Daily Mail Trophy Ranking Methodology

Ian Hamilton, University of Warwick

Professor David Firth, University of Warwick

March 1, 2019

Contents

1 Introduction

This report will analyse the current ranking methodology used in the Daily Mail Trophy and introduce some alternatives for consideration. The Daily Mail Trophy was inaugurated in the 2013/14 season. It met a desire to have an official competition based on the results of schools over the course of a season, akin to a league, rather than a knock-out, format. As the administrator for schools rugby fixtures in the UK, SOCS were given the challenge of devising the rules for this tournament. They have devised a system that seeks to take into account the differing schedule strengths of schools, in order to create a fair ranking. In this report we will look at the current methodology, examining in detail some of the known limitations, and seek to highlight some potential alternative approaches. A good place to start is to agree a set of criteria that we would like to be met by any ranking system. In discussion with SOCS we have identified ten points. Ideally a good ranking system should:

1. be such that the top-ranked team should not be obviously wrong in the opinion of a substantial proportion of stakeholders in the tournament - coaches, players, parents, administrators etc.
2. allow the identification of top-ranked teams in general and how far apart they are, so that a team may have an idea of how far from top position it is.
3. be such that all other relative rankings should not be perceivable as unreasonable by a substantial proportion of the tournament stakeholders.
4. be dependent on the match results of only the current season.
5. be such that any participating school could win.
6. be consistent in the sense that if they were applied to a full round robin they would achieve the same result as with standard rugby union league scoring rules.
7. be such that there is not a requirement for additional fixtures beyond the regular fixture list.
8. be transparent in the sense that it is readily explicable to a generalist audience.
9. allow for a ranking from early in the season.
10. take into account the relative strength of opposition faced, objectively and fairly.

It should be noted that it is likely to be impossible to meet all of these criteria. It remains however useful to be able to refer to them in order to identify strengths and weaknesses of alternative methodologies.

The report will proceed in three parts. We will initially investigate the current methodology, followed by an analysis of some alternative approaches, and finally we will provide a conclusion from our findings.

2 Current methodology

Currently the ranking is based on Merit Points, which are defined as the average number of League Points per match plus Additional Points, awarded in order to adjust for schedule strength.

League Points are awarded as:

4 points for a win
2 points for a draw
0 points for a loss
1 bonus point for losing by less than seven points
1 bonus point for scoring four or more tries

This is the standard scoring rule for rugby union leagues in the UK.

Additional Points in the Daily Mail Trophy are awarded based on the ranking of the current season's opponents in the previous season's tournament:

Rank 1 to 25:	0.3
Rank 26 to 50:	0.2
Rank 51 to 75:	0.1
Otherwise:	0

So, for example, a team with eight fixtures qualifying for the Daily Mail Trophy, with one of those against a top 25 team, three against 26-50th placed teams, and two against 51-75th placed teams, averaging 3.2 League Points per match, would get a Merit Points total of $3.2 + 1 \times 0.3 + 3 \times 0.2 + 2 \times 0.1 = 4.3$. We will refer to this methodology as DMT for the remainder of this report.

In the first part of this report we will look at this current ranking measure and specifically we will discuss three aspects: the use of the previous season's ranking for adjusting for schedule strength; the absolute size of the adjustments applied for the strength of opposition; and the theoretical soundness of the approach.

2.1 Using previous season's ranking

Using the previous season's ranking violates one of the stated criteria (no. 4 in our list) for the Daily Mail Trophy ranking method. However it also has advantages over using the current season to determine schedule strength. For example using the current season in the DMT methodology would mean that points earnt by Team A could be impacted by a match between Teams B and C, which is a concept unfamiliar, and possibly uncomfortable, to stakeholders. For example, the idea that Team A might find itself losing the title on the last day of the season due to the result of a match between Teams B and C where nothing is at stake for either team is not one that is familiar from more common round-robin tournaments (though the perhaps more concerning scenario where a Team A's ranking is dependent on the result of a match between Team B and Team C where only one of the teams has something at stake is a familiar one). Alternatively one could look at a team's record during the season only up to the point at which the match is played in order to address schedule strength. This would be complicated and would then bring greater advantages to particular orderings of matches. And so it is desirable to consider just how large the effect is of using the previous season and to what degree it can be justified.

We start by looking at the persistence of the ranking sections (Top 25, 26-50, 51-75, >75) from one season to the next. If they were maintained in their entirety, or something close to that, then it could be argued that using the previous season would be equivalent to using the current season and there would be no issue.

From Figure 1 we can see that fewer than half of teams maintain their position in the Top 25, with similar proportions maintaining their place in other segments. Since this is based on the current DMT methodology there is a bias towards maintaining the status quo ranking, as teams at the top have fixture lists that contain a higher proportion of fixtures against each other. Figure 1 might suggest that teams at the top of the table are benefiting from playing teams who were strong last

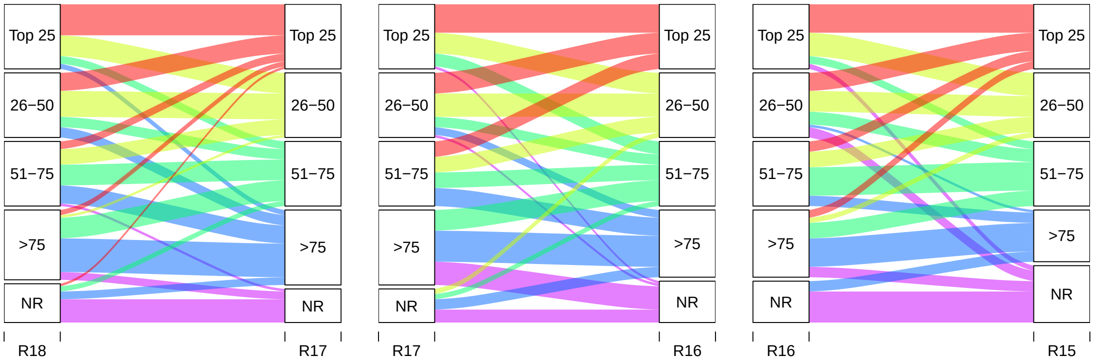

Figure 1: Alluvial diagrams showing persistence of ranking by Additional Points segment. NR - Not Ranked

year but weaker the next. On the other hand it could also be the case that teams at the top do not show much impact from using the previous season's results to determine additional points, as their opposition are as likely to rise in the rankings as fall.

We can investigate how this actually impacts the rankings by using something less dependent on the previous season's ranking in order to determine the current season's Additional Points. In Figures 2, 3, 4 we start with the current season's DMT ranking, subtract off the Additional Points and then re-add Additional Points based on the current season's DMT ranking. We repeat this iteratively a number of times and look at the changes to the top ten from this. Were we to propose this as a means of providing a ranking to be used it would be sensible to do this until some sort of convergence was reached, perhaps that the change in points for each team from iteration to iteration falls below some particular threshold. We do not propose this however as such a scheme would lack transparency but still not provide a principled statistical ranking. For the present purpose we therefore choose five iterations somewhat arbitrarily as a value that provides a long enough range that we would expect a decent degree of convergence, but short enough that it may still be clearly understood in the context of the initial ranking, and is readily presentable graphically.

This suggests there may be an effect. Overall, teams move an average of approximately five places from applying this adjustment. Looking solely at the top ten, we can see in Figures 2, 3 and 4 that in every season and under every iteration compared with the original DMT ranking, there is a net negative effect on the ranking of the top ten. However as we can see, this net effect is small (about half a place per team) and the absolute size of the ranking moves of the top ten is only

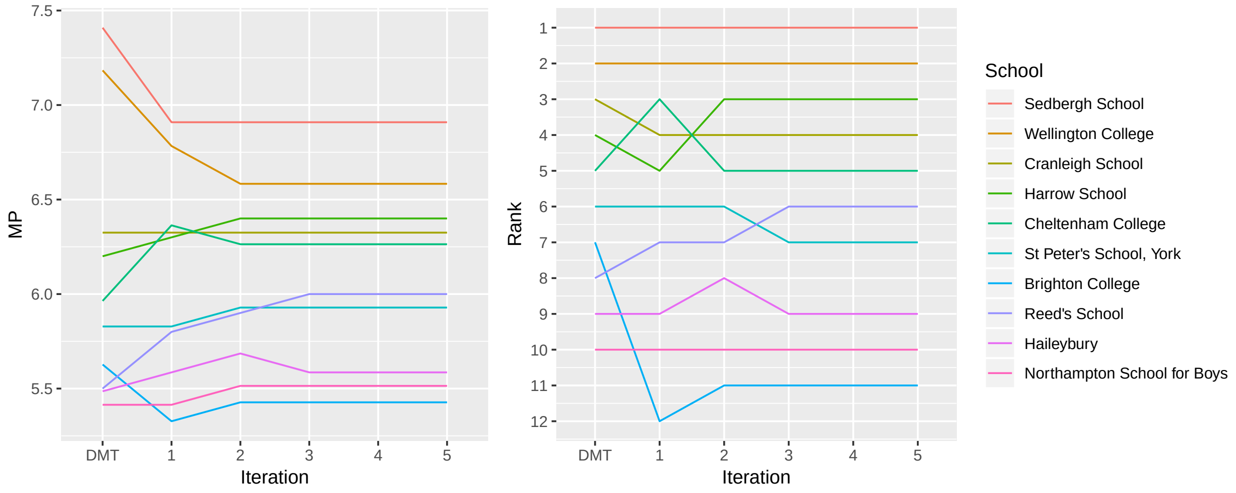

Figure 2: Top10 Merit Points and Rank variation with iterations of Additional Points based on current methodology Daily Mail Trophy 2017/18

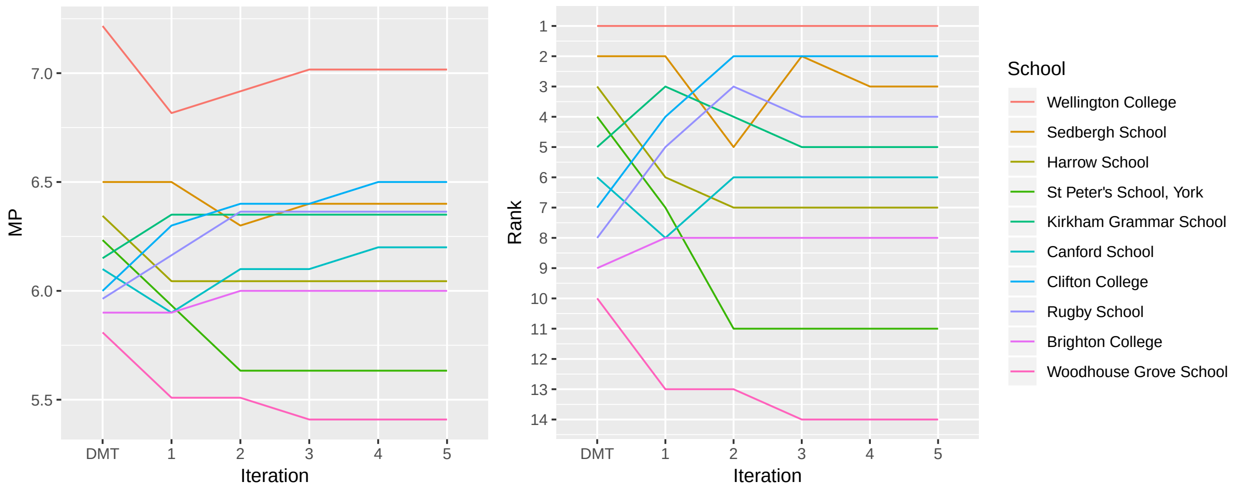

Figure 3: Top10 Merit Points and Rank variation with iterations of Additional Points based on current methodology for Daily Mail Trophy 2016/17

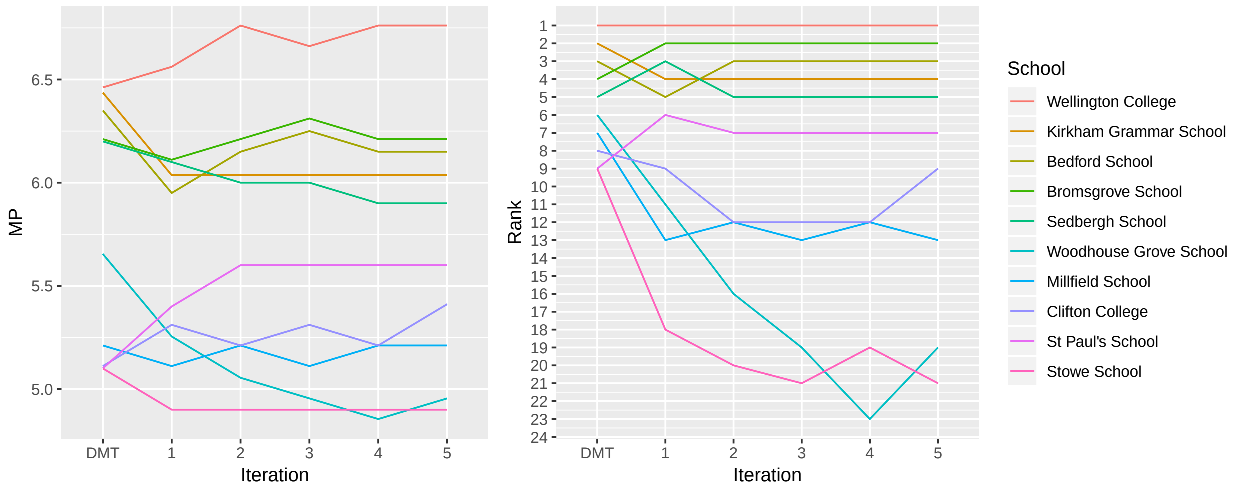

Figure 4: Top10 Merit Points and Rank variation with iterations of Additional Points based on current methodology for Daily Mail Trophy 2015/16

infrequently large. However the impact of the original Additional Points definition is persistent through these iterations, and so using this methodology does not enable us to disentangle totally the previous-season effect.

And so it makes sense to look at an alternative version where that is not the case. In order to do that we take as our first iteration Additional Points based on ranking teams by League Points per match in the current season. We then use the resultant Merit Points ranking to determine the Additional Points in the next iteration. By iterating we progressively account for schedule strength in a manner increasingly dependent on the current season's ranking.

As before, looking at Figures 5, 6 and 7 we continue to see effects of a similar size, with the overall average absolute effect continuing to be around five places, and with a small net negative effect of around half a place in the rankings of the top ten.

Introducing a second iterative method also allows us to compare these two in order to see the degree to which the difference in methodologies applied to the ranking in the first iteration persists over time, and thereby get an idea of how long the effect of using a previous-season methodology can last. The iterations in this context can be thought of as several consecutive seasons that happened to have exactly the same results. The two methods apply exactly the same process to the ranking and differ only in the methodology used to produce the first iteration. While, as discussed, it could

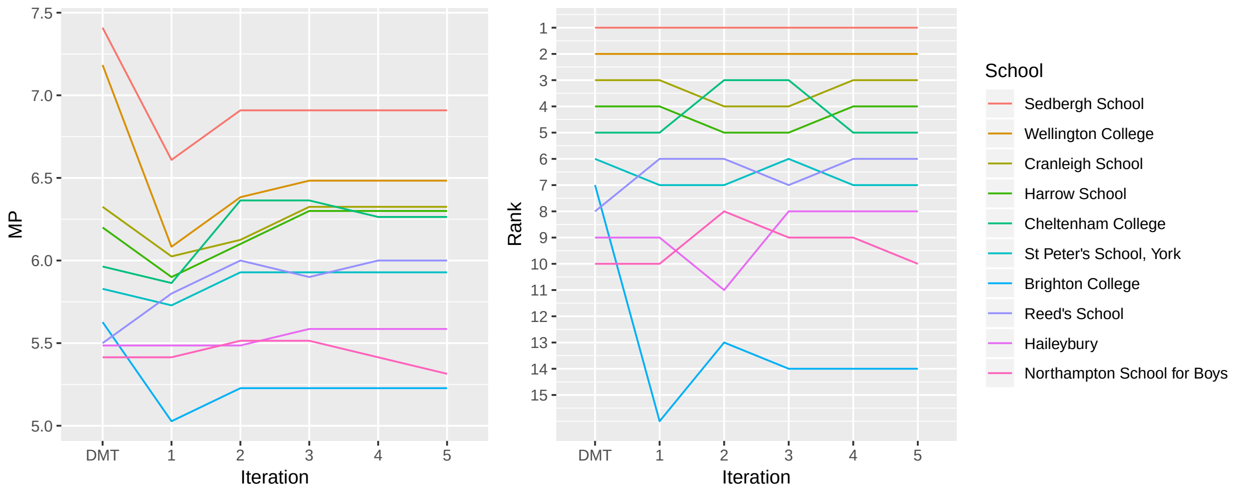

Figure 5: Top10 Merit Points and Rank variation with iterations of Additional Points based on LPPM for Daily Mail Trophy 2017/18

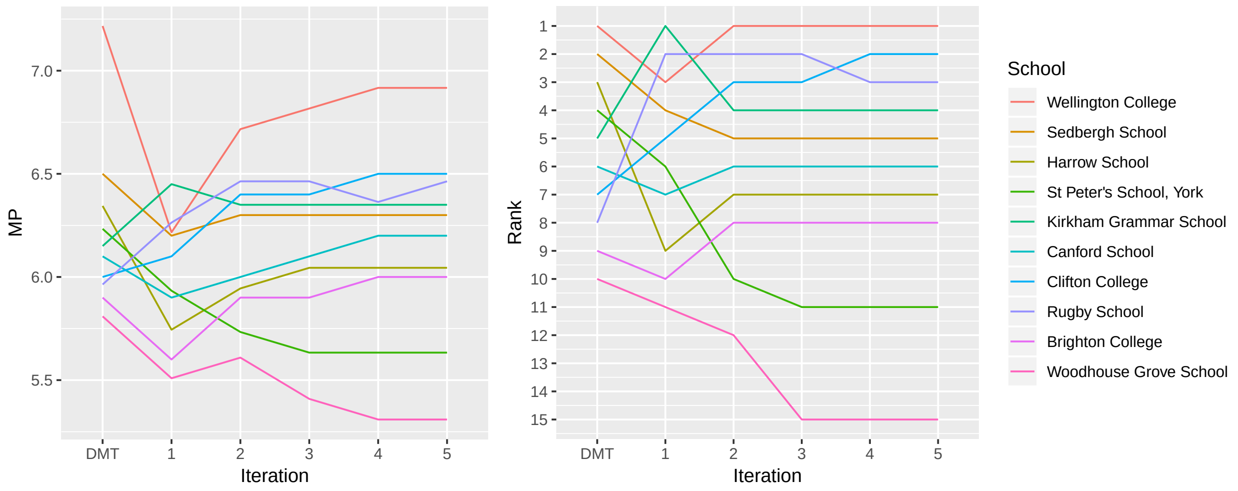

Figure 6: Top10 Merit Points and Rank variation with iterations of Additional Points based on LPPM for Daily Mail Trophy 2016/17

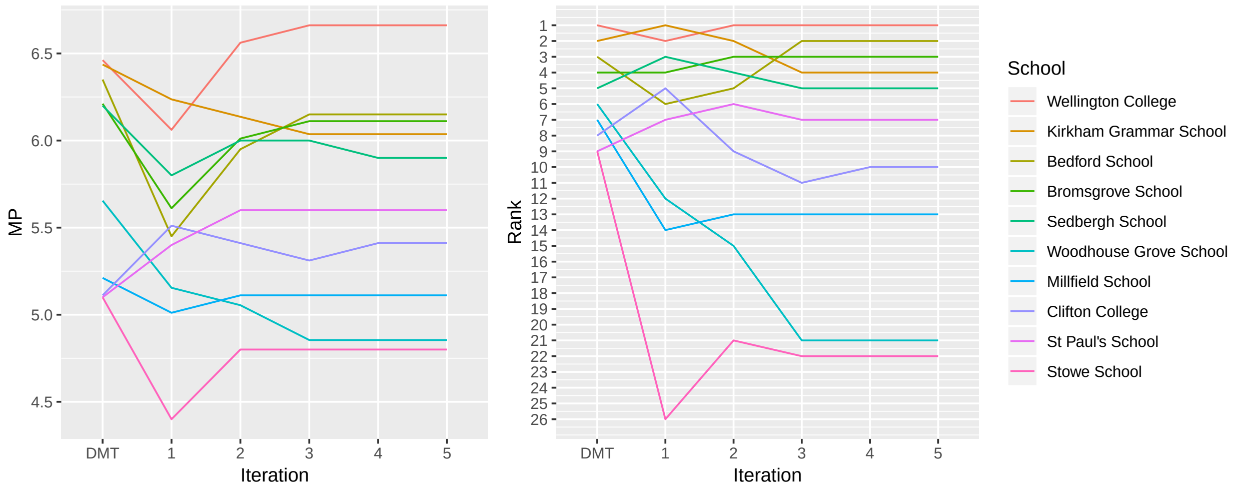

Figure 7: Top10 Merit Points and Rank variation with iterations of Additional Points based on LPPM for Daily Mail Trophy 2015/16

	Iterations				
Season	1	2	3	4	5
2017/18	0.21	0.11	0.04	0.04	0.04
2016/17	0.19	0.08	0.05	0.03	0.04
2015/16	0.18	0.05	0.05	0.00	0.04

Table 1: Average absolute difference in Additional Points of the top ten for DMT ranking created using DMT ranking itself, compared to using League Points per match for initial ranking

be an acceptable compromise to use the previous season's results if we believed that they gave a substantially similar outcome and avoided other undesirable features, it would become harder to justify if there were large persistent effects. For example, if a team's ranking were likely to be influenced by the results of three or four seasons ago then this would be troubling.

As we can see from Table 1 the average absolute difference between the methods persists even four seasons after the ranking methodology difference in the first iteration. Effectively 0.04 translates as four of the top ten having a difference of 0.1 in their Additional Points. While 0.1 is not large, the smallest non-zero amount it could be, four out of ten is notable. Given that the difference between teams in the top ten is frequently less than 0.1 (in 2015/16 the top three teams were separated by only 0.11 points) then the finding that on average four of the top ten are still impacted by something that happened four seasons previously is highly undesirable.

Overall there is a sound basis for questioning the maintenance of the dependence on the previous season's results in the context of the DMT methodology.

2.2 Absolute size of Additional Points

We begin investigating the absolute size of Additional Points by trying to understand the relative influence of the two components of the Merit Points methodology, namely League Points per Match and Additional Points. We do this by varying the weights of each component and observing the ranking produced. That is we recalculate rankings based on Merit Points using the formula

$$\text{Merit Points} = \left(\frac{\text{League Points}}{\text{Matches Played}} \times \text{LPPM Multiplier}\right) + (\text{Additional Points} \times \text{AP Multiplier})$$

	Rank										
LPPM Multiplier	1	1	1	1	**1**	0.75	0.5	0.25	0		
AP Multiplier	0	0.25	0.5	0.75	**1**	1	1	1	1	LPPM	AP
Sedbergh	1	1	1	1	**1**	1	2	2	2	4.9	2.5
Wellington	11	6	2	2	**2**	2	1	1	1	4.1	3.1
Cranleigh	4	2	3	3	**3**	3	3	5	28	4.6	1.7
Harrow	5	4	4	4	**4**	4	5	9	32	4.5	1.7
Cheltenham	8	7	5	5	**5**	5	4	3	27	4.4	1.8
St Peter's, York	7	8	7	6	**6**	6	8	21	43	4.4	1.4
Brighton	14	12	12	10	**7**	7	6	7	11	3.7	1.9
Reed's	2	3	6	7	**8**	15	25	53	73	4.8	0.7
Clifton	19	17	14	11	**8**	8	7	4	8	3.5	2.0
Haileybury	9	10	9	9	**10**	12	15	31	54	4.3	1.2
Mean Difference	10.4	7.4	4.7	2.2	**0.0**	2.8	5.9	11.5	20.0		
Mean Proportion	1.14	1.09	1.06	1.03	**1.00**	1.03	1.08	1.16	1.31		

Table 2: 2017/18 Ranking based on alternative LPPM and AP multipliers

Tables 2, 3, and 4, in and of themselves, cannot tell us if the weighting of Additional Points is reasonable or not. However it seems sensible to consider that League Points per Match should be the primary component, with Additional Points acting as a secondary corrective component to account for schedule strength. With this being the case we would expect to find that a good ranking measure would be more influenced by a change in the LPPM Multiplier than a change in the AP Multiplier. We see this to some degree, with the overall difference measures larger as we diverge from (LPPM Multiplier = 1) than as we diverge from (AP Multiplier = 1). However

					Rank						
LPPM Multiplier	1	1	1	1	**1**	0.75	0.5	0.25	0		
AP Multiplier	0	0.25	0.5	0.75	**1**	1	1	1	1	LPPM	AP
Wellington	6	1	1	1	**1**	1	1	1	1	4.4	2.8
Sedbergh	2	3	2	2	**2**	2	2	3	11	4.5	2.0
Harrow	5	4	4	3	**3**	3	3	6	17	4.4	1.9
St Peter's, York	8	7	5	5	**4**	4	4	7	16	4.3	1.9
Kirkham Grammar	1	2	3	4	**5**	6	11	18	39	4.8	1.4
Canford	10	9	9	6	**6**	5	6	8	15	4.2	1.9
Clifton	2	5	6	6	**7**	7	12	16	32	4.5	1.5
Rugby	7	8	8	8	**8**	9	10	14	30	4.4	1.6
Brighton	2	6	7	9	**9**	11	13	21	39	4.5	1.4
Woodhouse Grove	13	11	10	10	**10**	10	8	10	13	3.9	1.9
Mean Difference	9.0	6.4	4.1	2.0	**0.0**	2.4	5.6	11.3	19.9		
Mean Proportion	1.13	1.08	1.05	1.02	**1.00**	1.03	1.07	1.16	1.29		

Table 3: 2016/17 Ranking based on alternative LPPM and AP multipliers

					Rank						
LPPM Multiplier	1	1	1	1	**1**	0.75	0.5	0.25	0		
AP Multiplier	0	0.25	0.5	0.75	**1**	1	1	1	1	LPPM	AP
Wellington	16	8	4	3	**1**	1	1	1	1	3.5	3.0
Kirkham Grammar	2	2	2	1	**2**	2	4	5	15	4.6	1.8
Bedford	1	1	1	2	**3**	5	5	9	26	4.8	1.6
Bromsgrove	4	3	3	4	**4**	3	3	3	10	4.1	2.1
Sedbergh	11	7	5	5	**5**	4	2	2	2	3.7	2.4
Woodhouse Grove	17	13	9	6	**6**	6	6	4	7	3.5	2.2
Millfield	31	25	18	12	**7**	7	7	7	11	3.1	2.1
Clifton	4	5	6	7	**8**	15	26	37	66	4.1	1.0
St Paul's	13	14	13	11	**9**	11	14	18	26	3.5	1.6
Solihull	6	6	7	8	**9**	18	27	38	63	4.1	1.0
Stowe	33	31	24	15	**9**	9	9	8	9	3.0	2.1
Whitgift	37	34	27	19	**9**	8	8	6	5	2.9	2.2
Mean Difference	9.4	7.0	4.7	2.6	**0.0**	2.4	6.0	12.2	22.7		
Mean Proportion	1.16	1.12	1.09	1.05	**1.00**	1.04	1.09	1.15	1.30		

Table 4: 2015/16 Ranking based on alternative LPPM and AP multipliers

we might consider that it is the changes when the multipliers are close to 1 that are the most relevant to this question, since that represents the current calibration. Here we see that they are of comparable orders, with the AP Multiplier even having a larger effect in the 2015/16 season. Also we should be aware that there will be a tendency for historically strong teams to seek other historically strong teams for their fixture list causing a greater closeness between the DMT and AP ranking methodologies, which could be an explanation for why there is still good agreement even when the impact of the actual playing record becomes the smaller component.

We also look at the three most recent seasons and consider the Additional Points of the ten teams who have accrued the most.

	AP		
AP Rank	2017/18	2016/17	2015/26
1	3.1	2.8	3.0
2	2.5	2.6	2.4
3	2.4	2.6	2.3
4	2.1	2.3	2.2
5	2.1	2.2	2.2
6	2.1	2.2	2.2
7	2.0	2.2	2.2
8	2.0	2.2	2.1
9	2.0	2.1	2.1
10	1.9	2.0	2.1

Table 5: Ranking by Additional Points

As we can see in Table 5 the Additional Points of the team ranked highest is substantial, typically around 3 points in every match. This equates roughly to a Draw and a Try Bonus in every match in addition to the League Points they manage to accrue. The advantage enjoyed by the top team varied, but last season was as much as a try bonus in every match over the team ranked fifth by Additional Points. This would seem to place a very large, and in many circumstances unachievable, burden of outperformance on the vast majority of teams with fewer Additional Points in order to prove their superiority and win the tournament, effectively confining the winner of the tournament to a group as small as three or so before the season has even started. Despite the undoubted persistence of quality in schools' performances, the natural variation of team ability from the enforced turnover of players due to the academic cycle makes this highly questionable. Indeed the winner of the Daily Mail Trophy has been the team with the highest number of Additional Points in three of the four seasons under the DMT methodology, and in the other season it was won by the team ranked second by Additional Points. While it is true that there is likely to be a correlation between the strength of opposition and the performance of a team itself, this is suggestive of allowing the strength of opposition to be causative rather than correlated in the determination of a team's performance.

Finally we compare pairs of playing records under different values of the maximum Additional Points that may be earned, assuming that Additional Points for playing lower ranked teams occur in the same ratios as in the present methodology. Under the current methodology Max AP is taken to be 0.3. In each case, given the value taken for maximum Additional Points, we determine which of the two playing records would be ranked higher under the Merit Points methodology.

Some of the preferences implied under the current weighting of Additional Points seem unintuitive. For example, it weights a 6-0 record equivalent to an 8-2 record. While a subjective matter, it seems reasonable to say that there should be caution about rating a team playing four extra matches and losing two of them equally to one with a perfect record from six matches.

The evidence here suggests that the weighting of Additional Points is too large and is having a disproportionate influence on the outcome.

2.3 Additional notes

The Daily Mail Trophy ranking measure was conceived to meet a practical need of fairly ranking teams based on a system of matches with varying schedule strengths, and so it is in the success of this practical application that it needs to be assessed. However theoretical limitations need to be understood as these are likely to manifest over time. Some of the limitations of the current DMT methodology may be made clear by considering the most extreme cases. It is possible for example for a team to win the Daily Mail Trophy despite having lost all their matches and earnt no league points. Equally it would be possible for a team to be the only team with a 100% record of wins and bonuses, including victories against all the other top teams, and still not win the tournament. In practice both of these are extremely unlikely, but they highlight genuine reasons for concern about the current methodology.

		Max AP						
Record 1	Record 2	0	0.05	0.1	0.15	0.2	0.25	0.3
5-0	6-1	1	1	1	1	1	1	1
5-0	7-1	1	1	1	1	1	=	2
5-0	8-1	1	1	1	1	2	2	2
5-0	9-1	1	1	1	2	2	2	2
5-0	10-1	1	1	2	2	2	2	2
5-0	11-1	1	1	2	2	2	2	2
5-0	7-2	1	1	1	1	1	1	1
5-0	8-2	1	1	1	1	1	2	2
5-0	9-2	1	1	1	1	2	2	2
5-0	10-2	1	1	1	2	2	2	2
5-0	11-2	1	1	1	2	2	2	2
5-0	17-2	1	2	2	2	2	2	2
6-0	7-1	1	1	1	1	1	1	1
6-0	8-1	1	1	1	1	1	2	2
6-0	9-1	1	1	1	=	2	2	2
6-0	10-1	1	1	1	2	2	2	2
6-0	13-1	1	1	2	2	2	2	2
6-0	17-1	1	2	2	2	2	2	2
6-0	8-2	1	1	1	1	1	1	=
6-0	9-2	1	1	1	1	1	2	2
6-0	10-2	1	1	1	1	2	2	2
6-0	11-2	1	1	1	2	2	2	2
6-0	14-2	1	1	2	2	2	2	2
6-0	20-2	1	2	2	2	2	2	2

Table 6: DMT: Preferred playing record for different values of Max AP, assuming five points for win (i.e. win + try bonus) and one point for loss (i.e. losing bonus). Additional Points per match taken to be two-thirds of Max AP.

3 Alternative methodologies

In this section we will consider alternative approaches. The first to be introduced (named 'RASR') will be the unique statistical model that results from the League Points scoring rule and a crucial criterion, highly desirable in the production of a ranking. While from a statistical perspective this might be argued to be the 'correct' model to use, it necessarily lacks transparency and is a significant departure from the current methodology. We will therefore also consider a family of more familiar-looking alternatives (named 'Dapper') that combine some of the insights from the first part of this document with those from the development of our statistical model.

3.1 RASR — Ranking Algorithm for Schools Rugby

RASR (pronounced 'razor') is a statistical ranking model, designed to respect the standard scoring rules in rugby union (what we have referred to as League Points in this document) and to adjust for schedule strength. By schedule strength we mean strength of opponents, number of matches played, and venue (home or away). RASR is based on the long established Bradley-Terry model, which is also the foundation for, amongst other things, the PageRank algorithm used to prioritise internet searches, the Elo rating system used to rank chess players, and the KRACH ranking model used in US College sports. By looking at the full set of results from an individual season the model assigns to each team a performance score that properly takes into account the various aspects of each team's schedule strength. From this we may produce a Projected Points Per Match (PPPM) for each team, which is the projected League Points per match were it possible for each team in the tournament to play each other team in the tournament home and away. This gives us an intuitive measure of the relative performance of the teams, as measured by match results.

Unlike the other models considered here, the model is coherent in the sense that once we have the performance scores then were we to be told the fixtures but not the results of the matches, and we then simulated these matches using the performance scores we had derived, then the ranking we would expect to end up with is the ranking we started with i.e. the expected outcome is the same as the actual outcome. This so-called 'retrodictive' criterion describes how the model is connected directly and coherently to the match outcomes, and their probabilities, a feature that is not available with the other, more heuristic ranking measures we consider in this document.

It is important to note that by design RASR is not a predictive model (though its predictive ability is likely still to be superior to many other ranking methodologies), and only uses the information that would either be used were a full double-header round robin to be played i.e. League Points per match, or those elements that are controlled for by the structure of a typical full double-header round robin i.e. number of matches and venue. It deliberately does not therefore, for example, take account of scores, nor things like the previous season's performance, the order of current season's performances, whether players are rested, time of season, weather etc. as one might were one to attempt to build a full predictive model and were one to have the data available. In this sense it is consistent with a typical league ranking methodology.

It is also important to note that RASR effectively encompasses a family of models which may be calibrated in various ways.[1] While not sensitive to many of these calibration choices, the ranking is sensitive to the selection of the 'prior' within it, a matter of subjective judgment rather than statistical veracity and a topic to which we will return.

[1] The RASR model takes the form of modelling the probabilities of the result outcome and try bonus point as

$$\begin{aligned}
P(\text{team } i \text{ beats team } j \text{ by wide margin}) &\propto \tau^4\pi_i^4 \\
P(\text{team } i \text{ beats team } j \text{ by narrow margin}) &\propto \kappa\tau^3\pi_i^4\pi_j \\
P(\text{team } i \text{ draws with team } j) &\propto \nu\pi_i^2\pi_j^2 \\
P(\text{team } j \text{ beats team } i \text{ by narrow margin}) &\propto \frac{\kappa\pi_i\pi_j^4}{\tau^3} \\
P(\text{team } j \text{ beats team } i \text{ by wide margin}) &\propto \frac{\pi_j^4}{\tau^4}
\end{aligned}$$

and

$$\begin{aligned}
P(\text{team } i \text{ and team } j \text{ both gain try bonus point}) &\propto \theta\pi_i\pi_j \\
P(\text{only team } i \text{ gains try bonus point}) &\propto \tau\pi_i \\
P(\text{only team } j \text{ gains try bonus point}) &\propto \frac{\pi_j}{\tau} \\
P(\text{neither team gains try bonus point}) &\propto \phi
\end{aligned}$$

where π_i is a performance measure for team i and is used for ranking, and τ, κ, ν, θ, ϕ are structural parameters relating to home advantage, prevalence of wide results, prevalence of draws, prevalence of both teams earning try bonuses, and prevalence of neither team earning a try bonus respectively. The values of structural parameters are determined based on averages over available seasons' results. Within the range of reasonable potential values, the ranking is largely insensitive to these structural parameters.

Additionally in RASR a dummy 'team 0' is introduced, against which each team notionally achieves a win (gaining four points) and a loss (gaining no points). This acts as a prior within the model. The resultant ranking is meaningfully dependent on the points awarded for a win against this dummy 'team 0' (or alternatively the number of matches against 'team 0') in a manner similar to that described in Tables 6, 13, and 14. The recent Warwick Statistics MSc dissertation of Ian Hamilton provides a more thorough description of the model, and is available on request.

3.2 Dapper — Damped and Adjusted Points per Match

The 'Dapper' family of models may be described by the formula

$$\text{Merit Points} = \frac{\text{League Points} + \text{Additional Points} + (3 \times n)}{\text{Matches Played} + n},$$

where League Points are as previously defined, and the value of the parameter n is to be determined. Leaning on the benefits of familiarity we employ the same structure for Additional Points, basing them on the ranking of the opposition faced, but the specifications are to be determined.

This provides a formula that benefits from being somewhat familiar from the current DMT methodology, and is driven by insights derived in the first part of this document as well as the development of RASR. It encompasses two key differences of form when compared to the current Daily Mail Trophy methodology. First, the Additional Points are added on a per-match basis. This avoids some of the more extreme hypothetical outcomes, for example where a team with a losing record could win the tournament. Perhaps more importantly it also avoids disproportionately advantaging those teams who play a large number of highly ranked opponents, over those who have played fewer matches but with a clearly better playing record against the same average quality of opposition, a problem with the current methodology. It also significantly reduces the potential for gaming the system through adding or carefully selecting fixtures.

Second, we add in what we will refer to as a 'prior', n. This allows us to control more explicitly what value we put to, for example, a 5-0 record compared to a 11-2 or a 8-1 record, assuming equivalent strength opposition. It may be thought of as being equivalent to every team in the tournament playing and earning three points against a hypothetical average team n times, in addition to the actual fixtures they play. In practice it is a useful device whereby each team starts from an equivalent position, just as they would conventionally, but we are able to better calibrate the ranking given different numbers of matches played. This is likely to be particularly key to the Daily Mail Trophy where the number of matches played can vary from five to fourteen. Using the multiplier of three allows for a more symmetric differentiation at the bottom as well as the top of the table in comparing, for example, an 0-5 and a 1-8 record. This is the case since on average teams will gain two points for the result (Win, Lose, Draw) and approximately one extra from bonus points and Additional Points.

This leaves several elements of the model to be determined. We must decide on what ranking to use as our basis for Additional Points, how large to make the Additional Points, and what value to assign to n.

3.2.1 Additional Points

The benefits of familiarity mean that we choose to use the same structure for Additional Points as that used in DMT. We determine the size of these Additional Points based on the output of the methodology with different calibrations. That is we take Additional Points as:

Rank 1 to 25:	Max AP
Rank 26 to 50:	$2/3 \times$ Max AP
Rank 51 to 75:	$1/3 \times$ Max AP
Otherwise:	0

where we determine an optimal calibration by varying values of Max AP along with n and determining which provide the greatest agreement with RASR and also minimise violations, a concept introduced later in this document in Section 4.3, and an n that is in line with our understanding of how records of different lengths should be compared.

Before determining these however we should determine on what ranking these Additional Points will be based. As we noted in the first part, in the context of the current Daily Mail Trophy there are good grounds for questioning the use of the previous season's ranking to provide the basis for the next season's Additional Points. However in the Dapper family the influence of Additional Points is likely to be reduced and so it is again reasonable to consider using the previous season's ranking. The most obvious alternatives to this would be to use a ranking from the current season such as RASR itself, an iterated version of the chosen Dapper model, or League Points Per Match. These three alternatives all benefit from meeting the criterion of having the ranking solely based on the current season's results. However in the case of the first two, the crucial element of transparency that we are seeking in using the Dapper family rather than RASR is lost and so we reject these.

Instead we compare using the previous season's ranking to using League Points per match from the

	RASR(-1)	LPPM
Absolute Difference	20.9	8.9
Proportional Difference	0.784	0.254
Top25	14	20
26-50	9	14
51-75	11	14

Table 7: 2017/18: Comparison of using prior season RASR and current season LPPM. Table shows difference measures and the number of each ranking sector preserved, when compared to current season RASR.

	RASR(-1)	LPPM
Absolute Difference	21.0	7.5
Proportional Difference	0.824	0.221
Top25	11	21
26-50	8	15
51-75	7	14

Table 8: 2016/17: Comparison of using prior season RASR and current season LPPM. Table shows difference measures and the number of each ranking sector preserved, when compared to current season RASR.

current season. Since we consider RASR to be the best of the rankings from a statistical perspective we use that as the basis for this analysis. We compare League Points per Match from the current season and RASR from the previous season, each to the current-season RASR ranking.

In all rankings, teams playing fewer than five matches as part of the tournament are excluded from the ranking. When comparing to the previous season for the purposes of the overall difference measures, teams are excluded from the calculation if they do not appear in both seasons' rankings, but the ranking before exclusions will be maintained. Since RASR is not available for 2014/15 season then we look just at 2016/17 and 2017/18 seasons only.

As we can see from Tables 7 and 8, the current-season-LPPM method performs very clearly better than the previous-season-RASR method, with substantially lower difference scores, and much higher proportions of the segment rankings matching those of current season RASR. Of particular importance for the top of the table, where opposition is likely to be stronger, the current-season-LPPM method seems to be particularly good in identifying the top 25.

Beyond what we see in the table, using previous season's results comes with the benefits described in Section 2.1. On the other hand, using League Points per match meets the explicit criterion of allowing the ranking to be based purely on this season's matches. In this case the evidence from previous seasons is overwhelming and we strongly recommend using the current-season League Points per match.

For the determination of Max AP and n, the comparisons described at the beginning of this section were employed. The analysis is lengthier than we wish to present here, but details of the measures used and the analysis are given in the Appendix.

3.2.2 Dapper(2.25,3)

As suggested by this analysis, we apply the Dapper formula with $n = 3$ and Max AP $= 2.25$. That is

$$\text{Merit Points} = \frac{\text{League Points} + \text{Additional Points} + 9}{\text{Matches Played} + 3}$$

with Additional Points taken to be

Rank 1 to 25:	2.25
Rank 26 to 50:	1.5
Rank 51 to 75:	0.75
Otherwise:	0

with the rank here determined by the ranking based on League Points per Match in the current season.

As described in more detail in section 6.4 it is crucial to understand that the selection of n is a subjective rather than statistical judgment and is dependent on the relative preference that one wishes to assign to different performance records. For example, as shown in Table 13, the value $n = 3$ roughly equates a 6-0 record with somewhere between 8-1 and 9-1, or somewhere between 10-2 and 11-2, when thinking about the relative merits of 'perfect' and 'near perfect' performance records.

3.3 DMT2

Before coming to a comparison of the newly introduced methodologies with the current DMT methodology it would be reasonable to seek to optimise the parametrisation within the current methodology as we have with the other methodologies, especially given that we explicitly identified the absolute size of Additional Points as an issue in Section 2.2. Again details of this are included in the Appendix but we consequently use a methodology whereby

$$\text{Merit Points} = \frac{\text{League Points}}{\text{Matches Played}} + \text{Additional Points}$$

where Additional Points are earnt based on the ranking of opposition faced with

Rank 1 to 25:	0.15
Rank 26 to 50:	0.1
Rank 51 to 75:	0.05
Otherwise:	0

and this ranking is based on previous season performance as in the current methodology. For the 2015/16 season in the analysis that follows we use the current methodology DMT ranking as the previous season ranking, on which Additional Points are based.

4 Comparison

We therefore have four proposal methodologies under consideration: DMT(Max AP = 0.3), RASR, Dapper(n = 3,Max AP = 2.25), DMT2(Max AP = 0.15). As the most statistically sound methodology, RASR will be used as the benchmark for the other methodologies in looking at the overall ranking and the top ten. We will also compare all four methodologies using violations, an intuitive objective measure. In all cases, we maintain the requirement that a team must play at least five matches against Daily Mail Trophy registered opposition in order to be included in the ranking.

4.1 Top ten similarity to RASR

First we simply compare the ranking of the top ten by each of these methods. We do this by taking the top ten as given by RASR and compare the ranking of those teams with their ranking under the other methodologies.

Team	RASR	DMT	DMT2	Dapper
Sedbergh School	1	1	1	1
Reed's School	2	8	6	5
Cranleigh School	3	3	3	2
Harrow School	4	4	5	3
Wellington College	5	2	2	7
St Peter's School, York	6	6	7	4
Northampton School	7	10	8	9
Cheltenham College	8	5	4	6
Haileybury	9	9	10	8
QEGS, Wakefield	10	13	11	11

Table 9: 2017/18: Top ten comparison

Team	RASR	DMT	DMT2	Dapper
Kirkham Grammar School	1	5	2	1
Sedbergh School	2	2	3	2
Wellington College	3	1	1	5
Brighton College	4	9	9	10
Clifton College	5	7	6	4
Harrow School	6	3	4	8
Rugby School	7	8	7	3
St Peter's School, York	8	4	5	6
Canford School	9	6	8	7
St John's School, Leatherhead	10	12	11	9

Table 10: 2016/17: Top ten comparison

We see the results presented in Tables 9, 10, and 11. As we might expect, over the three seasons, Dapper shows the closest relationship with RASR, with DMT2 performing better than DMT.

4.2 Overall similarity to RASR

Second we look at the overall similarity of the methodologies to RASR. In fact we do this by measuring the dissimilarity. In Figure 8 we use two measures described in detail in the Appendix

Team	RASR	DMT	DMT2	Dapper
Kirkham Grammar School	1	2	2	1
Bedford School	2	3	1	3=
Bromsgrove School	3	4	3	2
Sedbergh School	4	5	5	3=
Seaford College	5	12	7	6
Wellington College	6	1	4	7
Clifton College	7	8	6	5
QEGS, Wakefield	8	17	10	13
Tonbridge School	9	18	15	9
Solihull School	10	13	9	10

Table 11: 2015/16: Top ten comparison

to show the similarity between the rankings. They look at the ranking of each team under each methodology and compare the ranking of each team to that from RASR. An average is then taken of those differences. Since we may consider those differences in an absolute sense (eleventh is one place different from tenth) or a proportionate sense (eleventh is ten percent different to tenth) we consider two measures. Hence a lower score corresponds to greater agreement. The proportionate measure will give greater weight to absolute differences at the top of the table, and in this context is therefore the one on which we would choose to place greater weight.

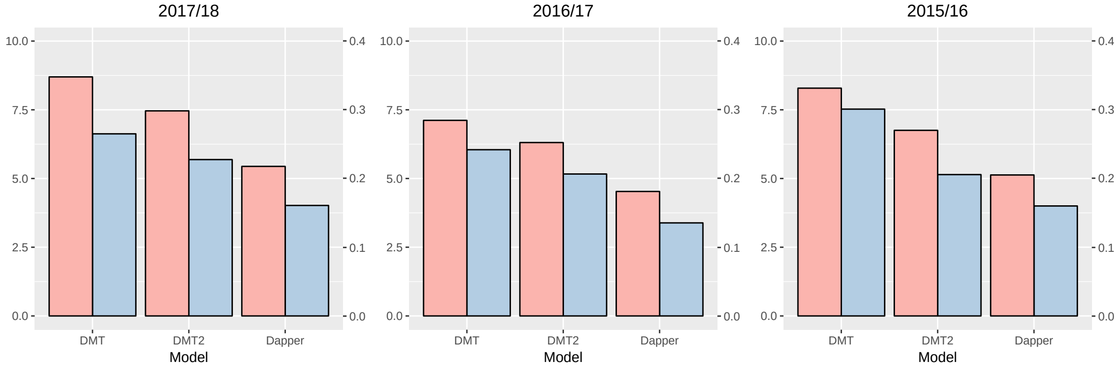

Figure 8: Comparison of models by average absolute difference to RASR (pink, left-hand axis) and average proportionate difference to RASR (blue, right-hand axis)

The two measures provide the same result. Again we see a clear progression with DMT2 showing smaller differences than DMT, and Dapper showing smaller differences still.

4.3 Violations

Finally we look at two closely-related measures that may more objectively assess the relative success of each of the methodologies. First we take each match and consider it a violation of the ranking if the match is won by the lower-ranked team. We consider this for all matches (referred to as 'violations' in this document) and calculate the proportion of matches where there is a violation. In the second measure we consider just those matches that end in a wide result, i.e., where no losing bonus point was earnt (referred to as 'gross violations' in this document). These are not perfect measures as they do not, for example, differentiate between being beaten by a team that is slightly lower-ranked and one that is much lower-ranked. In general though we would expect a more successful overall ranking methodology to be consistent with the outcome of a higher proportion of matches than in the case of a less successful overall ranking methodology, and therefore to show lower violations, and gross violations scores.

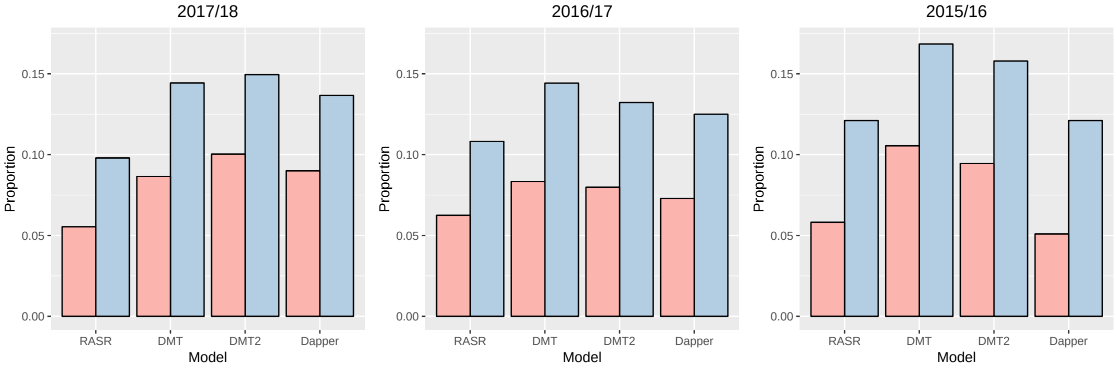

Figure 9: Comparison of models by violations (blue) and gross violations (pink)

Here the picture is slightly less clear-cut, but again we see the same general pattern of DMT2 outperforming DMT, though not in 2017/18, and Dapper outperforming both. Interestingly Dapper outperforms RASR in 2015/16 on gross violations. This is somewhat surprising, but does not seem to be a consistent pattern.

4.4 Assessment against criteria

We summarise here the four models against the criteria described in the introduction. As a reminder, the criteria for the methodology were that it should:

1. be such that the top-ranked team should not be obviously wrong in the opinion of a substantial proportion of stakeholders in the tournament — coaches, players, parents, administrators etc.
2. allow the identification of top-ranked teams in general and how far apart they are, so that a team may have an idea of how far from top position it is.
3. be such that all other relative rankings should not be perceivable as unreasonable by a substantial proportion of the tournament stakeholders.
4. be dependent on the match results of only the current season.
5. be such that any participating school could win.
6. be consistent in the sense that if they were applied to a full round robin they would achieve the same result as with standard rugby union league scoring rules.
7. be such that there is not a requirement for additional fixtures beyond the regular fixture list.
8. be transparent in the sense that it is readily explicable to a generalist audience.
9. allow for a ranking from early in the season.
10. take into account the relative strength of opposition faced, objectively and fairly.

Criterion	RASR	DMT	DMT2	Dapper
1	✓✓✓	✗	✓	✓✓
2	✓✓	✓✓✓	✓✓✓	✓✓
3	✓✓✓	✗	✓	✓✓
4	✓✓✓	✗	✗	✓✓✓
5	✓✓	✗	✓	✓✓
6	✓✓✓	✓✓✓	✓✓✓	✓✓✓
7	✓✓✓	✓✓✓	✓✓✓	✓✓✓
8	✗	✓✓✓	✓✓✓	✓✓
9	✓✓	✓✓✓	✓✓✓	✓✓✓
10	✓✓✓	✗	✓	✓✓

Table 12: Measures assessed against criteria

The summary expressed in Table 12 is clearly subjective, and it would be reasonable to dispute the exact categorisation of a number of these. However we believe the relative ranking against each criterion to be reasonable with many of these assessments backed by analysis or commentary elsewhere in this document. The overall picture we believe to be sound. RASR comes out outright top in three of the ten criteria and is equal top in another four, with transparency being the one criterion where it notably lags the other ranking measures. Dapper is the most consistent performer, never falling below a good mark in any criteria. DMT does well in a number of criteria, but falls down on its use of a previous season's results and the predetermined small set of teams that could realistically win the tournament each season, and never outperforms DMT2.

4.5 Other approaches

A number of other methodologies were considered including RPI [2], Expected Net Score [3], and where the strength of opposition was accounted for in a multiplicative rather than additive manner[4]. But these were not based on any particular learning from the analysis of DMT or of RASR, and their performance overall was below that of the alternatives presented here. We therefore do not include them, though we are open to the possibility that there are other methodologies that could be worth analysis.

A more radically alternative approach could be proposed based on noting that a similar challenge as faced by SOCS in the Daily Mail Trophy is faced by NCAA in US college sports in selecting teams to go through to their final knock-out competitions each year. Here too schedule strengths of competing teams vary. In the case of US college sports a variety of measures are used and discussed in the public domain, with a subset of these officially mandated to be taken into account, but the ultimate decision is made by a poll of a committee. A recent change in the primary ranking

[2] RPI - Rating Percentage Index - is a rating that has historically been popular in american sport. In that context RPI = $0.25 \times \text{WP} + 0.5 \times \text{OWP} + 0.25 \times \text{OOWP}$, where WP is win percentage, OWP is opposition's win percentage, and OOWP is opposition's opposition's win percentage. In rugby union an analogous rating can be produced by using average points per game in place of win percentage

[3] Expected Net Score is used to describe a model whereby the match scores rather than the result outcomes are used. It is assumed that the score difference in any match has a normal distribution with a mean equal to the difference in quality between the two teams. These qualities may then be calculated in such a way to minimise the aggregated errors, and teams then ranked on the basis of these qualities

[4] Under this schema a family of rankings were considered where rather than adding Additional Points to the League Points earnt, those League Points were multiplied by a factor, which depended on a measure of the strength of opposition

measure used in college basketball appears to be towards one more similar to RASR, though in the case of that tournament they have explicitly judged it as a predictive model, and have allowed some non-result elements, for example score, though not others, for example schedule order. However it is the idea of a poll, either of an informed and defined group, or of a self-selected wider group, that may also be one that SOCS might wish to consider. The idea of having public engagement in the outcome in this way may be somewhat appealing. We do not examine this further, though we would suggest that there may be greater tolerance of subjectivity in deciding finalists — with the winner determined through objective (knock-out) competition, as is the case in the US college tournament — than to deciding the ultimate winner of a tournament through subjective judgements.

5 Concluding remarks

In the first part of this report we established considerable evidence for revising the current methodology. Under the current structure, a very large proportion of teams would be practically unable to win the tournament. It is highly likely that the winner will continue to be one of a small group of predetermined teams. In addition to the issues of fairness this raises, potential sponsors could be put off by these systemic biases, which are likely to become more apparent over time as the same teams continue to win. The zombie impact of matches played multiple seasons ago is also a considerably troubling aspect of the methodology.

In Section 3 we introduced three alternatives, RASR, Dapper, and DMT2, and in Section 4 these were compared. From a purely statistical perspective, RASR would be the preferred choice, encapsulating a coherence and a consistency with standard rugby union league scoring rules that no other methodology is capable of capturing. It could also have the attractive feature of garnering attention for the tournament. The use of objective as opposed to subjective ratings measures is a point of active discussion amongst the large body of college sports fans in the USA. Being the first tournament of its type globally to use the statistically sound Bradley-Terry based model (i.e., RASR) would likely attract interest.

Alternatively RASR showed similarities to the outcome of the Dapper model and the methodology behind Dapper would be considerably more comprehensible to the vast majority of stakeholders, being of the same format as the system with which they have become familiar. One note of caution would be that the testing and calibration of Dapper has been performed on only three seasons, so while we should continue to expect to see better performance than DMT (due to using a single season's results and a more reasonable quantum for Additional Points) the calibration of Dapper might benefit still from further refinement in the light of more evidence from future seasons. We certainly consider Dapper to be a viable, practicable alternative to DMT.

RASR is rooted in a highly principled statistical modelling approach, and as such can reasonably be regarded as the 'gold standard' by which other methods should be judged. But RASR relies on an implicit, iterative algorithm, in contrast to the relatively simple, explicit formula that describes Dapper. In the end, perhaps, the criterion of transparency might reasonably be deemed decisive, in favour of Dapper (not quite mathematically optimal, but readily understood) over RASR (mathematically 'perfect', but a black-box approach that could perhaps prove puzzling to at least some tournament stakeholders).

6 Appendix

6.1 Violations

We define violations to be the proportion of non-drawn matches that result in a win for the lower-ranked team. The decision to remove drawn matches is made since we consider draws to be an uninformative result given their low prevalence, and that the likelihood of two teams having exactly the same ranking measure will be dependent on the type of methodology, and so including would artificially but not meaningfully increase the proportion of violations and advantage certain methodologies for reasons other than their quality. One potential flaw to the violations measure is that it ignores the impact of home advantage. Based on the last three seasons, we see that the average impact of playing at home is of the order of between three and five points, and based on the fitting of the RASR model between three and twenty percent as applied to strength as it is defined within that model. This would suggest that models that account for home advantage are likely to be disadvantaged in this metric compared to those that do not.

We could consider adjusting for this, for example by adjusting scores by this average home advantage. However we choose not to here for a number of reasons: the average home advantage may be produced by the difference between wide home wins and wide away wins, rather than the narrow results that will be affected; the scoring rule recognises that there is a big difference between losing by one point and winning by one point, and adjusting the score in that way assumes a linearity of change with quality which is especially unlikely around the result cusp - good teams grind out wins in close matches. Despite these flaws we consider that the measure is still of utility but it encourages us to also introduce a second measure of gross violations.

6.2 Gross violations

We define gross violations to be the proportion of matches ending in a wide result that are won by the lower-ranked team. This enables us to avoid the situation where the victory could be claimed to be from a home advantage effect, although there may still be a home advantage effect at play in the pool of matches that are under consideration. However taken along with the other metrics we consider this still to be informative.

6.3 Ranking similarity measures

The first measure is the mean of the absolute difference between the ranks of the teams under the two methodologies.

$$\eta_1 = \frac{1}{m}\sum_{i=1}^{m} \mid {}_a r_i - {}_b r_i \mid$$

where ${}_a r_i$ and ${}_b r_i$ are the ranks of team i under methodologies a and b respectively.

The second represents the mean of the proportionate difference and is defined as the difference between the exponential of the mean of the absolute difference between the logs of the ranks under the two methodologies and one.

$$\eta_2 = \exp\left(\frac{1}{m}\sum_{i=1}^{m} \mid \log {}_a r_i - \log {}_b r_i \mid\right) - 1$$

with ${}_a r_i$ and ${}_b r_i$ having the same definition as before.

6.4 Dapper

In calibrating Dapper we have two parameters that we are seeking to optimise. We will look at the overall similarity to RASR, and also the violations.

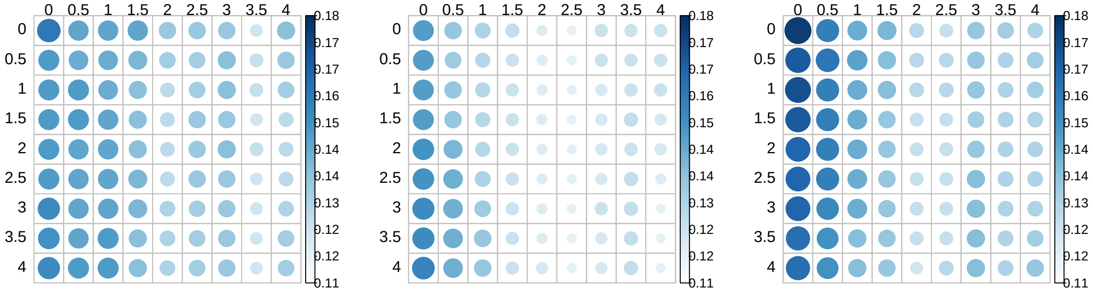

Figure 10: Dapper1: Change in violations with Max AP (x axis) and Prior Matches (y axis)

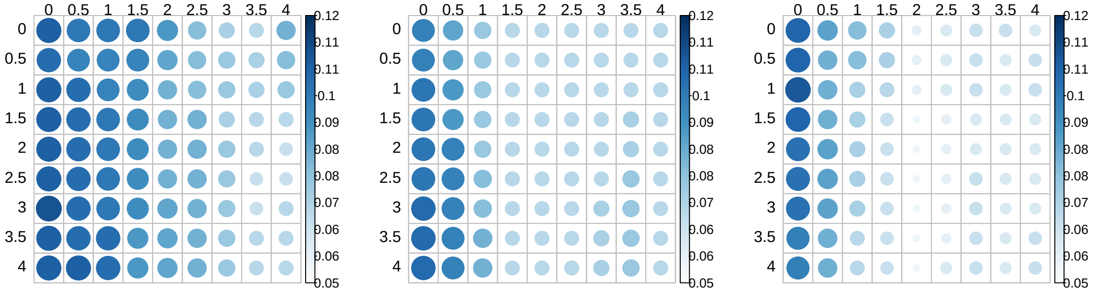

Figure 11: Dapper1: Change in gross violations with Max AP (x axis) and Prior Matches (y axis)

The analysis is somewhat inconclusive, and on violations there is some disagreement between the 2015/16 and 2016/17 seasons on the one hand and 2017/18 on the other. 2015/16 and 2016/17 suggest setting Max AP in the range 2-2.5. 2017/18 suggests a higher value for Max AP. None of the seasons is particularly differentiating to the number of Prior Matches. Comparison to RASR suggests Max AP in the range 2-3 and again is inconclusive on Prior Matches. Given that the proportional differences put more weight to the differences at the top of the table then we might consider the evidence of that analysis as being more influential. This suggests a range of 2-2.5. There remains a large amount of subjectivity in the decision. Given the structure of the use of

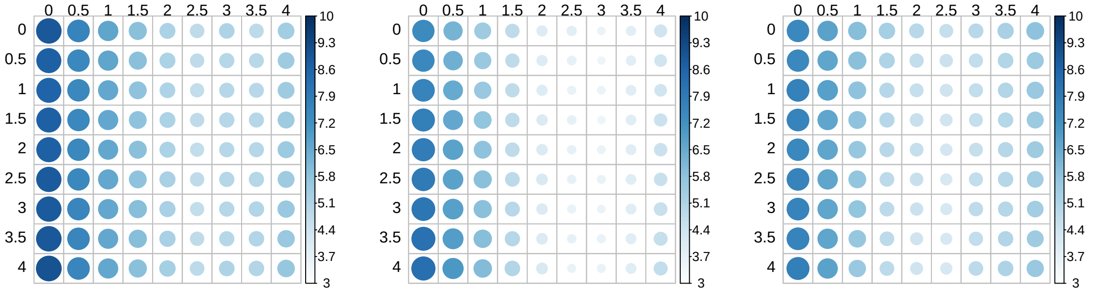

Figure 12: Dapper1: Change in absolute difference to RASR with Max AP (x axis) and Prior Matches (y axis)

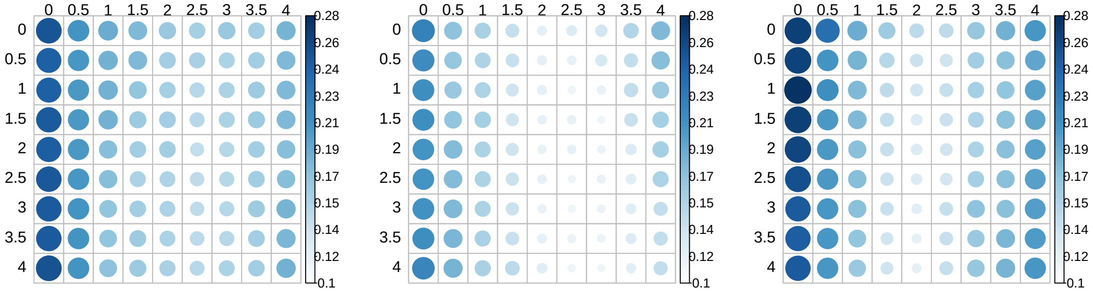

Figure 13: Dapper1: Change in proportional difference to RASR with Max AP (x axis) and Prior Matches (y axis)

the Max AP parameter it is convenient to take something nicely divisible by three. On balance we choose to take Max AP to be 2.25.

It is tempting in looking at the analysis to say that Prior Matches has an immaterial effect and can therefore be set to zero on the grounds of simplicity. However the concept that a 5-0 record is not as good as a 10-0 record against equivalent quality opposition is one that has been fundamental to the tournament from the start, and is particularly important in our differentiation of top teams. We must be cognisant here that this is a subjective judgment. Agreement with RASR, to the degree that there is differentiation, is largely a reflection of the subjective judgment that has been applied in the calibration of RASR. Having determined our value for Max AP, we may seek to understand this more explicitly. Looking at the last three seasons results, and using Max AP of 2.25 then the average Additional Points per Match for top teams is approximately 1.5. Using this assumption then one way to think about it is by comparing playing records and determining which value of n better matches the outcome that is preferred.

Some results from this are displayed in Tables 13 and 14. This is a highly subjective determination, but based on these tables, we choose to take n to be 3 for further analysis, as this value seems to provide a reasonable weighting to results.

6.5 DMT2

In order to calibrate Max AP within DMT2 we look at a range of Max AP and compare the similarity of the top ten and overall rankings with RASR, and look at violations. In doing this, for computational ease we take the current DMT methodology in providing the previous season ranking on which Additional Points are based. We also look at the comparative ranking under different values of Max AP as presented in Table 6 of Section 2.2 , since in the DMT framework it is the weighting of Additional Points that gives some differentiation in comparing teams with greater or fewer matches played.

The violations data in Figure 14 is notably inconclusive. Comparing the overall results to RASR in Figure 15 gives us a stronger indication and suggests a range of 1-2 for Max AP. Comparing the top tens is also inconclusive. The ranking of Reed's School in the 2017/18 season might suggest a value for Max AP of 0.1 or less. Similarly the rankings of Kirkham in 2016/17 and Berkhamstead

		n				
Record 1	Record 2	1	2	3	4	5
5-0	6-1	1	1	1	1	1
5-0	7-1	1	1	1	2	2
5-0	8-1	1	2	2	2	2
5-0	9-1	1	2	2	2	2
5-0	10-1	1	2	2	2	2
5-0	11-1	2	2	2	2	2
5-0	7-2	1	1	1	1	1
5-0	8-2	1	1	1	1	2
5-0	9-2	1	1	1	2	2
5-0	10-2	1	1	2	2	2
5-0	11-2	1	2	2	2	2
5-0	17-2	2	2	2	2	2
6-0	7-1	1	1	1	1	1
6-0	8-1	1	1	1	2	2
6-0	9-1	1	1	2	2	2
6-0	10-1	1	2	2	2	2
6-0	13-1	2	2	2	2	2
6-0	17-1	2	2	2	2	2
6-0	8-2	1	1	1	1	1
6-0	9-2	1	1	1	1	1
6-0	10-2	1	1	1	2	2
6-0	11-2	1	1	2	2	2
6-0	14-2	1	2	2	2	2
6-0	20-2	2	2	2	2	2

Table 13: Dapper: Preferred playing record for different values of n, assuming five points for win (i.e. win + try bonus) and one point for loss (i.e. losing bonus). Additional Points per Match taken to be 0.15.

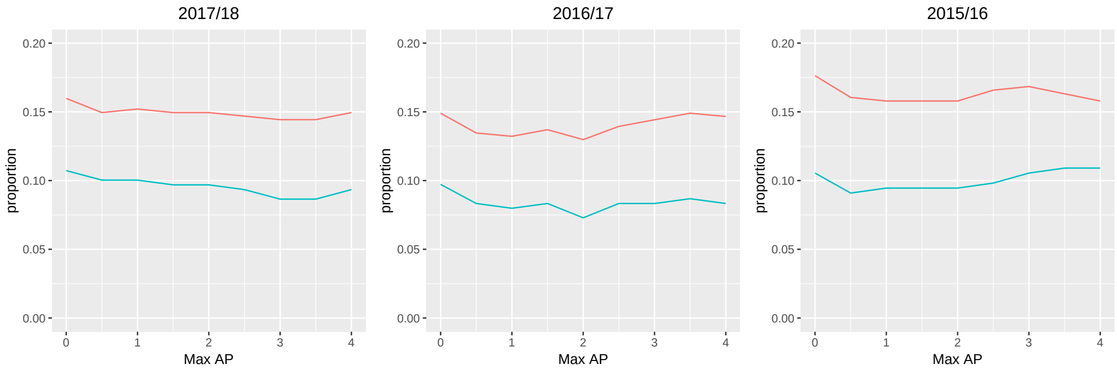

Figure 14: DMT2: Change in violations (red) and gross violations (blue) as Max AP is increased

				n		
Record 1	Record 2	1	2	3	4	5
5-0	6-1	1	1	1	1	1
5-0	7-1	1	1	1	1	1
5-0	8-1	1	1	1	2	2
5-0	9-1	1	1	2	2	2
5-0	10-1	1	2	2	2	2
5-0	11-1	1	2	2	2	2
5-0	7-2	1	1	1	1	1
5-0	8-2	1	1	1	1	1
5-0	9-2	1	1	1	1	1
5-0	10-2	1	1	1	1	2
5-0	11-2	1	1	1	2	2
5-0	17-2	1	2	2	2	2
6-0	7-1	1	1	1	1	1
6-0	8-1	1	1	1	1	1
6-0	9-1	1	1	1	2	2
6-0	10-1	1	1	2	2	2
6-0	13-1	1	2	2	2	2
6-0	17-1	2	2	2	2	2
6-0	8-2	1	1	1	1	1
6-0	9-2	1	1	1	1	1
6-0	10-2	1	1	1	1	1
6-0	11-2	1	1	1	1	1
6-0	14-2	1	1	2	2	2
6-0	20-2	1	2	2	2	2

Table 14: Dapper: Preferred playing record for different values of n, assuming four points for win (i.e. win, but no try bonus) and zero points for loss (i.e. no losing bonus). Additional Points per Match taken to be 0.15.

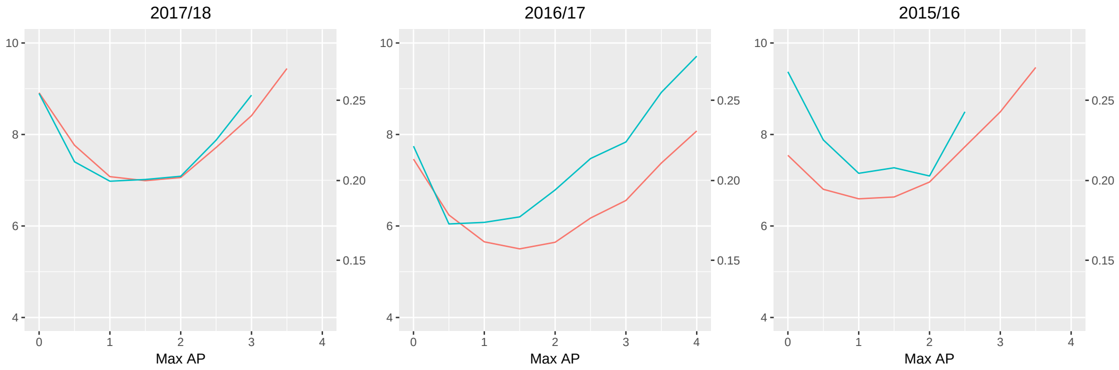

Figure 15: DMT2: Change in absolute difference (red, left axis) and proportional difference (blue, right axis) as Max AP is increased

Team	**RASR**	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4
Sedbergh School	**1**	1	1	1	1	1	1	1	1	1
Reed's School	**2**	2	2	5	6	7	7	8	10	15
Cranleigh School	**3**	4	3	2	3	3	3	3	3	3
Harrow School	**4**	5	5	4	4	4	4	4	4	4
Wellington College	**5**	11	9	3	2	2	2	2	2	2
St Peter's School, York	**6**	7	6	8	7	6	6	6	6	6
Northampton School for Boys	**7**	3	4	6	8	8	9	10	12	16
Cheltenham College	**8**	8	7	7	5	5	5	5	5	5
Haileybury	**9**	9	10	10	9	9	10	9	8	10
QEGS, Wakefield	**10**	10	11	11	11	12	13	13	17	17

Table 15: 2017/18: Top ten with varying Max AP

Team	**RASR**	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4
Kirkham Grammar School	**1**	1	1	2	3	4	4	5	6	6
Sedbergh School	**2**	2	3	3	2	2	2	2	2	2
Wellington College	**3**	6	2	1	1	1	1	1	1	1
Clifton College	**4**	2	5	5	6	6	7	7	7	7
Brighton College	**5**	2	6	6	7	8	9	9	9	11
Harrow School	**6**	5	4	4	4	3	3	3	3	3
St Peter's School, York	**7**	8	7	6	5	5	4	4	4	4
Rugby School	**8**	7	8	8	8	9	8	8	8	8
Canford School	**9**	10	9	9	9	7	6	6	5	5
Millfield School	**10**	20	15	14	12	11	11	11	11	9

Table 16: 2016/17: Top ten varying with Max AP

Team	**RASR**	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4
Kirkham Grammar School	**1**	2	2	2	2	1	1	2	2	2
Bedford School	**2**	1	1	1	1	2	2	3	3	5
Bromsgrove School	**3**	4	3	3	3	3	4	4	5	4
Sedbergh School	**4**	11	8	4	5	5	5	5	3	3
Berkhamsted School	**5**	7	7	9	11	12	19	24	27	29
Seaford College	**6**	3	4	7	7	8	8	12	18	19
Clifton College	**7**	4	5	6	6	7	7	8	14	17
Harrow School	**8**	13	16	19	22	26	26	28	29	27
Wellington College	**9**	16	11	5	4	4	3	1	1	1
Solihull School	**10**	6	6	8	9	9	10	13	20	22

Table 17: 2015/16: Top ten varying with Max AP

and Harrow in 2015/16 would also suggest a value less than 0.2.

In looking at the preferred record comparisons in Table 6 of Section 2.2 we look only at the scoring where a team gains five for a win and one for a loss, since under DMT this is the same as the scenario with four for a win and zero for a loss. As before this is a subjective matter, but it would seem likely that a value in the range 0.1-0.2 would be preferred.

On balance we choose to take a value of Max AP equal to 0.15 for our further analysis.