



Butler scoring: median or mean?

by David Stevenson



The following question was asked:

Which is better for Butler scoring, median or mean scoring?



This was my reply:

I put the question to a panel of experts on the [Bridge-laws mailing list \[BLML\]](#) and the answers were very long and varied.

The most important answer in my view came from David Burn of England who said that when the English selectors had discussed the problem of how to score their trials a few years ago he had taken the scores from the last trials and re-scored them using every known method he could think of. The order of the main results was unchanged in the various methods and only the most extreme methods altered the order in any way whatever.

According to David Grabiner of the USA, Robin Barker of England and Jeff Goldsmith of the USA the disadvantage of the median method of scoring [taking the middle result as the datum, or the arithmetic mean between the two middle results with an even number of frequencies] is primarily because of "polarised" boards. If there is a close 3NT that everyone will bid but might make or go off, then you might get a set of scores such as:

+1100 +600 +600 +600 +600 +600 +600 -100 -100 -100 -100 -100 -200

Using the median method the datum is +600 and the imp scores become

+11 0 0 0 0 0 0 -12 -12 -12 -12 -12 -13

Now, if one score is changed from +600 to -100 the datum becomes -100 and the imp scores become

+15 +12 +12 +12 +12 +12 0 0 0 0 0 0 -3

It does not seem reasonable that changing one score should make such an enormous difference. However, when the boards are not polarised this looks quite a fair method. It has the advantage that the datum score is a "real" score for people to imp against, and on non-polarised boards people quite like it.

According to the same three people the disadvantage of the arithmetic mean method of scoring is the losses and gains on ordinary scores and the fact that the datum is never a real score.

If you get a hand where slam is well against the odds but makes then you might get scores thus:

+230 +680 +680 +680 +680 +680 +680 +680 +1430 +1430 +2210

The arithmetic mean would be +915. However, it is generally accepted with this form of Butler that you drop the extreme scores: with 11 scores you would drop one from each end for the datum, so you are left with

+680 +680 +680 +680 +680 +680 +680 +1430 +1430

and the arithmetic mean of these is +847.

This would lead to imp scores of

-12 -5 -5 -5 -5 -5 -5 -5 +11 +11 +16

The trouble is that by playing in the "normal" contract players have lost 5 imps. The median scoring method would have given these players 0 imps, which feels instinctively correct.

If you do play arithmetic mean Butler you will get the problem of deciding how many scores at the end to drop. Probably one at each end up to 10 frequencies, two up to 25 frequencies and three at each end for more.

Jeff Goldsmith of the USA also mentions the mode method [taking the most frequent result as the datum, or the arithmetic mean between the most frequent results if they have the same frequency]. This has the same disadvantages with "polarised" boards. He feels it is an unstable method, but often the datum score will look fairer to the customers: instinctive feel suggests the most frequent score on a lot of boards. He also points out that with any method datum scores should be calculated across the field [not section by section].

He also had an idea that a mixed method depending on the type of scores [polarised or not] might have an attraction but had no idea what formula one could use.

Jeremy Rickard of England said that the disadvantage of the median method is that a small change in the set of scores can make a large difference in the median. The disadvantage of the mean method is that it distorts the IMP scale. In general, it exaggerates the difference between below-average scores and above-average scores.

Con Holzschereer of the Netherlands points out that the mean method can be thought of as a weighted mean with weightings thus:

0 0 1 1 1 1 0 0

He suggests that the weighting might be increased for very small numbers of tables. For example, with five tables, he uses a weighted average thus: multiply the scores by

1 2 2 2 1 respectively,

add them up and divide by ten.

Herman De Wael of Belgium says that rather than having limits for how many scores to delete using the mean method you should delete 10% of scores at each end. While he does not say, I believe 10% rounded up would be best with a maximum of three.

He also has an improvement that is small but complicated, and is primarily to avoid a particular circumstance whereby a small increase in a person's score can actually lead to a decrease in their imp score. The method is called Bastille and is described [on Herman's web site](#).

All people who provided an opinion were of the view that the mean method, dropping the extreme scores, was slightly better. In my view if you use Butler this is what you should do.



Robin Barker wrote: A method which I rather like is to chose par so the net NS score is zero; I thought this was Bastille (till I read Hermans web pages). It is not always possible to get the net NS score zero (and hence, the net EW score is zero) without using fractional scores and IMPS.

But you can make the score as close to zero as possible, although the calculations are more complicated than mean or median, only the computer has to do the complicated calculations -- contestants get a real par score to check their score against.

For comparison, I have run this method against the data sets in your article, plus and extra set which shows when it goes wrong.

Par: 290; net: 0
 -200 -100 -100 -100 -100 -100 +600 +600 +600 +600 +600 +600 +1100
 -10 -9 -9 -9 -9 -9 +7 +7 +7 +7 +7 +7 +13

Par: 230; net: 0
 -200 -100 -100 -100 -100 -100 -100 +600 +600 +600 +600 +600 +1100
 -10 -8 -8 -8 -8 -8 -8 +9 +9 +9 +9 +9 +13

Par: 830; net: 0
 +230 +680 +680 +680 +680 +680 +680 +680 +1430 +1430 +2210
 -12 -4 -4 -4 -4 -4 -4 -4 +12 +12 +16

Par: -60; net: -3
 -200 -100 -100 -100 -50 -50 -50 -50 -50 -50 +90
 -4 -1 -1 -1 +0 +0 +0 +0 +0 +0 +4



There is a method called cross-imping. You take a North/South score, and you calculate theimps obtained treating each of the East/West pairs separately as their team-mates. You add these scores up - and repeat for every other pair. Taking as an example the scores

+230 +680 +680 +680 +680 +680 +680 +680 +1430 +1430 +2210

For each pair with +680 they get

$$(1 * +10) + (6 * 0) + (2 * -13) + (1 * -17) = -33 \text{ imps}$$

Each pair who got +1430 will get

$$(1 * +15) + (7 * +13) + (1 * 0) + (1 * -13) = +93 \text{ imps}$$

Sometimes the organisers divide by the number of comparisons so that the above examples would come to -3.3 and +9.3 imps. This is to try to make the scores look more like normal imps.

Most good players and most of the people on BLML consider cross-imps is fairer. Very few have ever given a satisfactory reason - it seems to be more of a gut feeling. David Burn's analysis referred to above included cross-imps.

There are disadvantages. It is more difficult for players to check their scores, and it is more difficult for them to get a feel for why the results are as they are. Players who score +93 imps on a board find it difficult to relate to a real score, and when divided by the number of frequencies the scores seem "flattened": they come to less than seems right.

For this reason average plus should be less than +3 imps if using the scores divided by frequencies method: some use +2 imps. The English Bridge Union uses the square root of 8 times the number of frequencies before the scores are divided by the frequencies: with ten scores this comes to +9 imps, 0.9 after dividing: no-one knows where this formula came from and it seems wrong as well!

In a club or lesser competition I would always advise Butler with a mean, because the players like a datum score to compare with and to check their scores. For trials or major competitions cross-imps seems most acceptable to the players: you will have to decide what score you give for average-plus and whether to divide the results by any score.



Editor's note:

- If you want further help on this subject, why not write direct to [David Stevenson?](#) 
- I shall be happy to help!
- [The names of the various forms of Butler are discussed here](#)



Last
article

Laws
menu

Main
index

Top of
article

Local
menu

Next
article