Improved Scoring Methods for Unbalanced Athletic Conferences

Tomas, Amber          Yevak, Thomas
ant5j@virginia.edu    tyevak@hra.org

May 13, 2016

Abstract

There are many examples of athletic Cup competitions in which groups of high schools or colleges compete against each other in multiple sports. Typically points are awarded for performances in each sport, and the school with the most points at the end of the season wins the Cup. In this paper we analyze the scoring system of the Tidewater Conference of Independent Schools Headmaster’s Cup, and show that the current system strongly favors larger schools that are able to field teams in many sports. We describe the sources of bias in the scoring system and propose an alternative method of scoring. The revised scoring system is designed to “level the playing field” with regards to school size, and is shown to consistently reward schools that both perform well in the sports in which they participate and have a high participation rate relative to the size of their student body. The methods
of analysis and scoring presented in this paper are also applicable to other
athletic cups, particularly those where schools do not necessarily have teams
competing in every sport.

1 Background

Athletic competitions exist in schools and colleges for multiple reasons: to en-
courage physical activity, team work, and athletic excellence. Often a group of
high schools or colleges will compete against each other in multiple sports, such
as when competitions are arranged between a group of regional schools. In these
competitions there is often a trophy or Cup awarded to the school that shows
the best performance over all the sports contested, and winning this Cup is often
perceived as reflecting overall school-wide excellence. Different interscholastic
competitions have different aims, and are designed to reward particular features
of a school’s athletic program. For example, some competitions might value sim-
ply participating, whereas other competitions aim only to reward winning or out-
standing performance. The Learfield Director’s Cup in College Athletics “hon-
ors institutions maintaining a broad-based program, achieving success in many
sports, both men’s and women’s”\textsuperscript{1}. The Virginia High School League (VHSL)
Wells Fargo Cup “represents excellence”\textsuperscript{2}, and awards points for all sanctioned
sports as well as for “sportsmanship”. The Tidewater Conference of Independent
Schools (TCIS) Headmaster’s Cup awards points for athletic performance as well
\begin{footnotesize}
\begin{enumerate}
\item \url{www.thedirectorscup.com}
\item \url{www.vhsl.org/wells-fargo-cup-standings}
\end{enumerate}
\end{footnotesize}
as for the number of sports in which each school participates. Other similar programs include the CapitalOne Cup\textsuperscript{3} and the Virginia Preparatory League (VPL) Director’s Cup. These both award points to schools based on placings in a set of sports, and the school with the most points wins the Cup.

Most Cup competitions have the following features:

- A set of sports which count towards the Cup;

- A set of participating schools;

- Points are awarded within each sport such that the schools that do better in a sport are awarded more points.

The Cup competitions differ in the way that the winning school is decided, for example the winner might be:

- the school with the most points;

- the school with the most points per sport played;

- the school with the most points or points per sport after some adjustments are made for participation or sportsmanship.

If all schools participate in all sports then the first two bullet points above are equivalent. However, frequently some schools participate in more sports than others. We will see that this is a major source of bias in many of the scoring systems, as schools that play less popular sports are more likely to score highly.

\textsuperscript{3}www.capitalonecup.com
Previous research has looked at how financial resource allocation might influence final standings in college Cup competitions. A study by Katz et al. (2015) indicated that performance in the Learfield Directors’ Cup is positively associated with enrollment and high school GPA. However, we are not aware of any previous work that assesses whether the tendency for large schools to do well is due to intrinsic factors or to a scoring system that is biased in their favor. A commentary by Winthrop Intelligence discussed the different scoring systems used by the CapitalOne Cup and Learfield Directors’ Cup, and how this explains the observed disparity between the schools that tend to do well in each competition. This indicates that although the Cups claim to have similar goals in the characteristics they reward, each scoring system is biased towards particular features of a school’s program. The authors conclude “What can an AD do to climb the standings of either cup? Well, the easy answer is to add more sports.” In other words, the scoring systems are biased towards schools with more sports. Whilst adding more sports may be reasonable for some large, well-funded schools, it is not an option for smaller schools. The authors comment that “shrewd ADs at smaller schools can certainly still elevate their programs’ Capital One Cup scores by targeting Group B sport, or in sports with fewer programs fielding competitive teams”. This strategy exploits another source of bias in the scoring system – that

---

schools in less popular sports can expect to earn more points. We believe that a “fair” scoring system should reward the quality of athletic performance, and not be overly influenced by the resources of the school or the sports they choose to play.

In this paper we look in detail at the scoring system of the TCIS Cup (hereafter referred to as “the Cup”), although the ideas we present and the methods of analysis can easily be applied to the analysis of other Cup competitions and scoring systems. The schools competing in the TCIS Cup have substantial differences in enrollment size and ability to field teams, as shown in Table 1. We will show that these properties can lead to bias in the scoring system, and our analysis will be relevant to any Cup competition seeking to reward athletic excellence among schools with similar resource imbalances.

<table>
<thead>
<tr>
<th>School</th>
<th>Female</th>
<th>Male</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>10</td>
<td>11</td>
<td>21</td>
</tr>
<tr>
<td>BS</td>
<td>10</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>CHC</td>
<td>10</td>
<td>11</td>
<td>21</td>
</tr>
<tr>
<td>PC</td>
<td>8</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>NSA</td>
<td>10</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>NC</td>
<td>9</td>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>HRA</td>
<td>8</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>NCh</td>
<td>6</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>WA</td>
<td>7</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>GCA</td>
<td>4</td>
<td>6</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 1: Size of student enrollment and number of sports played by each school during the 2013-2014 TCIS Cup, for males and females.
Firstly, in Section 2, we describe the current scoring method used in the TCIS Cup and why it is biased in favor of large schools. In Section 3 we propose modifications to the scoring system to reduce bias and better align the scoring system with the goals of the Cup: to reward athletic excellence and encourage participation. The proposed scoring method is more fair in the sense that every school has a chance to win the Cup if their teams perform excellently and they have a high rate of athletic participation, regardless of the size and resources of the school. An example of how the proposed scoring system could have been used in the 2013-2014 season is presented in Section 4. In Section 5 we discuss further bias reduction techniques we considered, and then in Section 6 use simulation to evaluate and justify our recommendations.

2 Current Scoring Method

The current method used in the TCIS Cup to get the final score and rank for each school is as follows:

1. Schools are awarded points for every sport in which they participate. The first placed school in each sport gets 10 points, the runner-up gets 9 points, and subsequent placings are scored as shown in Table 2. If fewer than 10 schools participate, points are still awarded to placings as if all schools were playing, down to the placing of the final school. If there is a tie between two schools, the points available for those places are averaged and awarded to the schools that tied. Note that the TCIS consists of 10 schools, so points
are awarded for every possible placing in a sport.

2. Schools are ranked each season according to the number of sports they participated in that season. Points are then awarded to schools each season as if participation were any other sport.

3. The points awarded to each school are averaged to get the final score:

   \[ \text{final score} = \frac{\text{number of sporting points} + \text{number of participation points}}{\text{number of sports played} + 3} \]

   Note that the 3 on the denominator arises due to the fact that there are three seasons for the TCIS Cup and hence points for participation are awarded three times.

4. The school with the highest final score wins the Cup.

This system has three main sources of bias, all of which arise when some schools play more sports than other schools:

A. The average points awarded to a school that participates in a sport with few participating teams will be higher than the average points awarded to a school that participates in a sport with many teams (see Table 2). This advantages schools that play less popular sports. As shown in Figure 1, this is usually the bigger schools, since they have the numbers to field teams in more sports.

B. When averaging the points of a school to get the final score, the final score of schools that participated in fewer sports will have a bigger variance (since the
Table 2: The points awarded to schools in each sport based on their final placing in that sport. If schools are tied, each school gets the average of the available points for those places. As the number of schools participating in a sport decreases, the average points awarded to participating schools increases, and the variance decreases.

<table>
<thead>
<tr>
<th>Placing</th>
<th>Points</th>
<th>10 Schools</th>
<th>8 Schools</th>
<th>5 Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>5.5</td>
<td>6.5</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>9.2</td>
<td>6</td>
<td>2.5</td>
<td></td>
</tr>
</tbody>
</table>

average is over fewer sports). This means that even if two schools are equal in ability, the school that played fewer sports is more likely to win (or lose!) the Cup. This is illustrated in Figure 2.

C. The participation components (equivalent to three sports – one per season) reward schools that play many sports. This strongly favors the larger schools that have the student numbers and resources to field more teams.

In general, points A and C dominate and so there is a strong bias towards large schools winning the Cup. Since the goals of the TCIS Cup are to reward performance and encourage student participation, changes clearly need to be made.
Figure 1: In the TCIS, the number of sports played by a school increases as enrollment increases (dashed line). The bigger schools are the ones who tend to field teams in the less popular sports, as seen by the solid line that shows the average number of teams in the sports played by each school (2013-2014 season).

In the remainder of this document we describe our proposed changes and give evidence that it leads to a fairer scoring system.

3 Proposed Scoring Method

Our proposed method consists of two components: a performance component and a participation component. The performance component accounts for how well a school does in the sports in which they participated, and the participation compo-
Figure 2: Suppose all teams have equal ability, so the placings within each sport are decided by chance. The dashed curve shows the distribution of the average points awarded to a school that plays 5 sports; the solid line shows the distribution of the average points awarded to a school that plays 10 sports (ignoring the points for participation). This shows that both schools have the same score on average, but that the variance of the final score is bigger for the school that plays fewer sports.

Vcent rewards schools with a high participation rate, i.e. schools that participate in many sports relative to the size of their student body. The final score of school $i$ is calculated as follows:

$$\text{Final Score}_i = \text{Performance}_i \times \text{Participation}_i$$

In the following subsections we discuss how the performance and participation factors are calculated.
3.1 Performance

The following method is used to calculate the performance component of the final score:

1. Schools are awarded points for every sport in which they participate. The first placed school in each sport gets 60 points, the runner-up gets 30 points, and subsequent placings are scored as shown in Table 3. If fewer than 10 schools participate, points are still awarded as in Table 3 down to the placing of the final team. If there is a tie between two teams, the points available for those places are averaged and awarded to the teams that tied.

<table>
<thead>
<tr>
<th>Placing</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Points</td>
<td>60</td>
<td>30</td>
<td>20</td>
<td>15</td>
<td>12</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3: The points awarded to schools in each sport depend on their final placing in that sport.

2. For each school, the final performance score is calculated as a weighted average of the points awarded to it in each sport. The intuitive explanation for using a weighted average is that points obtained in sports in which fewer schools participate should be given less influence in determining the final score, since the results for those sports don’t reflect performance against all the other TCIS schools.

The final performance score for school \( i \) is calculated as follows:

\[
\text{Performance}_i = w_{i1} \times s_{i1} + w_{i2} \times s_{i2} + \ldots + w_{iN} \times s_{iN},
\]
where \( N \) is the total number of sports, \( s_{ij} \) is the number of points awarded to school \( i \) in sport \( j \), and \( w_{ij} \) is the weight of sport \( j \) for school \( i \). Mathematically, the weights are calculated as

\[
 w_{ij} = \begin{cases} 
  \frac{n_j}{\sum_{k \in s(i)} n_k} & \text{if school } i \text{ plays sport } j, \\
  0 & \text{otherwise}
\end{cases}
\]

where \( n_j \) is the number of schools that play sport \( j \), and \( s(i) \) is the set of sports played by school \( i \).

There are two main reasons for switching to the “60-30” scale shown in Table 3. Firstly, it is intuitively reasonable since it gives proportionally more points to schools the better they do, rewarding and motivating excellent performance; schools which are pretty good but rarely win will not be ranked as highly as schools that are outstanding in a few sports and average in others. Secondly, as described in Section 2, bias arises when there are fewer than the full number of schools playing a sport, since the average score of teams playing that sport will be higher than for full participation sports. The proposed 60-30 scale alleviates this somewhat, as shown in Table 4. For example, if there are only 4 schools participating in a sport, then under the 10-9 scale the average school receives 85\% of the points won by the first-placed school; if using the 60-30 scale, the average school receives just over 50\% of the points of the first-placed school. This means that winning is more important with the 60-30 scale, and schools can’t expect to do well just by participating in less popular sports. Similarly, under the current 10-9
Table 4: Points and average points awarded to teams when there are 10 (full participation), 8 or 4 schools participating in a sport, for two point scales.

<table>
<thead>
<tr>
<th>Number of Schools</th>
<th>10-9 scale</th>
<th>60-30 scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10  8  4</td>
<td>10  8  4</td>
</tr>
<tr>
<td>Best Team</td>
<td>10 10 10</td>
<td>60 60 60</td>
</tr>
<tr>
<td>Second-best Team</td>
<td>9   9   9</td>
<td>30 30 30</td>
</tr>
<tr>
<td>Second-Worst Team</td>
<td>2   4   8</td>
<td>4   8   20</td>
</tr>
<tr>
<td>Worst Team</td>
<td>1   3   7</td>
<td>2   6   15</td>
</tr>
<tr>
<td>Average Team</td>
<td>5.5  6.5  8.5</td>
<td>16.7  20.1  31.3</td>
</tr>
</tbody>
</table>

scale the team that comes last in a sport with 8 schools gets 3 points (30% of the winner’s score), but under the 60-30 scale the last place team gets 6 points (only 10% of the winner’s score).

A disadvantage of changing to a 60-30 scale is that it will increase the variance of the final score, so if a team does well in a few sports just by chance, they are more likely to win the Cup compared to when using the 10-9 scale. This trade-off is explored in more detail in Section 6.

Using a weighted average to calculate the final score helps to alleviate the bias mentioned in point A that occurs when not all schools participate in a sport. It also makes intuitive sense: since we want to rank all the schools based on performance, sports where all schools participated against one another should be more informative than those sports where only a few schools participated, so these sports should be given more weight when calculating the final score. The “weight” used to determine the importance of the $j$th sport is proportional to the number of schools that participated in that sport. For example, the weight of a sport in which
all 10 schools participated will be twice as big as the weight of a sport in which only 5 schools participated; thus the points awarded to a school in a sport where all 10 schools participated will have twice as much influence on the final score of that school as the points they were awarded in a sport where only 5 schools participated.

3.2 Participation

If only the performance component is considered, the school that has the best average performance across sports will win the TCIS Cup. However, this inherently favors large schools which have many more students from which to pick their teams. Therefore, the participation component is designed to reward schools that field many teams relative to their size, and encourage student participation.

The participation component for school \( i \) is calculated as follows:

\[
\text{Participation}_i = \frac{\text{Number of sports played by school } i}{\text{Number of students at school } i}.
\]

Note that the number of sports played is not necessarily limited to sports in the TCIS Cup, but may include other sports as well. We also chose to count football as two sports, since the team size is considerably larger for football than for other sports.
4 Example

In this example we repeat the scoring of the 2013-2014 TCIS Cup using the proposed new method. We consider there to be two separate competitions: one for girls and one for boys.

Firstly, we just look at the Fall season, and only for girls. The rankings of each of the schools for the sports played are shown in Table 5.

<table>
<thead>
<tr>
<th>School</th>
<th>G. VB</th>
<th>FH</th>
<th>G. Ten</th>
<th>G. CC</th>
<th>Num. sports</th>
<th>Num. students</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>247</td>
</tr>
<tr>
<td>BS</td>
<td>8</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>4</td>
<td>194</td>
</tr>
<tr>
<td>CHC</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>165</td>
</tr>
<tr>
<td>PC</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>139</td>
</tr>
<tr>
<td>NSA</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>134</td>
</tr>
<tr>
<td>NC</td>
<td>9</td>
<td>6</td>
<td>4</td>
<td></td>
<td>3</td>
<td>121</td>
</tr>
<tr>
<td>HRA</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>114</td>
</tr>
<tr>
<td>NCh</td>
<td>7</td>
<td>5</td>
<td></td>
<td></td>
<td>2</td>
<td>109</td>
</tr>
<tr>
<td>WA</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>97</td>
</tr>
<tr>
<td>GCA</td>
<td>10</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>76</td>
</tr>
</tbody>
</table>

Table 5: Results for girl sports from Fall 2013-2014 (ranking of each school) and participation statistics.

1. First of all, the performance component is calculated for each school. Within each sport points are awarded to schools based on the 60-30 scale given in

\footnote{Although the proposed system can handle draws in the same way as the old system, for simplicity we made slight adjustments to the results for the season so as to remove draws within each sport for this example.}
Table 3. Then the weighted average of the points for each school is calculated. This can be done at the end of each season to give a “running score”.

In Fall the sports offered for girls were volleyball, field hockey, tennis, and cross country. Norfolk Academy (NA) played all four sports and achieved rankings of 6, 1, 2 and 4 respectively. They hence were awarded 10, 60, 30 and 15 points. To calculate the weighted average first the weight of each sport needs to be calculated. The four sports had 10, 7, 8, and 6 teams participating, so the weights of the four sports used to calculate NA’s performance components are \[ \frac{10}{10+7+8+6} = 0.32, \frac{7}{10+7+8+6} = 0.23, \frac{8}{10+7+8+6} = 0.26, \text{and} \frac{6}{10+7+8+6} = 0.19. \] Thus the performance score for NA at the end of the Fall season is

\[
\text{Performance}_{NA} = 0.32 \times 10 + 0.23 \times 60 + 0.26 \times 30 + 0.19 \times 15 \\
= 27.4
\]

Notice that the weight for volleyball (0.32) is bigger than the weight for cross country (0.19), since 10 schools played volleyball but only 6 teams participated in cross country.

As another example consider Norfolk Christian (NCh) who in the Fall only participated in volleyball and tennis and achieved rankings of 7 and 5 respectively. They receive 8 points for 7th and 12 points for 5th. Since 10 schools played volleyball and 8 schools played tennis the weights used to average the points from each sport are \[ \frac{10}{10+8} = 0.56 \text{ and} \frac{8}{10+8} = \]
Thus the performance score for NCh at the end of the Fall season is

\[
\text{Performance}_{NCh} = 0.56 \times 8 + 0.44 \times 12 = 9.8
\]

2. Next, the Participation component is calculated. This is simply the ratio of the number of sports to the number of students. For example, NA played 4 sports in the Fall season and has 247 students, so their participation score is \(4/247 = .016\). NCh played 2 sports and has 109 students, so their participation score is \(2/109 = .018\). Even though NA played more sports than NCh, NCh gets a higher participation score since they played more sports per student.

3. Finally, the final score for a school is calculated as the product of their performance and participation scores. For example, NA had a performance score of 27.4 and a participation score of .016, so their final score is \(27.4 \times .016 = 0.44\). NCh had a performance score of 9.8 and a participation score of .018, so their final score is \(9.8 \times .018 = 0.18\).

The performance, participation, and final scores and rank of all schools for the girls Fall competition are shown in Table 6.

We also calculated the final scores and rankings for girls and boys at the end of the whole year. These results are shown in Tables 7 and 8 respectively. Note that although intermediate scores can be calculated at the end of the Fall season
School Performance Participation Final Score

<table>
<thead>
<tr>
<th>School</th>
<th>Performance</th>
<th>Participation</th>
<th>Final Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>27.4 (4)</td>
<td>.016 (9)</td>
<td>.44 (4)</td>
</tr>
<tr>
<td>BS</td>
<td>14.3 (6)</td>
<td>.02 (7)</td>
<td>.30 (7)</td>
</tr>
<tr>
<td>CHC</td>
<td>30.6 (1)</td>
<td>.024 (5)</td>
<td>.7 (2)</td>
</tr>
<tr>
<td>PC</td>
<td>29.5 (2)</td>
<td>.022 (6)</td>
<td>.64 (3)</td>
</tr>
<tr>
<td>NSA</td>
<td>28.5 (3)</td>
<td>.030(2)</td>
<td>.85 (1)</td>
</tr>
<tr>
<td>NC</td>
<td>9.2 (9)</td>
<td>.025 (4)</td>
<td>.23 (8)</td>
</tr>
<tr>
<td>HRA</td>
<td>16.6 (5)</td>
<td>.026 (3)</td>
<td>.43(5)</td>
</tr>
<tr>
<td>NCh</td>
<td>9.8(8)</td>
<td>.018 (8)</td>
<td>.18 (9)</td>
</tr>
<tr>
<td>WA</td>
<td>10.8 (7)</td>
<td>.03 (1)</td>
<td>.33 (6)</td>
</tr>
<tr>
<td>GCA</td>
<td>2.0 (10)</td>
<td>.013 (10)</td>
<td>.026 (10)</td>
</tr>
</tbody>
</table>

Table 6: Results for girl sports at the end of the Fall season. Ranks are shown in parentheses.

(as above) and the Winter season, the final score is not equal to the sum of the three seasonal scores. This is because fewer sports might be offered in some seasons compared to others, so it is not “fair” to treat scores from each season equally. However, the intermediate scores do give a clear indication of how schools are doing and motivation to improve or maintain their performance and/or participation when possible.

Nansemond-Suffolk (NSA) win the girls competition due to a solid performance component and the highest ratio of sports to students in the competition. NA and CHC both did better than NSA in performance, but had relatively low participation scores.

In the boys competition Norfolk Academy (NA) would win due to a very high performance score – this is just enough to keep them ahead of CHC in second, even though NA has the worst participation component in the competition.
Table 7: Results for girl sports using the proposed scoring method at the end of the 2013-2014 season. Ranks are shown in parentheses.

<table>
<thead>
<tr>
<th>School</th>
<th>Performance</th>
<th>Participation</th>
<th>Final Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>33.9 (1)</td>
<td>0.04 (10)</td>
<td>1.37 (3)</td>
</tr>
<tr>
<td>BS</td>
<td>12.8 (7)</td>
<td>0.052 (9)</td>
<td>0.6 (9)</td>
</tr>
<tr>
<td>CHC</td>
<td>27.1 (2)</td>
<td>0.061 (6)</td>
<td>1.64 (2)</td>
</tr>
<tr>
<td>PC</td>
<td>16.1 (5)</td>
<td>0.058 (7)</td>
<td>0.93 (5)</td>
</tr>
<tr>
<td>NSA</td>
<td>25.1 (3)</td>
<td>0.075 (1)</td>
<td>1.87 (1)</td>
</tr>
<tr>
<td>NC</td>
<td>11.9 (8)</td>
<td>0.074 (2)</td>
<td>0.88 (6)</td>
</tr>
<tr>
<td>HRA</td>
<td>16.98 (4)</td>
<td>0.07 (4)</td>
<td>1.19 (4)</td>
</tr>
<tr>
<td>NCh</td>
<td>9.9 (10)</td>
<td>0.064 (5)</td>
<td>0.64 (10)</td>
</tr>
<tr>
<td>WA</td>
<td>10.8 (9)</td>
<td>0.072 (3)</td>
<td>0.78 (8)</td>
</tr>
<tr>
<td>GCA</td>
<td>15.9 (6)</td>
<td>0.053 (8)</td>
<td>0.84 (7)</td>
</tr>
</tbody>
</table>

5 Other Adjustments Considered

The previous sections describe the final method we decided upon. The final method is designed to be as fair and transparent as possible. In this section we describe some other variants that were considered, and give our rationale for deciding not to use them for the TCIS Cup.

5.1 Standardizing and Smoothing

As described in Section 2, bias arises when there are fewer than the full number of schools playing a sport, since the average score of teams playing that sport will be higher than for full participation sports. The proposed 60-30 scale reduces but does not remove this bias. In this section we describe another adjustment we considered that further reduces this bias. However, for the TCIS Cup it was
Table 8: Results for boy sports using the proposed scoring method at the end of the 2013-2014 season. Ranks are shown in parentheses.

ultimately decided that the added complexity of these methods (and associated loss of transparency) outweighed the benefit.

The adjustment has two parts:

1. (Standardize) Within each sport, adjust the points so that the points awarded have the same mean and variance regardless of how many schools are participating.

2. (Smooth) “Smooth” (or “shrink”) the final score of each school towards the mean.

5.1.1 Standardization

The standardization step is designed to ensure that if all schools are of equal ability, the points they can expect to get from participating in a less-than-full-participation sport (i.e. a sport that not all schools play), is the same as the num-
ber of points they can expect to get from participating in a full-participation sport (i.e. a sport that all schools play). This removes the bias described above, since now the expected score of all schools will be the same, regardless of which sports they participate in. We also adjust the points so that points awarded within a sport have the same variance. If we did not do this then it would advantage schools participating in sports where the points have a higher variance.

Specifically, suppose the mean and variance of points awarded in a full-participation sport are $\mu$ and $\sigma^2$. Let $\mu_{n_j}$ and $\sigma^2_{n_j}$ be the mean and variance of the (unadjusted) points awarded in a sport with $n_j$ schools participating, i.e. the mean and variance of the first $n_j$ values of the point scale. Then if $s_{ij}$ is the points awarded to school $i$ in sport $j$, the standardized score $s^*_{ij}$ is

$$s^*_{ij} = s_{ij} \frac{\sigma}{\sigma_{n_j}} - \mu_{n_j} \frac{\sigma}{\sigma_{n_j}} + \mu$$

Standardized points for the case of sports with $n_j = 5$ and $n_j = 8$ teams participating are shown in Table 9. Notice that standardizing ensures the points available for all sports have the same mean and variance, regardless of the number of teams participating. Although standardizing equalizes the average points on offer in each sport, it does have some disadvantages. If a school is truly outstanding in some sport, the number of points they receive for winning a less-than-full participation sport is less than if they were to win a full participation sport. Similarly, if a school is truly terrible, they will still get more points for coming last in a less-than-full participation sport than if the sport was full-participation. However, this
<table>
<thead>
<tr>
<th>Placing</th>
<th>10 Schools u/s</th>
<th>8 Schools u/s</th>
<th>8 Schools s</th>
<th>5 Schools u/s</th>
<th>5 Schools s</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60</td>
<td>60</td>
<td>55.4</td>
<td>60</td>
<td>45.7</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>30</td>
<td>26.3</td>
<td>30</td>
<td>19.0</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>20</td>
<td>16.6</td>
<td>20</td>
<td>10.1</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>15</td>
<td>11.7</td>
<td>15</td>
<td>5.7</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>12</td>
<td>8.8</td>
<td>12</td>
<td>3.0</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>10</td>
<td>6.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>8</td>
<td>4.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>6</td>
<td>3.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Variance</th>
<th>Variance</th>
<th>Variance</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16.7</td>
<td>300</td>
<td>20.1</td>
<td>318.4</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>16.7</td>
<td>300</td>
<td>379</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>27.4</td>
<td>300</td>
<td>300</td>
</tr>
</tbody>
</table>

Table 9: Points awarded to places if standardization is used (s), compared to using unstandardized points (u/s). Note that for full-participation sports standardizing has no effect – it only affects points awarded in sports where not all schools participate.

The last statement is true of the unadjusted scoring method too (see Table 9), and standardization substantially reduces this effect. These disadvantages will be partly offset by weighting, since more weight is given to points from full participation sports.

5.1.2 Smoothing

Although standardizing removes bias due to sports with less-than-full participation, if we were only to standardize then the scoring system would still be biased – in favor of schools that play fewer sports. This is because the final score is calculated as an average: the variance of the final score for schools that play few
sports will be larger than the variance of the final score for schools that play many sports (see Figure 2). If the final score of schools that play few sports has a larger variance, this means that they are likely to get high (or low) final scores more often than schools that play more sports. Smoothing is a variance reduction technique that helps to equalize the variance of the average across schools. The idea of smoothing is to adjust the points awarded to a school so that the points are dragged a little bit towards the mean – the further the points are “smoothed” towards the mean, the larger the reduction in variance. To try and equalize the variance of the final scores for each school, points awarded to schools that participated in few sports are smoothed more than points awarded to schools that participated in many sports. Specifically, if $s^*_i$ is the final score for school $i$ (assuming that standardization and weighted averaging were used), the adjusted points awarded to that school, $\tilde{s}_i$, can be written

$$\tilde{s}_i = \lambda_is^*_i + (1 - \lambda_i)s^*, \tag{23}$$

where $s^*$ is the average points awarded in each sport after standardizing, and the smoothing parameter $\lambda_i$ is a number between 0 and 1 that controls how much the score is smoothed for school $i$. We considered several options for choosing the value of $\lambda_i$, all of the form

$$\lambda_i = \left( \frac{\text{number of sports played by school } i}{\text{total number of sports}} \right)^\alpha$$

and found that $\alpha = 0.4$ worked well for the TCIS data we had available.
The main drawback of smoothing is that if a smaller school is truly above average, smoothing will drag their score down a little way towards the average (and similarly it will slightly advantage small schools that are below average by dragging their final score up a little way towards the average). However, we expect the advantages of smoothing to outweigh the disadvantages, particularly in competitions where there is a large disparity in the number of sports played by different schools.

5.2 Points Scales

We considered the following four point scales:

A. 10, 9, 8, 7, 6, 5, 4, 3, 2, 1 (current point scale)

B. 20, 10, 5, 3, 1, 0, 0, 0, 0, 0

C. 60, 30, 20, 15, 12, 10, 8, 6, 4, 2 (proposed point scale)

D. 20, 12, 10, 8, 6, 5, 4, 3, 2, 1

Scale A is a “one-point” or “regular” scale. Using A is equivalent to any regularly spaced scale, for example the 50-45-40-35-...-5 scale used in the VHSL Wells Fargo Cup. The other scales above are examples of “increasing” scales; these give increasing rewards for each additional placing. Using an increasing scale has the following consequences compared to a one-point scale:

1. A school that excels and tends to win many sports is more likely to win the cup;
2. Schools that are outstanding in a few sports and average in others are more likely to win compared to schools that are pretty good in many sports but rarely win;

3. The variance of the final score for each school will increase since the variance of scores for each sport is bigger.

The performance component of the TCIS Cup is designed to reward excellence, which is a strong reason for proposing an increasing scale. Scales B and C reduce bias from sports in which not all schools participate compared to scales A or D (see Section 3.1). We decided to use scale C rather than scale B because it gives all schools who participate in a sport some reward, which makes for more interesting competition between schools at the lower end of the score range as well as between those at the top.

It should also be noted that it is the relative interval between the points for each place that determine the properties of the scale. For example, using a 60-30-20-15-... scale is no different from a 120-60-40-30-... or 1-0.5-0.33-0.25-... scale, in the sense that the final placings in the Cup will be the same for each.

6 Simulations

In this section we present results from simulations that are designed to provide additional justification for the claims and proposed changes presented in previous sections. The simulations show the proportion of times each school would win the TCIS Cup under the following conditions:
• The sports in which schools participate and the number of students enrolled at each school is the same as the 2013–2014 season.

• The season is repeated 10,000 times.

For each simulation we consider two scenarios regarding the athletic ability of the schools:

1. All schools and teams are equal in ability. In this scenario every team that competes in a sport has an equal chance of winning, i.e. the ranks of schools in each sport are assigned at random. If the scoring system is fair and based only on performance, in this situation each school should win the Cup 10% of the time.

2. Teams from the schools Greenbrier Christian Academy (GCA) and Bishop Sullivan CHS (BS) are more able than teams from other schools. Both GCA and BS are randomly assigned a top-5 placing in each sport in which they participate, and then the remaining placings are randomly assigned to other schools. We chose GCA and BS as they participate in the smallest and largest number of sports among TCIS schools respectively. If the scoring system is fair and based only on performance, in this situation both GCA and BS should win the Cup 50% of the time.

The results of each simulation are displayed as a bar chart: the height of a bar is equal to the proportion of times (in 10,000 simulated seasons) that school won the TCIS Cup. The bars are ordered the same in every plot – the school participating...
in the fewest number of TCIS sports is on the left, and the school participating in the most TCIS sports is on the right (in the case of a tie, the school with the lowest enrollment is further left). This allows for easy observation of bias due to the number of sports played, and also approximately orders the bars by size of school enrollment.

6.1 Simulation Results

Simulation results using the current scoring method of the TCIS Cup are shown in Figure 3. This shows a clear bias in favor of bigger schools. Even after removing the seasonal participation components, Figure 4 shows that the bias in favor of big schools remains. As discussed earlier, this is largely because bigger schools participate in sports in which fewer other schools participate. We ran our simulations assuming that every school plays every sport (not shown), and in this case the results were practically free of bias – the bias in the scoring system arises only because not all schools play every sport.

Figure 5 shows the effect of switching from the 10-9 scale (evenly spaced) to the 60-30 scale (increasing). Compared to the results in Figure 4, it can be seen that simply changing the point scale results in significant improvements. The proposed points scale reduces the bias in favor of schools participating in sports with fewer teams, for reasons discussed in Section 3.1. However, the 60-30 scale does result in a decrease in performance for Case 2: although GCA and BS both have an equal chance of winning (around 0.45 and 0.46 respectively), spreading out the point scale means that other schools also have a small chance of winning
because of the increase in the variance of points available for each sport.

The results of simulating our proposed method for calculating the performance component are shown in Figure 6. This is the same method as used in Figure 5 but additionally using weighted averaging to calculate the final score. Since points from sports with less than full participation are given less weight, this helps to reduce the bias further.

Compared to Figure 5, the results of the proposed scoring method shown in Figure 6 slightly favors schools that play fewer sports. This is due to the average score of these schools having a larger variance than the average for schools that play more sports. Figure 7 shows how standardizing and smoothing can help to reduce this source of bias. However, this figure also shows that in the case that some schools are better than others, smoothing increases the bias towards schools that play more sports: the smaller school GCA has less of a chance of winning
Figure 4: Using the scoring system as implemented in 2013-2014, but with the seasonal participation components removed.

than the bigger school BS because smoothing drags their points more strongly towards the mean. However, if the points in every sport were standardized to have the same mean and variance regardless of the number of teams playing but smoothing was not used, the results would look like those displayed in Figure 8. This displays a large bias towards schools that play fewer sports because the variance of their final score is larger than for schools that play more sports.

Finally, Figure 9 shows the simulation results of our proposed method, multiplying the performance component by the participation component. This looks quite different to the results of the current scoring system shown in Figure 3. In Figure 9 the schools that have the highest chance of winning the TCIS Cup are those that both have a good performance and a high rate of participation amongst their students – exactly what our scoring revisions set out to achieve.
Figure 5: Using simple averaging as in Figure 4 to calculate the performance component, but using the 60-30-... point scale. Note that the 60-30 scale helps to remove bias due to less than full participation.

6.2 Summary of Simulation Results for Performance

Table 10 shows the average absolute deviation from the “ideal” situation for simulations that only included performance and not participation. For Case 1, the ideal situation is every school has an equal chance of winning (so the table shows the average absolute deviation from 0.1 for each bar). For Case 2, the ideal situation is that both BS and GCA win 50% of the time. It is not sensible to compare Case 1 with Case 2, but for each case the scoring system that gives the lowest number should be preferred.

In Case 1, when all teams are of equal ability, the proposed method to calculate the performance component has an average absolute deviation of 0.007 compared to an absolute average deviation of 0.018 for the current method. However, due to the increase in variance of points awarded in each sport due to switching to the 60-
Figure 6: Using the 60-30 scale as in Figure 5 to calculate the performance score, but using weighted averaging. This is the proposed system to calculate the performance score.

30 scale, the average absolute deviation for Case 2 increases under the proposed system. However, comparing the bar plots in Figures 4 and 7, in both cases both GCA and BS have approximately equal chance of winning, and a much higher chance of winning than any other school. Standardizing and smoothing further improves the performance on Case 1, but does worse than the proposed method on Case 2. These results are consistent with the characteristics of the adjustments considered. It is also interesting that standardizing the points available in each sport to have the same mean and variance regardless of the number of teams participating is almost as biased as the current method on Case 1, and much more biased than the current method on Case 2.
Figure 7: Using the proposed method to calculate the performance component, but with the addition of standardizing and smoothing.

7 Discussion

In this paper we propose a method that can be used to improve the current scoring method for the TCIS Cup. The proposed final score of school $i$ is of the form

$$\text{Final Score}_i = \text{Performance}_i \times \text{Participation}_i$$

and consists of a performance and a participation component. Depending on the needs of the particular Cup, conceivably other components could be added as well, such as sportsmanship or academic performance. However, in our opinion qualitative performance measures such as sportsmanship should not be combined with quantitative factors, because subjective measures work against the fair and transparent scoring system we are aiming for. In such cases separate trophies could be considered to reward sportsmanship or other attributes.
Figure 8: Using the proposed method to calculate the performance component, but standardizing the points available in each sport to have the same mean and variance, regardless of the number of teams participating.

We showed that current methods of Cup scoring tend to be biased because of the fact that not all schools participate in every sport, and proposed ways to reduce the bias in the performance component due to this phenomenon. Our proposed method is a compromise between reducing bias and maintaining some degree of transparency and simplicity in the scoring method. This method uses an increasing scale and weighted averaging to reduce the bias, but it does not remove the bias completely – if the discrepancy between the number of teams playing each sport is very large, then the bias in favor of schools that play the less popular sports could still be substantial. For Cup competitions where there is a large discrepancy between the number of sports played by different schools, administrators might like to give more consideration to the standardization and smoothing adjustments described in Section 5, or perhaps to consider dividing the schools into more equally
matched groups, each with their own Cup competition. For example, problems may arise when some schools eligible for the Cup are co-ed and others are boys or girls schools. This causes problems for the performance component because it increases the disparity between the number of sports that schools participate in, and hence the bias associated with this phenomenon. In such cases it may be best to have separate competitions for boys and girls.

Most of this paper discussed methods to reduce bias and level the playing field in the *performance* component of the proposed final score. However, there are several practical details that need to be given appropriate consideration for determining the *participation* component, in particular:

1. The best way to account for disparate opportunities to participate for girls and boys;
Table 10: Average absolute deviations of performance components of the score from the ideal situation. These numbers correspond to the bar plots shown in Section 6.1.

2. Whether participation in non-Cup sports should contribute.

Considering the sports offered and the number of students enrolled, girls may have more or less opportunity to participate in a sports team than boys at the same school. One way to account for this would be to modify the participation component to consist of two parts: the participation rate of boys, and the participation rate of girls.

Another issue to consider is which sports should contribute to the Cup competition, and whether other sports should be counted in the participation component. Since Cups generally aim to reward the “best” school among a group of schools, it makes sense to only include sports played by a majority of the schools. Other sports should not count towards the Cup because they can not accurately be used to compare performance of all the schools in the Cup. Such a rule would also help to reduce bias due to sports in which only a few schools participate. We feel it is up to the athletic directors and Cup officials to decide exactly how to define participation, depending on if they are trying to motivate any athletic activity (in
which case a broader set of sports could be used for the participation component
than for the performance component), or specifically participation in the sports
that make up the Cup competition.

A central tenant of sport is fairness. Our research has led us to believe that
many Cup scoring systems are not fair, and in this paper we have proposed ways
to level the playing field and give all schools a chance to fight for victory. It is
our hope that Cup administrators will reflect on the ethics of Cup scoring, and
move towards a fairer system that truly reflects the strength of a school’s athletic
program.