.JAMSPORT.ORG

The Bradley-Terry Model in Binary Outcome Driven Rankings: An Application in Amateur Hockey

Connor Meissner Jeremy Losak

Syracuse University

Ranking systems serve critical roles in sport settings, most notably in determining playoff participants and seeding. Many ranking methodologies exist that are flexible enough to incorporate many input measures and produce models that are highly predictive of game outcomes. However, there are circumstances—especially for amateur sport leagues—in which more complex inputs are either unavailable or not desirable, as they may lead to adverse performance incentives. Therefore, the goal of this paper is to highlight a ranking methodology that only considers binary game outcomes, i.e., wins and losses. Specifically, we consider the efficacy of the Bradley-Terry Model to efficiently rank sport teams for playoff consideration. We apply this method as a case study to the New England Prep School Ice Hockey Association (NEPSIHA), and compare the accuracy of their current ranking system to the Bradley-Terry model using simulation methods. We show that Bradley-Terry significantly outperforms NEPSIHA's current method, especially when teams face unbalanced strengths of schedule. This result holds under various league competitive balance distributions

Keywords: Bradley-Terry; Rankings; Simulations; Amateur Athletics

Rankings and ranking systems serve critical roles in numerous practical contexts. Applied to sports, there are a multitude of prominent ranking systems, such as the Massey method (Massey, 1997), the Markov method (Govan, 2008; Vaziri et al., 2018), the Colley method (Colley, 2002), and the Elo method (Elo, 1978; Ingram, 2021; Kovalchik, 2020), in addition to the simple method of ranking teams by win percentage or point accumulation.

Most sport leagues, from youth to professional, implement ranking systems to determine champions or playoff participants (Stefani, 2011). In most leagues and divisions, wins and losses are tallied, and the teams are ranked based on win percentage. Some leagues rank teams by awarding points for winning, tying, or reaching overtime in a game. In many situations, scheduling limitations can cause simple systems like these to be inequitable, prompting a more complex, sometimes even subjective, system to rank teams (Vaziri et al., 2018). Examples include iterations of the FIFA World Ranking system and selection committees for college basketball and football (Stefani, 2011). In these examples, the number of teams participating in the competition far exceed the number of games played per team, resulting in unbalanced schedules. Teams facing more difficult schedules may have lower win percentages than some lower quality teams, despite being superior in skill—hence the need to rely on alternative ranking methodologies.

Ranking methods have various applications outside of sport as well. Google uses advanced algorithms to rank the importance and popularity of websites to decide which links to show their users (Evans, 2007). US News ranks the quality of universities and colleges across the country, utilizing metrics that include graduation rates, selectivity, and reviews (Standifird, 2005). Applications and websites use rankings and algorithms to provide consumers with recommendations, such as which restaurant to go to (Zhang et al., 2020).

In this paper, we focus on a particular ranking system: the Bradley-Terry model (Bradley and Terry, 1952; Bradley, 1954, 1955). The Bradley-Terry (BT) model is a probability model designed to predict pairwise comparisons of a sample set. It is estimated by maximizing likelihood functions, using a series of observed pairwise comparisons as inputs. Applications of the model are particularly relevant when pairwise comparisons of every possible combination within the sample set are impossible to complete. An example is the ranking and comparison of the thousands of wines that exist in the world (Agresti, 2007, p. 265). While it would be impossible for one person to try every single wine and be able to rank them effectively, the BT model can consider various one-on-one pairwise comparisons of wines from various reviewers to produce an aggregate ranking.

The BT model has obvious applications in sport settings. When a sporting contest is played, we consider this a pairwise comparison. While two teams may not play directly, we can use the BT model to rank them. In the simplest of terms, if Team A defeats Team B, and Team B defeats Team C, then BT would allow us to conclude that Team A is better than Team C, despite the fact they did not play one another. As the web of teams increases and the schedule dynamics become more intricate, the BT model is flexible enough to adapt and produce rankings of teams and hypothetical win probabilities if two of the teams were to face off. Applications of the Bradley-Terry model in sport include college ice hockey (Whelan and Wodon, 2020), European field hockey (Looijen, 2019), and international cricket (Islam et al., 2017; Dewart and Gillard, 2019). In the media, the BT model is used to make Ken's Ratings for American College Hockey (KRACH), which College Hockey News endorses "as the best system to objectively rank teams."

Traditionally, academic papers have focused less on the practical ranking of teams and more on the predictive modeling that these sorts of models allow for (Barrow et al., 2013; Dabadghao and Vaziri, 2021; Lasek et al., 2013; Leitner et al., 2010; Leung and Joseph, 2014; Stewart et al., 2022; Williams et al., 2020, for example). Occasionally, papers will consider biases or look for improvements in current ranking systems (Servien, 2022; Szczecinski and Roatis, 2022) for playoff seeding purposes (Burer, 2012; Bigsby and Ohlmann, 2017), especially as it relates to the NCAA March Madness College Basketball Tournament (Coleman et al., 2010; Dutta and Jacobson, 2018; Paul and Wilson, 2012; Sanders, 2007; Stocks-Smith, 2021; Stone and Arkes, 2018, among others). The BT model, especially, has received minimal consideration in the literature, as there are typically more advanced predictive models available that utilize a wide variety of variables to consider the strengths of teams. The inputs of these predictive models tend to involve more variables, ranging in complexity from margin of victory (Szczecinski, 2022) to expected goal metrics (Eggels, 2016; Kovalchik, 2020). But what happens in sport settings where these variables may be unavailable, such as in many amateur sport settings, or where they create adverse performance incentives that are counter to league objectives? A more simplistic methodology (maximizing likelihood functions is much simpler than alternative approaches) that relies on fewer inputs may be desirable in certain settings, including a methodology that just focuses on binary game outcomes (wins and losses).

We provide a case study illustrating the applicability of the BT model in sport settings desiring binary outcome-driven rankings. Specifically, we assess the BT model's ability to rank teams in the New England Prep School Ice Hockey Association (NEPSIHA). NEPSIHA currently implements its own unique ranking system that objectively selects and seeds teams for playoff tournaments. We benchmark the efficiency of the BT model to their current ranking system by observing each model's ability to identify and rank the "best" teams in the league. Utilizing a simulation approach with hypothetical team skill competitive balance distributions, we identify biases and inefficiencies in the current NEPSIHA ranking system and show how implementation of the BT model can address these biases.

The rest of the paper is structured as

follows. First, we discuss the intricacies of NEPSIHA and why we have chosen this league in particular as our case study. Next, we introduce the BT model, followed by an overview of the empirical strategy to compare the BT model to NEPSIHA's current ranking system. Then, we illustrate simulation results, highlighting critical inefficiencies in the current NEPSIHA ranking system and how the BT model corrects for them. Finally, we provide concluding and summarizing remarks.

NEPSIHA: A CASE STUDY

The New England Prep School Ice Hockey Association (NEPSIHA) is a group of 54 high school hockey programs across the six states of New England and New York. Known for elite academics, these schools are commonly referred to as NEPSAC schools for their membership in the New England Preparatory School Athletic Council. Hundreds of New England Prep alumni go on to play college sports, and many move on to play professionally, especially in hockey. The NEPSIHA has alumni currently playing in the NHL, such as All-Stars Max Pacioretty, Chris Kreider, and Conn Smythe Trophy winner Jonathan Quick, among others. Abbreviations for all school names referenced throughout the paper are available in Appendix A.

At the end of each season, 24 NEPSIHA teams take part in three different postseason tournaments: the Stuart/Corkery, Martin/Earl, and Piatelli/Simmons tournaments. The Stuart/Corkery tournament is commonly referred to as the "Open" tournament and consists of the eight best teams, with the winner of this tournament considered the NEPSIHA champion. The Martin/Earl tournament consists of the eight best "Large" schools not selected for the Open tournament. Similarly, the Piatelli/Simmons tournament consists of the eight best "Small" schools not selected for the Open tournament. The Large class vs. Small class distinction is made based on the enrollment totals of each school. Given its de facto status as the NEPSIHA playoffs, selection for the Open tournament will be our main focus.

To determine the top eight teams that make the Open tournament, NEPSIHA uses a customized approach called the Jeremy S. Philipson Rating (JSPR) system. JSPR starts by calculating the Rating Percentage Index (RPI) of every NEPSIHA team. RPI is a commonly used rating system that was most notably used in NCAA Basketball before it was replaced in 2018 by the NCAA Evaluation Tool (NET). For many applications, its simplicity and transparency can be very appealing. The standard RPI formula is

$RPI = (WP \times 0.25) + (OWP \times 0.5) + (OOWP \times 0.25)$,

where *WP* is Win Percentage, *OWP* is Opponents' Win Percentage, and *OOWP* is Opponents' Opponents' Win Percentage. However, NEPSIHA tweaks the coefficients so that the OWP section has far less weight. The resulting formula is

$RPI_{NEPSIHA} = (WP \times 0.25) + (OWP \times 0.21) + (OOWP \times 0.54)$.

It is believed that this tweak was made to avoid teams loading their schedule with difficult teams to artificially inflate their RPI.

From the RPI ranking, the top 16 teams are classified as the Teams Under Consideration (TUC). These teams are put through a pairwise comparison. This comparison, made between every pair of top 16 schools, is where the JSPR ranking system is very unique. There are four inputs to each pairwise comparison: head-to-head record, RPI, record against common opponents, and record against TUC.

Table 1
Hypothetical JSPR Pairwise Teams Under Consideration Example

	Team A	Team B
Head-to-Head Record	**1-0-0**	0-1-0
RPI	**0.632**	0.611
Record Against Common Opponents	2-1-0	**5-0-0**
Record Against TUC	**4-1-0**	3-3-1

If a team is better than the other team in more categories, they are awarded a JSPR point. Table 1 shows a pairwise example in which Team A is better in three (head-to-head record, RPI, record against TUC) out of the four categories, earning it the JSPR point. In the event each team takes the same number of categories, head-to-head record is used as the first tiebreaker, followed by the second tiebreaker of RPI. If teams are tied in a particular category, that category is not counted. For example, had Team A been 5-0-0 against common opponents, Team A would have won the JSPR point three categories to zero instead of their actual three categories to one. The maximum number of JSPR points a team can earn is 15, as that is a team that earns a JSPR point in each comparison against all other TUC. The team with the most JSPR points gets the top seed in the Open tournament, second-most gets the second seed, and so on until all eight playoff seeds are determined. When teams are tied in JSPR points, their comparison against each other is the tiebreaker. As an example, Table 2 shows the full comparison matrix for all 16 TUC from the 2017–18 season, ordered by their final seeding. Note how WES earned the eighth and final playoff spot over NMH since they won their JSPR pairwise comparison due to the head-to-head tiebreaker.

The JSPR ranking system has a number of nuances that, a priori, led us to believe there may have been inefficiencies and biases, hence the motivation for considering BT as an alternative ranking system. First, games carry varying levels of importance under JSPR. A game against an RPI top 16 opponent will impact the Head-to-Head, RPI category, and record against TUC components, while non top 16 games would only impact the RPI category (either game type could potentially impact the record against common opponents category). As a result, there is a clear discontinuity in terms of game importance, as losing to the 16th best team in RPI carries significantly more consequence compared to losing to the 17th best team in RPI. Also, JSPR systematically punishes a loss to the 16th best team in RPI more than a loss to the worst team in the league, despite the fact that such a loss would likely be a better indicator of true team ability. Lastly, RPI as a whole is a very outdated and inaccurate measurement of team quality. This is made evi-

Table 2

JSPR Pairwise Points 2017–18 Season

	SAL	RIV	KUA	DEX	MIL	EXE	CSH	WES	NMH	THA	SEB	BRU	GUN	WIL	LAW	WIN	Total
SAL		2-0	2-0	3-0	4-0	3-0	4-0	4-0	3-0	3-0	3-0	4-0	4-0	4-0	3-0	3-0	15
RIV	0-2		4-0	3-1	3-0	4-0	3-0	3-0	3-0	4-0	4-0	4-0	3-0	4-0	3-0	3-0	14
KUA	0-2	0-4		3-0	2-1	4-0	4-0	3-1	4-0	3-0	3-0	4-0	3-0	3-0	3-0	3-0	13
DEX	0-3	1-3	0-3		2-1	3-0	3-0	3-0	3-0	4-0	3-0	3-0	3-0	3-0	3-0	3-1	12
MIL	0-4	0-3	1-2	1-2		3-0	3-0	2-1	3-0	3-0	4-0	3-0	3-0	3-0	4-0	4-0	11
EXE	0-3	0-4	0-4	0-3	0-3		3-1	1-1**	4-0	4-0	3-1	3-0	4-0	2-1	3-0	3-0	10
CSH	0-4	0-3	0-4	0-3	0-3	1-3		2-1	2-1	3-1	2-1	3-0	3-0	4-0	3-0	3-0	9
WES	0-4	0-3	1-3	0-3	1-2	1-1	1-2		2-2*	3-0	2-1	3-1	1-3	4-0	3-0	2-1	7
NMH	0-3	0-3	0-4	0-3	0-3	0-4	1-2	2-2		3-1	2-1	3-0	2-1	3-0	3-0	3-1	7
THA	0-3	0-4	0-3	0-4	0-3	0-4	1-3	0-3	1-3		2-1	2-0	2-1	2-0	3-1	1-2	5
SEB	0-3	0-4	0-3	0-3	0-4	1-3	1-2	1-2	1-2	1-2		2-1	2-1	2-2	3-0	3-1	4
BRU	0-4	0-4	0-4	0-3	0-3	0-3	0-3	1-3	0-3	0-2	1-2		2-2*	2-1	1-1	2-1	3
GUN	0-4	0-3	0-3	0-3	0-3	0-4	0-3	3-1	1-2	1-2	1-2	2-2		3-1	2-0	2-2	3
WIL	0-4	0-4	0-3	0-3	0-3	1-2	0-4	0-4	0-3	0-2	2-2*	1-2	1-3		2-1	2-2*	3
LAW	0-3	0-3	0-3	0-3	0-4	0-3	0-3	0-3	0-3	1-3	0-3	1-1**	0-2	1-2		2-1	2
WIN	0-3	0-3	0-3	1-3	0-4	0-3	0-3	1-2	1-3	2-1	1-3	1-2	2-2*	2-2	1-2		2

*First tiebreaker of head-to-head used **Second tiebreaker of RPI used dent by the NCAA phasing out the use of the metric in many of their sports, replacing it with more accurate measures, such as NET in basketball.

When considering an alternative ranking system, there is an important consideration to take into account, specifically for NEPSIHA. Many ranking systems use margin of victory as an input because it has been proven to be more predictive of actual team ability than just wins and losses. For an amateur hockey association, however, it is preferable that only wins and losses are used as inputs in its ranking system. From a competitive standpoint, leagues would prefer teams to be win maximizers rather than goal differential maximizers. For example, if margin of victory affects the ranking, teams could be reluctant to pull the goalie late in the game and keep the margin of the loss at one instead of attempting to tie the game. From a sportsmanship standpoint, amateur leagues may not want teams to "run up the score," or hamper participation by limiting the opportunities for benchwarmers to play. The focus on just wins and losses is a major requirement being imposed for any alternative ranking system, similar to what is done in JSPR, and hence why more intricate ranking systems are not under consideration.

THE BRADLEY-TERRY MODEL

The BT model uses pairwise comparisons to estimate the strength of each subject (team) relative to each other. The following, known as Zermelo's iteration, is used to obtain maximum likelihood estimates (see Hunter, 2004 for more information) that are then compared to form a ranking

$$p_i = \frac{W_i}{\sum_{j \neq i} \frac{w_{i,j} + w_{j,i}}{p_i + p_j}},$$

where p_i is the parameter for Team *i* in the sample, W_i is team *i*'s total wins, and $w_{i,j}$ is the number of Team *i* wins over Team *j*. For the initial iteration, all p parameters are set to one. Then, the algorithm is calculated producing new parameter values, which are normalized and replaced in the initial equations and recalculated. For example, after the first iteration, it can be shown that

$$p_i = \frac{WP_i}{\sum WP},$$

where *WP* is a school's overall win percentage. These iterations are repeated until the parameters converge. After convergence, it can be said that the probability that any Team *i* beats any Team *j* is given by

$$p_{ij} = \frac{p_i}{p_i + p_j}.$$

The following is a simple example of the usage of the BT model in a league of three teams. Suppose Team 1 beat Team 2 twice and lost to Team 2 once; Team 3 beat Team 2 three times and lost to Team 2 four times; Team 1 and Team 3 never played each other. Below shows the first iteration plugging initial values into Equation 1:

$$p_1 = \frac{2+0}{\frac{2+1}{1+1} + \frac{0+0}{1+1}} = 1.333,$$

$$p_2 = \frac{1+4}{\frac{1+2}{1+1} + \frac{4+3}{1+1}} = 1,$$

and $$p_3 = \frac{0+3}{\frac{0+0}{1+1}+\frac{3+4}{1+1}} = 0.857$$.

These parameters are then normalized by dividing them by the sum of the parameters, which is 3.19 in this case. The resulting parameters are [0.418, 0.313, 0.269], which are equivalent to what can be derived by plugging in initial win percentages into Equation 2. These parameters are then used in the second iteration:

$$p_1 = \frac{2+0}{\frac{2+1}{0.418+0.313}+\frac{0+0}{0.418+0.269}} = 0.487,$$

$$p_2 = \frac{1+4}{\frac{1+2}{0.313+0.418}+\frac{4+3}{0.313+0.269}} = 0.310,$$

and

$$p_3 = \frac{0+3}{\frac{0+0}{0.269+0.418}+\frac{3+4}{0.269+0.313}} = 0.249.$$

After the second iteration, the normalized parameters are [0.466, 0.296, 0.238]. This sequence is repeated until all three parameters converge. In this example, after 12 iterations, the values converge to [0.533, 0.267, 0.200]. As we can see, the BT model concludes that Team 1 is better than Team 3, and predicts that Team 1 would have a 72.7% win probability if facing Team 3, despite them never playing one another.

While we focus our analysis on simple binary game outcomes (win/loss), BT can be extended in numerous ways, as summarized by Butler & Whelan (2004). For instance, general and team-specific order effects, such as the home field advantage, can be incorporated into the BT model (Davidson and Beaver, 1977). Suppose team *i* is always the home team. Modifying Equation 3, the probability that Team *i* beats Team *j* is given by

$$p_{ij} = \frac{\gamma p_i}{\gamma p_i + p_j},$$

where γ is a non-negative parameter, which can be solved for when maximizing likelihood functions. While Equation 4 only accounts for a generic home field advantage, the model can also be expanded to allow for individual team home advantages. In addition, ties can be incorporated if that is a desirable outcome (see Davidson, 1970), although in practice, treating a tie as half a win and half a loss typically works if the end goal is to produce a ranking of teams (as opposed to predicting game outcomes).

COMPARING AND EVALUATING RANKING SYSTEMS

The goal of most ranking systems, JSPR among them, is to identify a "true" hierarchical ordering of team ability. This is typically done using observed game data, which could include game result, score differential, production measures, expected scoring measures, and more (see Govan et al., 2009, among others). Of course, relying on observed outcomes, which are inherently random and come with high degrees of variance, will never perfectly identify "true" team skill level. The best team does not always win the game, and as long as that holds true, there is going to be variation in results that might cloud how good a team really is. Finite schedules involving between 20 and 30 games per school exacerbate the problem. So, in the end, a ranking system is only an estimate of a team's quality, relative to the competition.

Table 3

RPI Versus BT Hypothetical Example

	Team A	Team B	Team C	Team D
Record Against				
Team A	X	3-4	1-1	1-6
Team B	4-3	X	2-4	0-2
Team C	1-1	4-2	X	3-5
Team D	6-1	2-0	5-3	X
Ranking Measures				
Win Percentage	0.688 (1)	0.600 (2)	0.500 (3)	0.235 (4)
RPI	0.565 (2)	0.581 (1)	0.436 (3)	0.407 (4)
Adjusted RPI	0.567 (1)	0.526 (2)	0.490 (3)	0.404 (4)
BT Rating	0.401 (1)	0.354 (2)	0.168 (3)	0.077 (4)

Diverse systems can yield distinct team rankings, despite analyzing the same game outcomes. Consider the following hypothetical example comparing rankings produced by two systems: RPI and BT. Table 3 considers a simplistic scenario involving four teams in which teams play unbalanced schedules consisting of between 15 and 17 games. According to win percentage and BT, Team A emerges as the top-ranked team, while RPI designated Team B as the highest ranked. This discrepancy arises due to variations in schedule quality. Team A is penalized in the RPI rankings due to having a substantial proportion of their games being played against the bottom-ranked Team D. Although JSPR's adjusted RPI somewhat mitigates this scheduling penalty by assigning less weight to opponent win percentage, it nonetheless underscores the influence of methodological choices on the final rankings.

The BT model possesses an additional advantage over RPI, JSPR, and other points-based systems, as it facilitates hypothesis testing. In our example, there is not enough evidence to reject the null hypothesis that Team A and Team B are of equal strength, while there is sufficient evidence to suggest that both Teams A and B are of greater strengths than Teams C and D.

In terms of evaluating ranking systems, it is difficult to do ex-post analysis given the sample size (there are few seasons of data) and lack of reliable validation tools. One could consider analyzing the relationship between final rating and subsequent playoff performance, yet this only considers a subset of teams and relies on a small number of overall matchups. Conversely, the ranking system could be evaluated based on its predictive ability of regular season matchups. However, this relies on roster quality homogeneity, an assumption that likely falls apart when considering injuries (and other game-missing scenarios) and general team improvement during a season. Instead, we utilize a simulation approach to identify and compare the accuracy of ranking systems using actual schedules and hypothetical team skill distributions. Specifically, we compare NEPSIHA's current JSPR system with the Bradley-Terry model.

Our main methodological tool is simulation. We utilize actual schedules over six seasons, from the 2013–14 season to the 2018–19 season. This approach has two key benefits. First, working with six different seasons allows us to analyze ranking system quality under six different team skill level distributions. Second, employing real schedules offers a more comprehensive depiction of the intricacies within the New England prep hockey calendar. We simulate regular season outcomes and utilize both JSPR and BT to rank teams for playoff consideration in the Open tournament. The better model is the one that 'gets it right' more often—greater skilled teams should be selected for the playoff field more frequently than lesser-skilled teams.

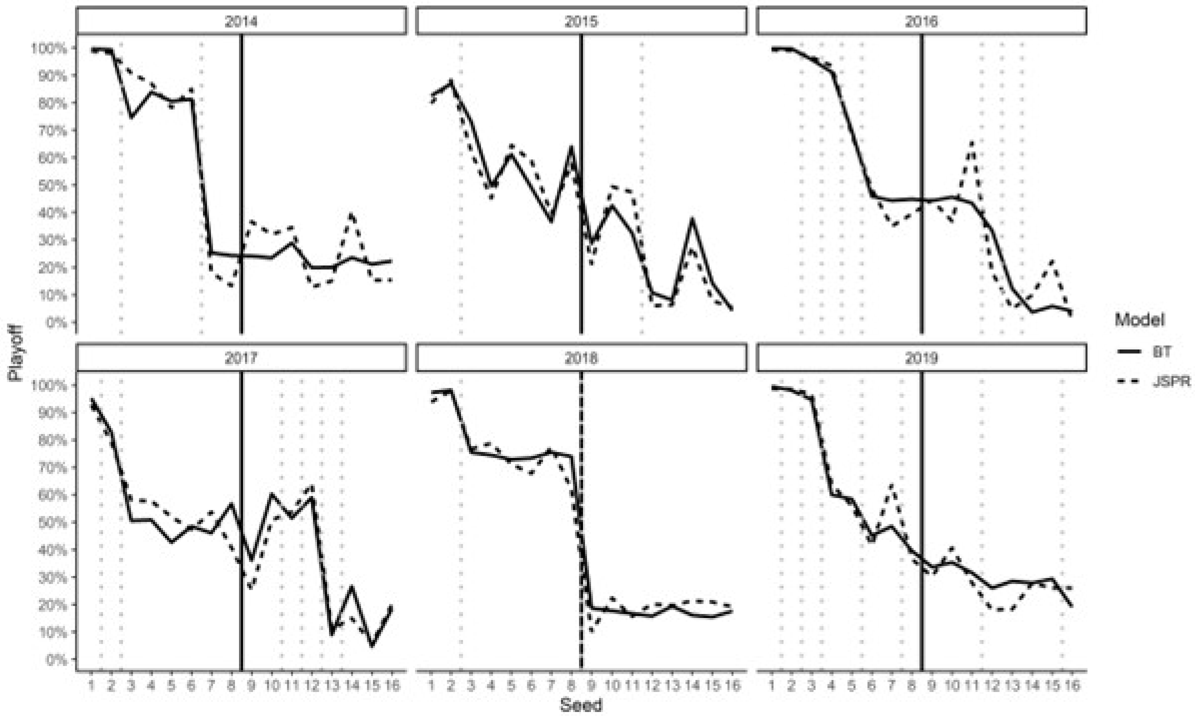
We start by assigning each NEPSIHA team in each season a hypothetical skill level, thus providing a "true" team ability benchmark. The best ranking systems should, on expectation, place teams in the order of their true ability levels. For illustrative purposes, we select hypothetical skill levels that are meant to somewhat mimic the real-life distribution of team skill levels. That said, we tweaked certain team rankings to introduce various competitive balance features that would enrich final analysis. For example, we may consider a competitive balance structure where there is a clear number one team, six teams bunched up at the playoff cut line, etc. In some settings, we also manipulate certain skill levels to be equivalent across any number of teams. This is especially powerful because an accurate ranking system should treat them the same on average across the simulations, while a biased system might systematically favor one of the teams more than the other. Ratings for the 16 best teams each season are available in Appendix B.

Next, we utilize those team skill levels and simulate each team's schedule. We trained a probit model to provide individual team win probabilities given team ratings. For simplicity, we ignore home advantage, rest, and other factors that may impact individual team win probability. Following each season simulation, we use the aforementioned JSPR methodology and BT methods to rank teams. After 5,000 simulations per season, we analyze features of the distributions of individual team rankings when comparing the two methodologies. For apples-to-apples comparison, we utilize both JSPR and the BT model for the same set of simulations, rather than run a different set of simulations to be analyzed by each model. Ranking systems that provide playoff team ordering closest to the hypothetical team skill distribution are considered "better."

When analyzing final results, it is critical to understand that these are simulations with assigned hypothetical skill levels. While they are designed to roughly reflect the real world team levels of that season, they intentionally do not mimic

Figure 1

Simulation Top-16 Finishes by Year



Note. The proportion of simulations in which either the BT (solid) or JSPR (dashed) models rank each of the top 16 teams (by hypothetical rating) in a top-eight playoff position. The solid vertical line indicates the playoff cut point and the dotted vertical lines separate teams of comparable ratings.

them perfectly. So, the statement "JSPR was biased against Team A in favor of Team B" does not always imply those exact teams received those biases in the real world in that given season. Ultimately, we could have assigned any skill levels and the practical use of the results would still be valid. Having a pre-established "true" ranking allows us to benchmark the accuracy and effectiveness of the two ranking systems, while mimicking actual schedules and team abilities allows for practical and real-life exploration of any biases.

RESULTS

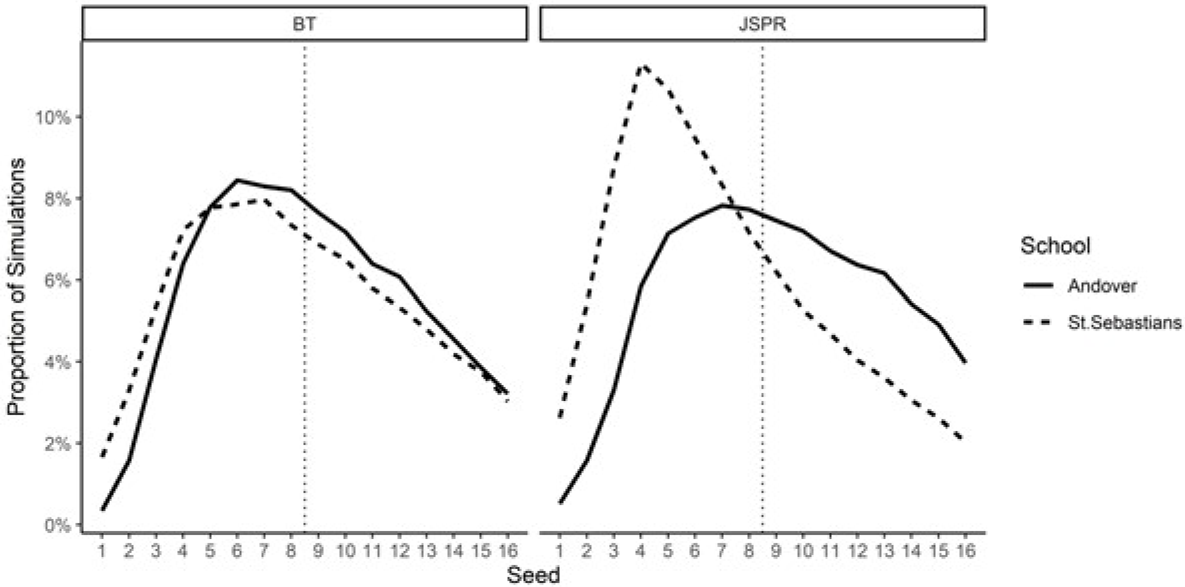
The majority of our analysis revolves around the percentage of simulations in which a school was rated in the top eight and selected for participation in the Open tournament. When considering rating system effectiveness, we expect the better system to put the better rated school in the tournament more frequently. While random variation in game outcomes may result in a school being rated lower or higher in a single simulation, over a large sample of simulations, the better schools should ultimately be rated more frequently than lower rated schools. Sizeable deviation from that is indicative of a potential bias in the ranking system.

Overall Model Performance

Figure 1 illustrates the percentage of simulations in which each of the top 16 rated teams made the Open tournament

Figure 2

Ranking Simulation Comparisons, AND vs SEB 2018–19



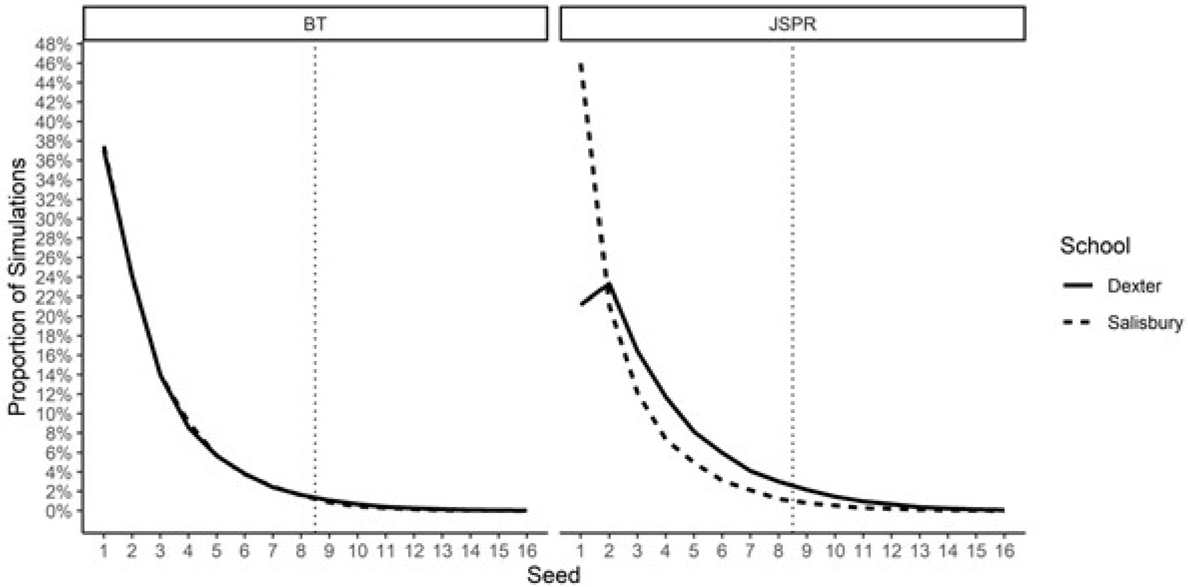
Note. Ranking distribution for 2018–19 sixth rated AND (solid) and seventh rated SEB (dashed) under the BT and JSPR systems. The vertical dotted lines represent the playoff cut point.

under the JSPR and BT models in each season. An efficient and equitable ranking system would have higher rated teams with higher playoff frequencies, and it is apparent that the BT model outperforms JSPR significantly in this department. Full schedule, rating, and playoff percentage breakdowns are available in Appendix B.

After analyzing Figure 1, it is apparent that BT's playoff proportions much more strongly trend with school rating levels compared to JSPR. Consider competitive balance distributions illustrated in 2014, 2016, and 2018. In each of those cases, similarly rated teams, separated by dashed vertical lines in Figure 1, have more comparable playoff probabilities under BT compared to JSPR. Visually, this is represented by the smooth solid lines for BT that follows the true skill level distributions, while the JSPR model zig-zags and varies greatly. For example, in 2014, BT provides similar playoff probabilities for similarly rated teams ranked seven through sixteen (13% to 23%), while JSPR sees significantly more variation (8% to 31%). In 2016, there were six equal teams ranked sixth to eleventh, three on either side of the playoff cut line. Using BT, these teams made the playoffs between 34–37% of simulations, while JSPR had a significantly wider range of 25–57%. In 2018, teams similarly ranked from three to eight made the playoff field 64–68% of simulations under BT, while between 59–72% under JSPR, and teams similarly ranked between ninth and sixteenth made the playoff field 10–15% of simulations under BT while between 6–16% under JSPR. These three years provide clear evidence that BT outperforms JSPR.

Figure 3

Ranking Simulation Comparisons, DEX vs SAL 2017–18



Note. Ranking distribution for 2017–18 first rated DEX (solid) and second rated SAL (dashed) under the BT and JSPR systems. The vertical dotted lines represent the playoff cut point.

Scheduling Biases in JSPR

Another takeaway from analyzing simulation results is the clear JSPR scheduling bias, especially its treatment of matchups against top 16 opponents. A JSPR point is awarded based on teams' records against top 16 opponents (TUC), yet not all games against top 16 opponents are truly of the same difficulty. Figure 2 highlights one obvious example of this bias from 2019. Andover (AND) and St. Sebastian's (SEB) were comparably rated teams ranked sixth and seventh, respectively. JSPR, however, was much more favorable to SEB in terms of making the playoffs (57.78%) compared to AND (33.18%), relative to BT's placement of the two schools (36.66% for AND and 41.72% for SEB). AND had 13 games against top 16 teams with four against top three opponents. Meanwhile, SEB had six games against top 16 opponents, none of which were ranked in the top 8, making it very easy for them to have a stronger record against TUC than an AND team that went through the gauntlet of difficult opponents. In other words, under JSPR, teams with fewer top eight opponents are rewarded, while teams playing difficult schedules are penalized.

Another example of scheduling bias is presented in Figure 3, which compares the playoff seed simulation outcomes for the top two equally rated schools from 2018, Dexter (DEX) and Salisbury (SAL). As illustrated in Figure 3, JSPR was putting SAL as the top seed twice as often as DEX (44.8% versus 22.2%), despite the two teams being of identical skill, while BT was much more equitable (37.7% and 36.6%, respectively). All of DEX's seven games against TUC were top nine opponents, five of which were top eight. SAL had three out of eight TUC games against top eight opponents, a lower rate than what DEX played. The BT model also had DEX or SAL as the top team around 74% of the time, compared to JSPR's rate of 67%. This is to say the BT model correctly identified one of these two top teams as the best team more often than JSPR did.

A third example of scheduling imbalances impacting JSPR can be seen in the three equally rated 14^{th} ranked teams from 2016. Simulations under BT had the three teams in the playoffs between two and four percent of the time, while JSPR saw wild swings with Rivers (RIV) in the playoffs in 16.16% of simulations, Noble and Greenough (NOB) in the playoffs in 5.68% of simulations, and Deerfield (DEE) in the playoffs in just 0.38% of simulations. RIV played only four games against TUC, with only one being in the top five. Facing easier TUC opponents inflates that JSPR component. Also, playing only four of these types of games increases the variation of their record against TUC; it is not difficult for them to snag two (25.5% likelihood), three (6.3%), or even four (0.55%), of those games and have that part of JSPR be very inflated and helping them secure a spot in the playoffs. Meanwhile, DEE played an abnormally high fourteen games against top 16 opponents (over half their schedule!), including seven games against the top four. Their record against TUC, along with their playoff chances under JSPR, did not stand a chance.

Where Bradley-Terry Produces Suboptimal Results

Despite the clear advantages of BT over the incumbent JSPR system, BT is not necessarily without flaw. In two seasons, 2015 and 2017, BT and JSPR perform similarly, both providing non-optimal playoff placement outcomes. In 2015, we incorporated a team skill distribution, such that teams between three and ten were all equally rated. With such parity, both ranking systems struggled. The group of teams from three to ten showed zig-zag like patterns in Figure 1, with a playoff proportion range of 23–68% under BT and 17–59% under JSPR. In 2017, tied-for-ninth rated Loomis (LOO) had higher playoff percentages than seven (under BT) and four (under JSPR) comparable-or-better rated schools, respectively. Fifteenth ranked Belmont (BEL) had probabilities significantly lower than similarly ranked Taft (TAF) at 14th and much-lower ranked SEB at 16th. More analysis is needed to better understand the scheduling and competitive balance circumstances in which the BT model deviates from its more ideal performances in the other season simulations.

DISCUSSION AND CONCLUSION

While ranking systems are of critical importance in numerous sport settings, the Bradley-Terry model has received minimal attention in the literature, mostly due to its sole reliance on binary game outcomes. However, this feature makes the BT model an ideal ranking system in settings where competitive incentives may make the more commonly used score differential input a less-than ideal tool. When teams play balanced schedules, win-loss records provide an unbiased estimate of team rankings. For leagues like the New

England Prep School Ice Hockey Association, balanced schedules are not possible due to numerous logistical considerations (travel costs, schedule length, number of schools in competition). The BT model provides a more equitable ranking even in the presence of unbalanced schedules.

In highlighting NEPSIHA, we illustrate a flaw in their current ranking system and highlight BT's effectiveness in minimizing the bias's effect. Specifically, while JSPR systematically punishes schools that face exceptionally tough competition, it rewards schools that play more games against lower-quality TUC and fewer games against TUC in general. Via simulated examples of real-world competitive balance setups, it was shown that BT does a much better job of minimizing this bias and ranking teams according to their true skill level. While BT was not necessarily perfect in each of the season simulations—especially in situations where there was significant parity among the top teams—it generally outperformed the JSPR system in terms of selecting the better teams to participate in the playoffs.

While our focus was on high school hockey, there was no sport-specific element incorporated in the methodology, making it extendable to other sport settings. This includes other high school, college, and amateur sport leagues in which teams play unbalanced schedules and binary game outcomes are the desired sole input when ranking teams. The BT model is also flexible enough to incorporate additional inputs if desired. For instance, we do not consider home ice advantage, a common input in many ranking systems (NET in NCAA Basketball, for example) that can be incorporated in the BT framework.

Further work needs to be done to understand how sample size of games (teams in NEPSIHA play between 20 and 30 games each season, typically) and other scheduling quirks impact the general accuracy of BT. As illustrated in two of the six simulated seasons, while BT is an improvement over JSPR, there may still be prevalent biases that impact the system's ability to identify the "best" teams. For instance, this methodology ignores matchup-specific tendencies. If a team plays additional games against an opponent in which it has a matchup and schematic-specific advantage relative to other teams, BT will underestimate the team's true win probability in that matchup, and will systematically overvalue the team's true ability against a generic opponent, leading to an inflated ranking.

A major consideration of the BT model is its need for a reasonable amount of data before rankings become meaningful. Maximum likelihood produces estimates of team strengths, which may be far from their true values for noisy or small data sets. Leagues that implement the BT model will likely not have reliable rankings early in a season, which may be a desirable feature in a ranking system. League administrators should be aware of this tradeoff (better final rankings but noisy early-season rankings) when choosing BT. Also, it is important to emphasize that the BT model is not perfect, as we illustrated in our simulations. But the "perfect" ranking system does not exist, and the lack of perfection should not come at the expense of exploring obvious improvements in current techniques and methodologies.

While implementation of the BT model for playoff ranking purposes would surely be an improvement over JSPR for NEPSIHA, other amateur league administrators should understand that every potential ranking system comes with empirical nuances that should be fully considered. BT is relatively simple to implement and allows the data—the web of completed matchups—to holistically speak together to create a final ranking. Although such a system may not be the most predictive, it balances the need to reward game outcomes (winning and losing), while controlling for scheduling imbalances that can inflate those outcomes. League administrators must identify what they prioritize in a ranking system, a decision that is specific to league objectives and missions.

REFERENCES

Agresti, A. (2007). *An Introduction to Categorical Data Analysis* (2nd ed.). John Wiley & Sons.

Barrow, D., Drayer, I., Elliott, P., Gaut, G., & Osting, B. (2013). Ranking rankings: an empirical comparison of the predictive power of sports ranking methods. *Journal of Quantitative Analysis in Sports*, *9*(2), 187–202. doi:10.1515/jqas-2013-0013

Bigsby, K. G., & Ohlmann, J. W. (2017). Ranking and prediction of collegiate wrestling. *Journal of Sports Analytics*, *3*(1), 1–19. doi:10.3233/jsa-160024

Bradley, R. A. (1954). Incomplete Block Rank Analysis: On the Appropriateness of the Model for a Method of Paired Comparisons. *Biometrics*, *10*(3), 375–390. doi:10.2307/3001592

Bradley, R. A. (1955). Rank Analysis of Incomplete Block Designs: III Some Large-Sample Results on Estimation and Power for a Method of Paired Comparisons. *Biometrika*, *42*(3/4), 450–470. doi:10.2307/2333391

Bradley, R. A., & Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, *39*(3/4), 324–345. doi:10.2307/2334029

Burer, S. (2012). Robust Rankings for College Football. *Journal of Quantitative Analysis in Sports*, *8*(2). doi:10.1515/1559-0410.1405

Butler, K., & Whelan, J. T. (2004). The existence of maximum likelihood estimates in the Bradley-Terry model and its extensions. arXiv:math/0412232. doi: 10.48550/arXiv.math/0412232

Coleman, B. J., DuMond, J. M., & Lynch, A. K. (2010). Evidence of bias in NCAA tournament selection and seeding. *Managerial and Decision Economics*, *31*(7), 431–452. doi:10.1002/mde.1499

Colley, W. N. (2002). *Colley's bias free college football ranking method*. PhD thesis, Princeton University.

Dabadghao, S. S., & Vaziri, B. (2021). The predictive power of popular sports ranking methods in the NFL, NBA, and NHL. *Operational Research*, *22*(3), 2767–2783. doi:10.1007/s12351-021-00630-9

Davidson, R. R. (1970). On Extending the Bradley-Terry Model to Accommodate Ties in Paired Comparison Experiments. *Journal of the American Statistical Association, 65*(329), 317–328. doi: 10.2307/2283595

Davidson, R. R., & Beaver, R. J. (1977). On Extending the Bradley-Terry Model to Incorporate Within-Pair Order Effects. *Biometrics, 33*(4), 693–702. doi: 10.2307/2529467

Dewart, N., & Gillard, J. (2019). Using Bradley-Terry models to analyse test match cricket. *IMA Journal of Management Mathematics*, *30*(2), 187–207. doi:10.1093/imaman/dpx013

Dutta, S., & Jacobson, S. H. (2018). Modeling the NCAA basketball tournament selection process using a decision tree. *Journal of Sports Analytics*, *4*(1), 65–71. doi:10.3233/jsa-170149

Eggels, H. P. H. (2016). *Expected goals in soccer: Explaining match results using predictive analytics*. Master's thesis, Eindhoven University of Technology.

Elo, A. E. (1978). *The Rating of Chessplayers, Past and Present*. Arco Pub.

Evans, M. P. (2007). Analysing Google rankings through search engine optimization data. *Internet Research*, *17*(1), 21–37. doi:10.1108/10662240710730470

Govan, A. Y. (2008). *Ranking theory with application to popular sports*. PhD thesis, NC State University.

Govan, A. Y., Langville, A. N., & Meyer, C. D. (2009). Offense-Defense Approach to Ranking Team Sports. *Journal of Quantitative Analysis in Sports*, *5*(1). doi:10.2202/1559-0410.1151

Hunter, D. R. (2004). MM algorithms for generalized Bradley-Terry models. The *annals of statistics, 32(1)*, 384–406. doi: 10.1214/aos/1079120141

Ingram, M. (2021). How to extend Elo: a Bayesian perspective. *Journal of Quantitative Analysis in Sports*, *17*(3), 203–219. doi:10.1515/jqas-2020-0066

Islam, M. M., Khan, J. R., & Raheem, E. (2017). Bradley-Terry Model for Assessing the Performance of Ten ODI Cricket Teams Adjusting for Home Ground Effect. *Journal of Data Science*, *15*(4), 657–668. doi:10.6339/JDS.201710_15(4).00005

Kovalchik, S. (2020). Extension of the Elo rating system to margin of victory. *International Journal of Forecasting*, *36*(4), 1329–1341. doi:10.1016/j.ijforecast.2020.01.006

Langville, A. N., & Meyer, C. D. (2013). *Who's #1?* Princeton University Press.

Lasek, J., Szlávik, Z., & Bhulai, S. (2013). The predictive power of ranking systems in association football. *International Journal of Applied Pattern Recognition*, *1*(1), 27–46. doi:10.1504/ijapr.2013.052339

Leitner, C., Zeileis, A., & Hornik, K. (2010). Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the EURO 2008. *International Journal of Forecasting*, *26*(3), 471–481. doi:10.1016/j.ijforecast.2009.10.001

Leung, C. K., & Joseph, K. W. (2014). Sports Data Mining: Predicting Results for the College Football Games. *Procedia Computer Science*, *35*, 710–719. doi:10.1016/j.procs.2014.08.153

Looijen, B. (2019). *A Comparison Between Poisson Regression And The Bradley-Terry Model To Predict The Rank Order Of The 2019 Eurohockey Championship*. Bachelor's project mathematics, University of Groningen.

Massey, K. (1997). *Statistical models applied to the rating of sports teams*. Honors project, Bluefield College.

Paul, R. J., & Wilson, M. (2012). Political Correctness, Selection Bias, and the NCAA Basketball Tournament. *Journal of Sports Economics*, *16*(2), 201–213. doi:10.1177/1527002512465413

Sanders, S. (2007). A cheap ticket to the dance: Systematic bias in college basketball's rating percentage index. *Economics Bulletin*, *4*(34), 1–7.

Servien, R. (2022). An alternative ranking system for biathlon pursuit races. *Journal of Sports Analytics*, *8*(2), 141–148. doi:10.3233/jsa-200598

Standifird, S. S. (2005). Reputation Among Peer Academic Institutions: An Investigation of the US News and World Report's Rankings. *Corporate Reputation Review*, *8*(3), 233–244. doi:10.1057/palgrave.crr.1540252

Stefani, R. (2011). The Methodology of Officially Recognized International Sports Rating Systems. *Journal of Quantitative Analysis in Sports*, *7*(4). doi:10.2202/1559-0410.1347

Stewart, C., Mazel, M., & Sadler, B. (2022). Application of Probabilistic Ranking Systems on Women's Junior Division Beach Volleyball. *SMU Data Science Review*, *6*(2), 1–19.

Stocks-Smith, J. (2021). College Basketball Rating (CBR): A new body-of-work metric for NCAA tournament selection. *Journal of Sports Analytics*, *7*(1), 47–55. doi:10.3233/jsa-200457

Stone, D. F., & Arkes, J. (2018). MARCH MADNESS? UNDERREACTION TO HOT AND COLD HANDS IN NCAA BASKETBALL. *Economic Inquiry*, *56*(3), 1724–1747. doi:10.1111/ecin.12558

Szczecinski, L. (2022). G-Elo: generalization of the Elo algorithm by modeling the discretized margin of victory. *Journal of Quantitative Analysis in Sports*, *18*(1), 1–14. doi:10.1515/jqas-2020-0115

Szczecinski, L., & Roatis, I.-I. (2022). FIFA ranking: Evaluation and path forward. *Journal of Sports Analytics*, 1–20. doi:10.3233/jsa-200619

Vaziri, B., Dabadghao, S., Yih, Y., & Morin, T. L. (2018). Properties of sports ranking methods. *Journal of the Operational Research Society*, *69*(5), 776–787. doi:10.1057/s41274-017-0266-8

Whelan, J. T., & Wodon, A. (2020). Prediction and evaluation in College Hockey using the Bradley–Terry–Zermelo model. *Mathematics for Application*, *8*(2), 131–149. doi:10.13164/ma.2019.09

Williams, L. V., Liu, C., Dixon, L., & Gerrard, H. (2020). How well do Elo-based ratings predict professional tennis matches? *Journal of Quantitative Analysis in Sports*, *17*(2), 91–105. doi:10.1515/jqas-2019-0110

Zhang, X., Luo, H., Chen, B., & Guo, G. (2020). Multi-view visual Bayesian personalized ranking for restaurant recommendation. *Applied Intelligence*, *50*(9), 2901–2915. doi:10.1007/s10489-020-01703-6

APPENDIX A: SCHOOL ABBREVIATIONS

Below are the abbreviations for all school names invoked in this paper. This is not a complete list of the schools incorporated in the simulation. These are also not necessarily the official school abbreviations.

- ALB (The Albany Academy)
- AND (Phillips Academy - Andover)
- AVO (Avon Old Farms School)
- BEL (Belmont Hill School)
- BER (Berwick Academy)
- BRO (Brooks School)
- BRU (Brunswick School)
- CHO (Choate Rosemary Hall)
- CSH (Cushing Academy)
- DEE (Deerfield Academy)
- DEX (Dexter Southfield School)
- EXE (Phillips Exeter Academy)
- GUN (formerly The Gunnery, now The Frederick Gunn School)
- HOT (The Hotchkiss School)
- KEN (Kent School)
- KUA (Kimball Union Academy)
- LAW (Lawrence Academy)
- LOO (Loomis Chaffee School)
- MIL (Millbrook School)
- NMH (Northfield Mount Hermon)
- NOB (Noble and Greenough School)
- PAU (St. Paul's School)
- PRO (Proctor Academy)
- RIV (The Rivers School)
- SAL (Salisbury School)
- SEB (St. Sebastian's School)
- TAB (Tabor Academy)
- TAF (The Taft School)
- THA (Thayer Academy)
- TIL (Tilton School)
- TPS (Trinity-Pawling School)
- WES (Westminster School)
- WIL (The Williston Northampton School)
- WIN (The Winchendon School)

APPENDIX B: SCHEDULE MATRIX

Table B1

Team Schedules and Playoff Percentages 2018–19

	SAL	DEX	KUA	TIL	GUN	AND	SEB	NMH	CSH	LAW	EXE	DEE	KEN	NOB	WES	PRO
Rank	1	2	3	4	5	6	7	8	8	10	10	12	13	14	15	16
Rating	0.898	0.88	0.85	0.773	0.767	0.752	0.751	0.731	0.731	0.728	0.728	0.717	0.715	0.713	0.71	0.687
BT Playoff %	99.06%	97.02%	92.22%	52.72%	50.18%	36.66%	41.72%	31.86%	27.68%	28.52%	24.76%	19.54%	22.36%	21.36%	22.68%	14.98%
JSPR Playoff %	97.82%	97.44%	94.92%	56.40%	47.68%	33.18%	57.78%	28.88%	23.48%	33.50%	21.20%	13.86%	14.20%	21.08%	19.66%	19.14%
SAL		0	1	0	1	1	0	0	1	0	0	2	2	1	1	0
DEX	0		0	0	0	2	0	1	2	1	2	0	0	1	0	1
KUA	1	0		3	0	1	0	1	2	0	1	2	0	1	0	3
TIL	0	0	3		0	1	0	1	0	0	1	0	0	0	0	2
GUN	1	0	0	0		0	0	0	0	0	1	0	2	0	1	0
AND	1	2	1	1	0		0	1	1	0	2	1	0	1	1	1
SEB	0	0	0	0	0	0		0	0	2	1	0	0	2	0	1
NMH	0	1	1	1	0	1	0		2	1	1	1	0	0	1	0
CSH	1	2	2	0	0	1	0	2		2	1	1	0	0	0	0
LAW	0	1	0	0	0	0	2	1	2		0	0	0	2	0	0
EXE	0	2	1	1	1	2	1	1	1	0		1	1	0	0	1
DEE	2	0	2	0	0	1	0	1	1	0	1		1	1	2	0
KEN	2	0	0	0	2	0	0	0	0	0	1	1		0	1	0
NOB	1	1	1	0	0	1	2	0	0	2	0	1	0		1	0
WES	1	0	0	0	1	1	0	1	0	0	0	2	1	1		0
PRO	0	1	3	2	0	1	1	0	0	0	1	0	0	0	0	
17–28	11	4	5	2	13	4	5	4	5	3	2	10	13	6	11	5
29–40	1	5	9	8	3	9	12	3	7	11	8	1	2	9	3	10
41–54	2	7	2	8	4	1	2	7	5	3	4	2	2	2	3	7

Table B2

Team Schedules and Playoff Percentages 2017–18

	DEX	SAL	KUA	RIV	MIL	EXE	WIN	CSH	THA	WES	SEB	GUN	TP	NMH	BRU	WIL
Rank	1	1	3	3	3	6	7	8	9	9	9	9	9	14	14	14
Rating	0.897	0.897	0.822	0.822	0.822	0.82	0.819	0.816	0.717	0.717	0.717	0.717	0.717	0.713	0.713	0.713
BT Playoff %	95.66%	96.24%	67.92%	66.86%	64.06%	64.70%	67.78%	65.44%	12.94%	11.92%	11.50%	11.46%	14.94%	11.28%	10.32%	12.54%
JSPR Playoff %	90.84%	96.56%	69.44%	71.90%	63.64%	58.94%	70.28%	52.64%	6.46%	16.46%	11.22%	15.06%	15.16%	15.50%	15.16%	14.20%
DEX		0	0	1	0	1	1	2	2	0	0	0	0	0	0	0
SAL	0		1	0	1	0	0	1	0	1	0	1	2	0	1	1
KUA	0	1		1	1	1	0	2	0	1	0	0	0	1	1	0
RIV	1	0	1		0	1	0	0	1	0	2	0	0	0	3	1
MIL	0	1	1	0		0	1	0	2	0	2	0	0	0	0	0
EXE	1	0	1	1	0		1	1	1	0	1	1	0	1	0	0
WIN	1	0	0	0	1	1		1	0	0	1	1	0	1	0	3
CSH	2	1	2	0	0	1	1		1	0	0	0	0	2	0	1
THA	2	0	0	1	2	1	0	1		0	2	0	0	1	0	0
WES	0	1	1	0	0	0	0	0	0		0	1	1	1	1	1
SEB	0	0	0	2	2	1	1	0	2	0		0	0	0	0	1
GUN	0	1	0	0	0	1	1	0	0	1	0		1	0	2	1
TP	0	2	0	0	0	0	0	0	0	1	0	1		0	2	0
NMH	0	0	1	0	0	1	1	2	1	1	0	0	0		0	0
BRU	0	1	1	3	0	0	0	0	0	1	0	2	2	0		1
WIL	0	1	0	1	0	0	3	1	0	1	1	1	0	0	1	
17–28	6	7	7	5	9	5	4	8	8	10	9	8	8	8	6	8
29–40	5	8	5	3	4	6	6	4	4	8	3	6	8	6	3	4
41–54	6	0	8	10	3	9	4	6	4	0	5	2	1	5	1	1

Table B3

Team Schedules and Playoff Percentages 2016–17

	SAL	THA	LAW	KUA	DEX	GUN	EXE	KEN	AVO	LOO	RIV	WIN	AND	TAF	BEL	SEB
Rank	1	2	3	3	5	6	7	7	9	9	11	11	13	14	15	16
Rating	0.84	0.809	0.794	0.794	0.789	0.783	0.779	0.779	0.776	0.776	0.75	0.75	0.719	0.707	0.705	0.681
BT Playoff %	93.02%	77.62%	43.20%	43.10%	36.76%	42.26%	39.94%	49.72%	30.74%	54.18%	45.98%	52.46%	7.52%	22.60%	2.90%	13.44%
JSPR Playoff %	90.24%	72.78%	50.88%	50.98%	44.66%	41.52%	47.40%	34.70%	21.24%	46.18%	46.90%	57.50%	8.76%	11.86%	3.90%	15.30%
SAL		0	0	1	0	1	0	2	1	1	0	0	0	2	0	0
THA	0		2	0	1	0	0	0	0	0	1	1	1	0	2	1
LAW	0	2		0	0	1	0	0	0	0	0	0	0	0	2	2
KUA	1	0	0		0	0	1	0	0	0	1	0	1	0	0	0
DEX	0	1	0	0		1	1	0	0	0	1	1	1	0	0	0
GUN	1	0	1	0	1		1	2	1	1	0	0	0	0	0	0
EXE	0	0	0	1	1	1		1	0	0	1	1	1	0	0	1
KEN	2	0	0	0	0	2	1		1	2	0	0	0	2	0	0
AVO	1	0	0	0	0	1	0	1		3	0	0	0	2	0	0
LOO	1	0	0	0	0	1	0	2	3		0	0	1	1	0	0
RIV	0	1	0	1	1	0	1	0	0	0		0	0	0	2	1
WIN	0	1	0	0	1	0	1	0	0	0	0		1	0	0	1
AND	0	1	0	1	1	0	1	0	0	1	0	1		0	1	1
TAF	2	0	0	0	0	0	0	2	2	1	0	0	0		0	0
BEL	0	2	2	0	0	0	0	0	0	0	2	0	1	0		2
SEB	0	1	2	0	0	0	1	0	0	0	1	1	1	0	2	
17–28	6	9	7	10	8	7	6	4	8	6	2	7	6	3	7	9
29–40	10	3	5	8	5	10	6	10	8	9	3	3	8	12	4	2
41–58	0	4	6	5	7	1	6	0	0	0	13	7	3	0	6	4

Table B4
Team Schedules and Playoff Percentages 2015–16

	SAL	AVO	KUA	EXE	BRU	MIL	CSH	GUN	TAB	ALB	THA	LOO	WES	NOB	RIV	DEE
Rank	1	1	3	4	5	6	6	6	6	6	6	12	13	14	14	14
Rating	0.934	0.934	0.9	0.878	0.834	0.799	0.799	0.799	0.799	0.799	0.799	0.778	0.729	0.694	0.694	0.694
BT Playoff %	99.44%	99.40%	94.02%	86.42%	61.00%	37.04%	30.98%	56.84%	34.32%	34.50%	33.70%	25.32%	8.00%	2.02%	3.90%	2.46%
JSPR Playoff %	98.66%	98.28%	94.32%	89.48%	59.16%	37.96%	33.40%	51.56%	33.94%	26.96%	56.90%	13.28%	3.02%	5.68%	16.16%	0.38%
SAL		1	1	0	1	1	1	1	0	1	0	1	1	0	0	2
AVO	1		0	0	0	0	1	1	1	1	0	2	2	0	0	2
KUA	1	0		1	0	2	2	0	1	0	0	0	1	0	0	2
EXE	0	0	1		1	0	2	1	1	1	0	0	0	0	0	1
BRU	1	0	0	1		0	0	1	0	3	0	1	1	0	1	0
MIL	1	0	2	0	0		0	0	1	0	2	0	0	2	0	1
CSH	1	1	2	2	0	0		1	2	0	1	0	1	0	0	1
GUN	1	1	0	1	1	0	1		1	0	1	1	1	0	0	0
TAB	0	1	1	1	0	1	2	1		0	1	1	0	1	0	0
ALB	1	1	0	1	3	0	0	0	0		1	0	0	0	1	0
THA	0	0	0	0	0	2	1	1	1	1		0	0	2	1	0
LOO	1	2	0	0	1	0	0	1	1	0	0		2	0	0	2
WES	1	2	1	0	1	0	1	1	0	0	0	2		1	0	2
NOB	0	0	0	0	0	2	0	0	1	0	2	0	1		1	1
RIV	0	0	0	0	1	0	0	0	0	1	1	0	0	1		0
DEE	2	2	2	1	0	1	1	0	0	0	0	2	2	1	0	
17–28	8	8	3	6	8	7	7	12	7	7	9	7	6	9	7	4
29–40	6	4	9	8	4	5	6	3	5	2	7	6	5	6	9	5
41–58	0	2	8	7	4	3	5	2	6	5	2	2	2	4	6	2

Table B5

Team Schedules and Playoff Percentages 2014–15

	CSH	EXE	SAL	GUN	NOB	KUA	BRU	LOO	AVO	DEX	BRO	HOT	CHO	WES	KEN	BEL
Rank	1	1	3	3	3	3	3	3	3	3	11	12	12	12	12	12
Rating	0.898	0.898	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.807	0.717	0.717	0.717	0.717	0.717
BT Playoff %	78.00%	84.08%	68.28%	42.80%	55.38%	42.24%	30.98%	56.84%	22.92%	37.22%	28.90%	8.26%	5.68%	30.84%	10.34%	2.72%
JSPR Playoff %	74.36%	84.98%	56.84%	39.30%	59.48%	52.00%	33.40%	51.56%	16.66%	43.60%	41.02%	4.68%	4.20%	22.28%	5.58%	3.82%
CSH		2	1	1	0	2	0	0	1	1	0	0	0	0	0	1
EXE	2		0	1	0	1	1	0	0	1	1	1	0	0	1	0
SAL	1	0		1	1	1	1	1	1	0	0	2	1	2	2	0
GUN	1	1	1		0	0	0	1	1	1	0	1	0	1	2	0
NOB	0	0	1	0		1	0	0	0	1	0	1	0	0	0	2
KUA	2	1	1	0	1		0	0	0	0	0	0	0	0	0	0
BRU	0	1	1	0	0	0		1	0	0	0	1	1	1	1	1
LOO	0	0	1	1	0	0	1		2	0	0	1	2	2	2	0
AVO	1	0	1	1	0	0	0	2		0	0	1	2	2	2	0
DEX	1	1	0	1	1	0	0	0	0		1	0	0	0	0	0
BRO	0	1	0	0	0	0	0	0	0	1		0	0	0	0	1
HOT	0	1	2	1	1	0	1	1	1	0	0		1	2	2	0
CHO	0	0	1	0	0	0	1	2	2	0	0	1		2	1	1
WES	0	0	2	1	0	0	1	2	2	0	0	2	2		1	0
KEN	0	1	2	2	0	0	1	2	2	0	0	2	1	1		0
BEL	1	0	0	0	2	0	1	0	0	0	1	0	1	0	0	
17–28	7	6	7	9	9	7	7	7	9	6	3	7	6	6	7	8
29–40	8	7	4	4	5	13	5	5	3	7	11	4	6	4	3	5
41–58	5	7	1	1	5	4	5	2	2	7	10	1	2	2	1	7

Table B6

Team Schedules and Playoff Percentages 2013–14

	KEN	SAL	THA	CSH	GUN	EXE	CHO	BER	NOB	PAU	DEX	WES	NMH	KUA	LOO	DEE
Rank	1	2	3	3	5	5	7	7	7	7	7	7	7	7	7	7
Rating	0.931	0.922	0.86	0.86	0.859	0.859	0.775	0.775	0.775	0.775	0.775	0.775	0.775	0.775	0.775	0.775
BT Playoff %	99.59%	99.33%	74.59%	83.92%	80.53%	81.33%	25.33%	24.29%	24.13%	23.49%	28.87%	19.87%	19.97%	23.58%	21.11%	22.27%
JSPR Playoff %	98.83%	98.03%	90.85%	86.97%	78.19%	84.95%	18.04%	13.32%	36.65%	32.03%	34.56%	12.72%	15.07%	40.37%	15.32%	15.42%
KEN		2	0	0	0	1	1	2	0	1	0	1	0	0	1	1
SAL	2		0	1	1	0	1	1	1	0	0	1	0	1	1	2
THA	0	0		1	0	0	0	0	2	2	0	0	0	0	0	0
CSH	0	1	1		0	2	0	0	0	0	3	0	2	2	0	1
GUN	0	1	0	0		1	0	2	0	1	1	1	1	0	1	0
EXE	1	0	0	2	1		0	0	0	1	1	0	2	1	0	1
CHO	1	1	0	0	0	0		1	0	0	0	2	2	0	2	2
BER	2	1	0	0	2	0	1		1	0	0	1	2	0	0	0
NOB	0	1	2	0	0	0	0	1		2	1	1	0	0	0	0
PAU	1	0	2	0	1	1	0	0	2		1	0	0	0	1	0
DEX	0	0	0	3	1	1	0	0	1	1		0	0	0	0	0
WES	1	1	0	0	1	0	2	1	1	0	0		1	1	2	2
NMH	0	0	0	2	1	2	2	2	0	0	0	1		1	1	1
KUA	0	1	0	2	0	1	0	0	0	0	0	1	1		0	2
LOO	1	1	0	0	1	0	2	0	0	1	0	2	1	0		2
DEE	1	2	0	1	0	1	2	0	0	0	0	2	1	2	2	
17–28	11	10	8	8	9	6	9	11	10	9	3	8	8	8	11	7
29–40	2	2	8	7	3	7	2	2	6	6	6	2	2	7	1	3
41–58	1	2	4	2	3	6	2	4	3	3	8	2	4	8	2	1