

What is the best season by a hitter in British top-tier baseball since 2001?

Joe Gray ★ 14 March 2009

Before attempting to answer the question posed in the title of this article, let us first explore its components, which we will do so in reverse order, starting with the chosen time-frame.

The question

Why have the beginning of 2001 as the cut-off?

Using 2001's status as the first season of the 21st century is a little too arbitrary a justification. Instead there is a pragmatic explanation and, coincidentally, a more satisfying baseball-related reason too. The pragmatic reason is that we do not have complete data for all teams in a top-tier league before 2001. The baseball-related reason is that 2001 saw a switch from metal bats to wooden bats in the top tier.¹

What leagues are considered as top tier?

The definition we shall use is that a league is considered top tier if its teams had a chance to qualify for the end-of-season national finals, which gives us the following:²

- ★ **2001** – the Rawlings National League (southern teams) and the Northern League;
- ★ **2002–2003** – the Rawlings National League (southern and northern teams);
- ★ **2004–2007** – the National League South and the National League North;
- ★ **2008** – the National Baseball League South and the National Baseball League North.

We do not have complete stats for the Northern League of 2001, the National League North of 2004–2007, and the National Baseball League North of 2008. Therefore, these leagues cannot be considered in answering the question. This still leaves us with data from each season between 2001 and 2008.

What hitting statistic shall we use?

The overall goal of baseball is to win games, and to do that you must score more runs than the opposing team. Therefore, we want a statistic that summarizes what a batter did at the plate in a given season with a number that is as close to the batter's contribution to team run scoring as possible.

Batting average would not be sufficient because it ignores extra-base power and the ability to draw walks. Instead, we want to incorporate all of the raw statistics such as singles and bases on balls and to weight them according to their contribution to run scoring.

The assigned weightings can be referred to as linear weights (the term “linear” is used because none of the raw statistics are raised to a power).³ There are several means of deriving the weightings, including an empirical method.⁴ The scoresheets for each game would provide us with the level of detail needed for the data analysis, but the number of games in a British season (peaking at 30 in 2004) is far too small to give us enough information on which to base meaningful weightings.⁵

An alternative approach to take is to choose a ready-made statistic that offers a good approximation of run contribution. The now widely accepted OPS (on-base average [OBA] plus slugging average [SLA]) is good, but a statistic known as gross production average (GPA; created by Aaron Gleeman in 2003) is better in two ways: firstly, it uses a weighting that more accurately reflects a hitter's contribution to offensive production; and secondly, to quote Alan Schwarz in a *New York Times* article from 2007,⁶ it has “the comfortably familiar scale of batting average, with figures [in the Major Leagues] generally ranging from .200 (horrible) to .265 (roughly average) to

around .360 (superior).” The equation for GPA is given below:

$$GPA = (0.45 * OBA) + (0.25 * SLA)$$

What do we mean by best?

Now that we have chosen our statistic – GPA – the simplest way to establish the best season a hitter has had is to look for the single highest value, among hitters with at least 60 plate appearances, which is the Great Britain Baseball Scorers

Association cut-off for single-season offensive average records. However, to compare all batters in this way would be to assume that the environment was equally conducive to hitting across the eight seasons under examination. But this assumption is inaccurate, not least because in 2006 and 2007 the top-tier teams played a good chunk of their schedule against the weaker pitching of second-tier teams.

One way to correct for inter-season differences is to calculate the mean GPA among all players with at least 60 plate appearances in a particular season and then calculate the mean GPA in the next year for the subset of these players who had at least 60 plate appearances in that season too. By stringing together multiplying factors, you can set one year as your baseline and then calculate a difficulty factor for each season relative to the baseline (see Figure 1). Unfortunately, these are based on too few players to be totally trustworthy: 16 for the gap between 2000 and 2001, and then 13, 20, 21, 11, five, and nine for subsequent gaps. So they will not form our primary means of correction. However, we will return to them later in this article.

A different way to approach the problem of the shifting hitting environment is to quantify the dominance of a hitter relative to the other hitters in a particular season. While using the other hitters as a yardstick in this way is not a fool-proof means of correcting for the inter-season variability in the environment, it will at least give us a reasonably solid method for tackling our question. To quantify

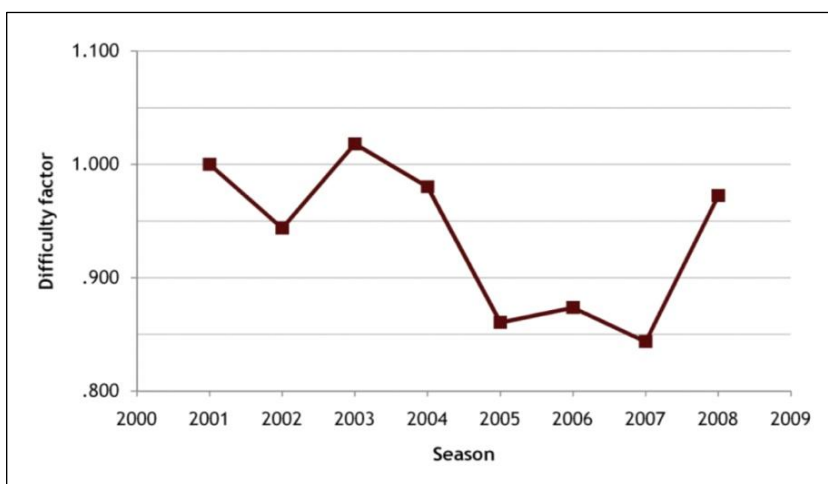


Figure 1. Difficulty factors for offensive production, relative to the 2001 season, for British top-tier baseball between 2001 and 2008. Scores above 1.000 indicate a season that was more difficult than 2001, while scores below 1.000 indicate an easier season.

dominance we can borrow something from statistics called the z-score.⁷

Reframing and answering our question

Building on the exploration of the question’s components, we can reframe it as:

What is the most dominant GPA that a player with at least 60 plate appearances has posted since 2001, as assessed by z-score?

The top-ten most dominant performances are shown in Table 1. Number one on the list is Simon Pole’s 2005 season with the London Warriors.

#	Player	Team	Year	GPA	z-Score
1	Simon Pole	London Warriors	2005	.549	2.444
2	Roddi Liebenberg	Bracknell Blazers	2005	.511	2.146
3	Dennis Grubb	Windsor Bears	2004	.487	1.908
4	Jon Miller	Bracknell Blazers	2003	.446	1.843
5	Ryan Bird	Richmond Flames	2008	.478	1.694
6	Phil Clark	London Mets	2008	.473	1.663
7	Cody Cain	London Warriors	2005	.446	1.621
8	Ryan Barwick	Croydon Pirates	2007	.531	1.552
9	Troy Kantor	London Mets	2008	.448	1.491
10	Mark Rigby	London Mets	2008	.447	1.480

Table 1. The top-ten most dominant single-season offensive performances, assessed using the z-score of GPA, in British top-tier baseball between 2001 and 2008.

Would the difficulty factor approach have given a different answer?

Next we will see what would have happened if we had chosen the difficulty factor correction approach. In this method, we use the difficulty factor for each season (as shown in Figure 1) as a multiplier to adjust all GPAs to a 2001 equivalent. Table 2 gives the top-ten performances as assessed by this method.

So the answer, then, is yes we would have got a different answer if we had chosen this method instead: Dennis Grubb leads the way with an adjusted GPA of .477.

While Grubb's GPA of .487 was adjusted down slightly, to .477, Pole's figure of .549 from 2005 was deflated substantially more, to .472. This is because 2005 was, according to the data we have, the second easiest for hitters among the eight seasons examined.

While we ruled out this method of correction as our primary means of answering the question on the grounds that the multiplying factors were based on too few players – opting, instead, for dominance – the adjustment between 2004 (when Grubb posted a .487 GPA) and 2005 (the year of Pole's .549 GPA) is based on 21 players, the most for any pair of years. So, in relative terms at least, we can be reasonably confident in what the correction is telling us in this case.

The problem of not correcting for pitching

A factor that has possibly the largest bearing on offensive production difficulty is the quality of pitching. This is not corrected for in the current analysis, which could be giving us misleading results.

To illustrate this, imagine a hypothetical scenario in which the hitting environment is constant across two seasons, with the exception that one of the teams in the league substantially upgrades its pitching quality between those two seasons. This will lead to a higher difficulty factor in the second season. However, for batters on the team that upgraded its pitching, the environment will be no more difficult (as they will not face their own pitchers); therefore, it is misleading to inflate these batters' GPAs at the end of the season based on a one-size-fits-all difficulty factor.

While not as extreme as the hypothetical example, the London Mets upgraded their pitching between the 2007 and 2008 seasons to have the best in the league.⁸

#	Player	Team	Year	GPA	GPA ₀₁
1	Dennis Grubb	Windsor Bears	2004	.487	.477
2	Simon Pole	London Warriors	2005	.549	.472
3	Ryan Bird	Richmond Flames	2008	.478	.465
4	Phil Clark	London Mets	2008	.473	.460
5	Jon Miller	Bracknell Blazers	2003	.446	.454
6	Ryan Barwick	Croydon Pirates	2007	.531	.448
7	Roddi Liebenberg	Bracknell Blazers	2005	.511	.440
8	Troy Kantor	London Mets	2008	.448	.436
9	Mark Rigby	London Mets	2008	.447	.434
10	Brett Tait	Croydon Pirates	2007	.510	.428

Table 2. The top-ten best single-season offensive performances, assessed using GPA adjusted to a 2001 equivalent (GPA₀₁), in British top-tier baseball between 2001 and 2008.

Because of the relative quality of their pitching in 2008, the Mets' batters, on average, enjoyed an easier hitting environment than those of the other three teams. The gap between the Mets' pitching and that of other teams was such that it is plausible for the difficulty factor to be less than 1.000 for the Mets' batters and greater than 1.000 for all other batters; that is, we should perhaps be deflating the GPAs of the Mets' batters and inflating those of other batters in 2008. To do this properly, it would be best to look at the average quality of pitching that each individual hitter faced and to adjust all of the GPAs accordingly. This will not be attempted here; instead, the lack of correction for pitching will simply be acknowledged as a weakness of the current analysis and is something that should be borne in mind when interpreting the results.

Other corrections that could be made

Quality of pitching faced is not the only factor that could be corrected for if our one-size-fits-all approach to adjusting GPAs in a given season was replaced with something more sophisticated. An obvious modification that, theoretically at least, could be made would be building in a park factor, so that the relative advantage or disadvantage of having a particular field as your home ground could be ironed out. There are only marginal differences in altitude between the fields in Britain, but wind patterns and dimensions have variation that cannot simply be dismissed as inconsequential.

If accurate park factors were to be calculated for British fields, it would be vital to remove the effect of the quality of the home team's pitching from that of the ballpark itself.

Why was 2005 so easy for batting?

Before the conclusion of this article, there is a question that arose from this analysis that has not yet been addressed. We saw that 2006 and 2007 appear to have been relatively easy seasons for hitting, which was to be expected based on the already noted fact that the top tier teams played some games against second-tier opponents. But this does not explain why the 2005 season appears to have also been a relatively easy season (less difficult, in fact, than 2006).

A significant occurrence between 2004 and 2005 was the withdrawal of the Windsor Bears from the top tier. Ryan Koback and Dean Stoka, the Bears' two main starters in 2004, were two of the best three pitchers that season (the other was Cody Cain).⁹ With Koback and Stoka gone – and Cain throwing only just over half the number of innings he did the year before – 2005 had relatively weak pitching.

The overall standard of pitching improved in 2006, aided by the folding of the London Warriors at the end of 2005, which saw several strong arms from the Warriors join the Richmond Flames, greatly improving that team's pitching strength. In addition, the Croydon Pirates got a full season out of Ben Percey, who had joined the team late in the 2005 season, and also added the arm of Adam Lemke – Percey and Lemke were both stand-out pitchers. The improvement seen in 2006 was possibly sufficient to offset the introduction of games against second-tier teams.

Conclusion

Statistically, the most dominant hitting performance by a player with at least 60 plate appearances in a season since 2001 was that of Simon Pole playing for the London Warriors in 2005. However, it does not necessarily follow that this was the outright best performance by a batter after correcting for the difficulty of the hitting environment. Using an adjustment based on a sample of 21 players who had at least 60 plate appearances in both 2004 and 2005, Dennis Grubb's performance for the Windsor

Bears in 2004 was arguably better than Pole's season the year after.



To comment on this article, please visit
<http://www.baseballgb.co.uk/?p=2120>

Notes

- 1 My source for this is an article titled “Real baseball’ is back” by John Boyd in the Summer 2001 issue of *Double Play*.
- 2 This is taken from Mark Tobin's British Baseball Data website (<http://www.britishbaseballdata.co.uk/>), the website of the British Baseball Federation (<http://www.britishbaseball.org/>), and Appendix 6 of *British Baseball and the West Ham Club* by Josh Chetwynd and Brian A Belton (ISBN-13: 978-0-7864-2594-5).
- 3 For instance, if “S” is singles and “W” is walks, the formula would contain $0.5*S + 0.3*W$ rather than $0.5*S\text{-squared} + 0.3*W\text{-cubed}$. The latter would not be linear as the raw statistics are raised to a power.
- 4 For details, see the section titled “Evaluating statistics in relation to wins” in my article *Does sabermetrics have a place in amateur baseball?* (<http://www.baseballgb.co.uk/?p=2063>).
- 5 Statistically, the numbers churned out would be spuriously precise, because any real information that they contained would be swamped by sampling error.
- 6 See <http://tinyurl.com/au5lb5>.
- 7 In technical terms, the z-score is the number of standard deviations that a particular value lies from the mean of a sample. All you really need to know, though, is that the higher the z-score the more dominant a player has been.
- 8 The London Mets' pitchers struck out 2.84 batters for every one they walked unintentionally or hit, compared with figures of 1.47 for the Richmond Flames, 1.39 for the Croydon Pirates, and an incredibly low 0.57 for the Bracknell Blazers.
- 9 These three pitchers each struck out over 4.00 batters for every one they walked unintentionally or hit, while no other pitcher with more than 5.0 innings logged exceeded 3.00.