The Australian Golf Handicap System

Executive Summary

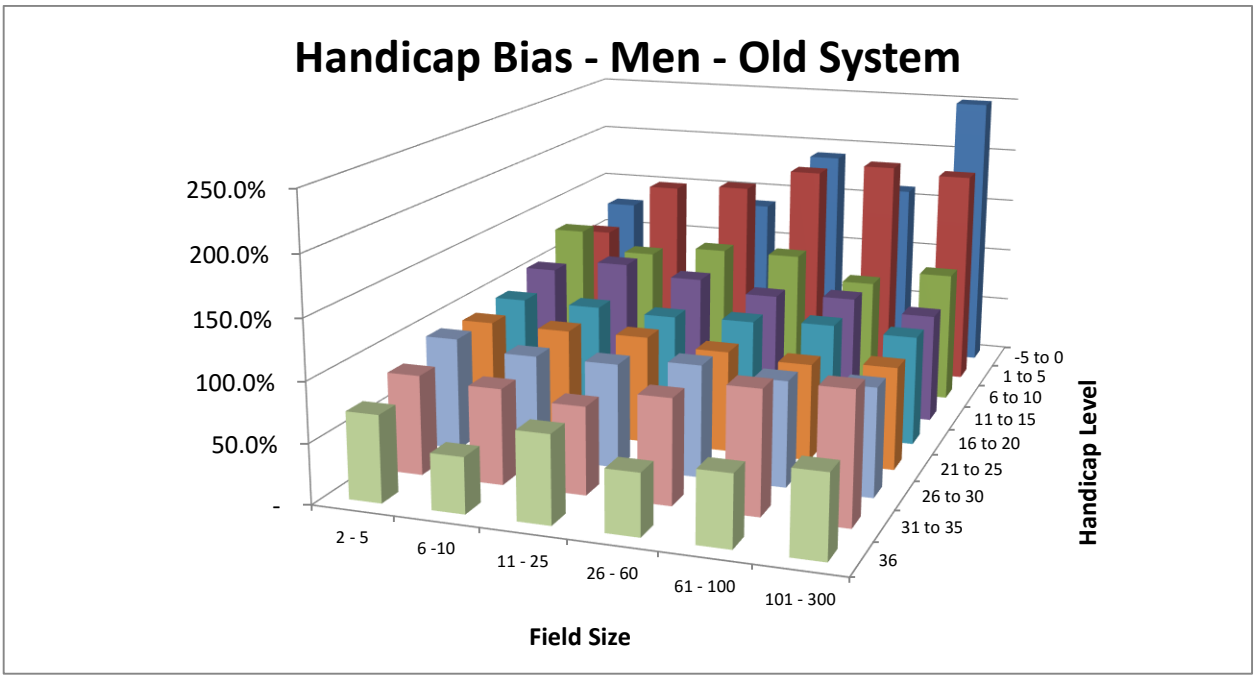
A number of theories and suggestions have been put forward in response to the first nine months of competition under the New handicap system.

Using a sample of one million rounds of golf from the GolfLink system, for over 27,000 competitions involving over 400,000 golfers, an analysis has been completed which provides answers to the following suggestions and questions.

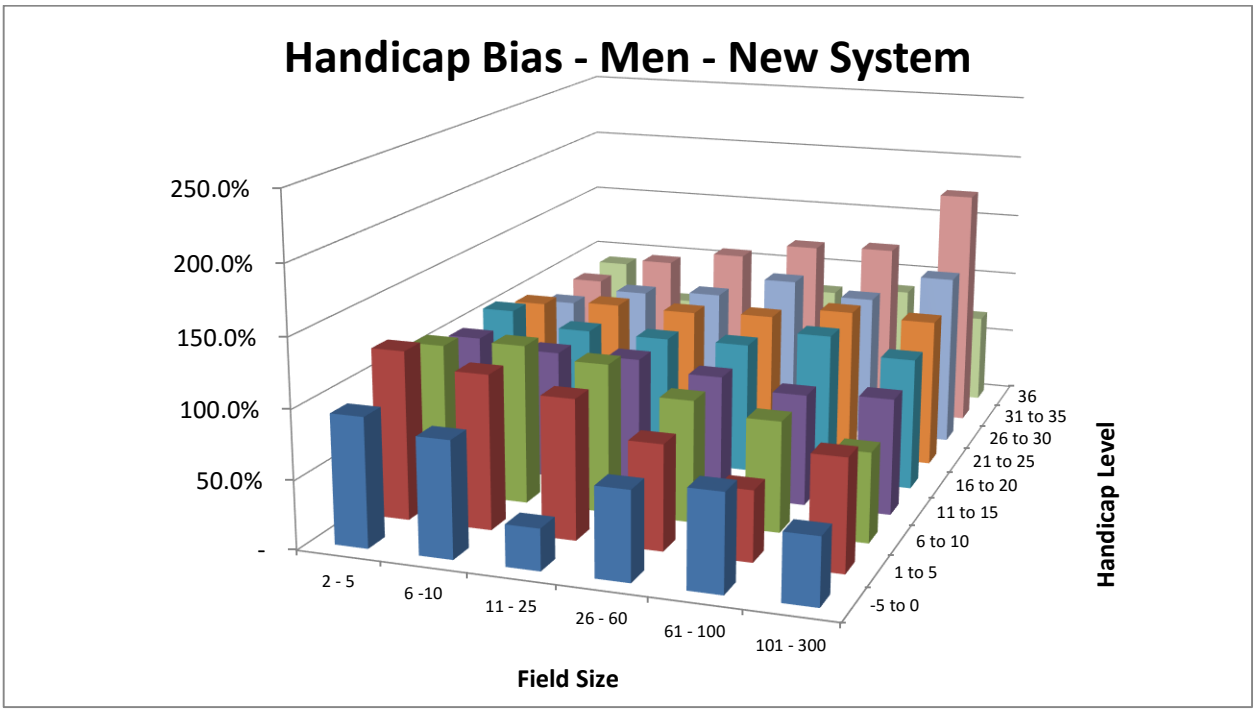
- Higher markers do indeed exhibit a greater scoring standard deviation than do low markers. As it is not uncommon for a score to be two standard deviations from the mean, it should not be uncommon for a scratch golfer to vary by 8-9 strokes from his mean, and for a high marker to vary by up to 16 strokes for men, 18 for women. But the low marker is having this deviation added to a higher mean than the high marker: it turns out that the "Top scores" only vary by two to three strokes across handicap groups.

- Under the Old system, a higher marker was less likely to play to their handicap on a given day than was a lower marker. Men: 7.4% chance v 35.5% chance; Women: 5.3% v 25.0%. This holds true under the New system but to a lesser degree. Men: 10.9% v 21.3%; Women: 7.6% v 17.6%.

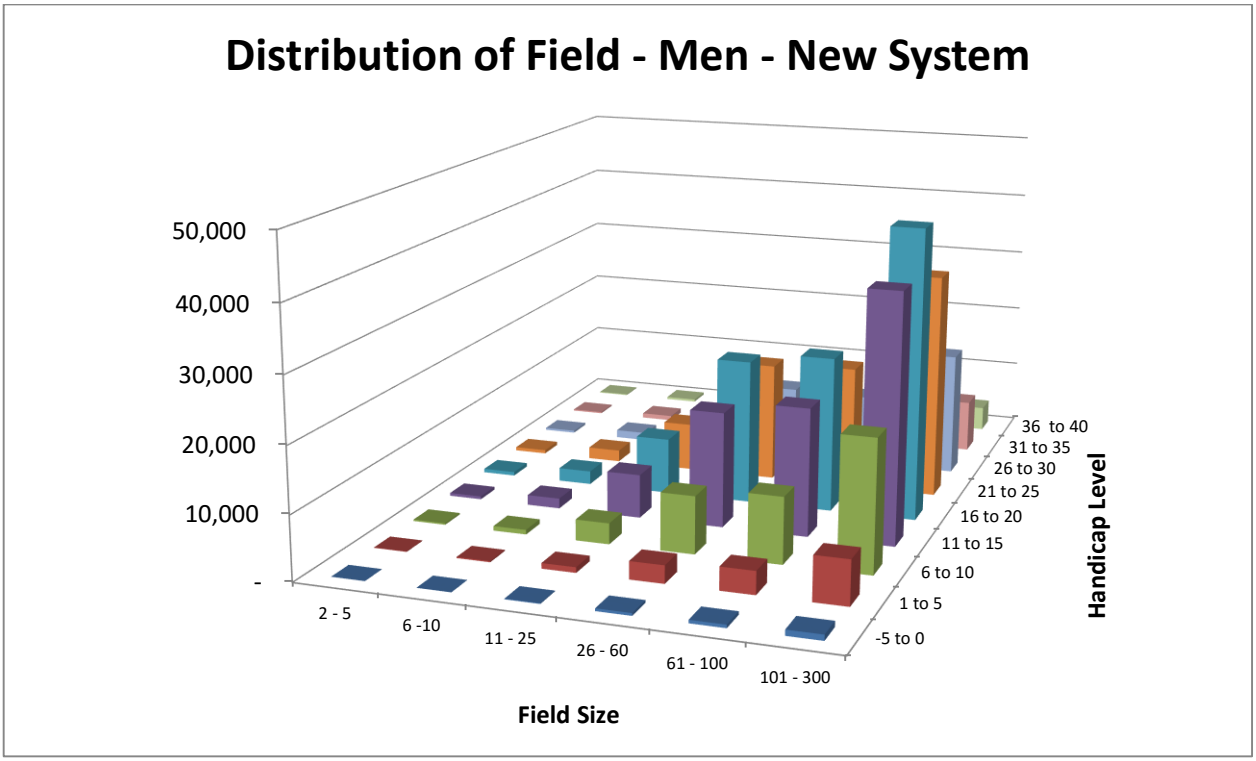
- It was suggested that under the Old system and the New system, the low marker will be advantaged in very small fields. In fact the data shows that there was no small field size advantage for the low marker under the Old system. There is however under the New system.
  - The low marker does find it harder to compete as the field size increases.
  - Under the New system, the field size value at which a low marker is disadvantaged is about 50 for men, 100 for women. They do still compete and are winning competitions at very high field sizes, but the chances of them winning will be lower than occurred with the Old system.
  - So prize patterns do vary depending on the field size, but differently at different handicap levels.
  - Under the Old system, there was a significant advantage to the low markers, which actually grew with field size. Consider first the Men's competitions under the Old system. A method of describing "advantage" is developed in the analysis that follows, and called "Handicap Bias". Under the Old system the "Playing Field" was significantly lifted at one end in favour of the low markers. In this style of chart, the 100% level is neutral, over 100% is a favourable bias, while under 100% is an unfavourable bias. The lowest markers are the dark blue bars rising at the back. But even the red bars rise significantly for the 1 to 5 markers.



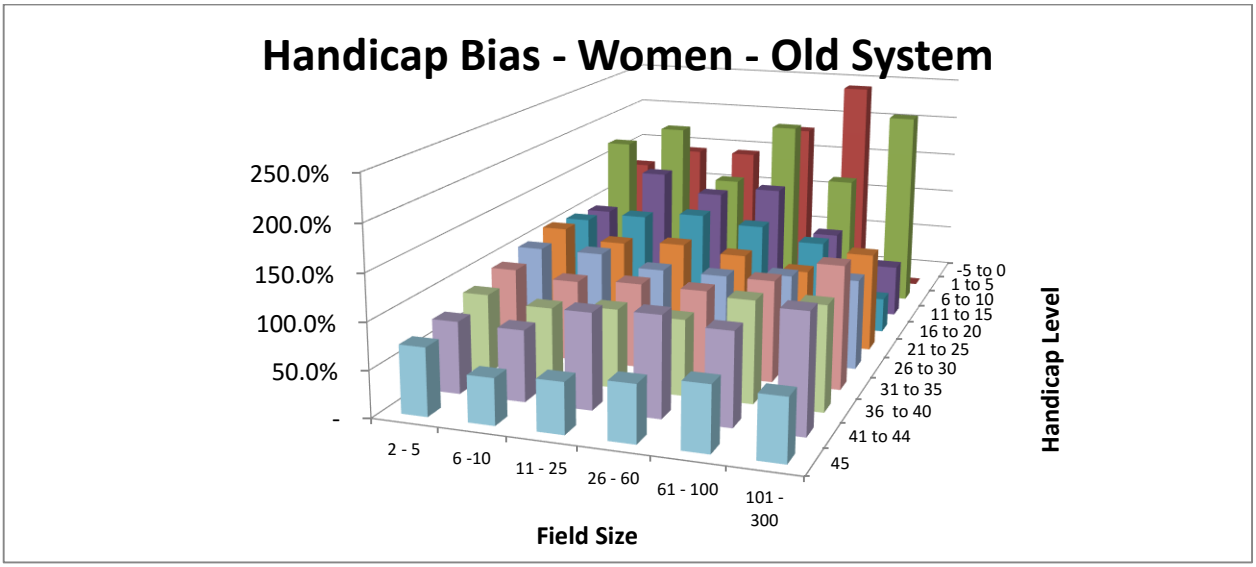
- Under the New system the “Playing Field” is more level, but there are some bumps and variations that disadvantage the lower markers and advantage the higher markers at higher field sizes. Note the low markers are in the foreground here so that they are better seen. One can see the high markers rising with field size at the back. The bumpy nature of the -5 to 0 markers can be attributed to the fact that they represent only 0.5% of the field, and the numbers in the sample are too small to give a smooth ride.



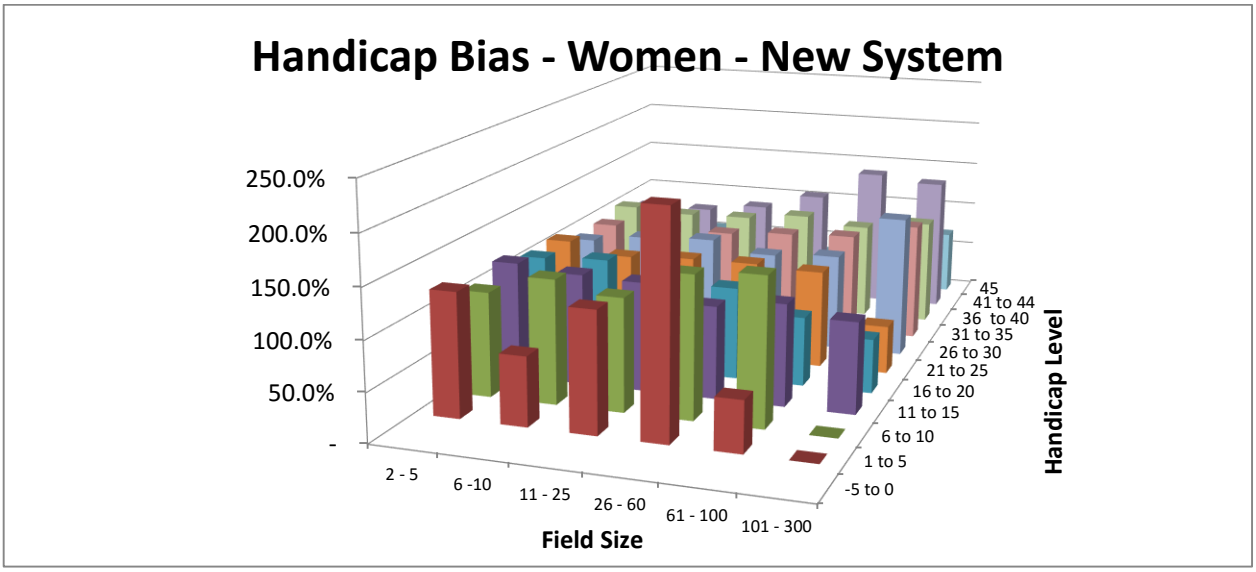
- The number of players in the above chart is as follows. It is clear that most are competing under larger fields.



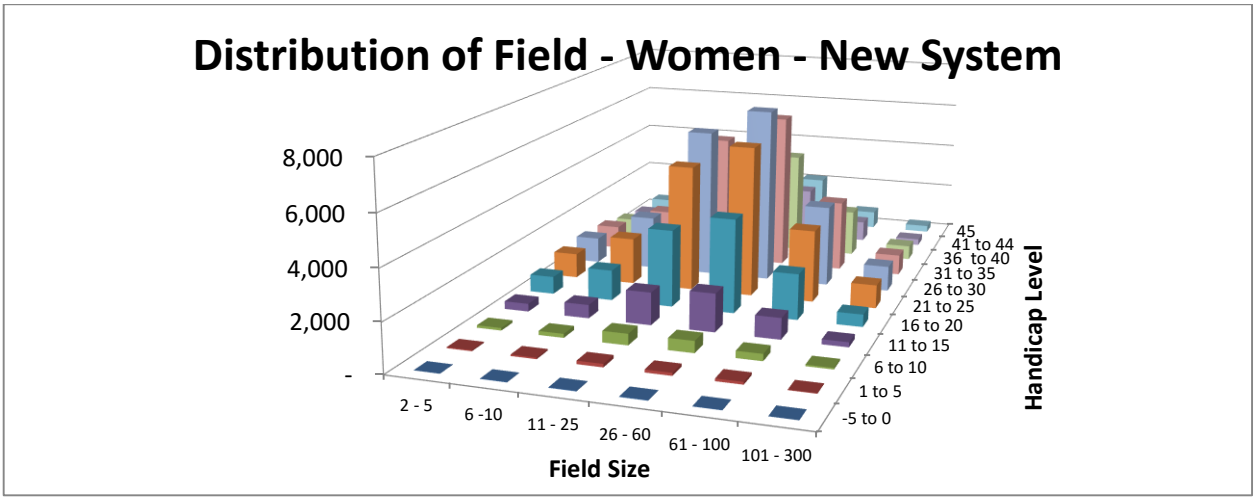
- For the Women's competitions, a somewhat similar pattern is evident. Under the Old system, there was a significant advantage to the low markers, 1 to 10, which actually grew with field size. From 11 to 20 the trend reversed, was neutral in the 20's, and again rose with field size in the 30's.



- Under the New system, the very low markers do not seem to be as disadvantaged as they were for the men. There is still a marked disadvantage to 16 to 25 handicap players at larger field sizes.



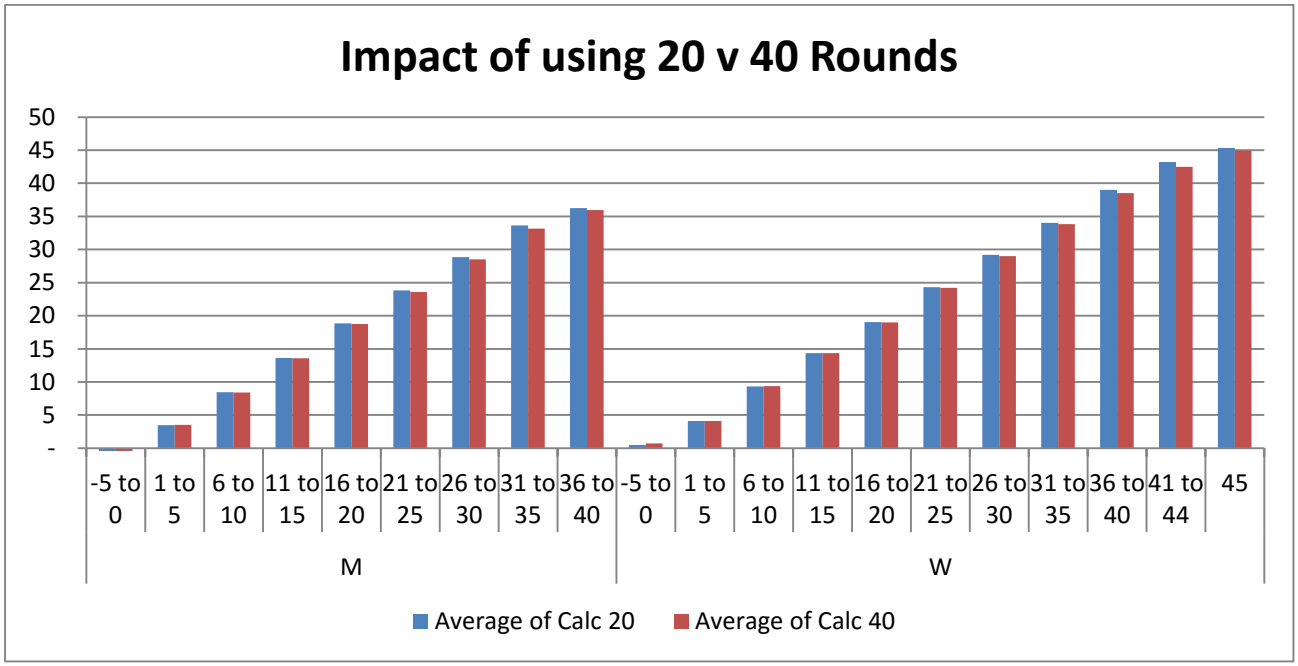
- The Field Size distribution for the chart above is as follow. It is clear that unlike the men, most women are playing in the smaller field size area, off higher handicaps, where the changes are not so extreme.



- So is the New system equitable for small fields but not for large fields? There is still a small favourable Handicap Bias towards the low markers at field sizes up to 50 for men, 100 for women. Over this the bias becomes more and more unfavourable for the low markers and favourable for the high markers.

- Is a 20-score sample large enough to accurately reflect the scoring potential of an inconsistent player?

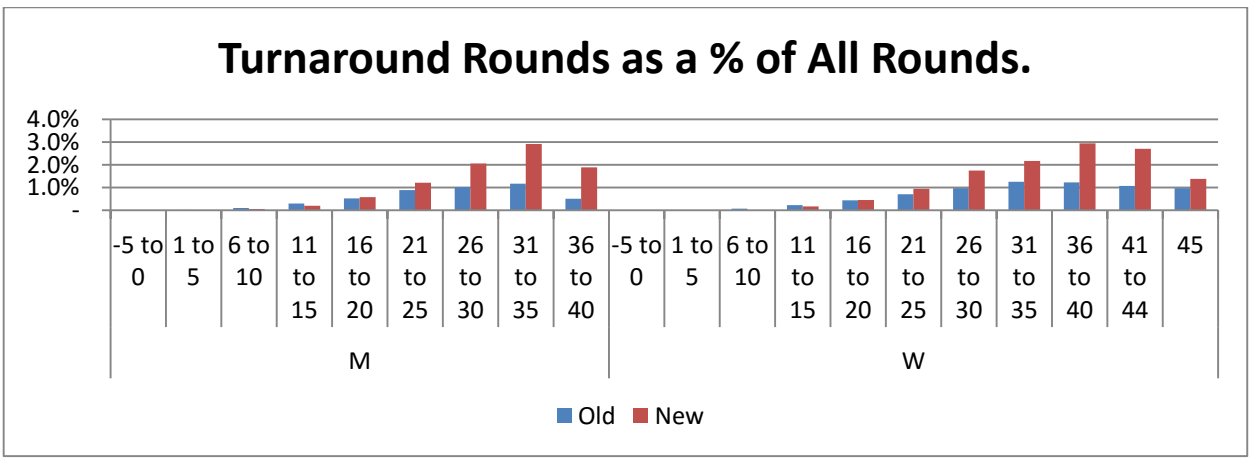
  If we were to use a 40 sample (taking the 20 best scores) vs the current 20 sample (taking the ten best scores), the average variation at low handicaps is negligible, rising to a one stroke reduction in handicap at the highest handicap levels:



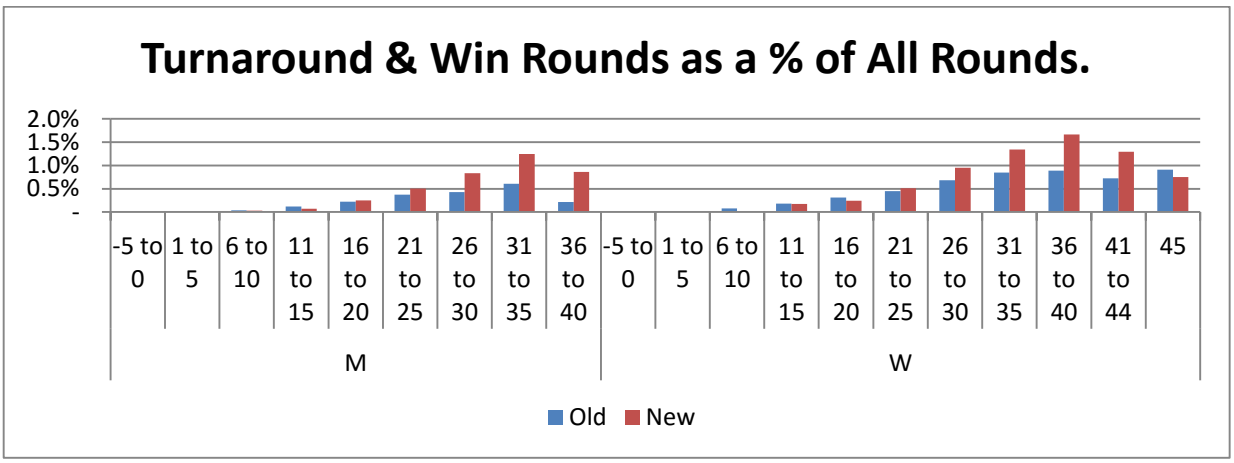
However what is essentially a 40 period moving average responds more slowly to changes in performance than does a 20 period moving average. This means that on a 40 period sample, handicap relief when form deteriorates will be less than for a 20. However the flip side of this coin is that after improvement, the 40 period sample will also respond more slowly to this improvement in form. So if there has been a slump from which there has been a recovery, the improved form will benefit from the higher handicap longer than it will under a 20 sample.

- Is it feasible that a high-marker will endure an extended run of poor scores resulting in a handicap which is not relevant to their scoring potential?

  It is possible that players can have a run of poor scores, which inflates their handicap, followed by an excellent score. Defining a Turnaround as achieving a score of 40 or more, where the average of their last 20 rounds is 30 or less, there has been some increase of these Turnarounds. Over all they represent 1% of all rounds, compared to 0.5% under the Old system.



The occasions when this Turnaround leads to a Win has also increased, overall from 0.25% to 0.46%.

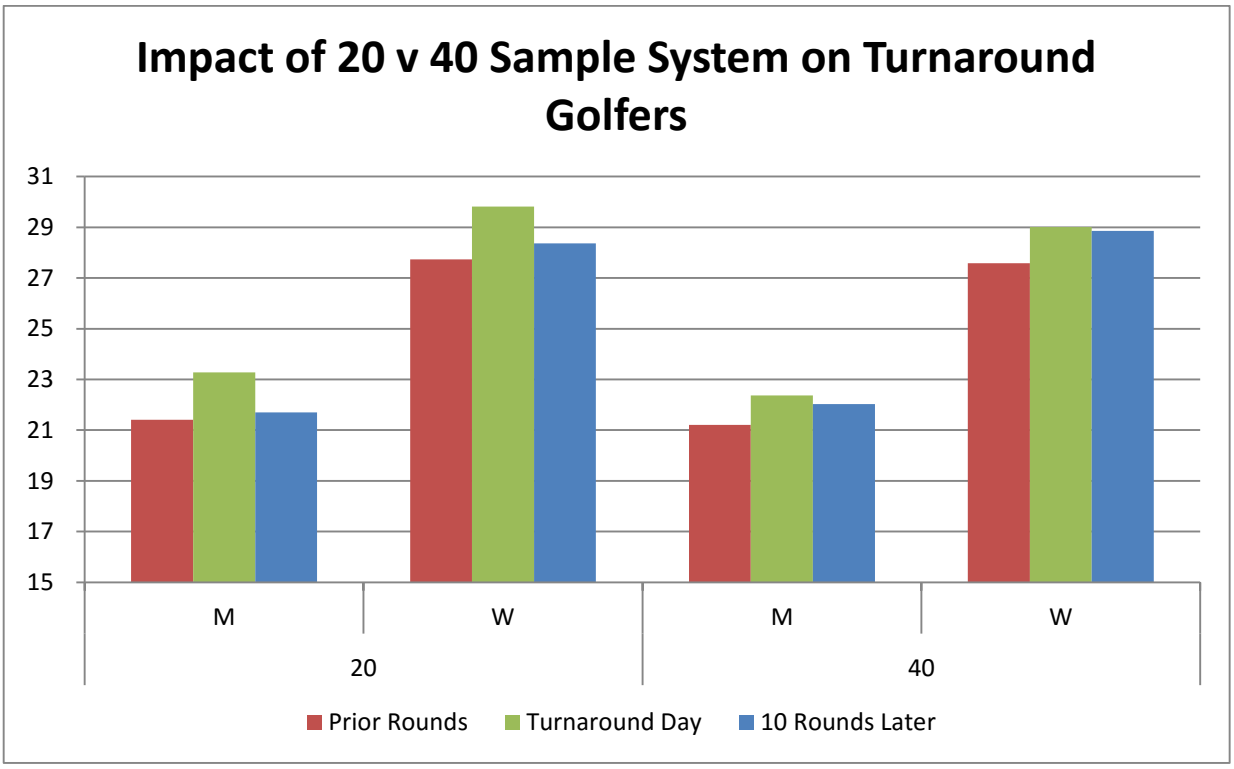


Thus, yes, markers over 26 can have a run of poor rounds and then a very good round. But the Turnaround and Win scenario accounts for less than one half of one percent of the rounds.

- Is it likely that large fields will exhibit enough players in such a situation that the challenge of the low marker becomes further extended to a material extent?

The challenge to low markers, as discussed above, already includes any impact of Turnarounds. In other words, the challenge is as already measured, and includes any impact of Turnarounds.

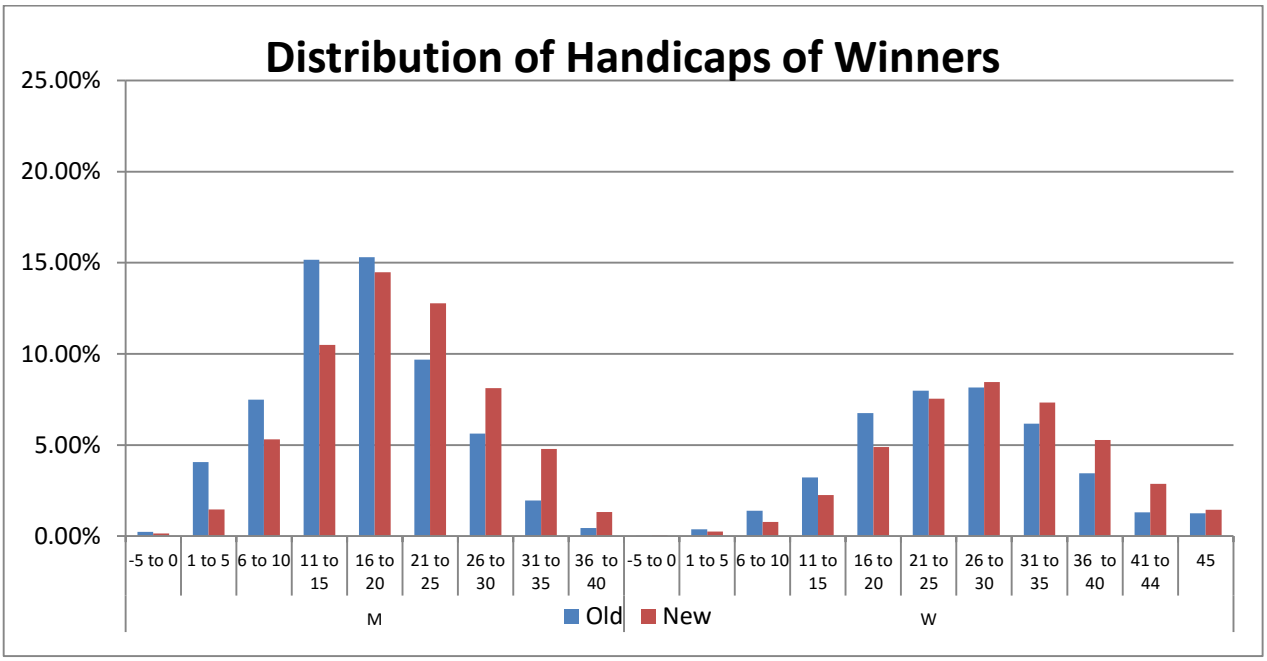
To what extent would using a forty rather than twenty sample impact Turnarounds?



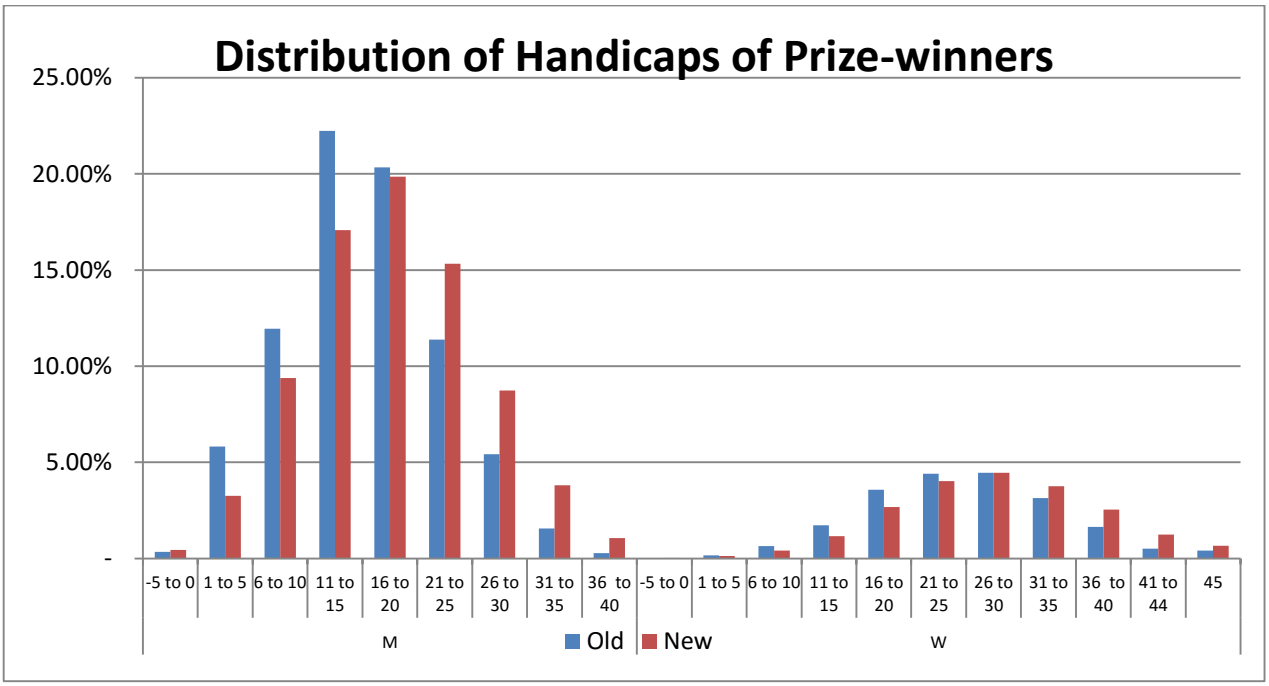
The golfers who have experienced this Turnaround were compared and their handicaps averaged. One can see how the handicap rose by about two strokes under the 20 system, and a little less, one to 1½ strokes under the 40 systems, on the day they had the Turnaround. In the case of the 20 system, the handicap has fallen, almost back to what it was on average, ten rounds later. But in the case of the 40 system, it is still one to two strokes higher than it was.

- What have been the changes in competition patterns between the Old system and the New system?

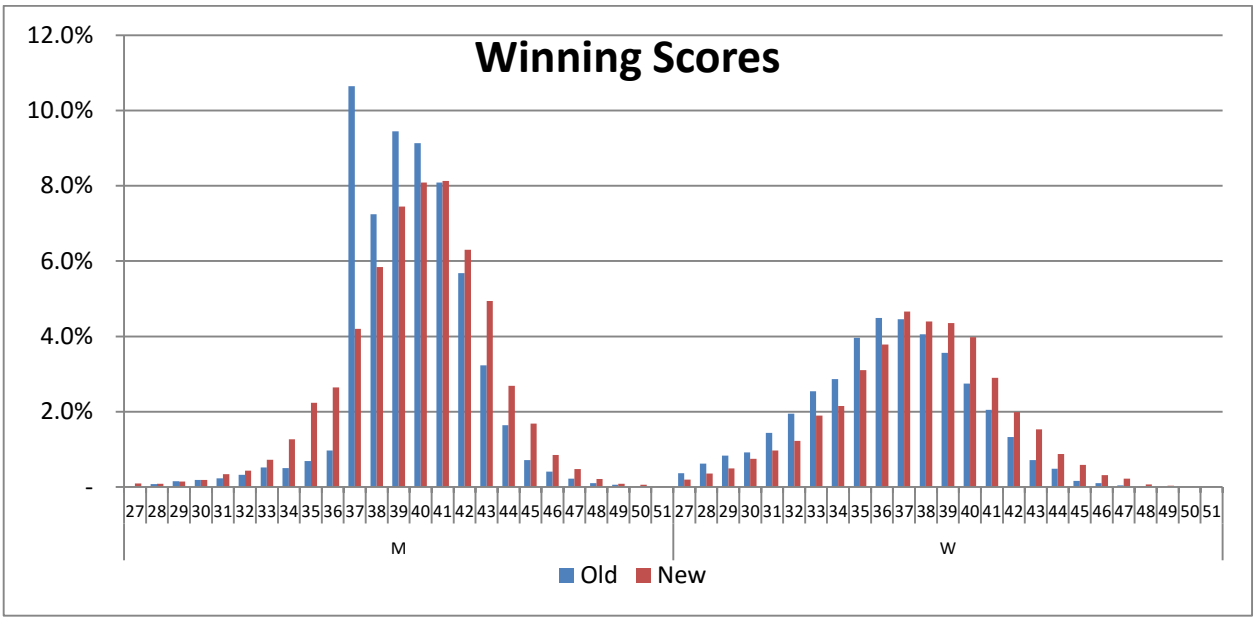
  - What handicaps are winning competitions? Fewer low handicaps and more higher handicaps. But the predominant winners are still in the 16 to 20 range for men, 26 to 30 for women. This has not changed.



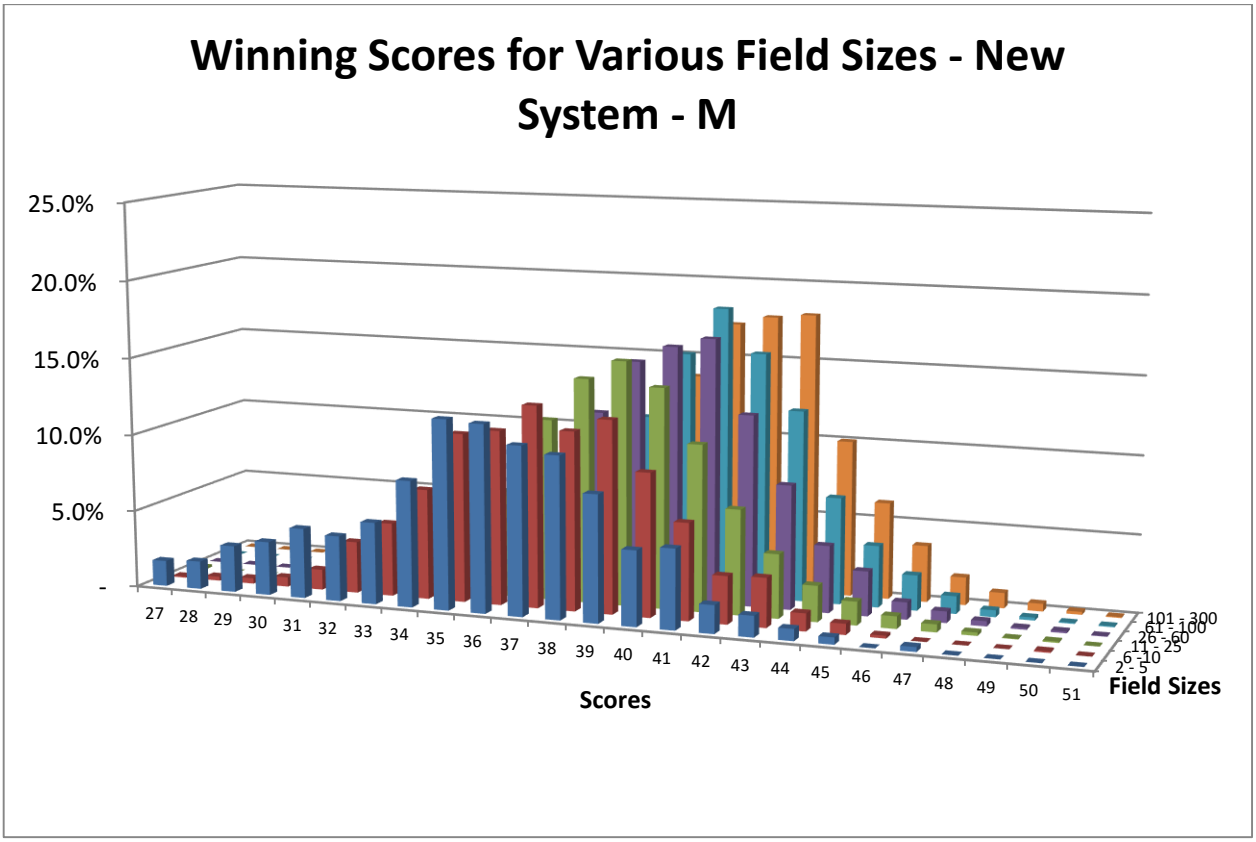
  - What handicaps are placing in competitions? The pattern is similar to the winners:



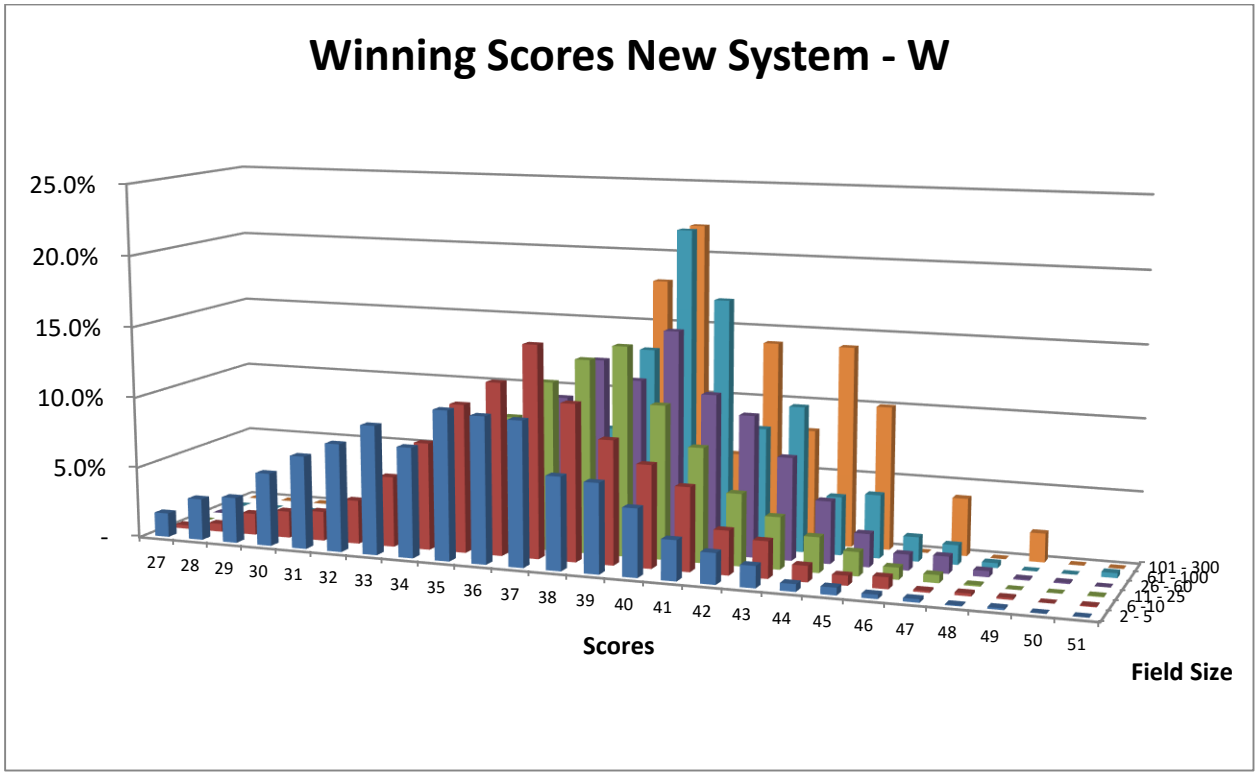
- What scores are required to win or place in competitions? All scores in this report are adjusted for the course rating, and converted to stableford equivalent. As with handicaps, there has been a shift to higher scores. For men, 41 is the predominant score compared to 37 and 39 under the Old system. For Women, 37 is still predominant under both systems, but there is still a clear shift to higher handicapped winners.



- Do these patterns vary depending on field size? Indeed they do. Under the New system the most frequent winning score for fields 6 thru 10 was 37 points. As the field size increased this rose to 39, 41, 41 and 43. Follow the 43 score through different field sizes, its rising share of the winning scores is clear.

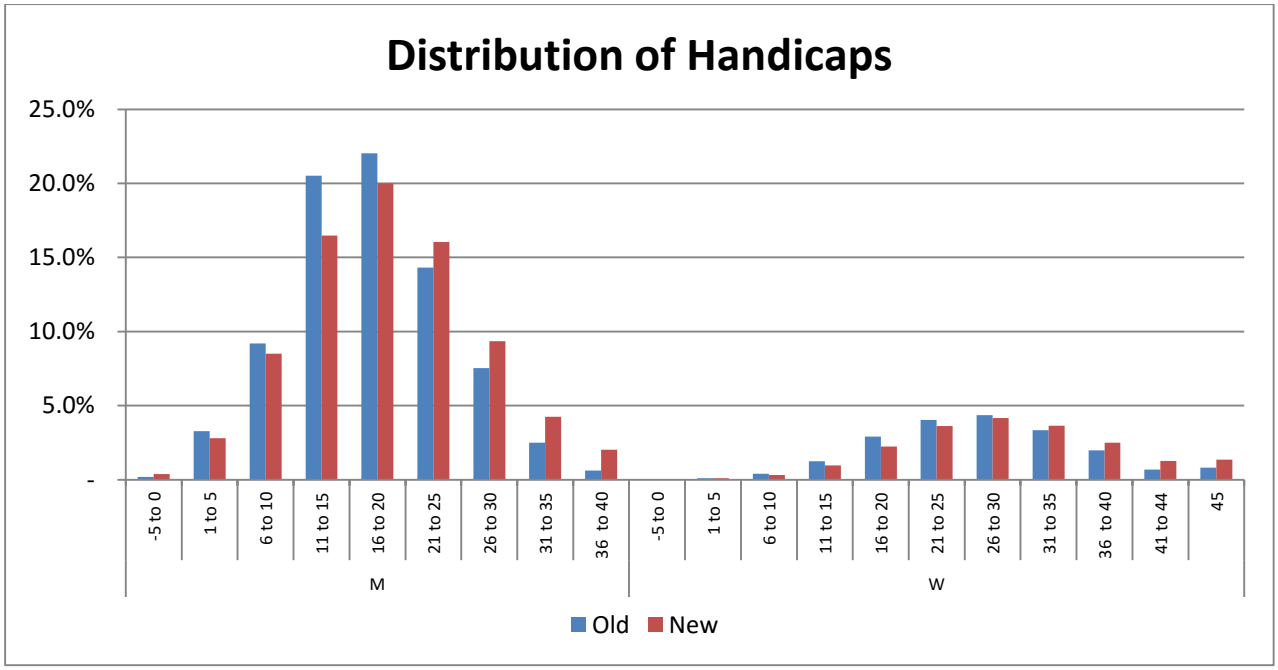


- For women, the pattern is similar.



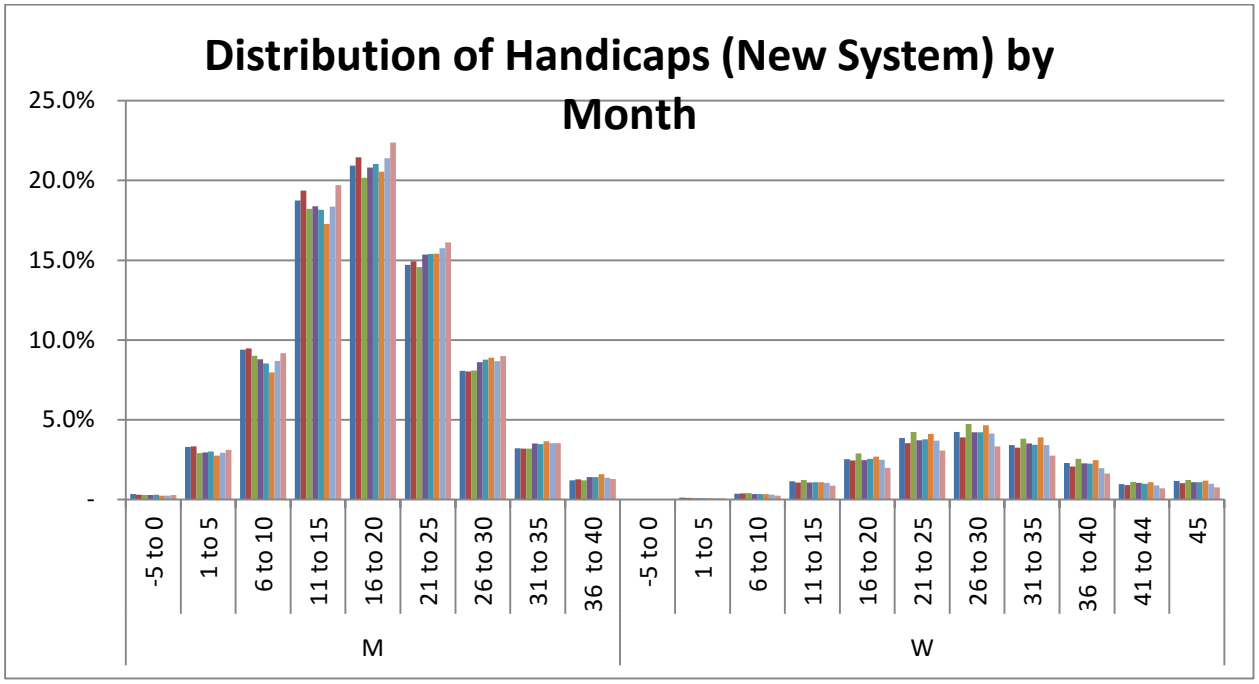
- However it should be noted that what we are seeing here is merely the impact of a larger field on the "standard deviation" effect. With more players in the field, there are more chances for one of them to play to an upper extreme of their performance distribution. The Old system evidences the same rise in winning scores as field size increases.

- How has handicap distribution changed from the Old system to the New system? There has been a distinct move to higher handicaps with the New system.



- Has there been any 'settling down' in the period since 9 April 2010? No, there is no significant variation between the month on month distribution of handicaps in the months following the introduction of the

New system. The "New" handicaps issued on Day One of the New system reflected the same distribution of handicaps as in evidence seven months later.



- What are the implications for an "Equitable" system?
  - Neither the Old nor the New systems are representative of a level playing field.
  - Under the Old, the high markers were disadvantaged so that they won far fewer competitions than their representation in the field. Many of them were chronically playing at handicap levels far in excess of their playing handicaps. Essentially this was due to the uneven way that increments were applied. Handicaps for high markers only "eased out" by 0.1 stroke for a poor round, no matter how poor, but they were tightened far more quickly if the player had the occasional good round.
  - Under the New system players are being handicapped closer to their recent form, but what with the greater inconsistency of the higher markers, when field sizes grow, there are enough chances for a high marker to put in one of his or her exceptional rounds, and the high markers are now winning more than their fair share of the competitions with larger field sizes.
  - It is clear that a reduction in BFE, or in the number of "best" scores taken, would tilt the playing field in favour of the lower markers. However there is likely to be some negative impact on other handicap groups. Modelling of specific solutions should be performed before introducing any change.

Michael Maher BSc (Syd) MBA (Harv)

Sydney 10th March 2011.

Contents

Executive Summary 1

Introduction 11

Higher markers exhibit a greater scoring standard deviation than do low markers: Confirmed. 11

Under the Old system, a higher marker was less likely to play to their handicap on a given day than was a lower marker. This should hold true under the New system but to a lesser degree: Confirmed. 11

Under the Old system and the New system, the low marker will be advantaged in very small fields: Yes and No. 13

- Handicaps Men -5 to 0. 17
- Handicaps Men 1 to 5 17
- Handicaps Men 6 to 10 & 11 to 15 17
- Handicaps Men 16 thru 30 18
- Handicaps Men 31 and over 18

Under the Old system and the New system, the low marker will be advantaged in very small fields: Yes and No. . 18

The low marker will find it harder to compete as the field size increases. Yes – but only in the New System. 18

Under the New system, the field size value at which a low marker cannot realistically compete (or win any sort of prize) will be lower than occurred with the Old system. Yes - But. 19

Hence, do prize patterns vary depending on the field size? 20

Is the New system equitable for small fields but not for large fields? 20

If so, at what field size does the New system become inequitable for the low marker? 25

How does this equate with the Old system? 25

Is a 20-score sample large enough to accurately reflect the scoring potential of an inconsistent player? 25

Is it feasible that a high-marker will endure an extended run of poor scores resulting in a handicap which is not relevant to their scoring potential? 26

Is it likely that large fields will exhibit enough players in such a situation that the challenge of the low marker becomes further extended to a material extent? 26

What have been the changes in competition patterns between the Old system and the New system? 28

What handicaps are winning competitions? 28

What handicaps are placing in competitions? 29

What scores are required to win or place in competitions? 30

Do these patterns vary depending on field size? 31

How has handicap distribution changed from the Old system to the New system? 33

Has there been any ‘settling down’ in the period since 9 April 2010? 33

Implications for an Equitable system 34

Conclusion 35

Appendix 1 – Comparison of Data sets 35

Appendix 2 – Distribution of Field and Winners 38

Appendix 3 – Distribution of Scores 39

Appendix 4 – An “Equitable” Performance System 47

Appendix 5 – One Golfer’s Tale 49

Appendix 6 – Three of the best 51

Appendix 7 – Competition Type by Handicap 53

Appendix 8 – About the Author 54

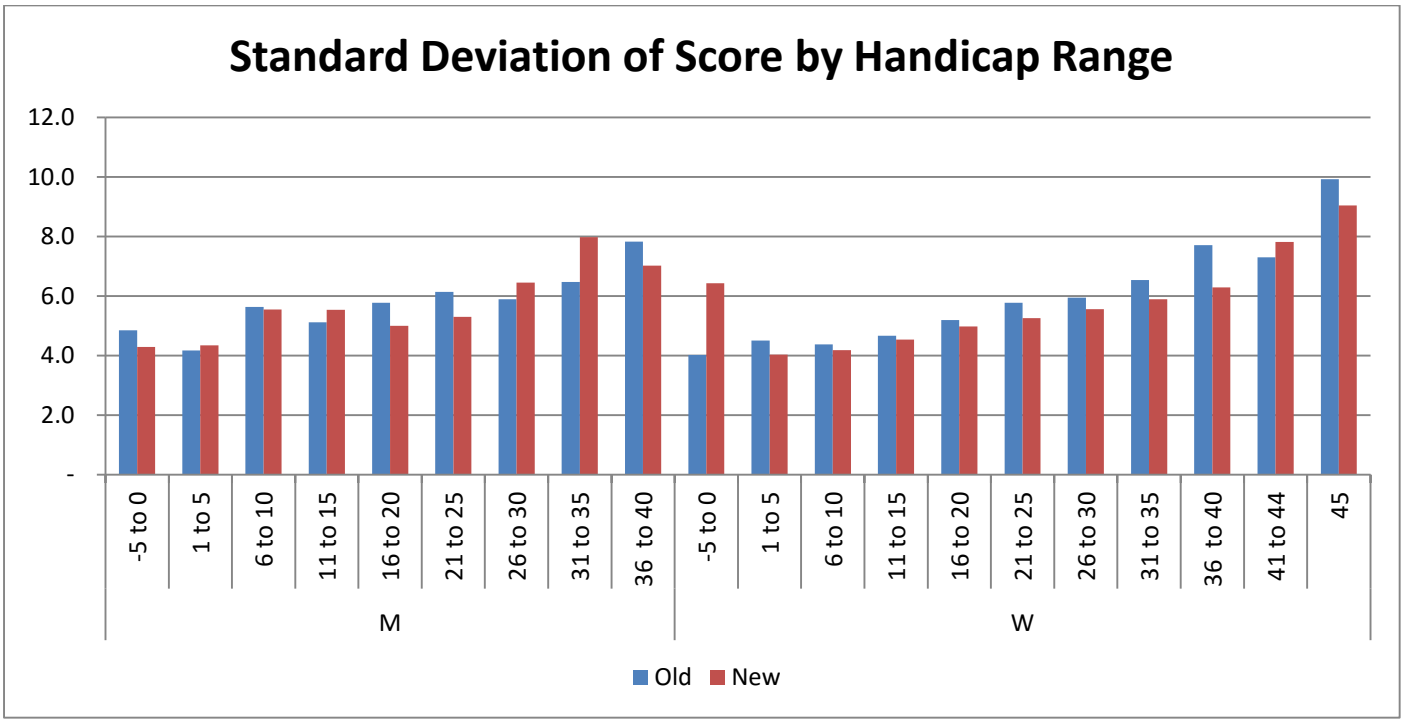
Introduction

A data subset was taken, comprising 500,000 rounds of golf under both the Old and the New systems, over the same time period of 9th April through 9th November, 2009 and 2010 respectively.

In the analysis following the "Score" is an Equivalent Stableford Score, thus enabling Stroke, Par and Stableford competitions to be on a comparable basis. Both men and women are included.

This data provided results for 27,600 competitions, and all players in a given competition were included. The total number of individual players was just over 400,000.

Higher markers exhibit a greater scoring standard deviation than do low markers: Confirmed.



The standard deviation clearly increases with the handicap range, under both the Old and New systems.

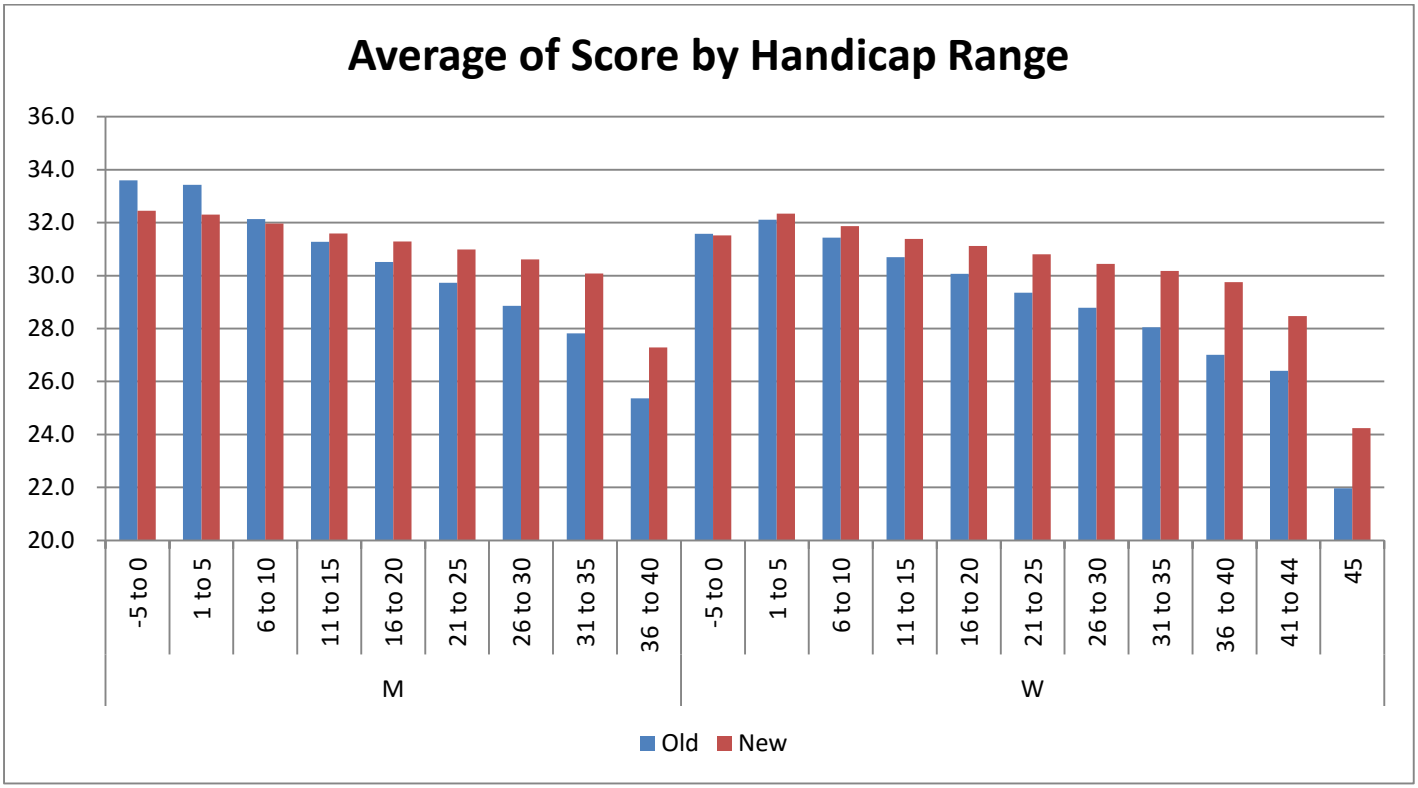
- Note that it is not uncommon for a score to be two standard deviations from the mean. So it should not be uncommon for a male scratch golfer to vary by 8-9 strokes from his mean, and for a high marker to vary by up to 16 strokes. For women the expectations would be 8-12 and 18 respectively. But the low marker is having this deviation added to a higher mean than the high marker: it turns out that the "Top scores" only vary by a couple of strokes across handicap groups. See Appendix 3 for more on Distribution of scores.

Under the Old system, a higher marker was less likely to play to their handicap on a given day than was a lower marker. This should hold true under the New system but to a lesser degree: Confirmed.

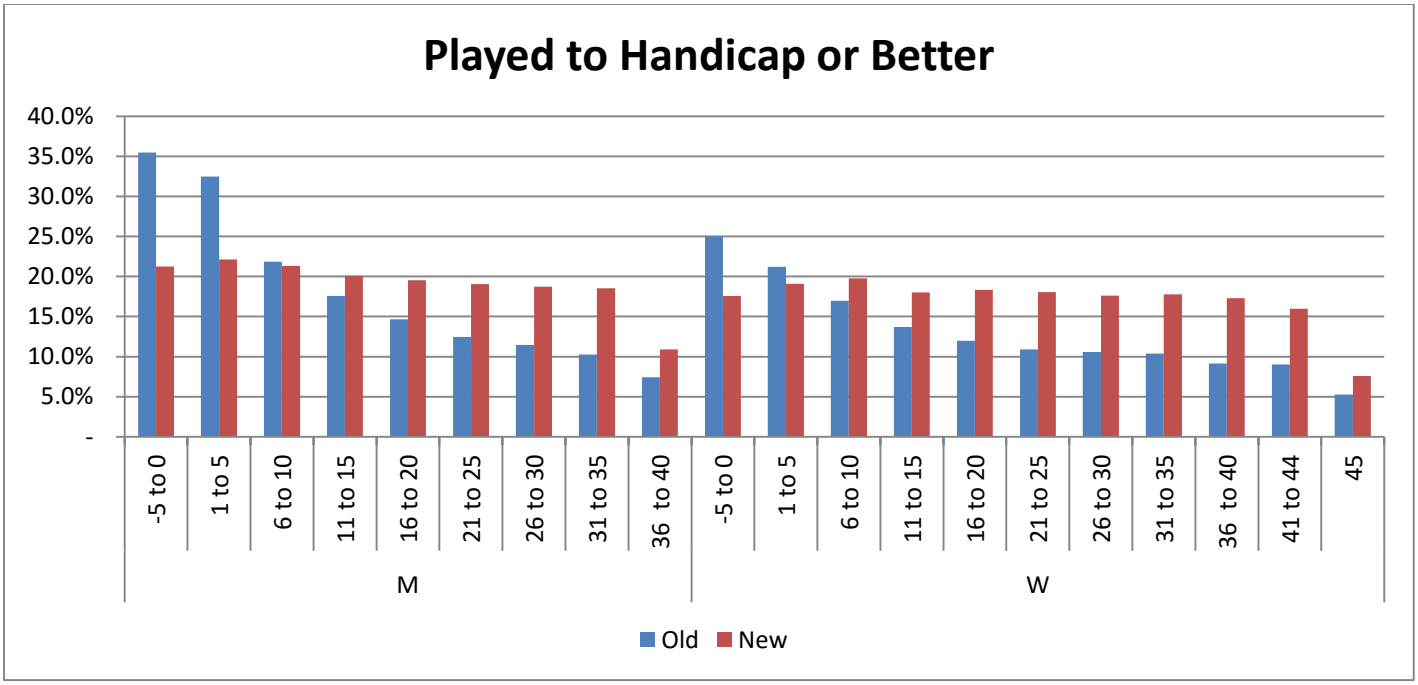
The first measure of performance is the average or mean score that players achieve at various handicap levels.

The following chart shows that under the Old system, the drop off of average performance, with increasing handicaps, from a Men's average of 33.6 to an average of 25.4 (or a drop of 8.2 points) was approximately "linear" or exhibited an even reduction.

Under the New system, the reduction was from 32.5 to 30.1 (only 2.4 points) at the 31 to 35 handicap level, with a further drop to 27.3 at a 36 handicap (a total of 5.2 points.). The women's competition is similar.



To specifically answer the question, one can calculate the percentage of rounds where a player played to their handicap or better.



On the Old system, the lowest marker men had a 35.5% chance of playing to their handicap or better, while this dropped away to 7.4% for the highest markers. Women also dropped off significantly from 25% to 9%. On the New, the inclination of drop off is much less, with men the low markers should only play to their handicap or better 21.3% of the time, the high markers (to 35), 18.5% of the time, and women low markers in the order of 20%, high markers 16%.

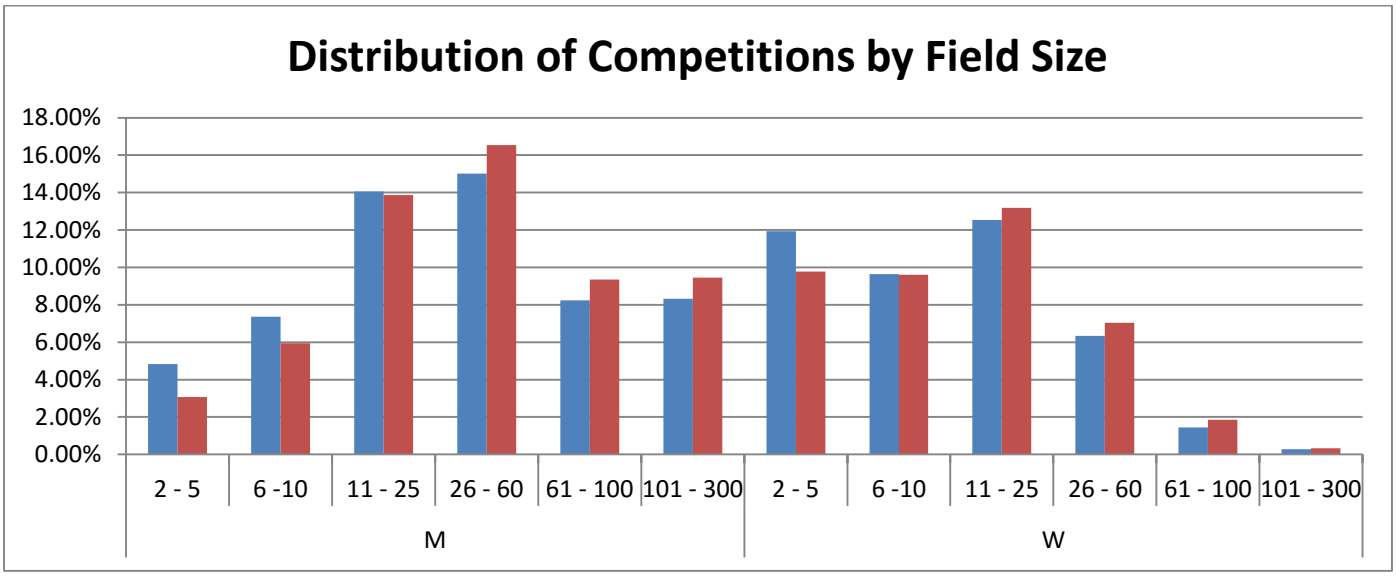
Detailed score distribution data are given in Appendix 3.

Under the Old system and the New system, the low marker will be advantaged in very small fields: Yes and No.

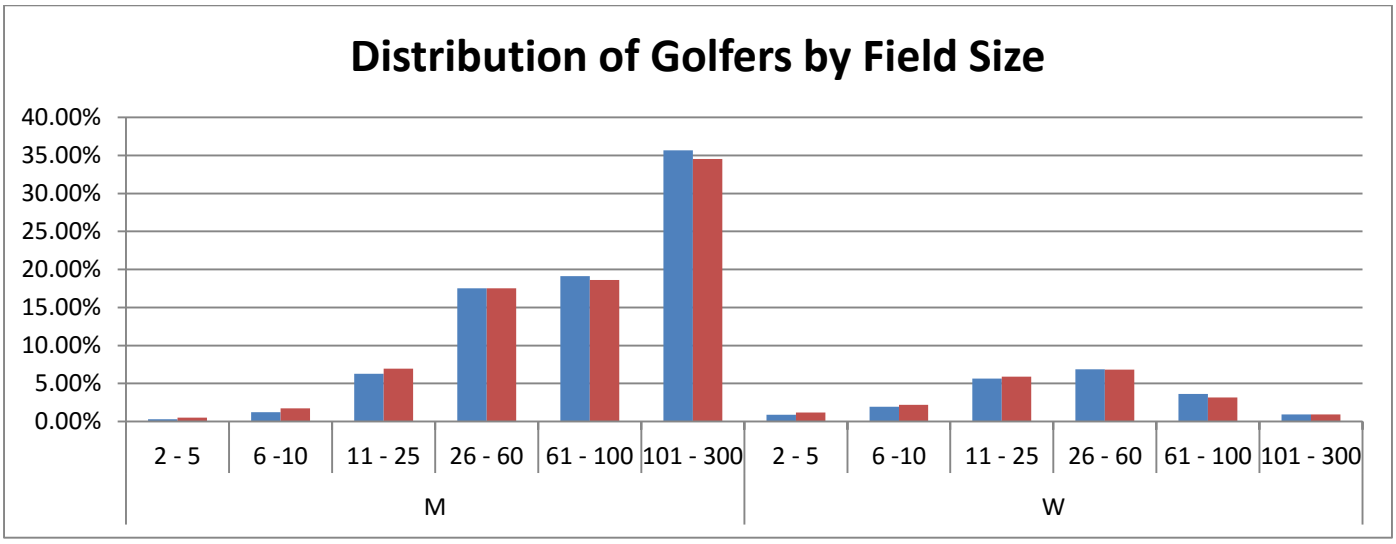
Let's take a detailed look at the pattern of men's competition under the Old and New systems.

In the data set there were over 27,000 competitions analysed.

Field size varied from 2 to 262.



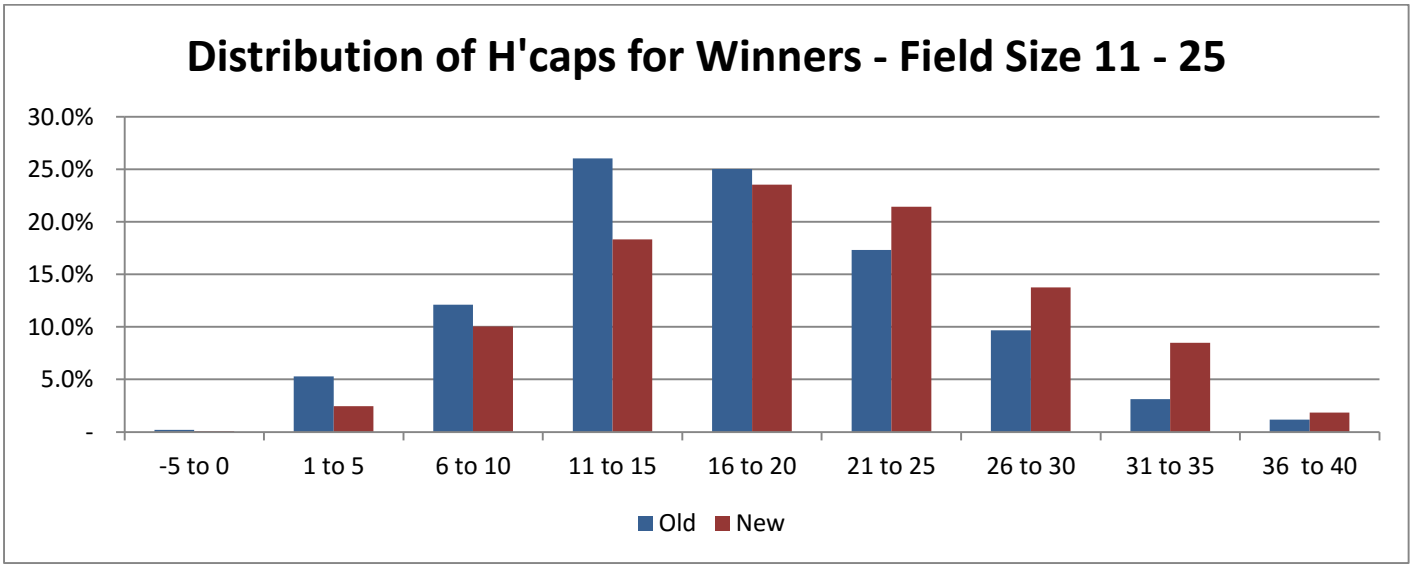
Most men golfers (nearly 70%) were taking part in the larger field size competitions, women's fields were smaller.



In order to ascertain how field size impacts on competition results, and on golfers of different handicaps, we need to look at the distribution of competition winning handicaps and compare them with the distribution of competitors, for varying field sizes.

For example, consider all the men's competitions in the sample with field size 11 to 25. This range is merely chosen as an example of the analysis. There were 3,777 competitions, with over 65,000 golf rounds. I have defined a "Prize-winner" as any golfer whose score is in the top 20% of the field in a competition. In the sample, there were just over 16,000 prize-winners. Finally the "Winners": there were 4,519 players who were ranked first or equal first in the competitions. They broke down to approximately half of each of these numbers in the Old system and New system.

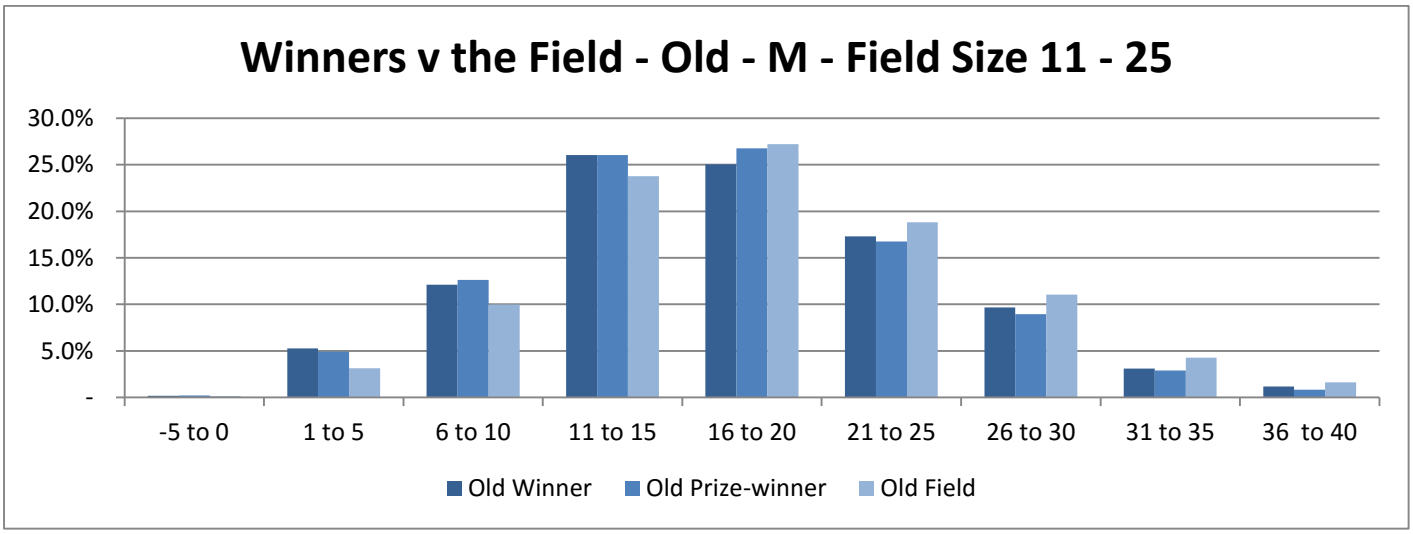
The distribution of handicaps of the Winners was:



Clearly the handicap distribution has "moved to the right", ie a greater number of competitions is being won by higher markers under the New system than under the Old. And the converse, fewer competitions are being won by low markers under the New system than under the Old.

But in isolation this does not give us the full story, as handicaps in general (it will be shown below) have also shifted to the right.

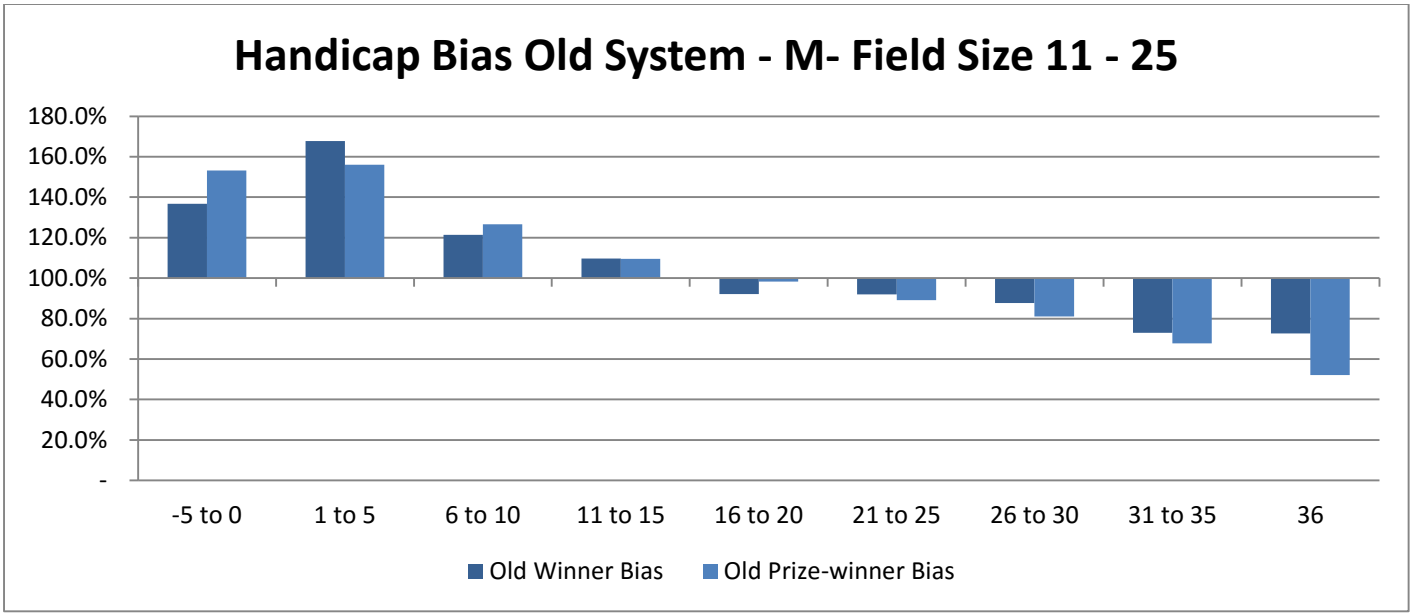
The starting point is to put the Winners, the Prize-winners, and the Field together and compare how each handicap group fared compared to the field. For the Old system, putting the three distributions together, one gets:



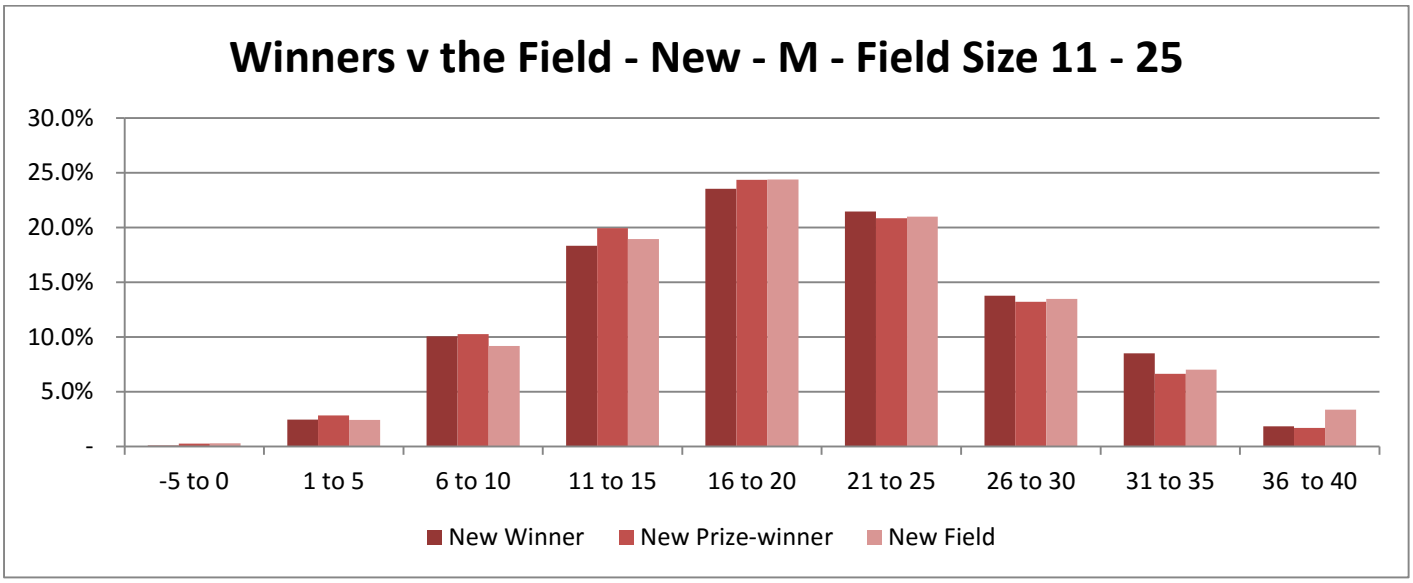
The interpretation here is that at handicaps under 16, the proportion of winners was greater than the representation of this group in the field. Not by a huge margin, the winners above in all handicap groups up to 15 represent 43.7%, compared to their representation in the Field of 37.0%. This can be represented as a ratio, 110.3%. Let's call it the "Handicap Bias". It is the over or under representation of a handicap group in the winner's circle compared to their representation in the field. 100% is neutral. Over 100% represents a positive bias, under a negative bias.

Clearly in the above chart the 1 to 5 handicap group the Handicap Bias is greater (it is a ratio of 167%), and at the 36 to 40 end it is less (it is a ratio of 72.6%.)

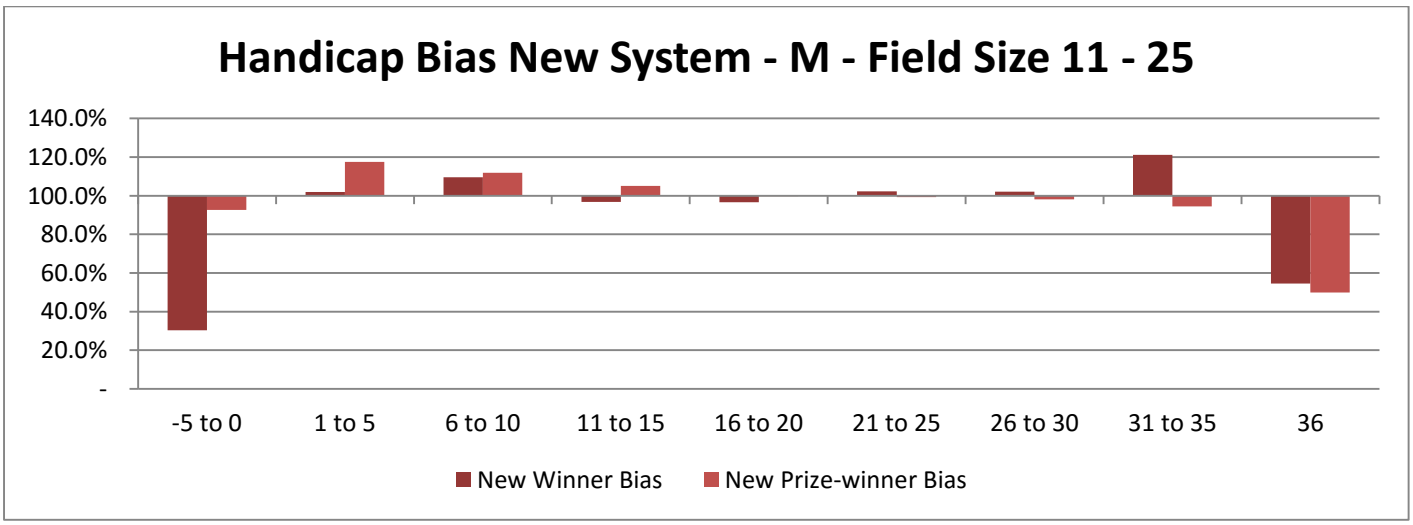
This concept of Handicap Bias gives us a simple way to compare the Old and New systems, and to compare the impact of handicaps at varying field sizes.



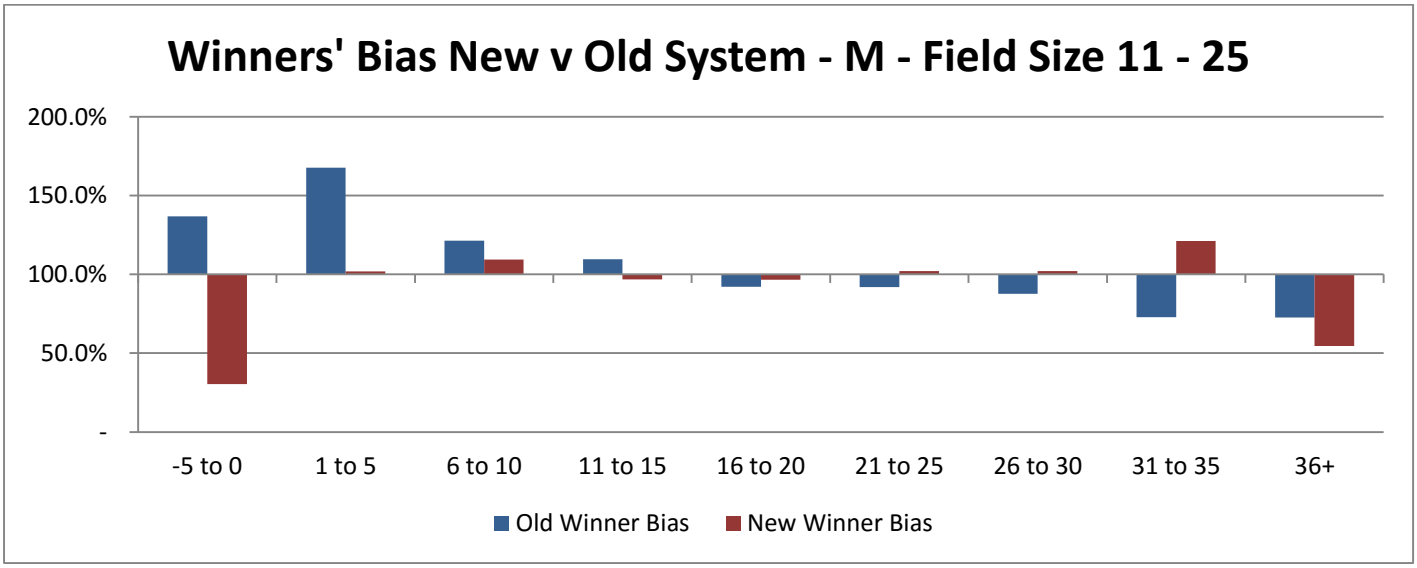
The Winners v the Field for the Old system, Field Size 11 to 25, is then calculated and shown in the next chart.



And the Handicap Bias is as below. Note that it is not unexpected that the 36 category would be negative under both systems, as the players in this group are limited to 36, and are often playing to much more.

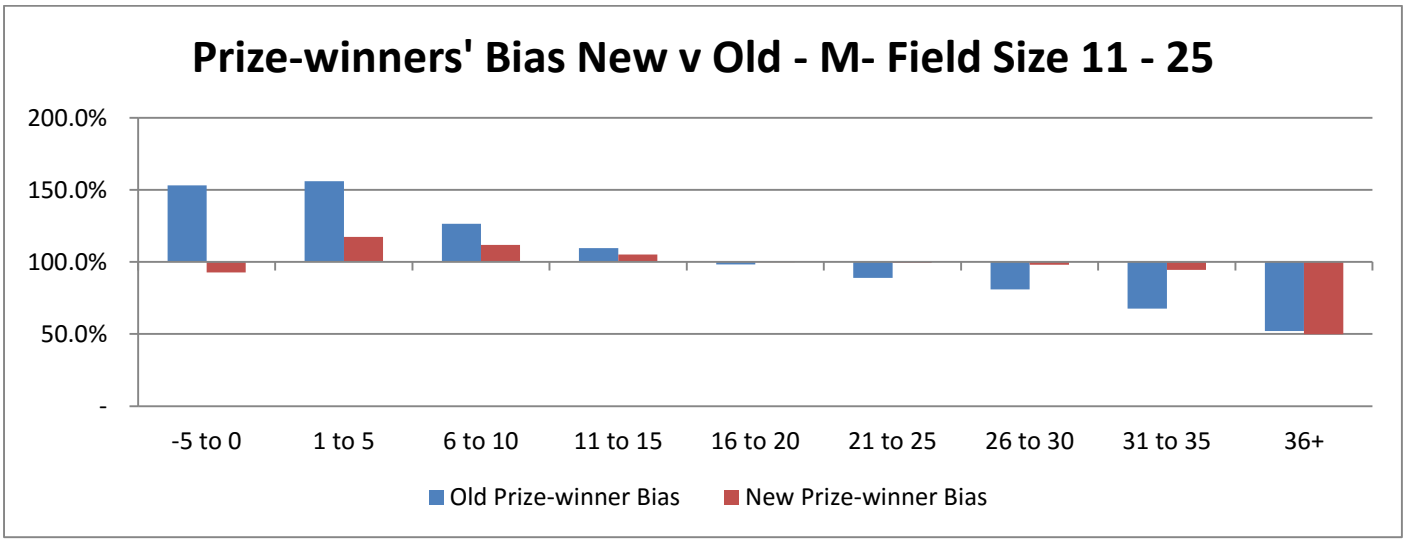


Comparing the Old and New Systems for Winners' Bias:



Clearly at this field size there has been a marked shift in the Winners Bias, especially at the very low end of the handicap range, from favourable (136%) to unfavourable (30%). For handicaps at the other end of the range, competitiveness has increased, eg at 31 to 35 the Bias has increased from 73% to 121%.

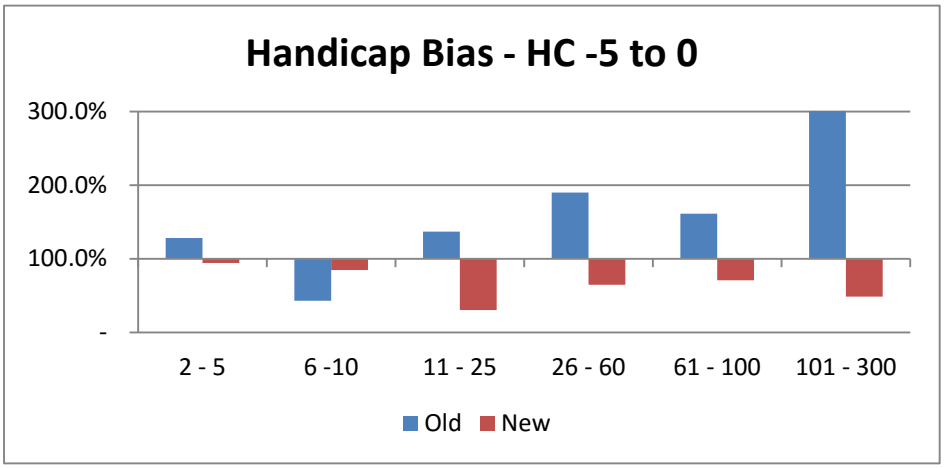
And for Prize-winners:



The Prize-winners' Bias has moved to a lesser extent. In the case of the low markers, from favourable to just slightly unfavourable, while the high markers, at least up to 35, have risen from unfavourable to neutral.

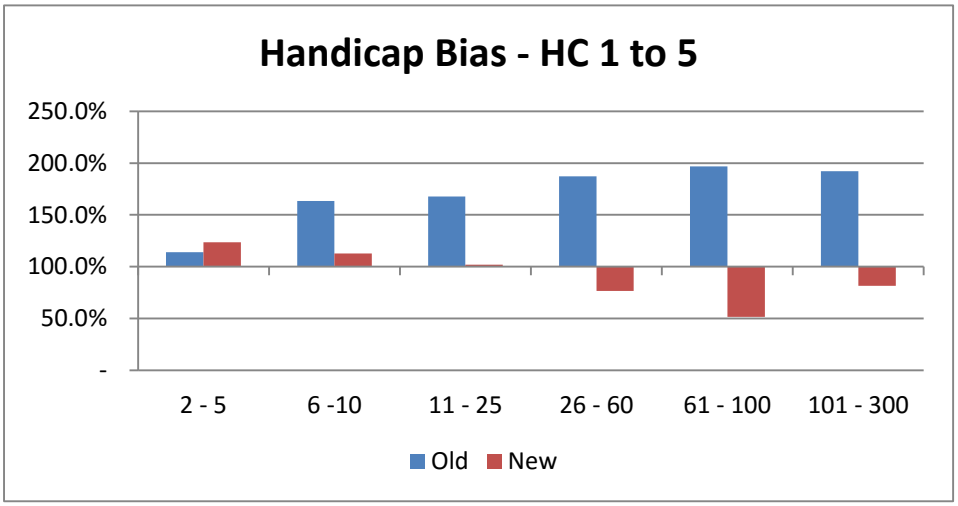
It is now possible to repeat the above analysis for all the varying field size groups.

Handicaps Men -5 to 0.



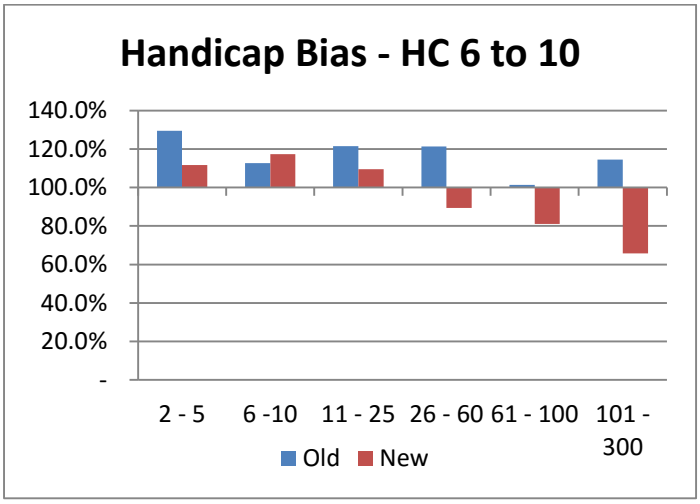
Due to the relatively small number of rounds in this group (2,811 compared to the full sample of one million), these results do not exhibit a smooth pattern. Whereas on the Old system there was an advantage with larger field sizes, the New system has removed this and in some ranges turned advantage to disadvantage.

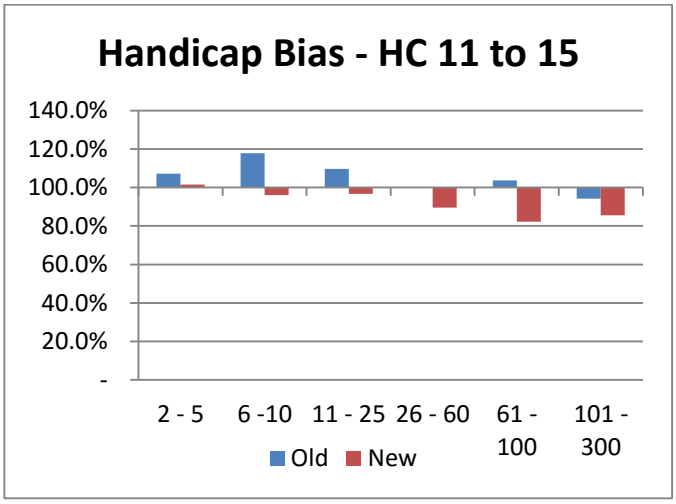
Handicaps Men 1 to 5



Here the impact of field size is very apparent. Once the field is over 50, say, the bias is in the 50% to 80% range. On the Old system, however, there was no disadvantage in large field sizes.

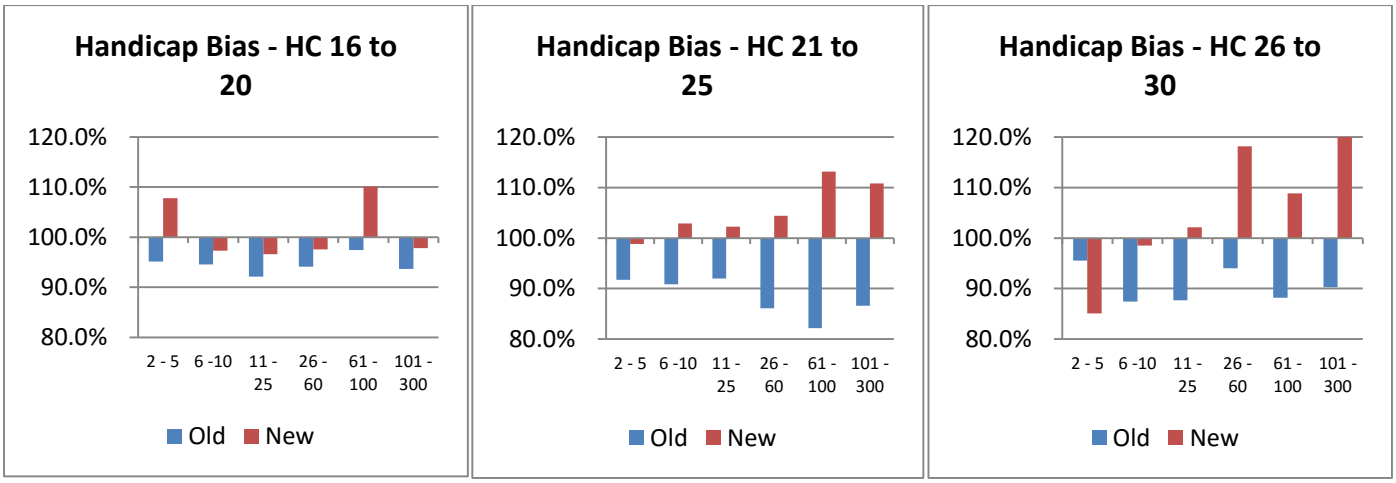
Handicaps Men 6 to 10 & 11 to 15





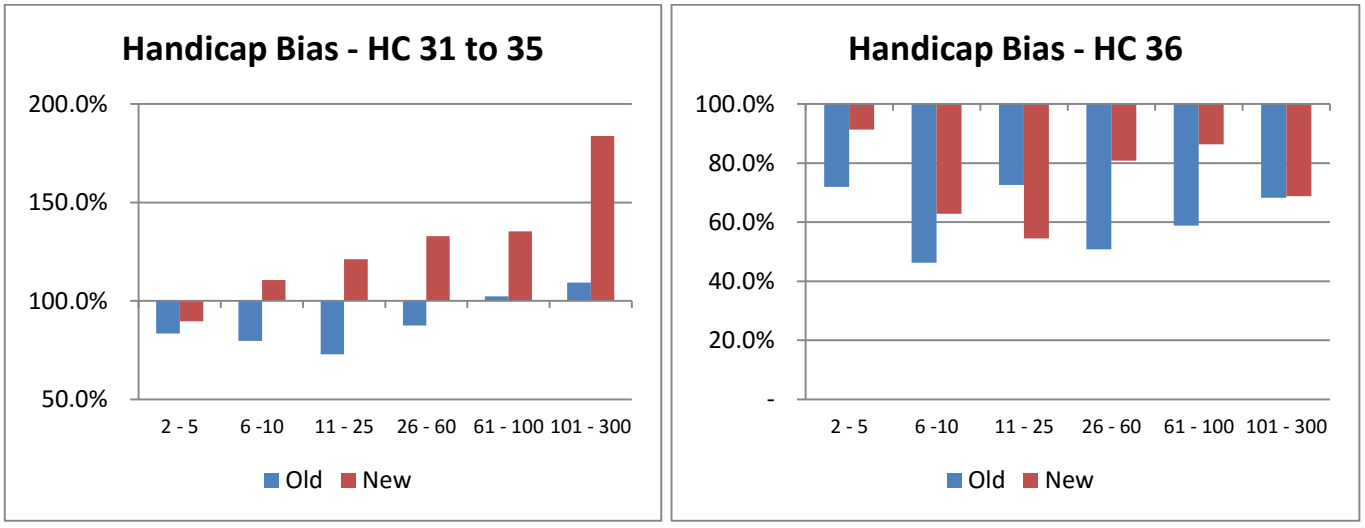
As handicaps rise, the pattern of change becomes less extreme. But there is clearly more disadvantage as field sizes rise.

Handicaps Men 16 thru 30



We are starting to see the advantage that is entering the equation for larger field sizes as handicaps increase. The Bias for players with handicaps 26 to 30 for the 100+ field size has risen from 90% to 130%.

Handicaps Men 31 and over



The Bias for players with handicaps 31 to 35 for the 100+ field size has risen from 109% to 184%.

It is now possible to address the specific propositions.

Under the Old system and the New system, the low marker will be advantaged in very small fields: Yes and No.

For handicaps up to 5, in the Old system these players were in fact advantaged more as field size rose. For handicaps beyond this up to 10, there was a positive bias (ie 100% or more) at all field sizes.

It is under the New system that the impact of field size is really marked.

The low marker will find it harder to compete as the field size increases. Yes – but only in the New System.

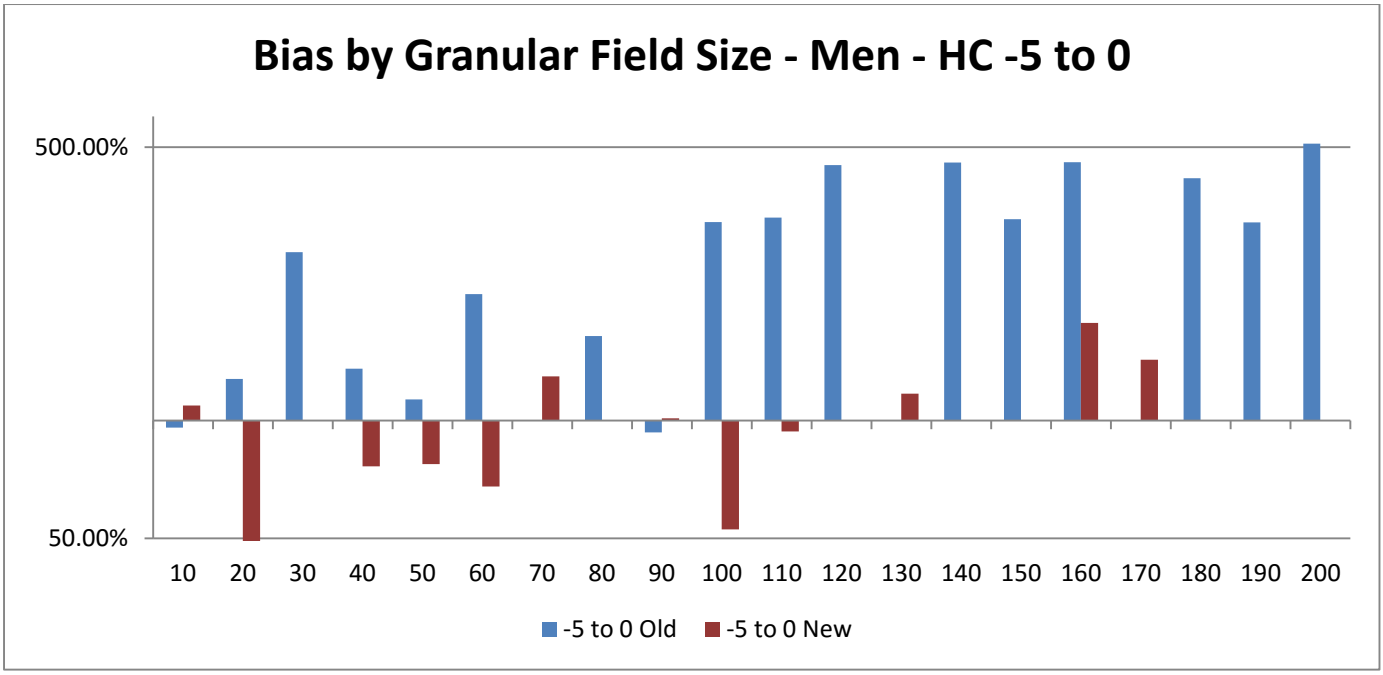
The above charts demonstrate this for the New system, with what used to be a positive or neutral bias for all handicaps up to 15 becoming neutral or negative in most handicap ranges.

Under the New system, the field size value at which a low marker cannot realistically compete (or win any sort of prize) will be lower than occurred with the Old system. Yes - But.

Given that it turns out that there was no deterioration in performance for larger fields under the Old system, but there is for the New, then this must be true.

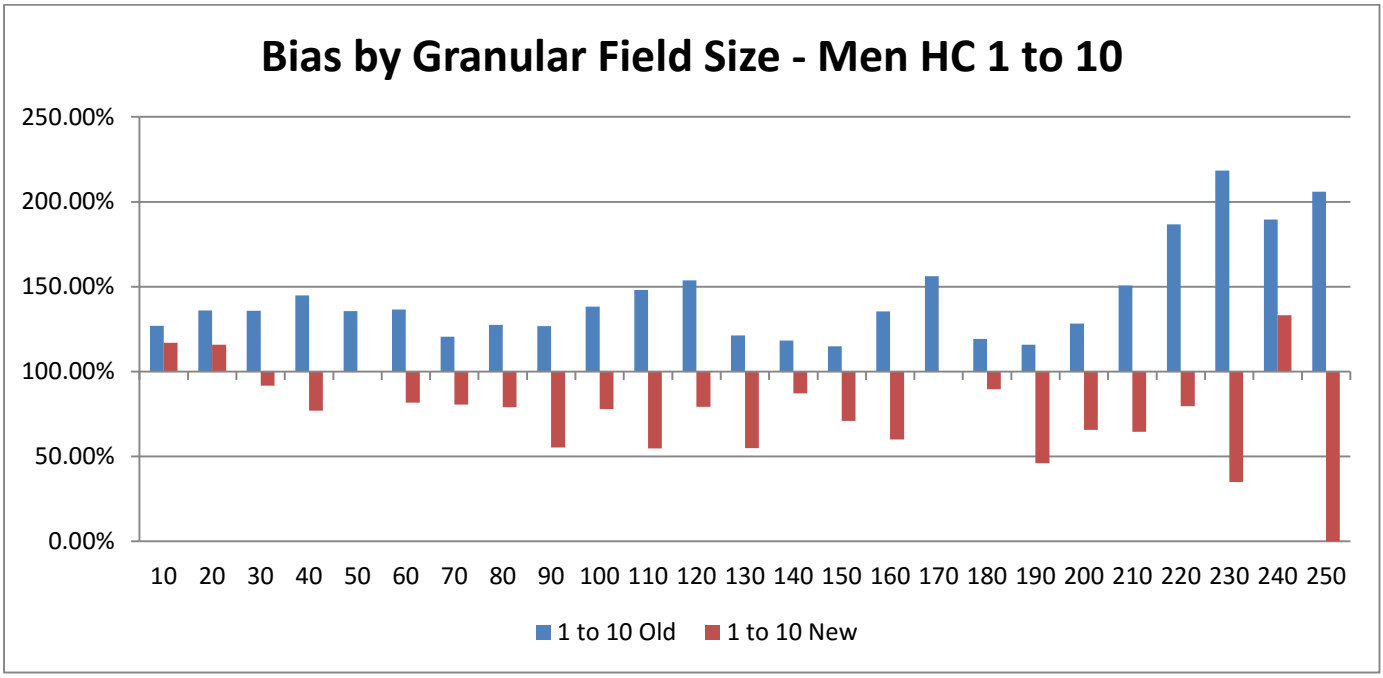
For the -5 to 0 markers, under the New system they are winning some prizes at high field sizes, but not as many as under the Old system.

Here is a chart with the field size grouping reduced to intervals of ten. When the data is reduced to this level of "granularity", one has to expect a more lumpy result, and accurate assessment needs to be modulated by allowance for statistical "noise".



Everywhere that there is a red bar, these lowest markers have won a competition under the New system. So both a field range of 151 – 160 and of 161 - 170 have seen lowest markers win on the New system. But the problem with this "small" grouping is that there were no lowest markers winning in the 180 group, or the 190, 200, etc. Thus the Handicap Bias in the 170 size group is overstated when one considers what it would be if the 180, 190 and 200 result was aggregated. So in this lumpy result it is hard to conclude that field size is relevant to the lowest markers. It is also likely that they are playing in competitions with their peers, eg a competition with a cut, and thus all the competitors may well be low handicappers.

Consider a grouping of all the men's handicaps from 1 to 10.



There is a strong positive bias at field sizes up to 20. Even though the bias is positive at 50 and 170, it is negative elsewhere over the 20 group. The next group is 21 to 30. But we know from an earlier chart (on page 14) that the bias was positive up to 25. So one could conclude that the bias starts to become negative shortly after 25.

This does not imply that competitions are unwinnable over 25. Eg the whole 41 to 50 field size group is neutral. But once over this the bias is in the order of 80% up through fields of 100. To recap, a bias of 80% doesn't mean a competition is unwinnable, only that a player will win only 80% as often as his representation in the field might suggest.

To recap, at every red bar above the low markers (1 to 10) are winning competitions. But less frequently than before.

Hence, do prize patterns vary depending on the field size?

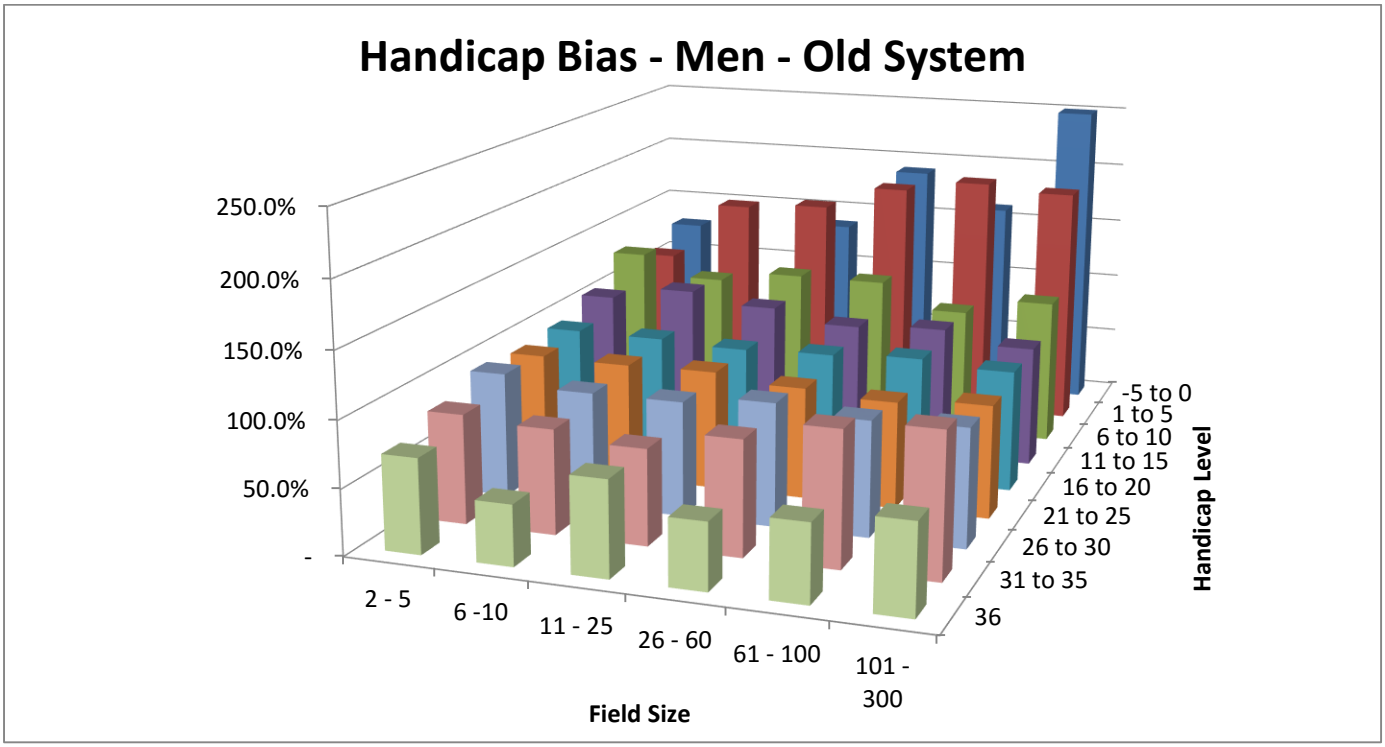
Yes but differently for different handicap levels.

Is the New system equitable for small fields but not for large fields?

It is possible to try to view all the field sizes and all the handicap levels in one 3-D chart. This gives a picture of the playing field.

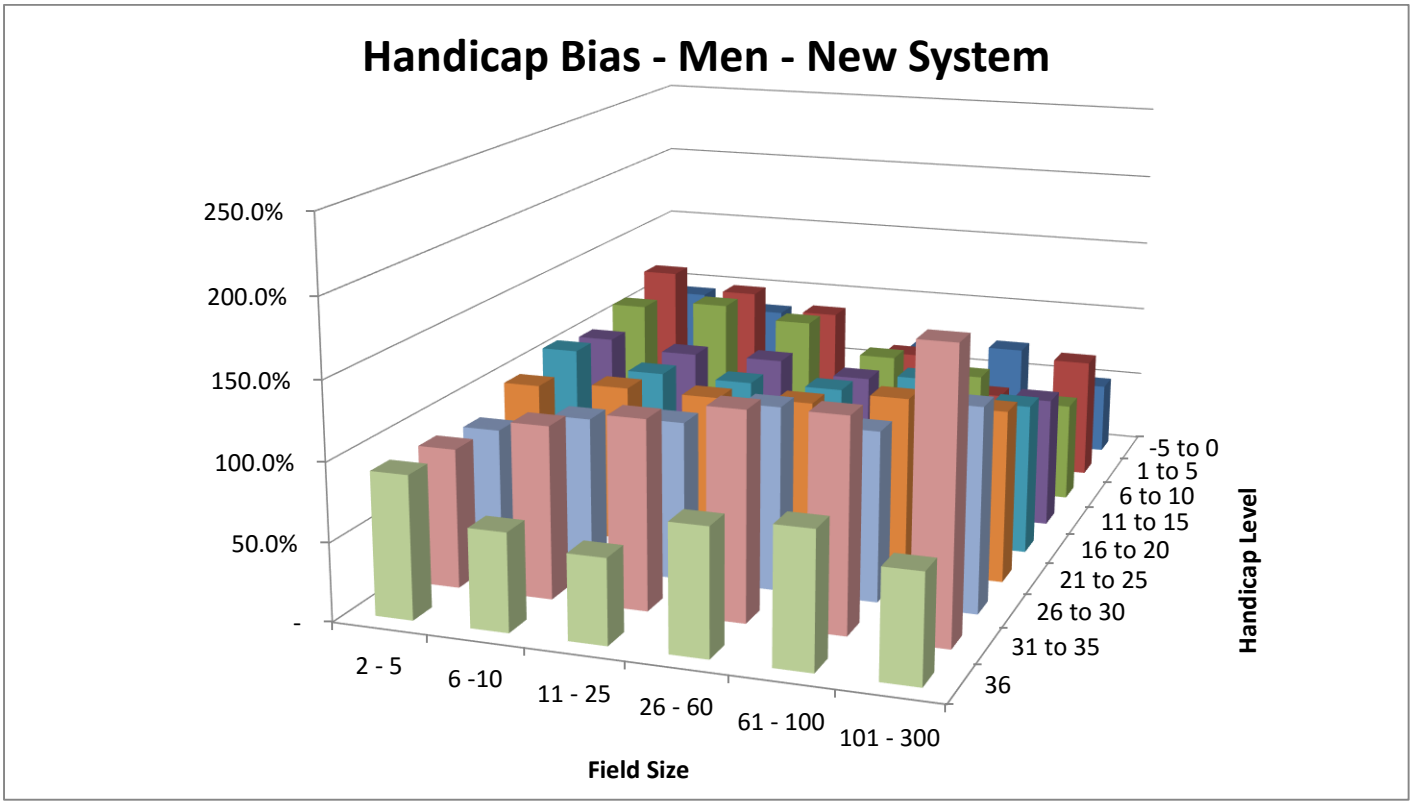
The women's competition will also be addressed when we look at this 3-D summary.

Under the Old system, there was a significant positive bias for the low markers.

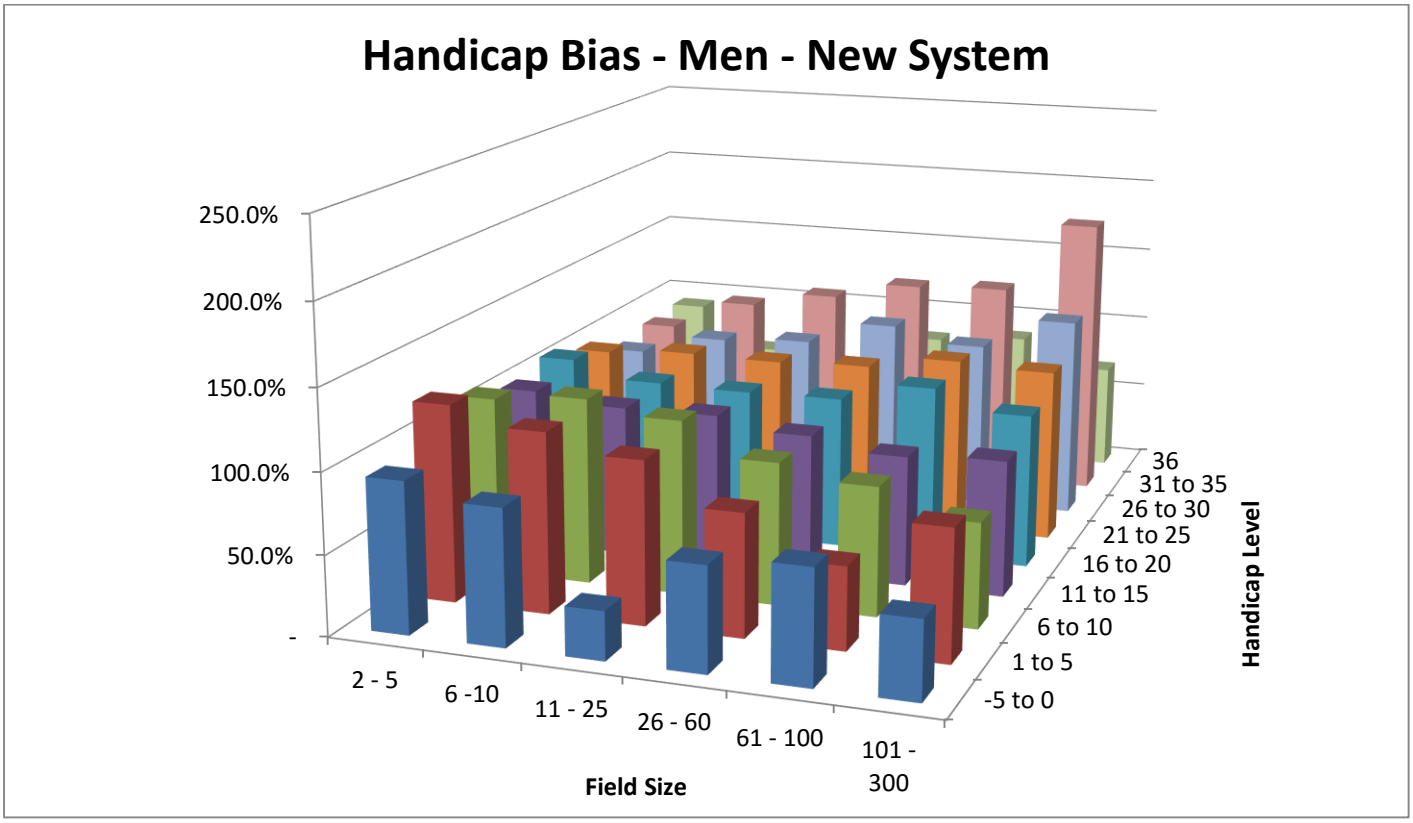


Low markers were favoured by larger field sizes; high markers were more neutral to size.

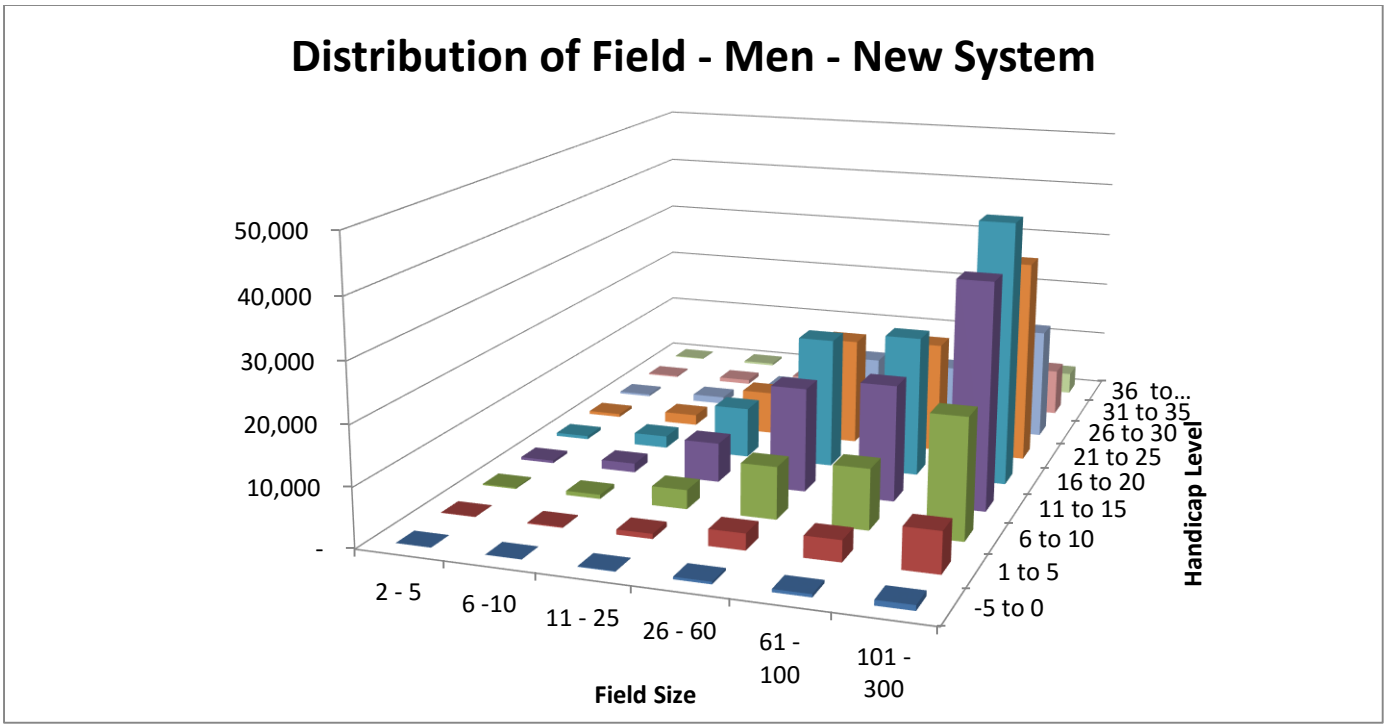
Under the New system, the low markers have had their advantage turned to disadvantage at higher field sizes, and higher markers have had their disadvantage turned to advantage.



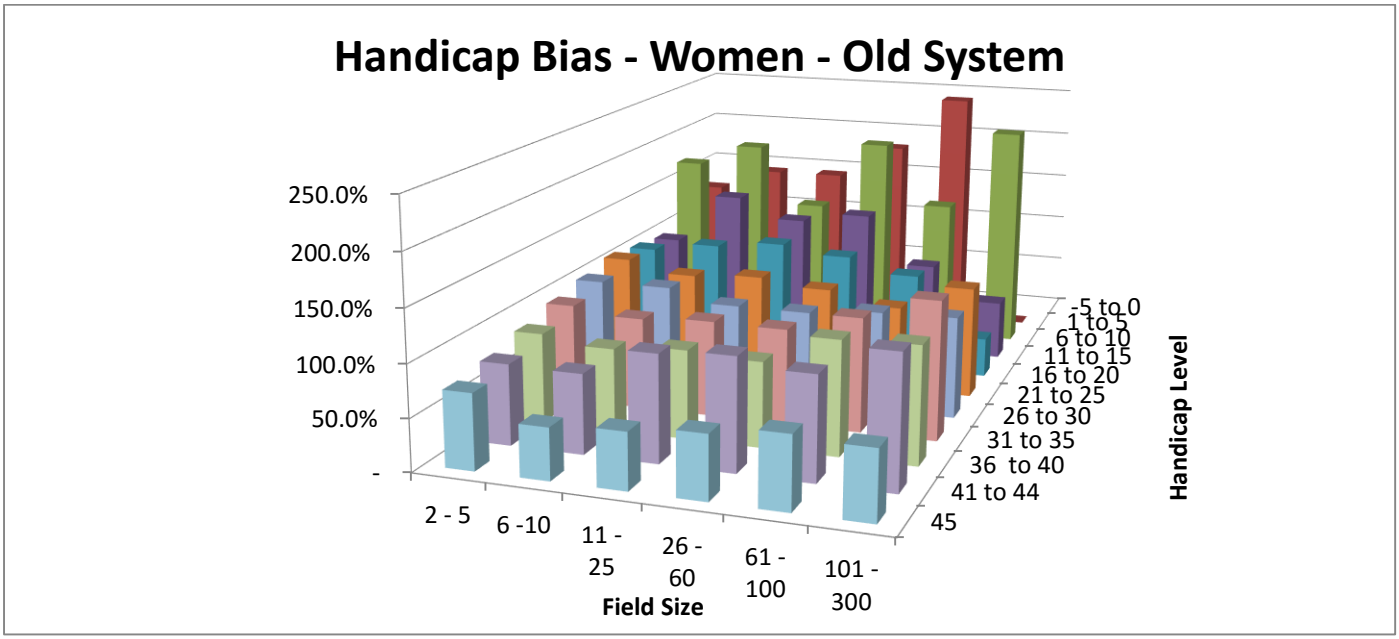
It is easier to see the impact on the low markers if we reverse the Handicap axis. Concentrate on the red bars, as the dark blue bars are a bit lumpy due to the small sample size.



It may be worth comparing the above playing field to the number of players on it. There is no significant difference in field distribution between New and Old. Here is the chart on the New:

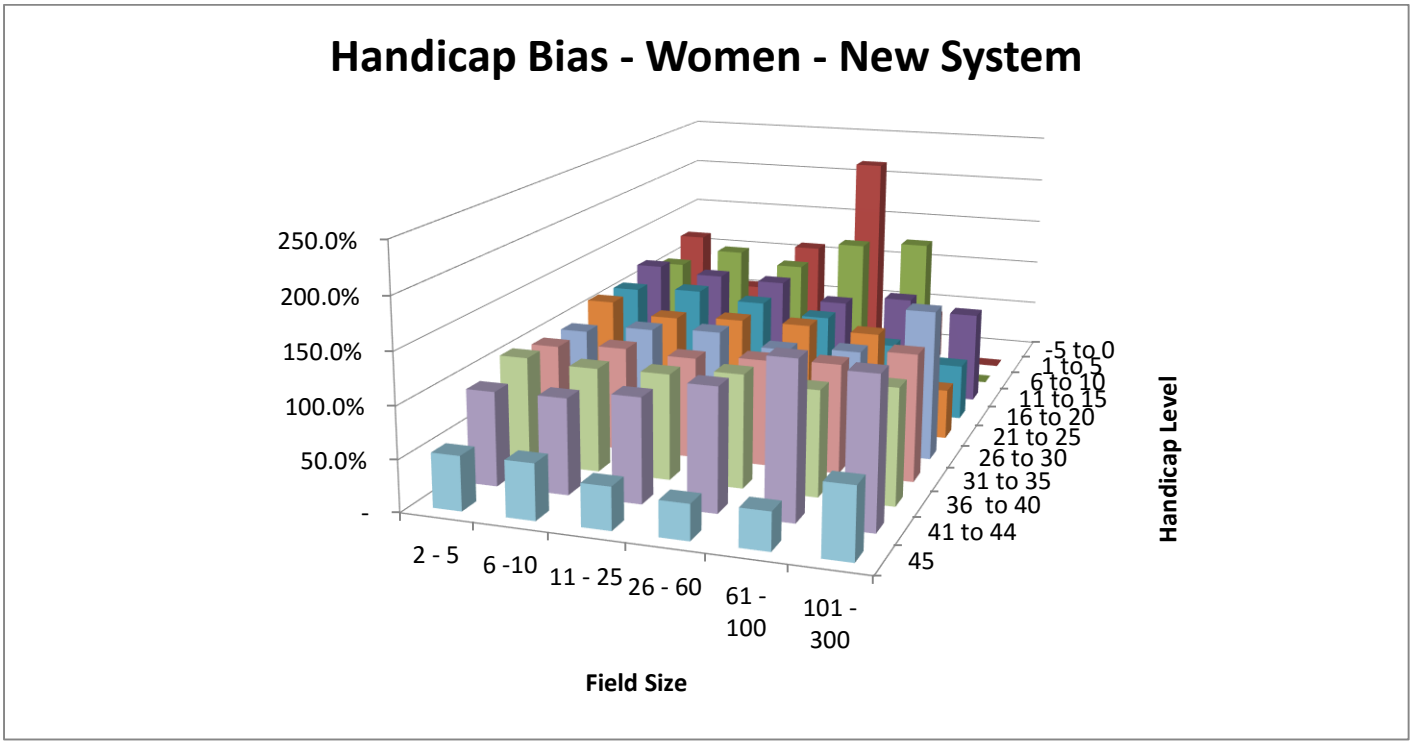


The women's competitions under the Old system had the following bias:

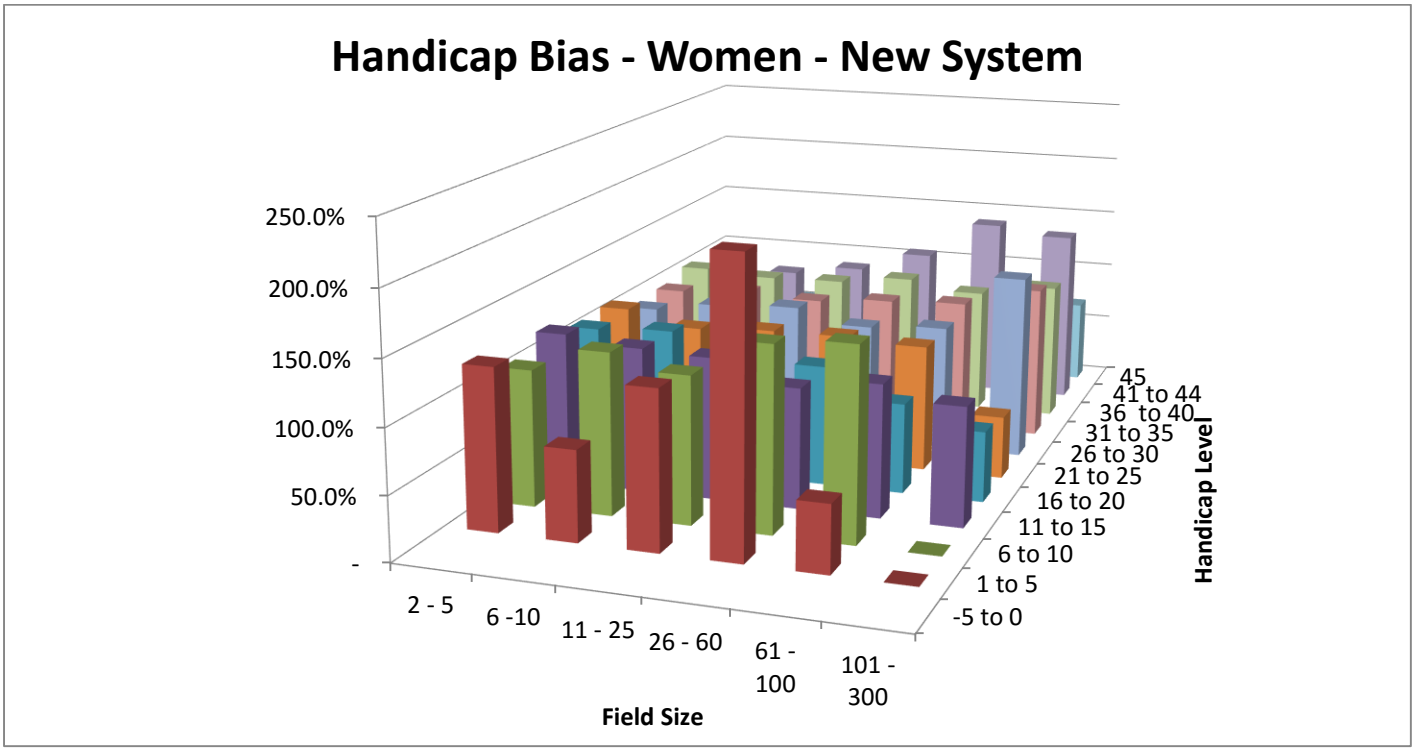


The very low handicaps evidence positive bias as field sizes increased, the players off 11 to 20 had a negative bias, and the others were more or less neutral.

Under the New system the field has changed to:

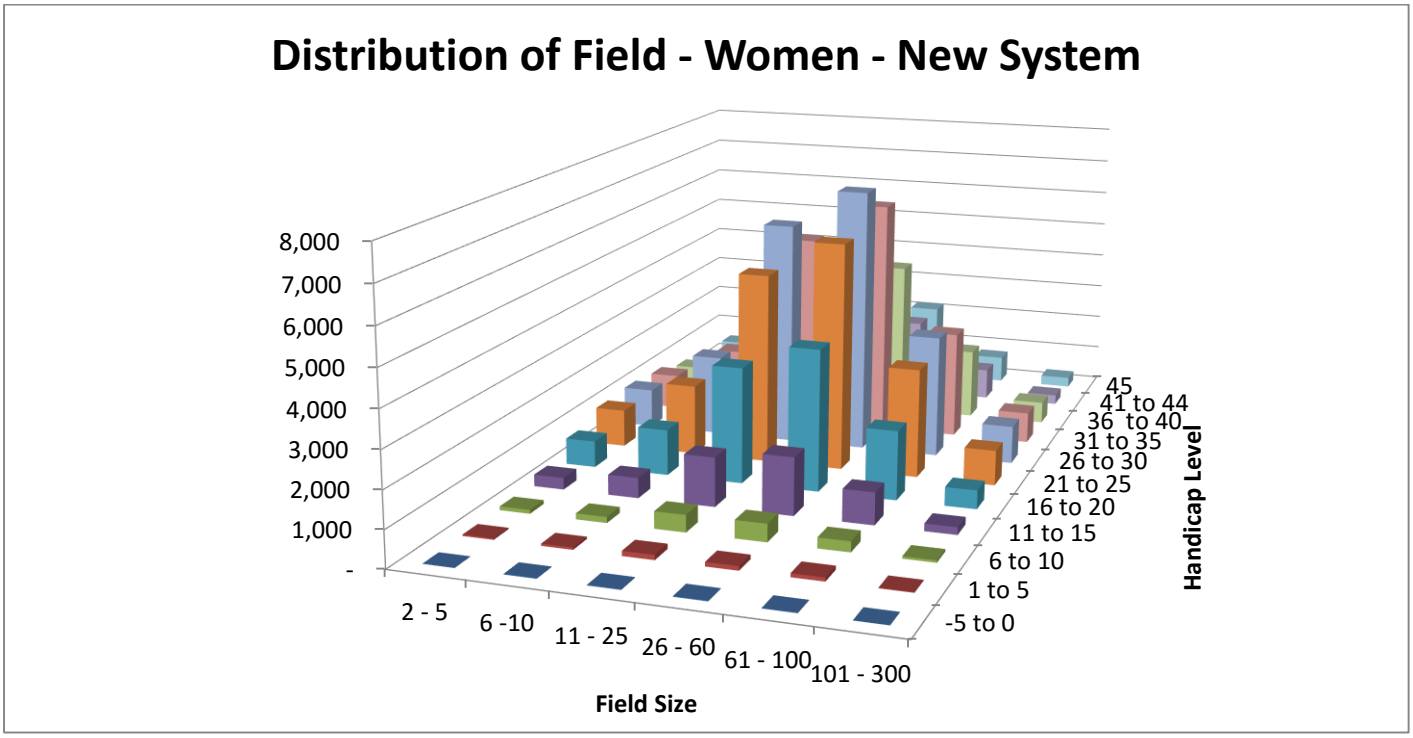


And turning the field around:



Once the field size approached 100, there is evidence of the highest handicappers having a positive bias. But for lower fields the field is remarkably flat, with even the 6 to 10 handicappers not evidencing the deterioration in competitiveness that was the case for the men. There is a drop off in bias for the 16 to 25 groups over 100.

What is relevant to the above is the distribution of field sizes in women's competitions. Essentially, there are not many women playing on that part of the field where it tips.



If so, at what field size does the New system become inequitable for the low marker?

Taking all of the above into consideration, a field size of about 50 would appear to be the crossover point for men, 100 for women. Inequitable seems a bit strong – what was previously a strong positive bias under the Old system for low markers is still positive to neutral for field sizes up through 50 (men and 100 (women), and there is a negative bias under the New system for fields over this.

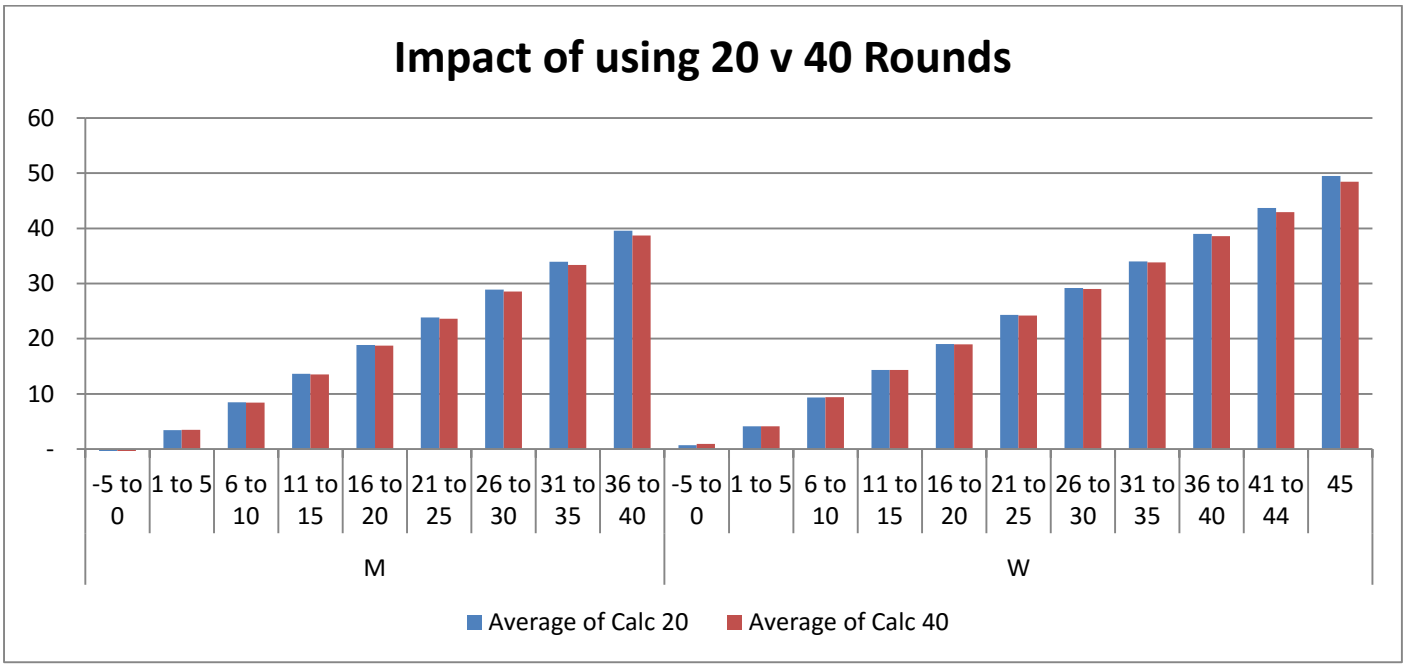
How does this equate with the Old system?

The Old system had no field size advantage except for the low markers who were advantaged with higher field sizes.

Is a 20-score sample large enough to accurately reflect the scoring potential of an inconsistent player?

To address the issues raised in this section, a second sample of one million rounds of golf was extracted. For approximately every 40th golfer in the GolfLink database, every round played over the last three years was taken. This represented nearly 15,000 golfers. Of these 6,400 had played more than 60 rounds in the time frame.

If we were to use a 40 sample (taking the 20 best scores) vs the current 20 sample (taking the ten best scores), the average variation at low handicaps is negligible, rising to a one stroke reduction in handicap at the highest handicap levels:



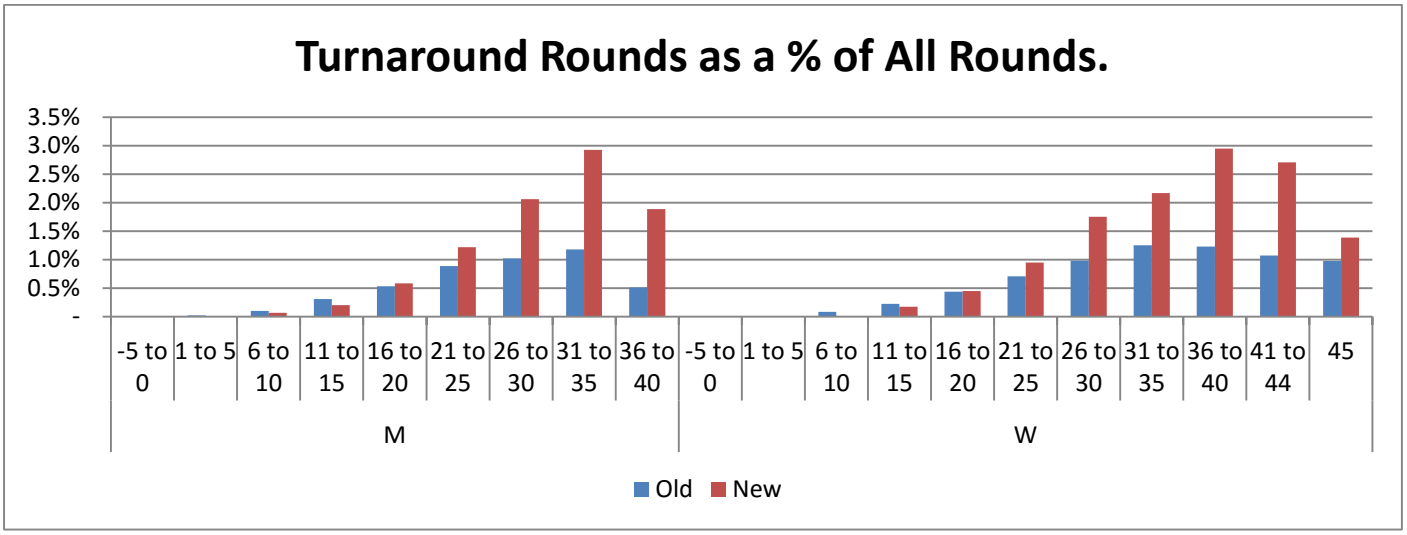
However what is essentially a 40 period moving average responds more slowly to changes in performance than does a 20 period moving average. This means that on a 40 period sample, handicap relief when form deteriorates will be less than for a 20. However the flip side of this coin is that after improvement, the 40 period sample will also respond more slowly to improvement in form. So if there has been a slump from which there has been a recovery, the improved form will benefit from the higher handicap longer than it will under a 20 sample.

A number of specific examples of this are given in Appendices 5 and 6.

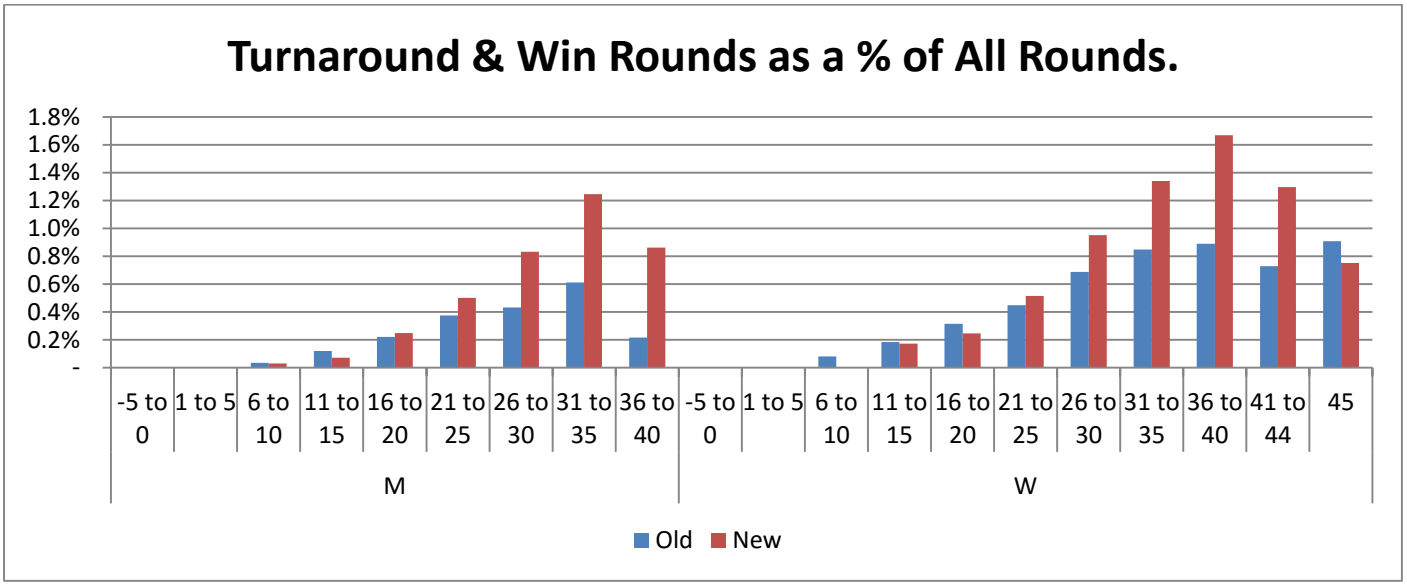
Is it feasible that a high-marker will endure an extended run of poor scores resulting in a handicap which is not relevant to their scoring potential?

To quantify this, it is necessary to define a "Turnaround". Let's say that a player has had a Turnaround if their score is 40 or more, where the average of their last 20 rounds is 30 or less. A "Turnaround and Win" is defined as winning the competition with one of these Turnaround scores.

The proportion of Turnaround rounds has increased under the New system. Over all they represent 1% of all rounds, compared to 0.5% under the Old system.



The occasions when this Turnaround leads to a Win has also increased, overall from 0.25% to 0.46%.

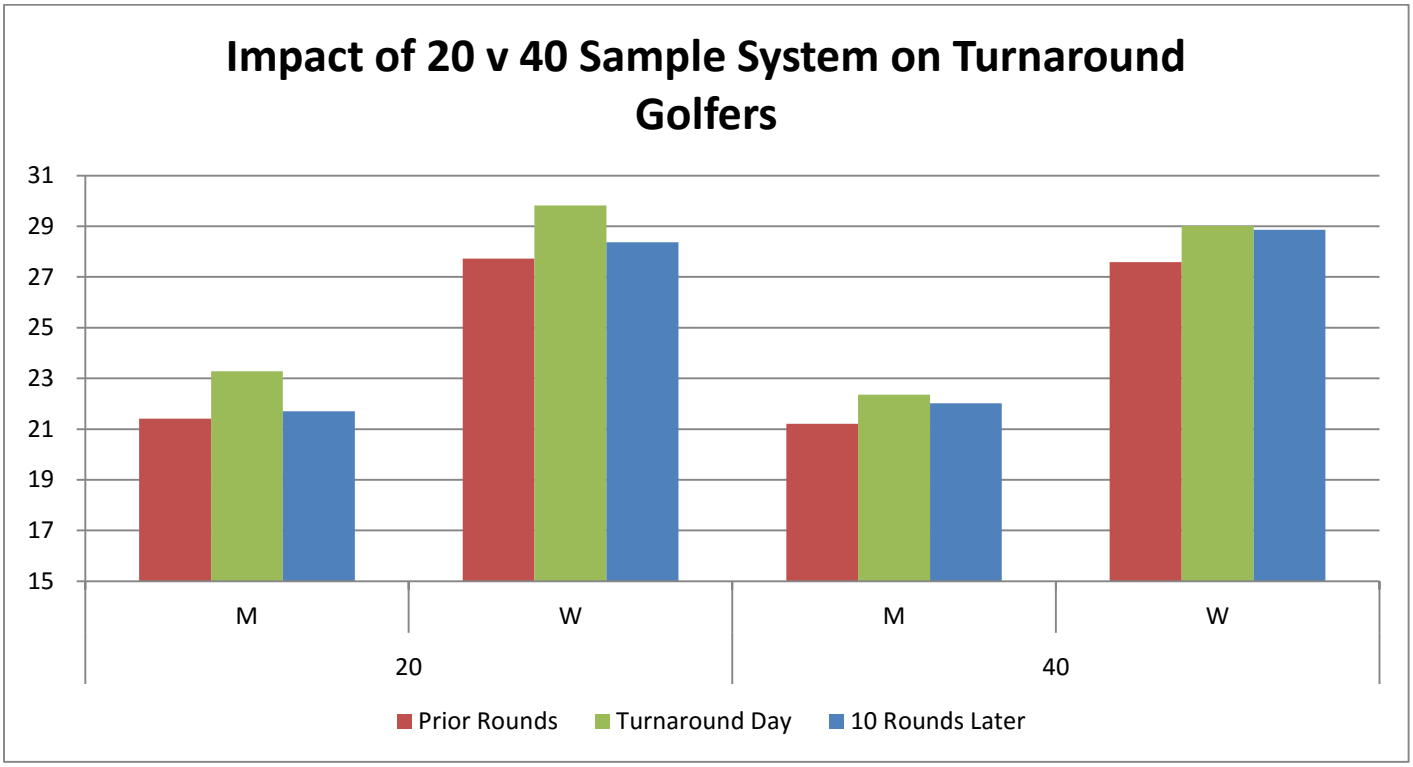


Thus, yes, markers over 26 can have a run of poor rounds and then a very good round. But the Turnaround and Win scenario only accounts for less than one half of one percent of the rounds.

Is it likely that large fields will exhibit enough players in such a situation that the challenge of the low marker becomes further extended to a material extent?

The challenge to low markers, as discussed above, already includes any impact of Turnarounds. In other words, the challenge is as already measured, and includes any impact of Turnarounds.

To what extent would using a forty rather than twenty sample impact Turnarounds?

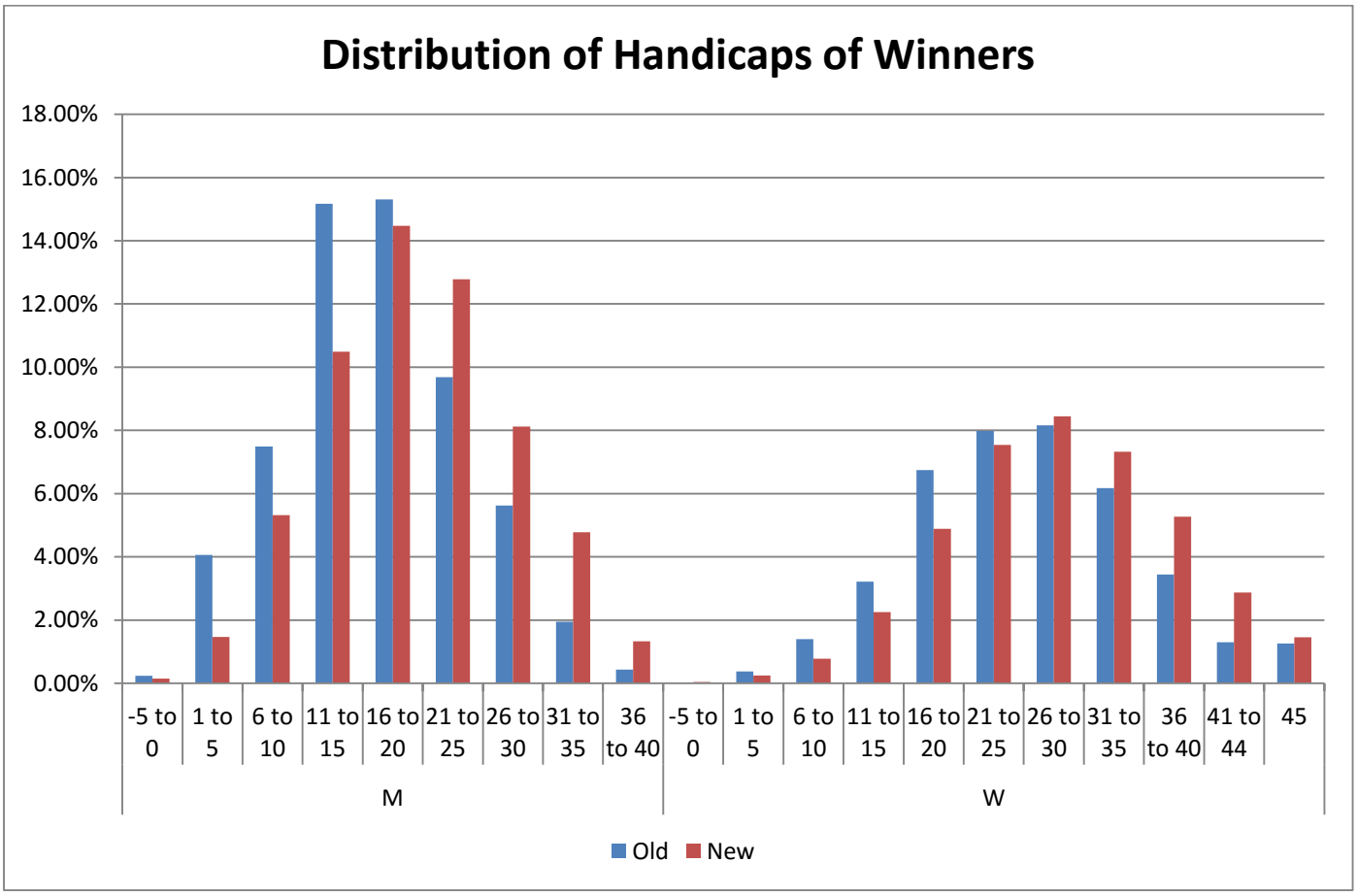


The golfers who have experienced this Turnaround were compared and their handicaps averaged. One can see how the handicap rose by about two strokes under the 20 system, and a little less, one to 1½ strokes under the 40 systems, on the day they had the Turnaround . In the case of the 20 system, the handicap has fallen, almost back to what it was on average, ten rounds later. But in the case of the 40 system, it is still one to two strokes higher than it was.

What have been the changes in competition patterns between the Old system and the New system?

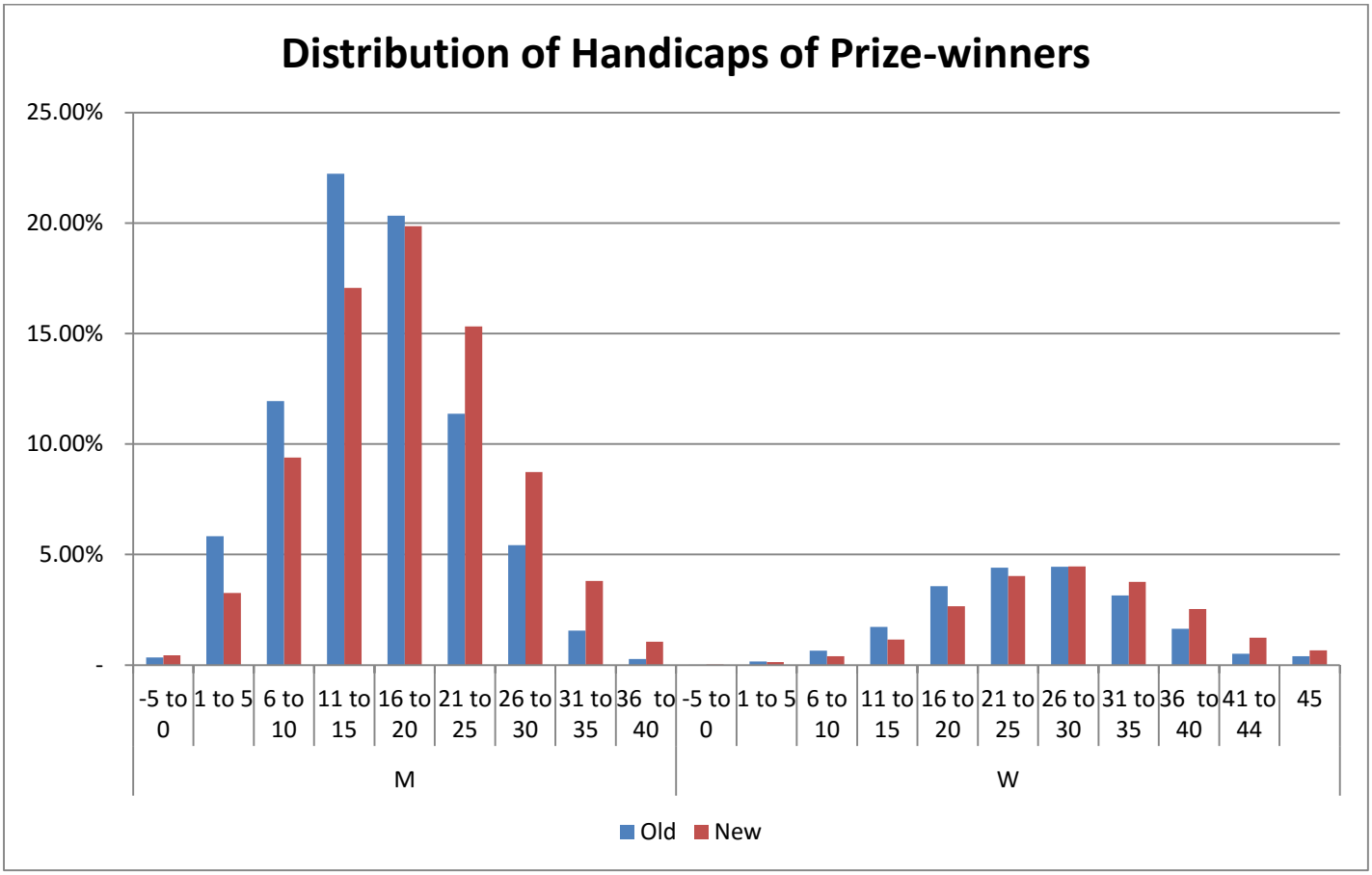
Much of this has been covered in the analysis above. But to respond to specific points:

What handicaps are winning competitions?



The lowest markers (-5 to 0) have had a small reduction in their wins. But the 1 to 5 men's group has had their wins more than halved. The men up through 15 have had a 25% reduction, and the high markers have increased their wins. The women have had more moderate changes, with those up through 25 experiencing a reduction in wins, the others an increase.

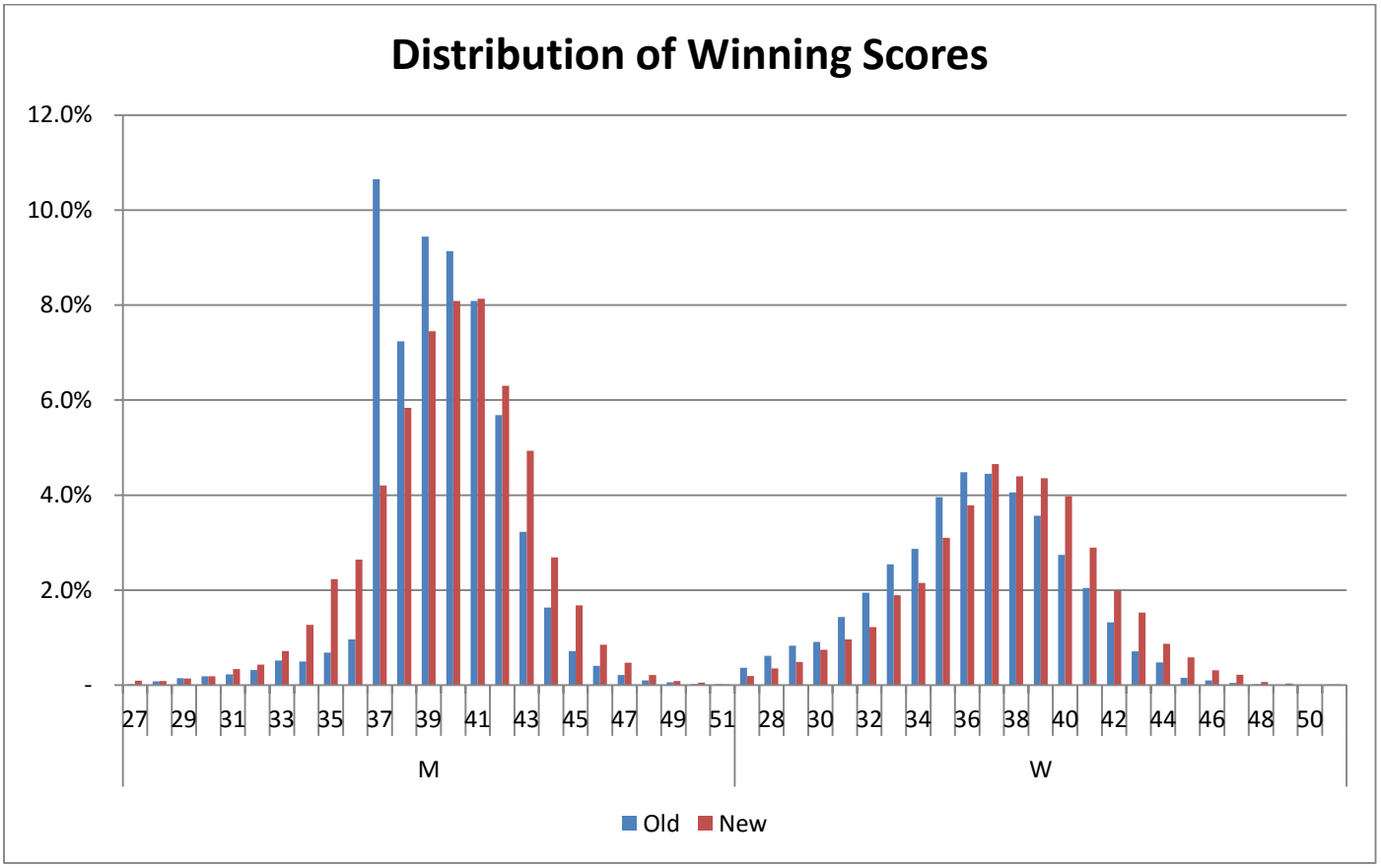
What handicaps are placing in competitions?



"Placing" is interpreted as being in the top 20% of the field.

Here the lowest men have had a 20% increase, but the 1 to 5 group has again been almost halved. The Men up through 15 have had a 25% reduction, and the high markers have increased their wins. Again the movement for the women is more moderate.

What scores are required to win or place in competitions?



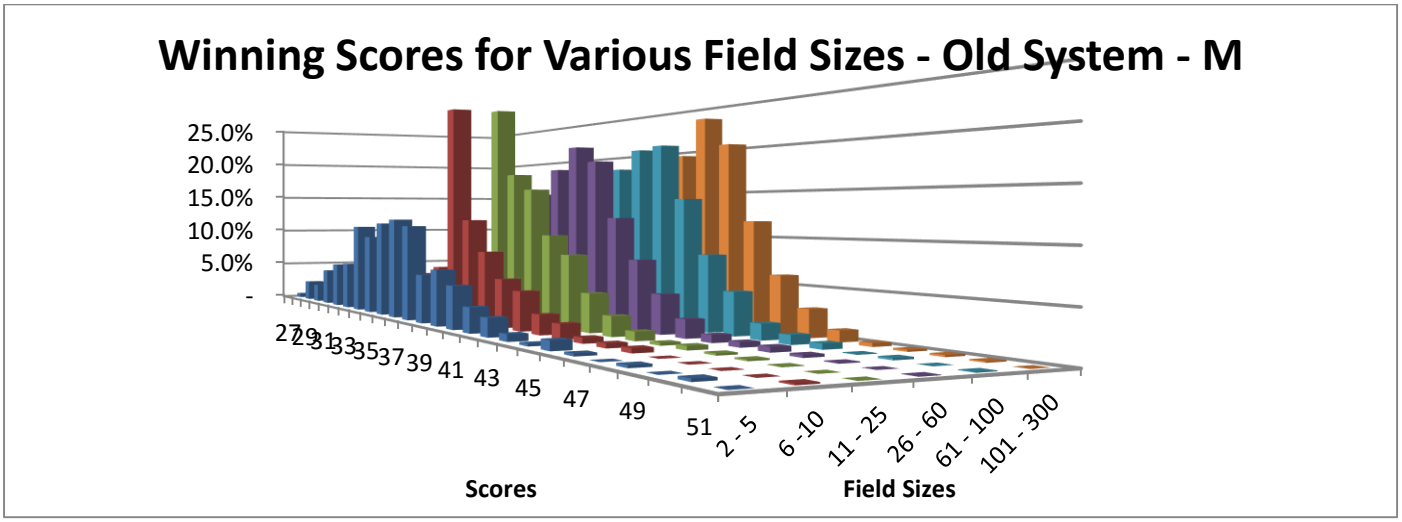
Under the Old system, 37 points was the most frequent winning score for men. This is now 41 under the New system. While the number of wining scores 44 and over has doubled, it only represents 8% of the wins. The women's most frequent score has risen from 36 to 37.

However to get a more accurate reflection of what is happening the charts on page 28 should be consulted as they factor in the number of players at each field size level.

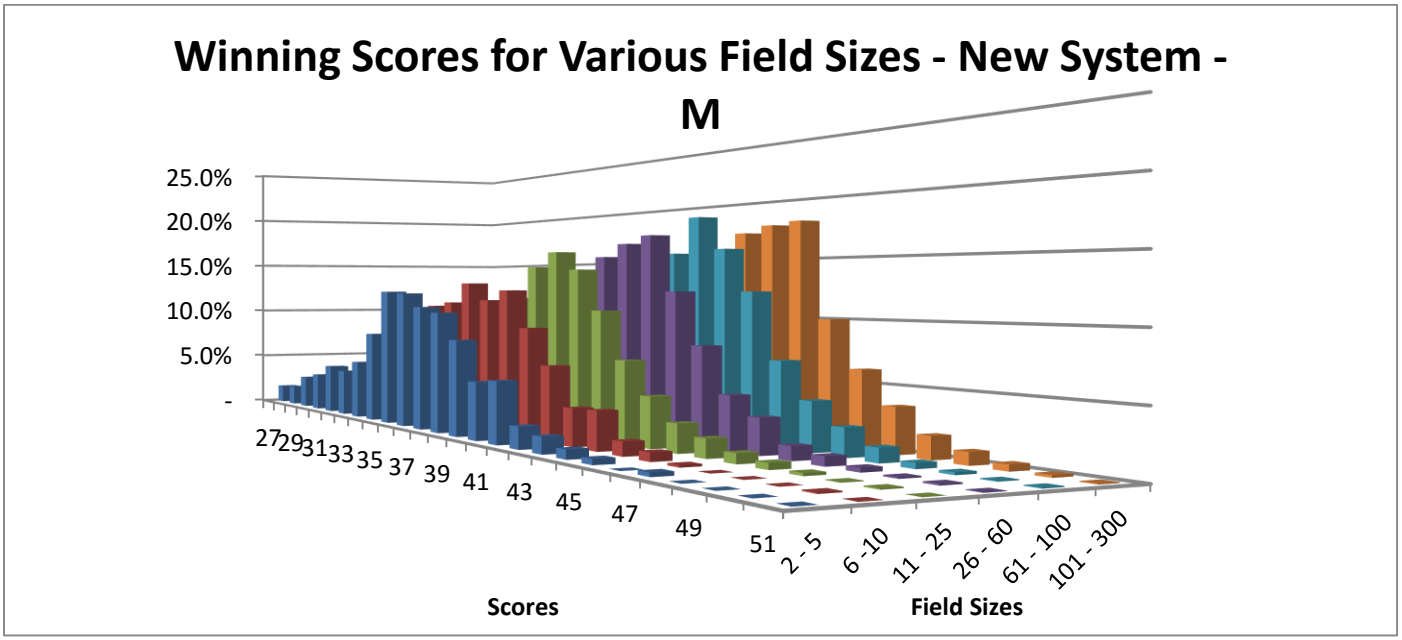
Do these patterns vary depending on field size?

Yes. As might be expected, under both the old and the New system, a larger field size provides opportunity for more outstanding performances.

For men's competitions, under the Old system the most frequent winning score for fields 6 thru 25 was 37 points. As the field size increased this rose to 39, and then 41.

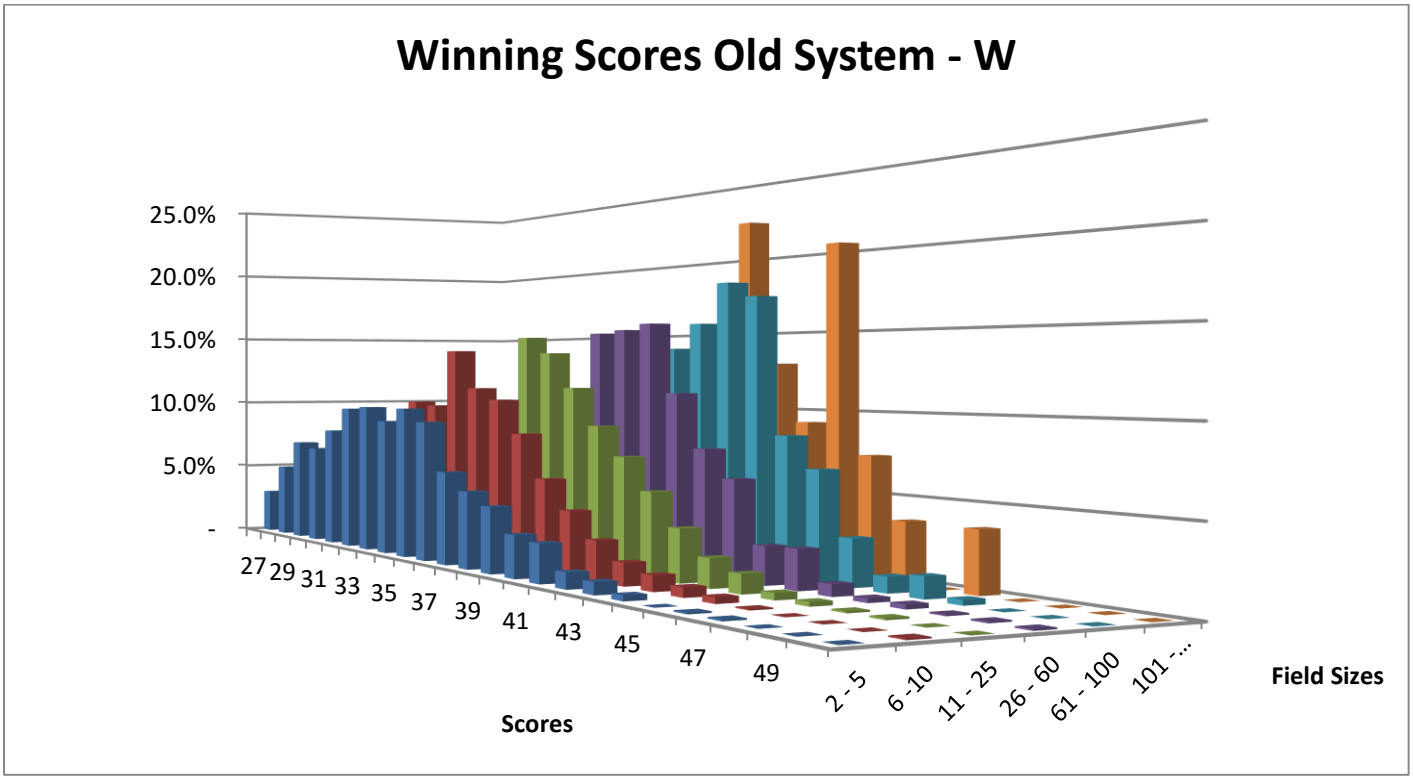


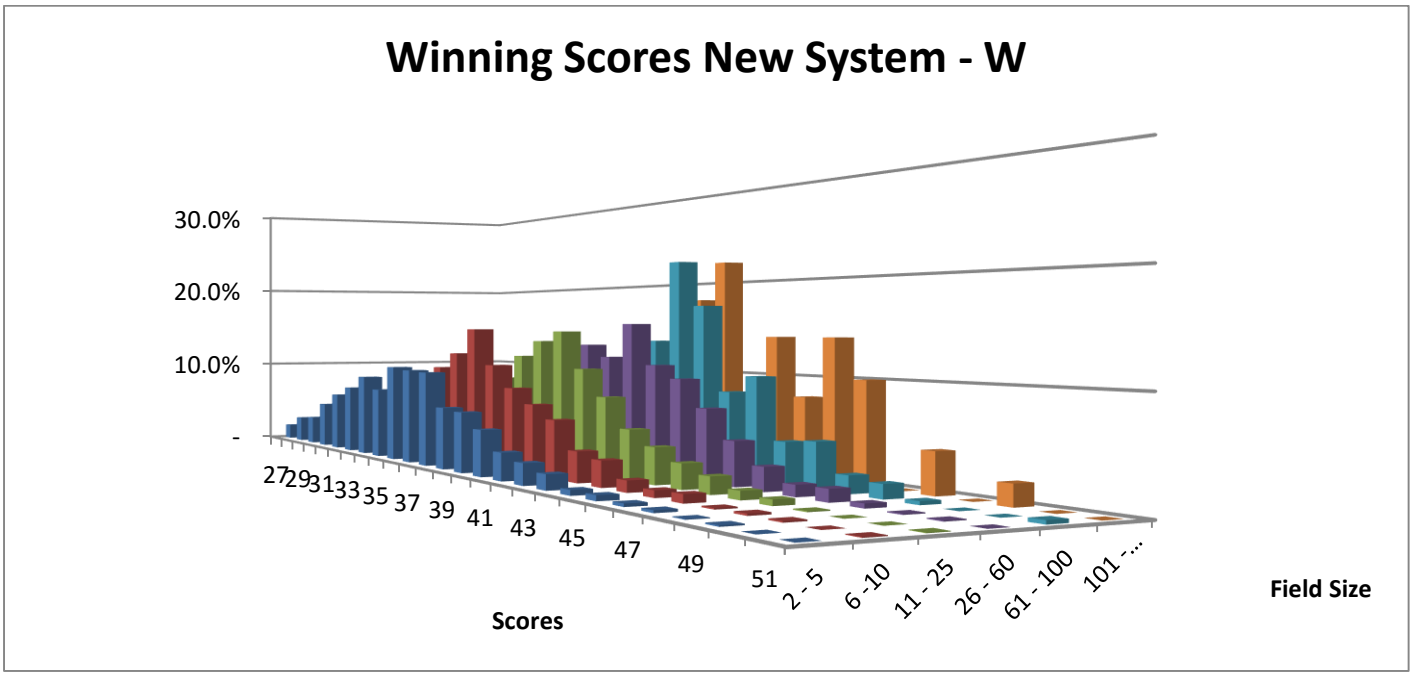
Under the New system the pattern is similar, but with a clear increase in the higher scores, especially at larger field sizes.



Under the New system the most frequent winning score for fields 6 thru 10 was 37 points. As the field size increased this rose to 39, 41, 41 and 43.

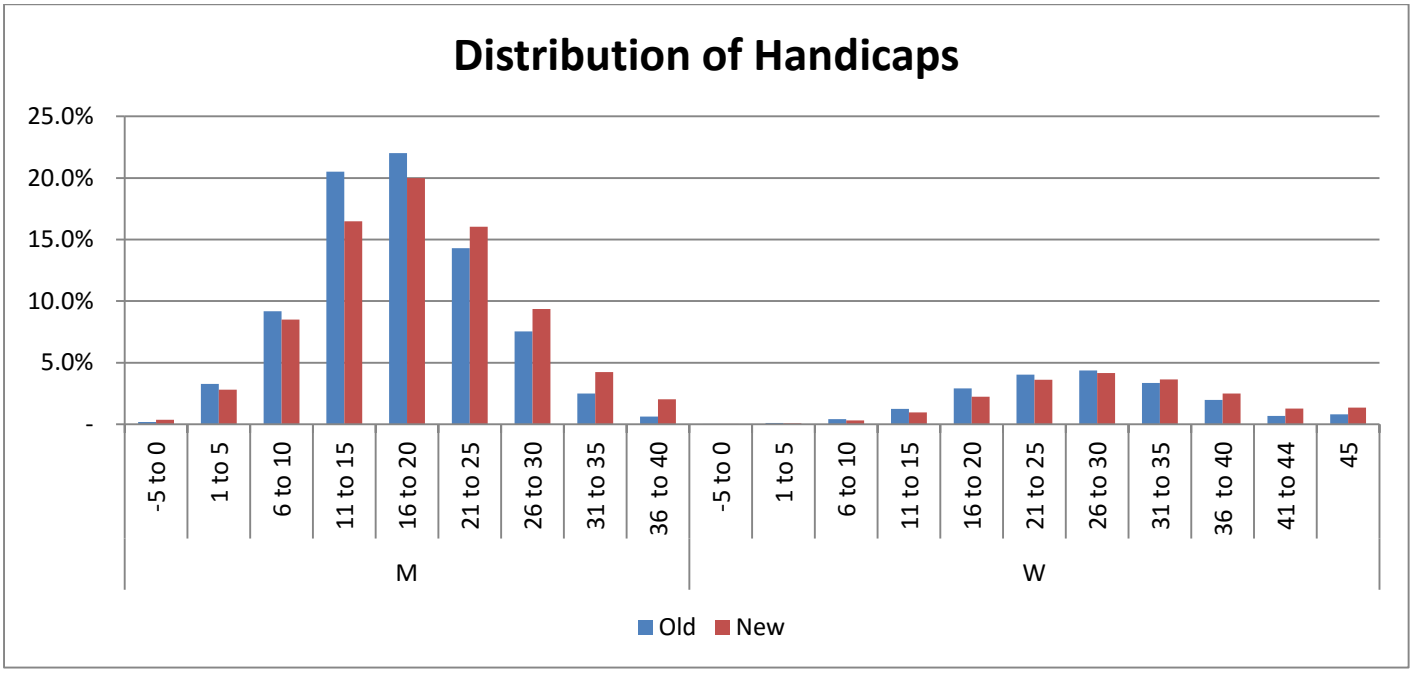
For women's competitions, as by now would have been expected, the differences between the Old and the New are moderate.





How has handicap distribution changed from the Old system to the New system?

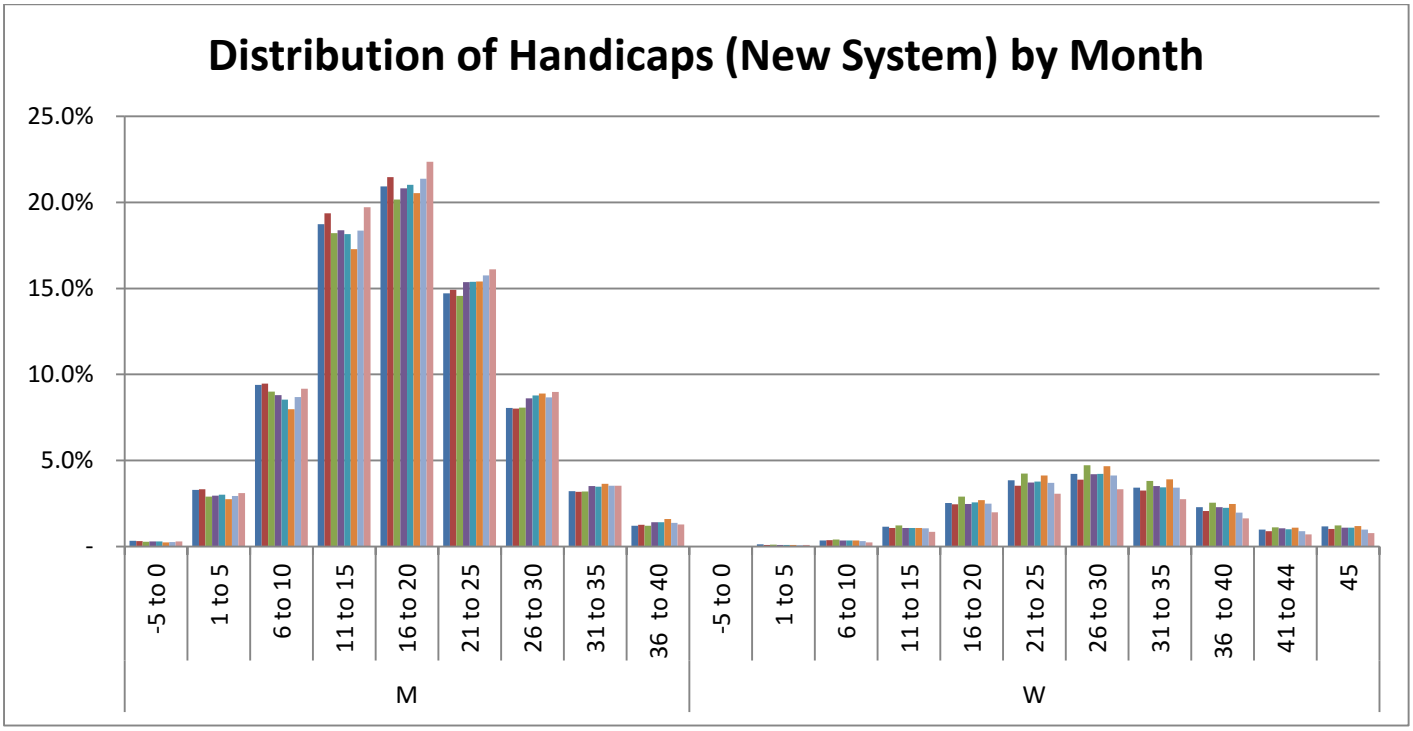
There has been a distinct move to higher handicaps with the New system.



Overall, players with handicaps up to 20 have declined from 59.9% of the sample to 51.5%. Obviously the higher marker groups have increased as a proportion (40.1% to 50.5%.)

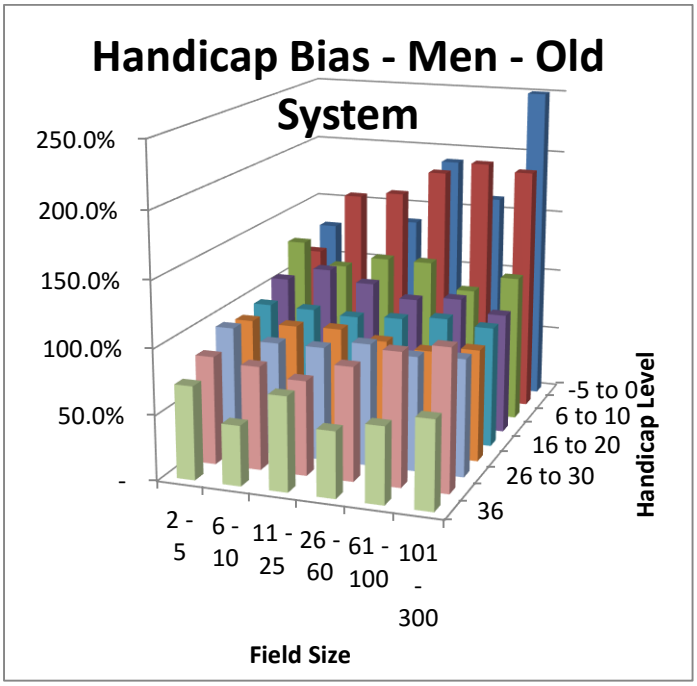
Has there been any 'settling down' in the period since 9 April 2010?

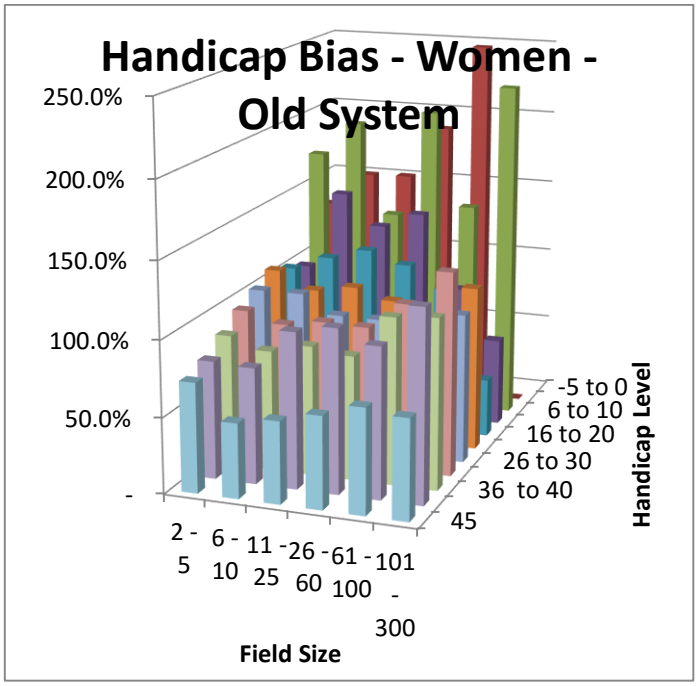
As the next chart shows, there is no evidence of a "Settling Down Period". In other words, there is no significant variation between the month on month distribution of handicaps in the months following the introduction of the New system. The "New" handicaps issued on Day One of the New system reflected the same distribution of handicaps as in evidence seven months later.



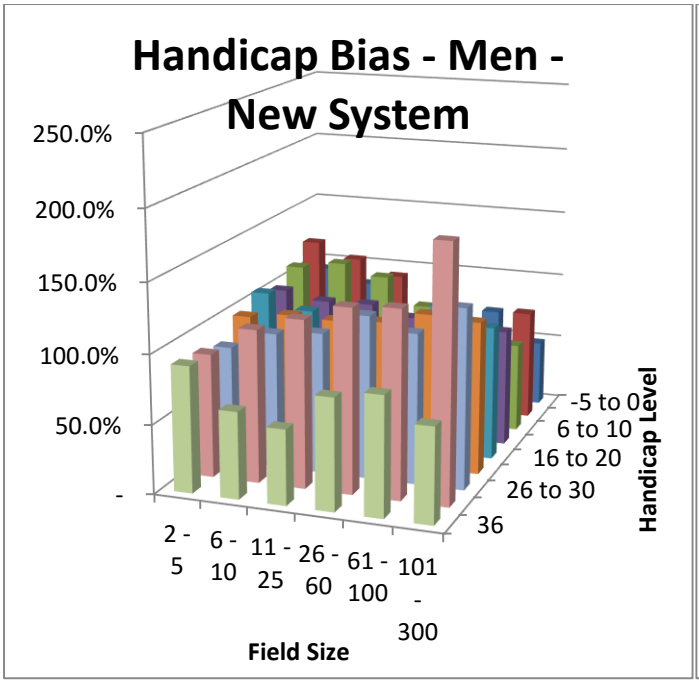
Implications for an Equitable system

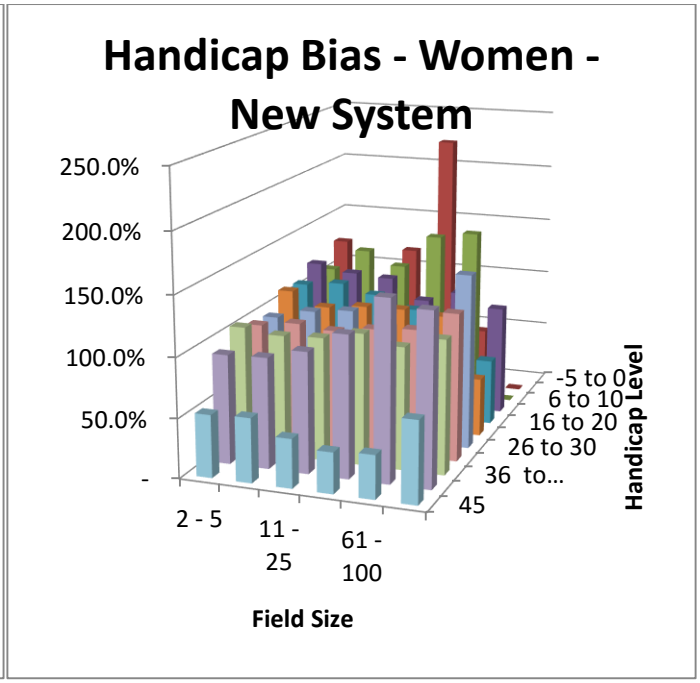
The charts shown above for the Old system, and repeated here, are quite telling description of the Old system:





The markers from -5 to 5 had a most significant Handicap Bias, with results 2 to 3 times better than "a level playing field". The higher markers, even from as low as 16 up, had an unfavourable bias, and were essentially playing off lower handicaps than they were, on average, capable of playing to.





With the introduction of the New system, although all players are handicapped to a level they can play to (times 0.96), the impact of the higher standard deviations of these high markers starts to be felt, and, especially in larger fields, some of them are scoring at levels which low markers can't easily replicate.

Clearly the New field is not level either, with pot holes for the low marker and a field size boost for the high marker.

Conclusion

- Neither the Old nor the New systems are representative of a level playing field.

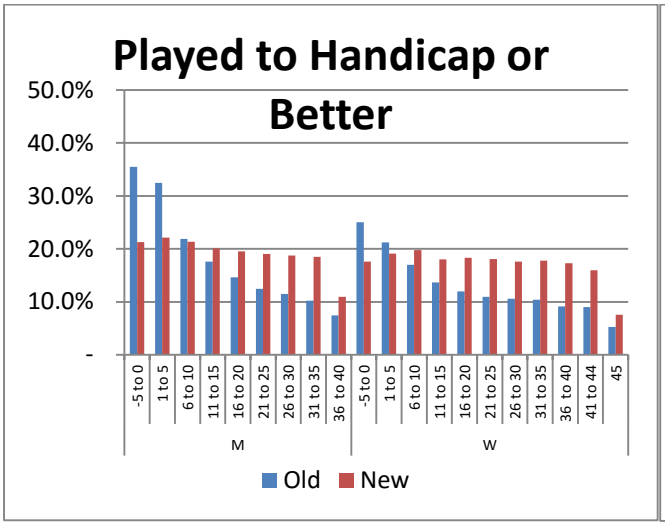
- Under the Old, the high markers were disadvantaged so that they won far fewer competitions than their representation in the field. Many of them were chronically playing at handicap levels far in excess of their playing handicaps. Essentially this was due to the uneven way that differentials were applied. Handicaps for high markers only "eased out" by 0.1 stroke for a poor round, no matter how poor, but they were tightened far more quickly if the player had the occasional good round.

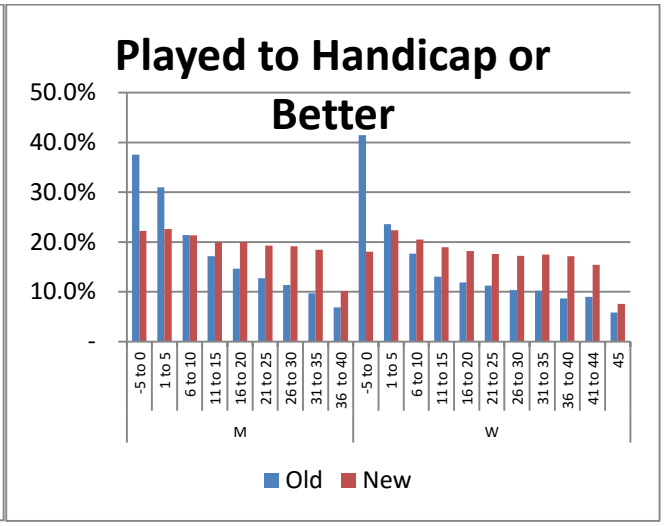
- Under the New system players are being handicapped closer to their recent form, but what with the greater inconsistency of the higher markers, when field sizes grow, there are enough chances for a high marker to put in one of his exceptional rounds, and the high markers are now winning more than their fair share of the competitions with larger field sizes.

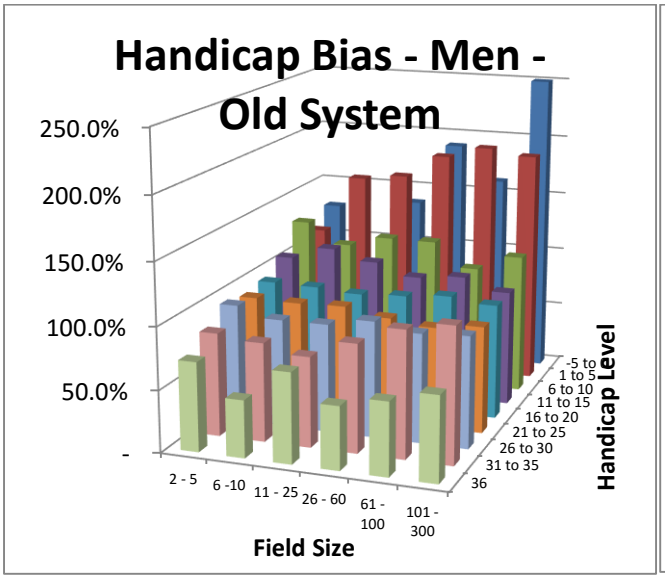
- It is clear that a reduction in BFE, or in the number of "best" scores taken, would tilt the playing field in favour of the lower markers. However there is likely to be some negative impact on other handicap groups. Modelling of specific solutions should be performed before introducing any change.

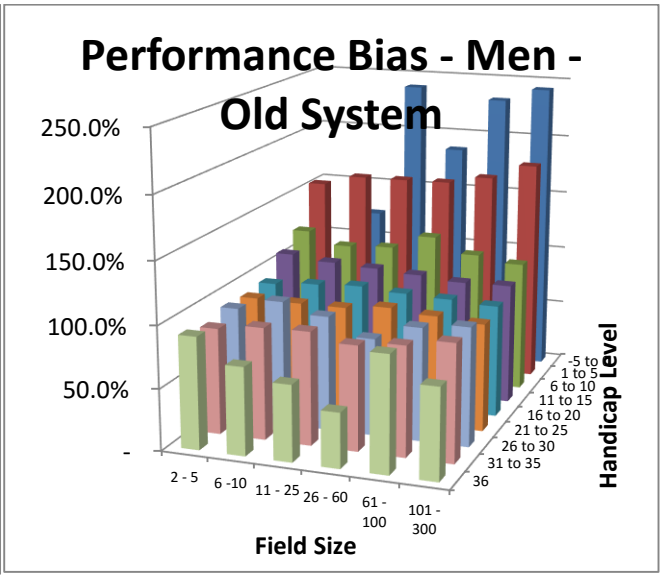
Appendix 1 – Comparison of Data sets

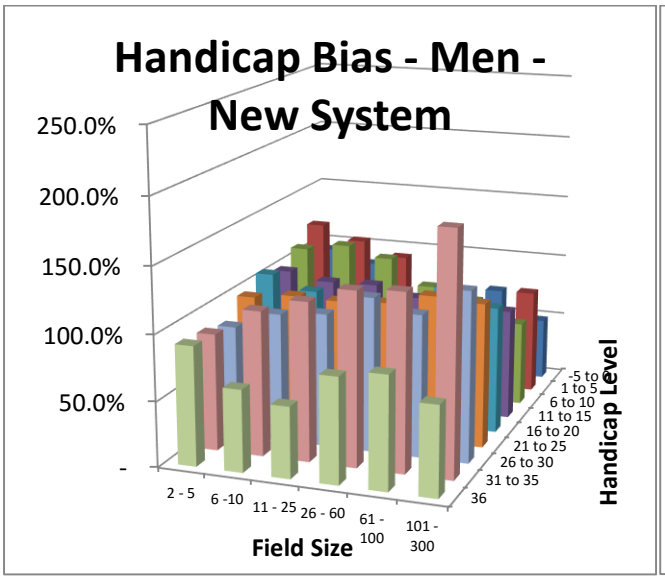
In order to ensure that the dataset chosen was representative of the whole golfer population, an additional sample was extracted. It too consisted of nearly 1 million rounds of golf. In the following comparison, the first dataset is used in the results on the left, the second in the results on the right. The conclusion is that the results are remarkably similar.

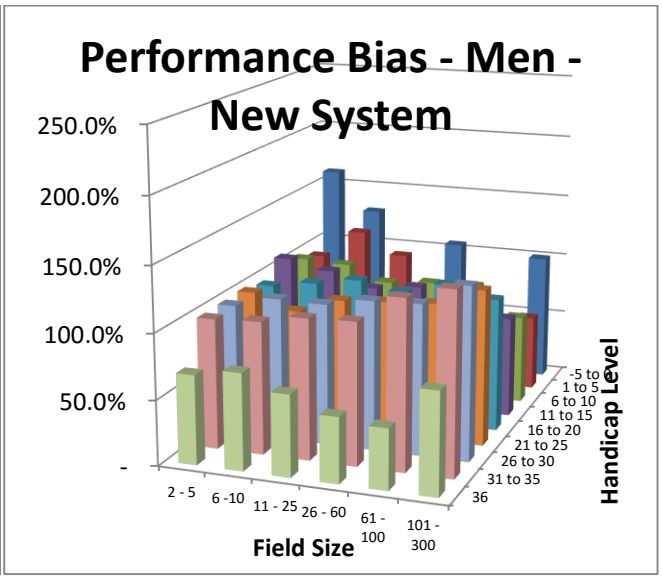


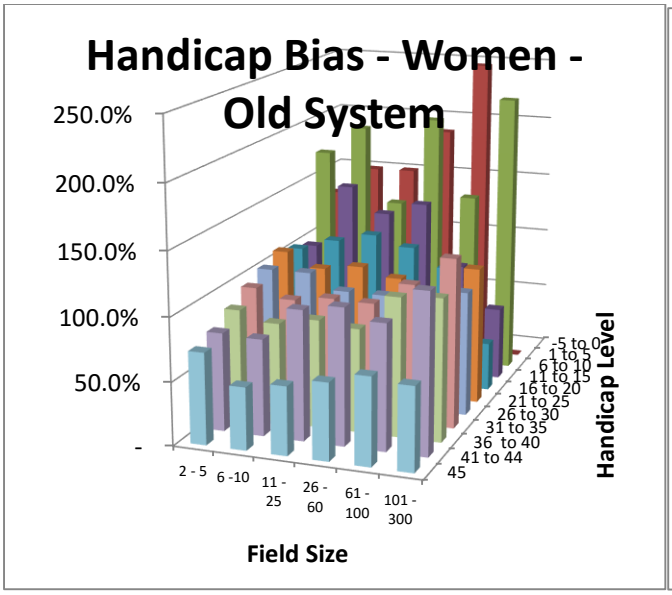


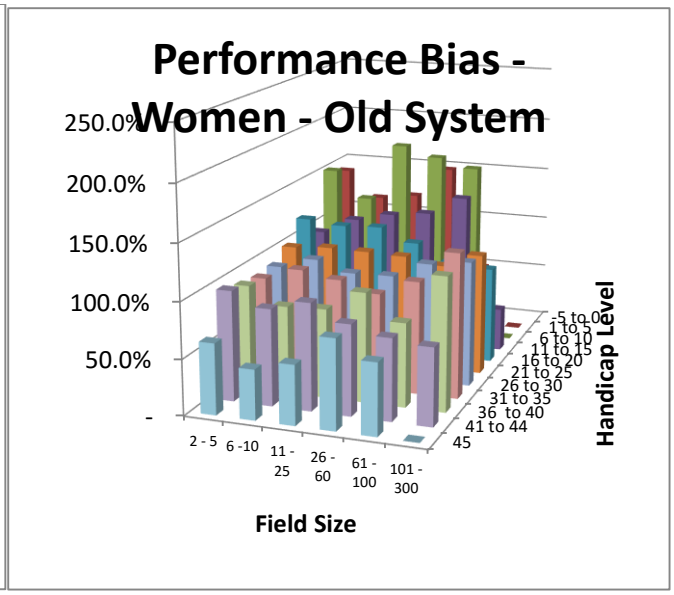


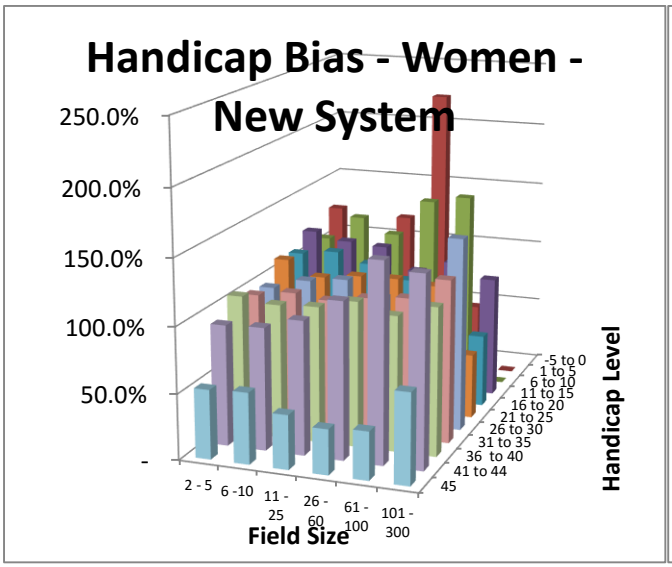


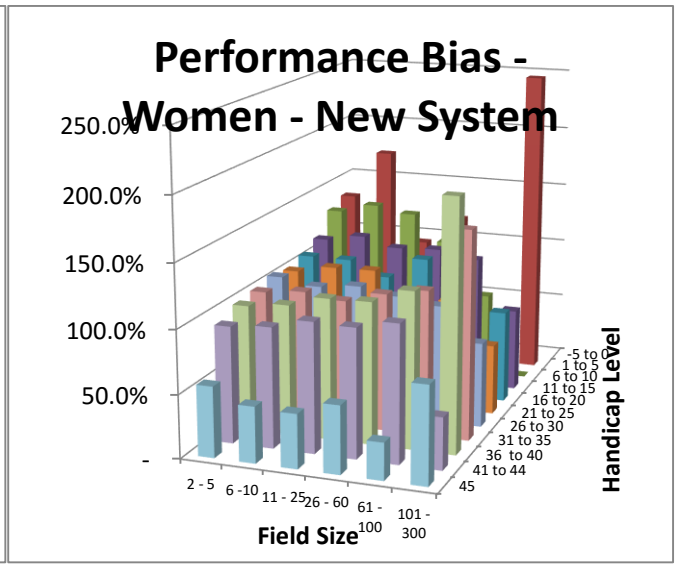


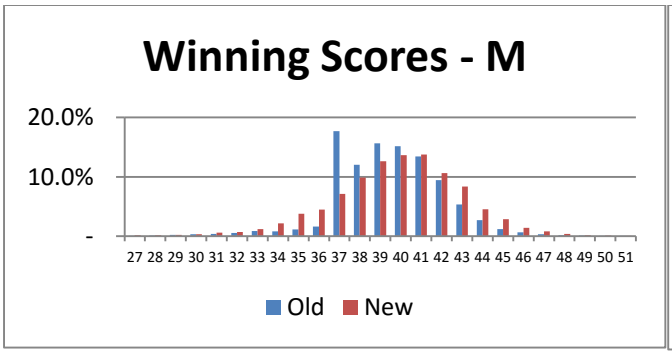


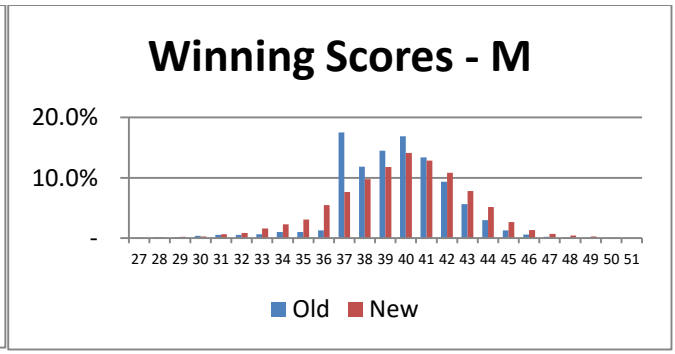

Handicap Bias - Women - Old System
250.0%
200.0%
150.0%
100.0%
50.0%
-
2 - 5
6 -10
11 - 25
26 - 60
61 - 100
101 - 300
Field Size
-5 to 0
1 to 5
6 to 10
11 to 15
16 to 20
21 to 25
26 to 30
31 to 35
36 to 40
41 to 44
45
Handicap Level

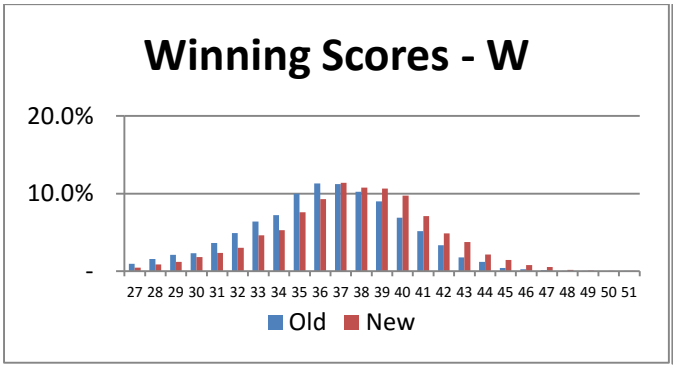

Performance Bias - Women - Old System
250.0%
200.0%
150.0%
100.0%
50.0%
-
2 - 5
6 -10
11 - 25
26 - 60
61 - 100
101 - 300
Field Size
-5 to 0
1 to 5
6 to 10
11 to 15
16 to 20
21 to 25
26 to 30
31 to 35
36 to 40
41 to 44
45
Handicap Level

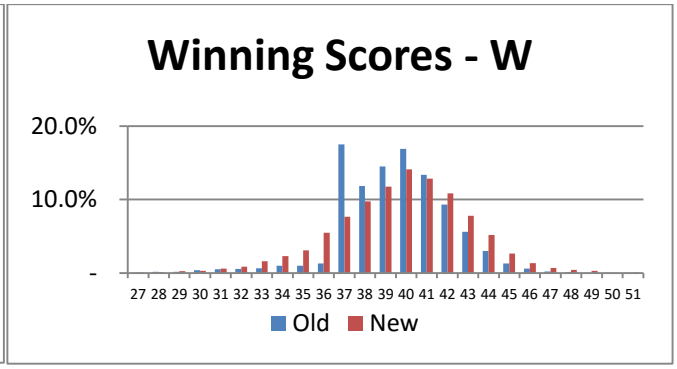

Handicap Bias - Women - New System
250.0%
200.0%
150.0%
100.0%
50.0%
-
2 - 5
6 -10
11 - 25
26 - 60
61 - 100
101 - 300
Field Size
-5 to 0
1 to 5
6 to 10
11 to 15
16 to 20
21 to 25
26 to 30
31 to 35
36 to 40
41 to 44
45
Handicap Level


Performance Bias - Women - New System
250.0%
200.0%
150.0%
100.0%
50.0%
-
2 - 5
6 -10
11 - 25
26 - 60
61 - 100
101 - 300
Field Size
-5 to 0
1 to 5
6 to 10
11 to 15
16 to 20
21 to 25
26 to 30
31 to 35
36 to 40
41 to 44
45
Handicap Level


Winning Scores - M
20.0%
10.0%
-
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51
Old
New

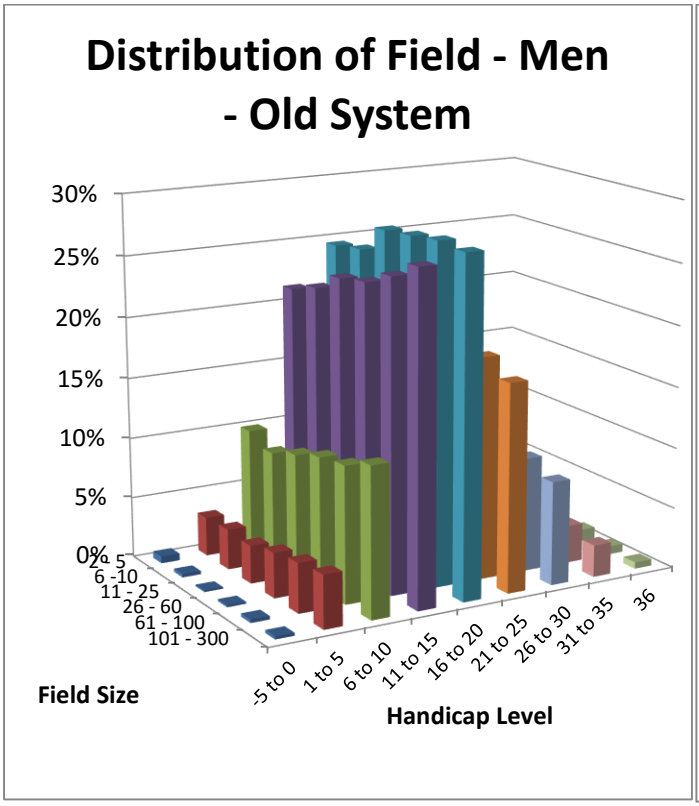

Winning Scores - M
20.0%
10.0%
-
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51
Old
New

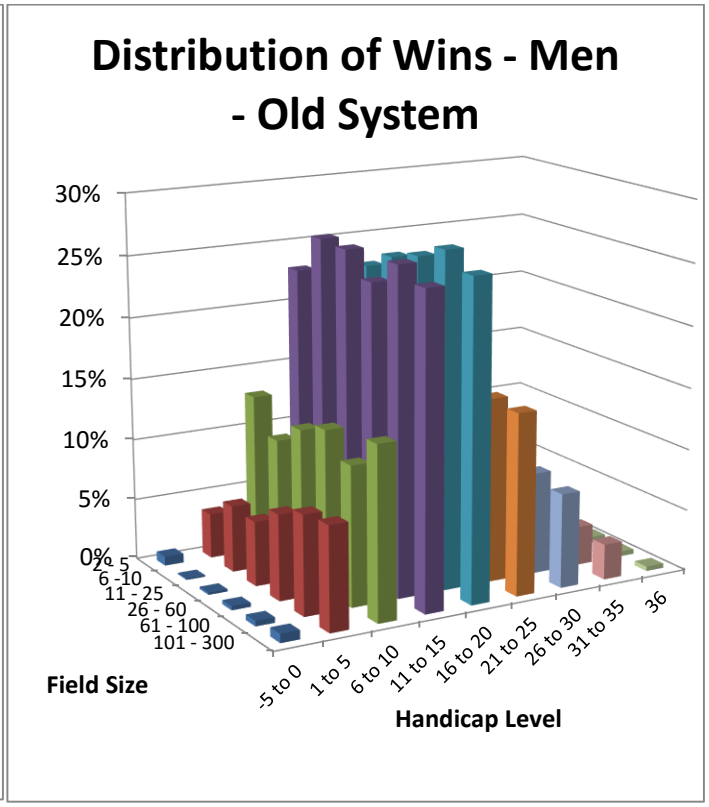

Winning Scores - W
20.0%
10.0%
-
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51
Old
New

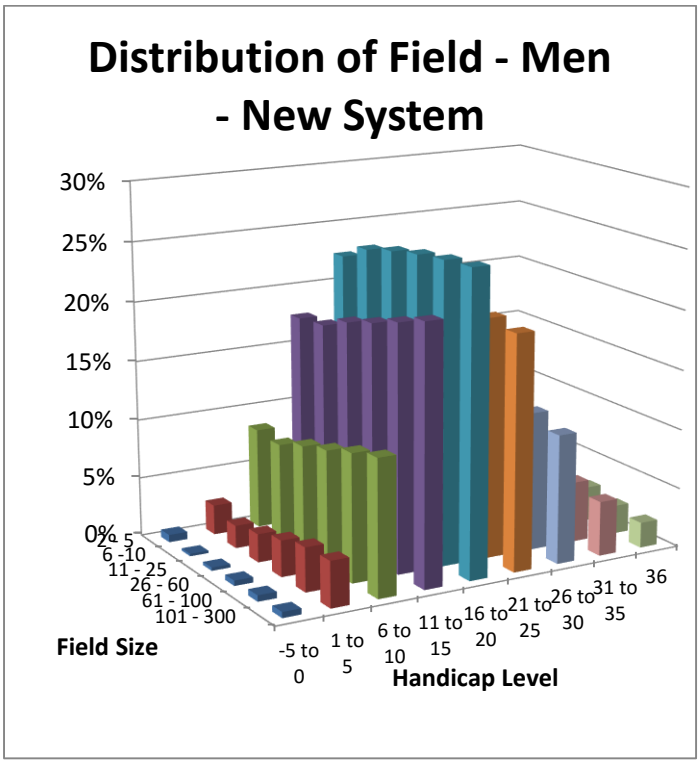

Winning Scores - W
20.0%
10.0%
-
27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51
Old
New

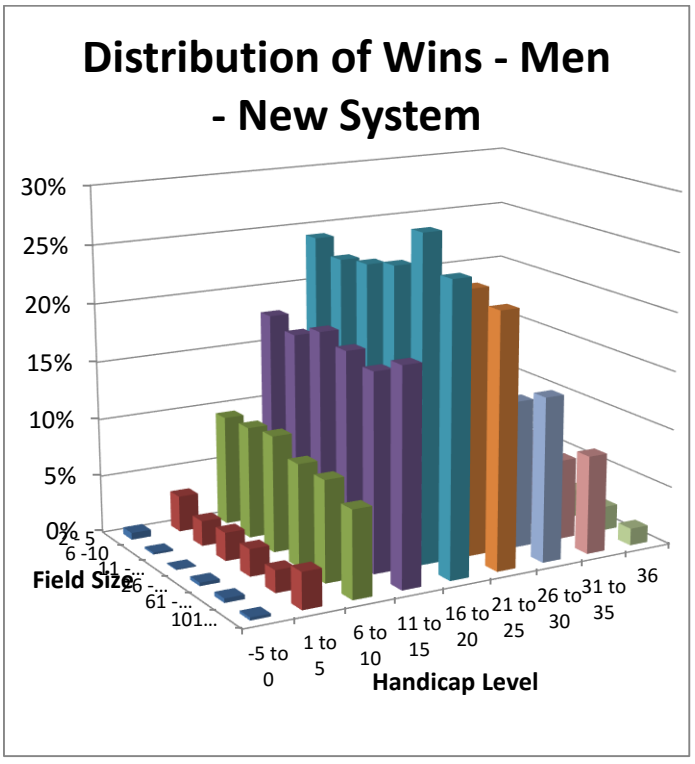
Appendix 2 – Distribution of Field and Winners

The "Raw distributions" from which the Handicap Bias is derived are presented below. The huge share of wins of the low handicappers in the Old system, compared to their proportion of the field, is the underlying cause of their high Handicap Bias in the Old system.



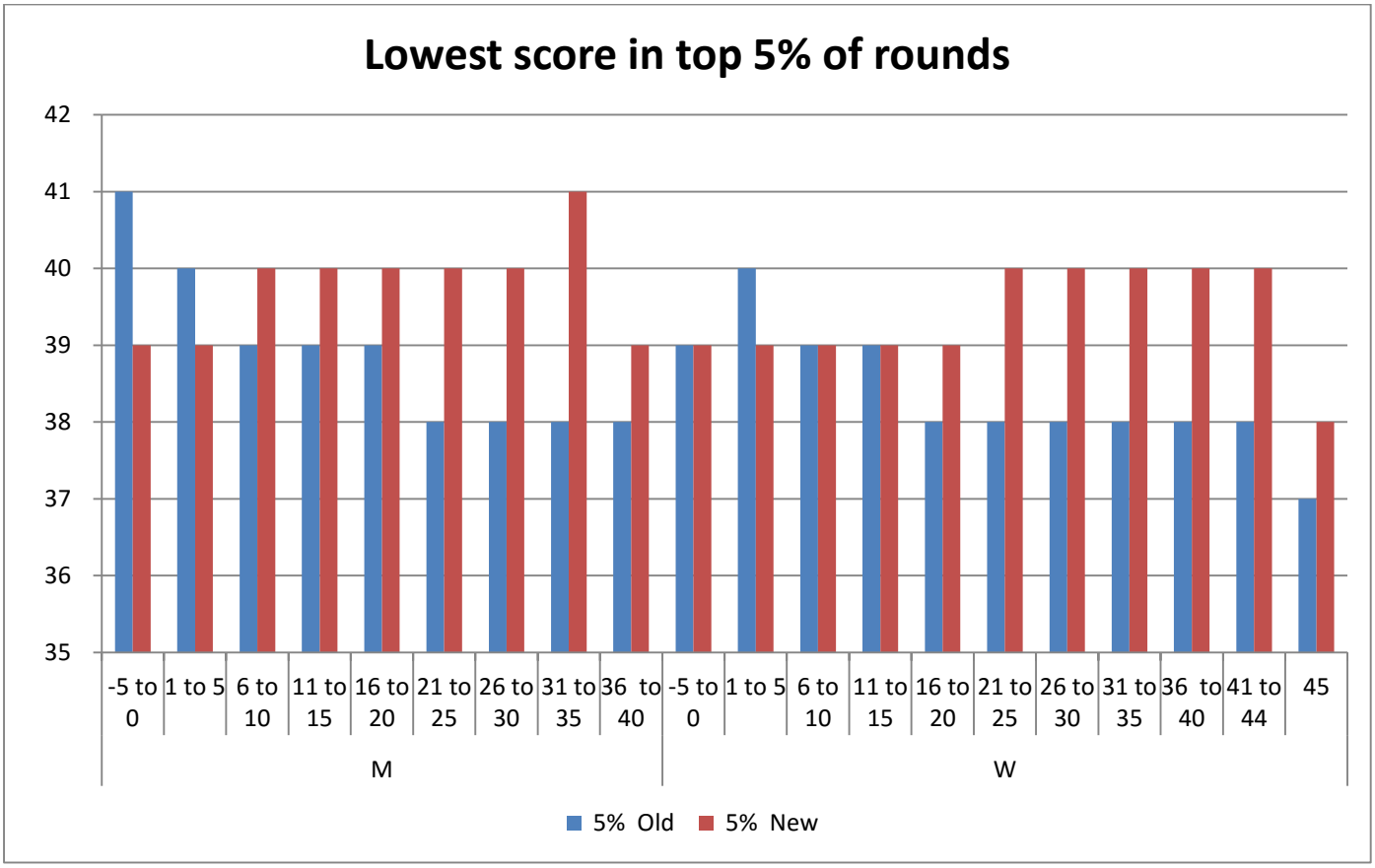


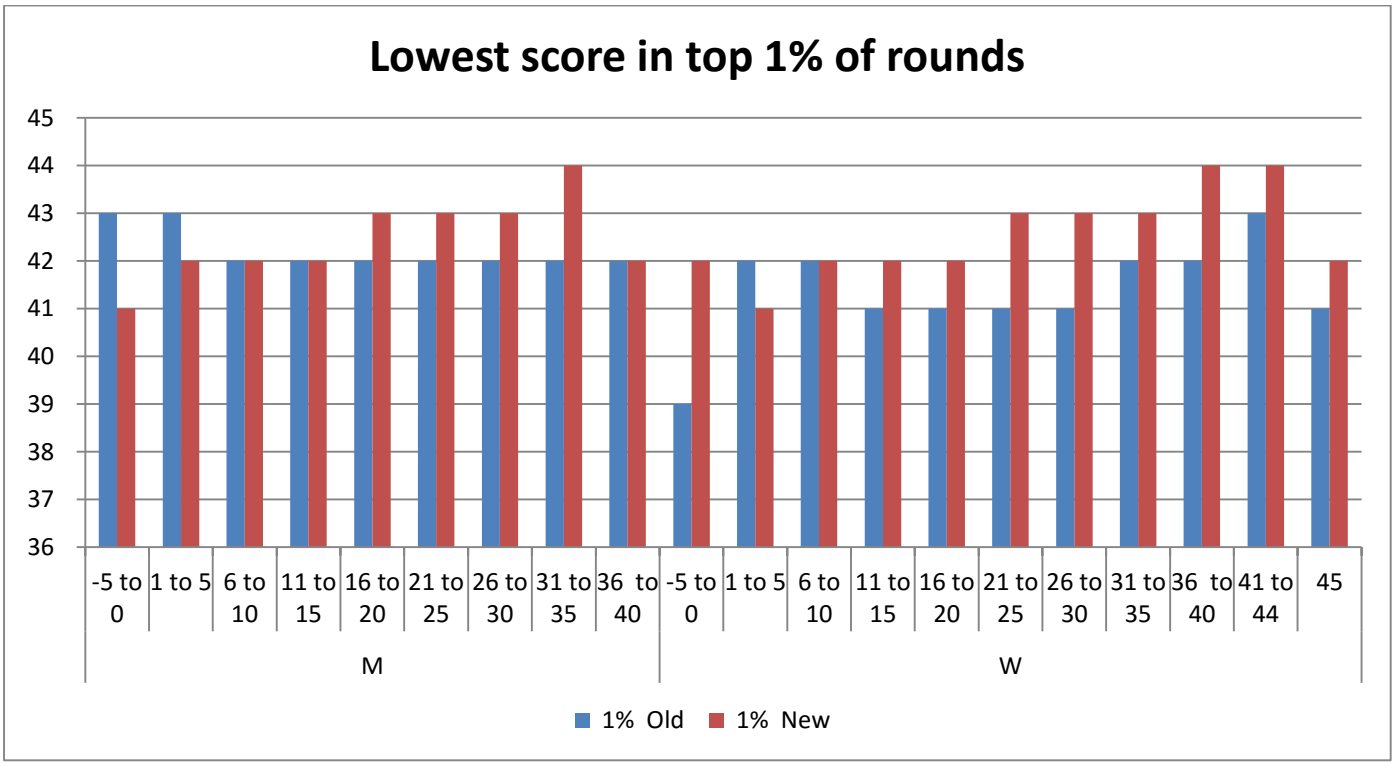


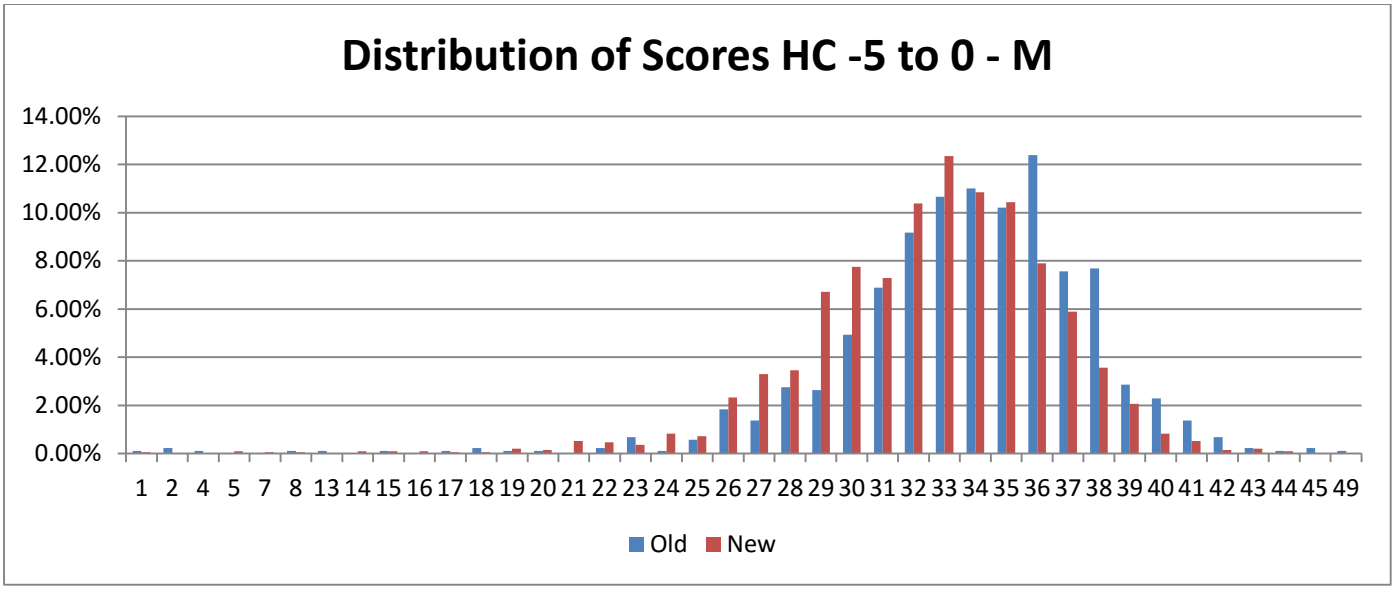


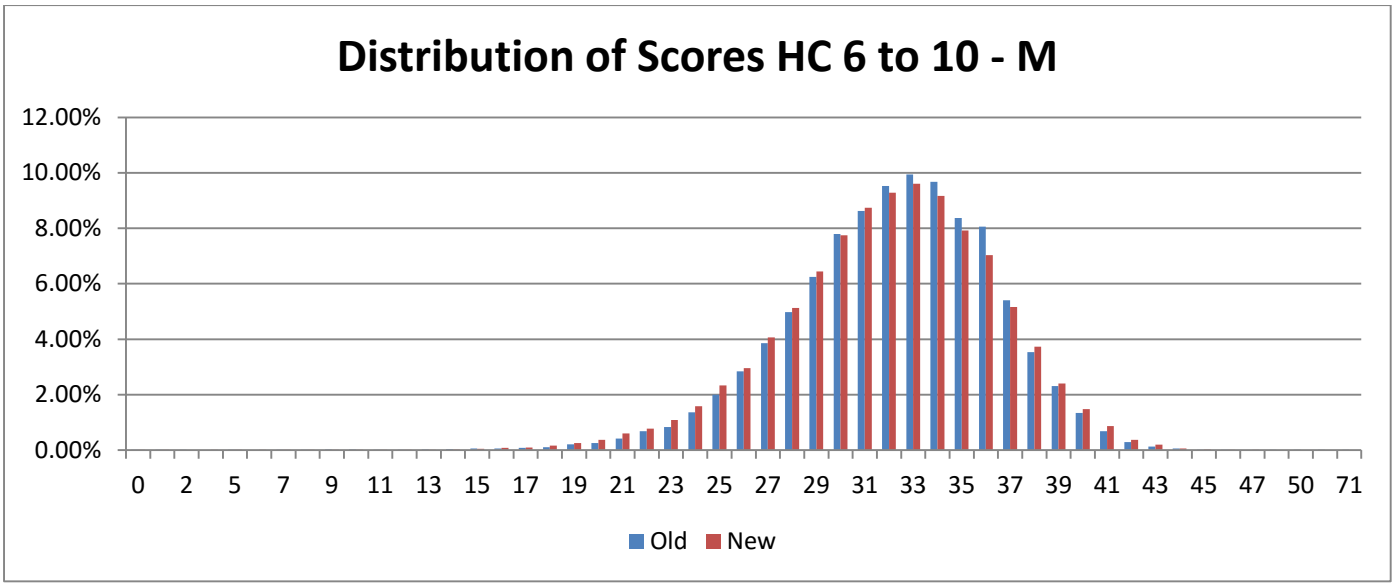
Appendix 3 – Distribution of Scores

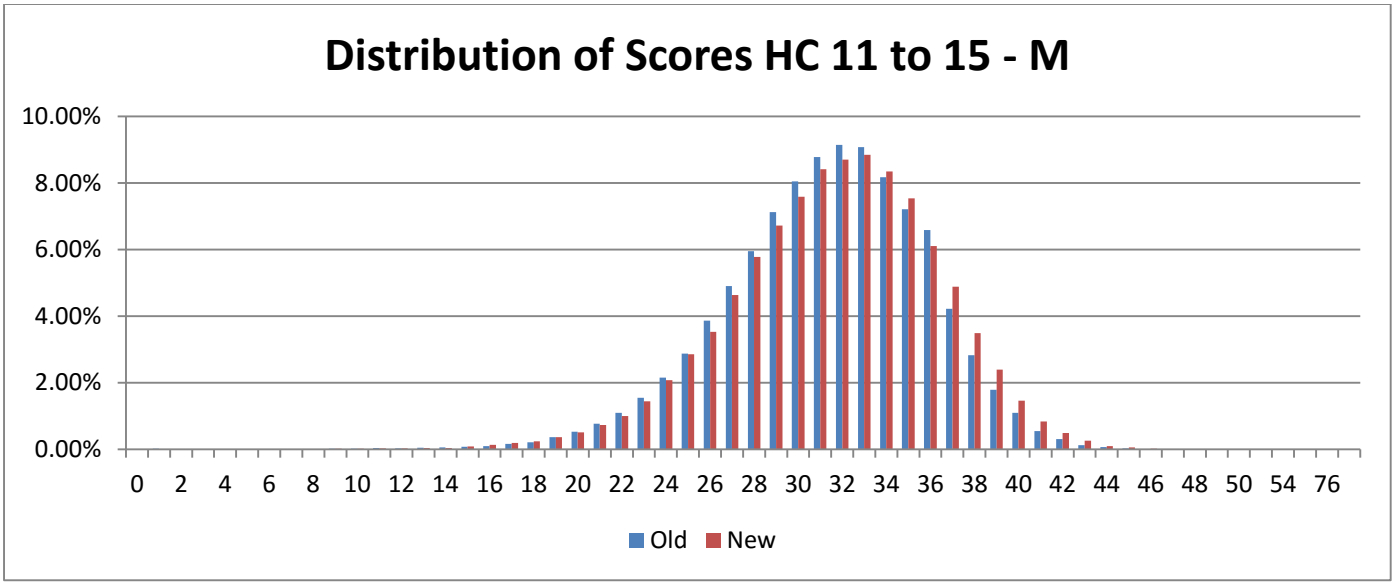
Specific distributions of scores for men and women are given below. We have already seen that low markers have a lower standard deviation, but a higher mean score. Because the low marker is having this deviation added to a lower mean than the high marker, it turns out that the "Top scores" only vary by two to three strokes across handicap groups.

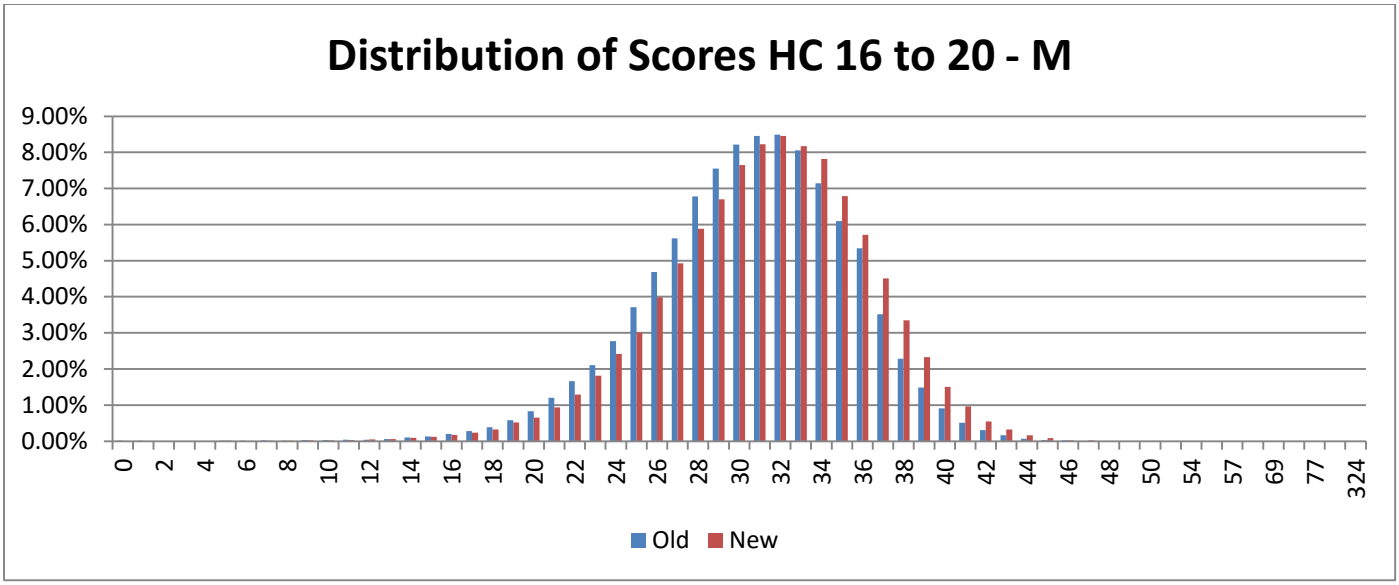


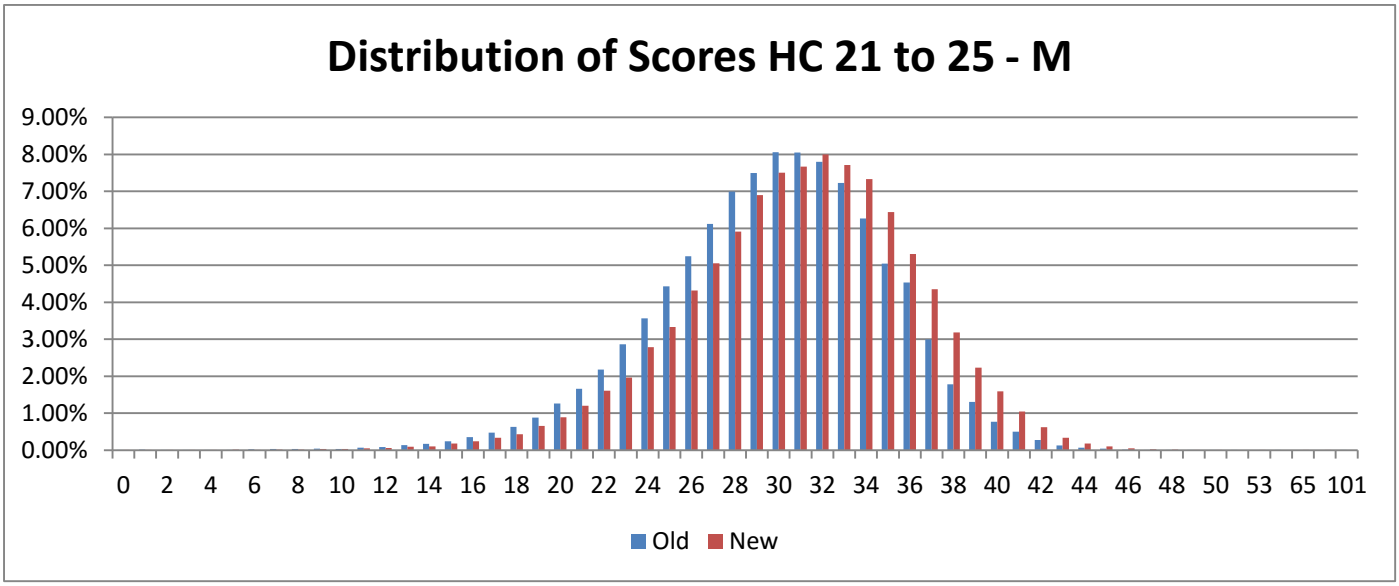


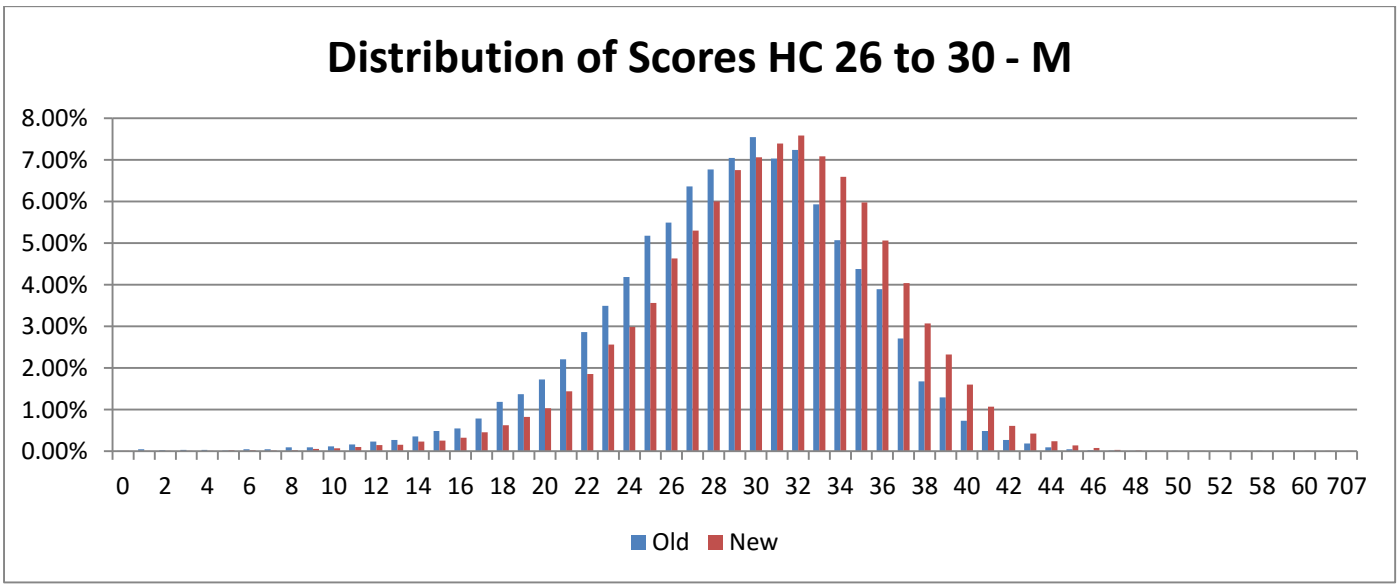

Distribution of Scores HC -5 to 0 - M
14.00%
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
1 2 4 5 7 8 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 49
Old
New

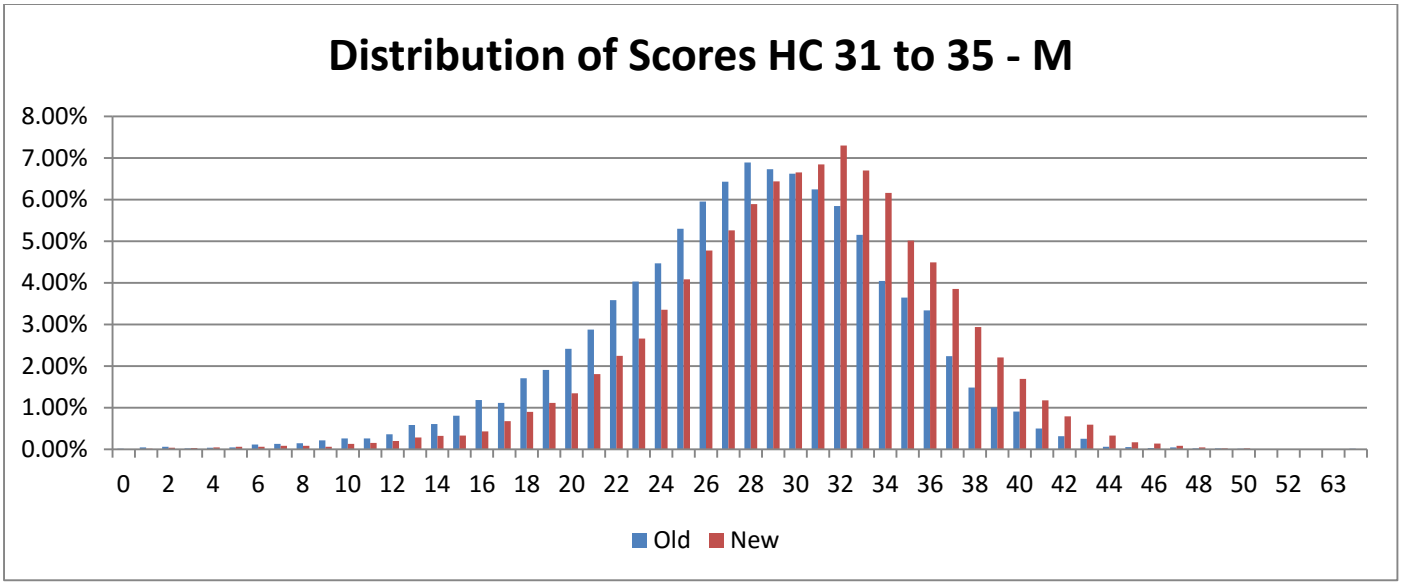

Distribution of Scores HC 6 to 10 - M
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
0 2 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 50 71
Old
New

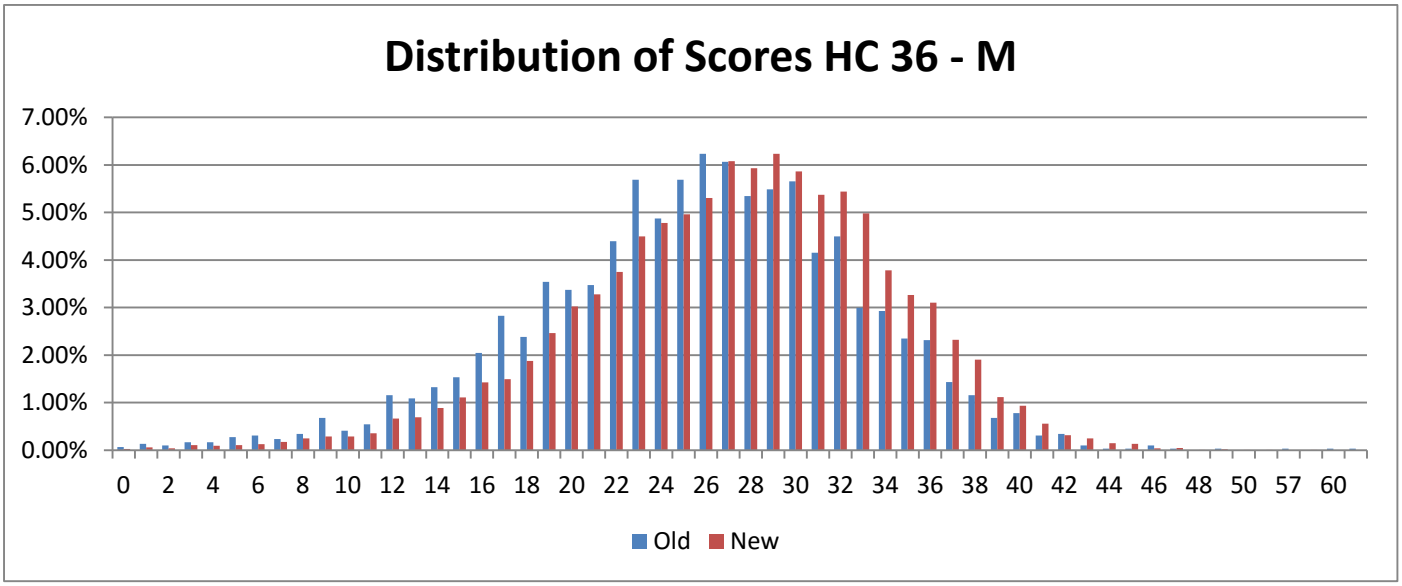

Distribution of Scores HC 11 to 15 - M
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 54 76
Old
New


Distribution of Scores HC 16 to 20 - M
9.00%
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 54 57 69 77 324
Old
New

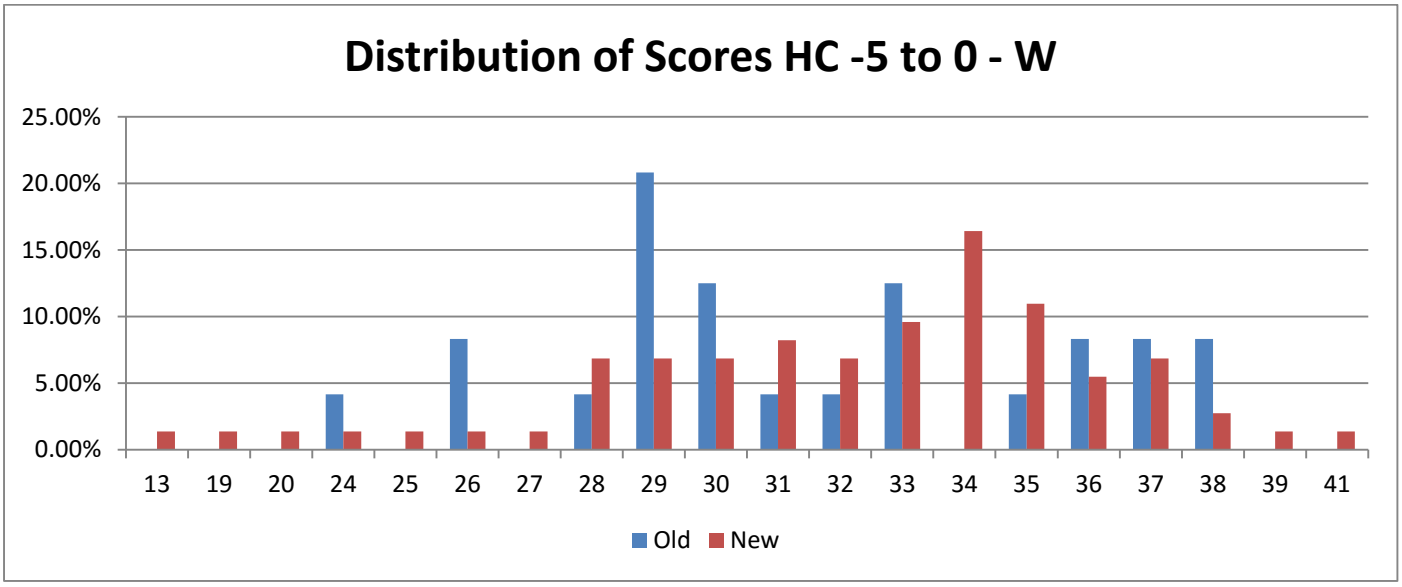

Distribution of Scores HC 21 to 25 - M
9.00%
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 53 65 101
Old
New

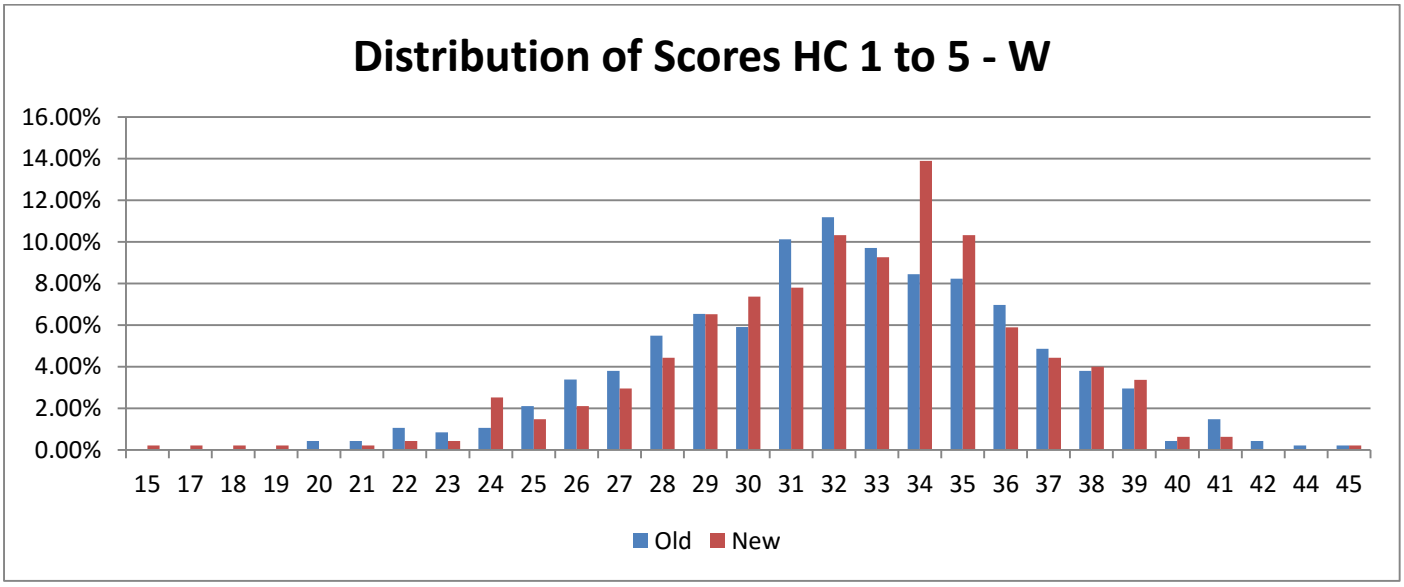

Distribution of Scores HC 26 to 30 - M
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 58 60 707
Old
New

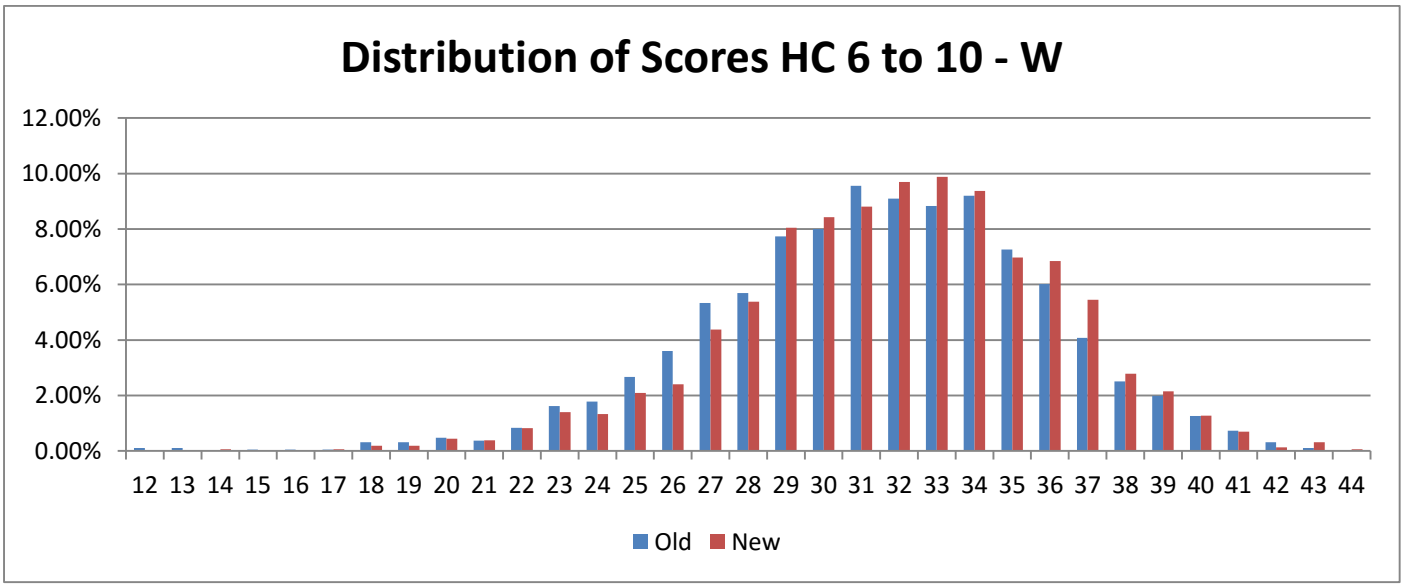

Distribution of Scores HC 31 to 35 - M
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 63
Old
New

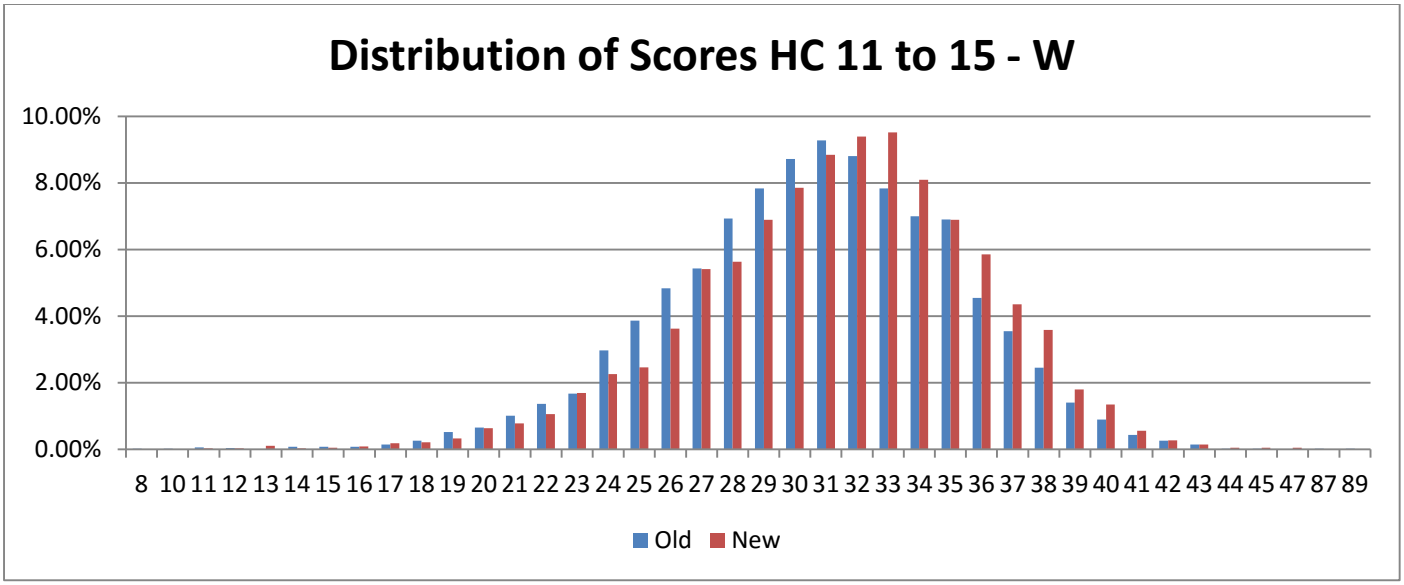

Distribution of Scores HC 36 - M
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 57 60
Old
New

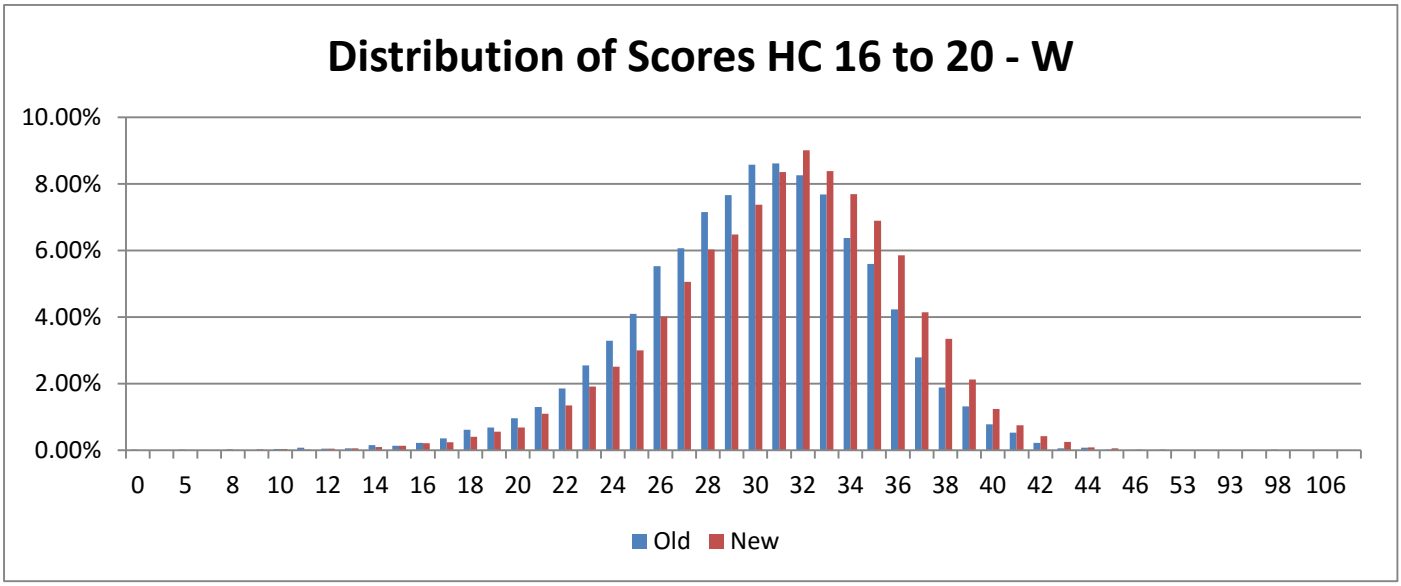
This chart is “Lumpy” because it has less than 100 rounds in the sample.

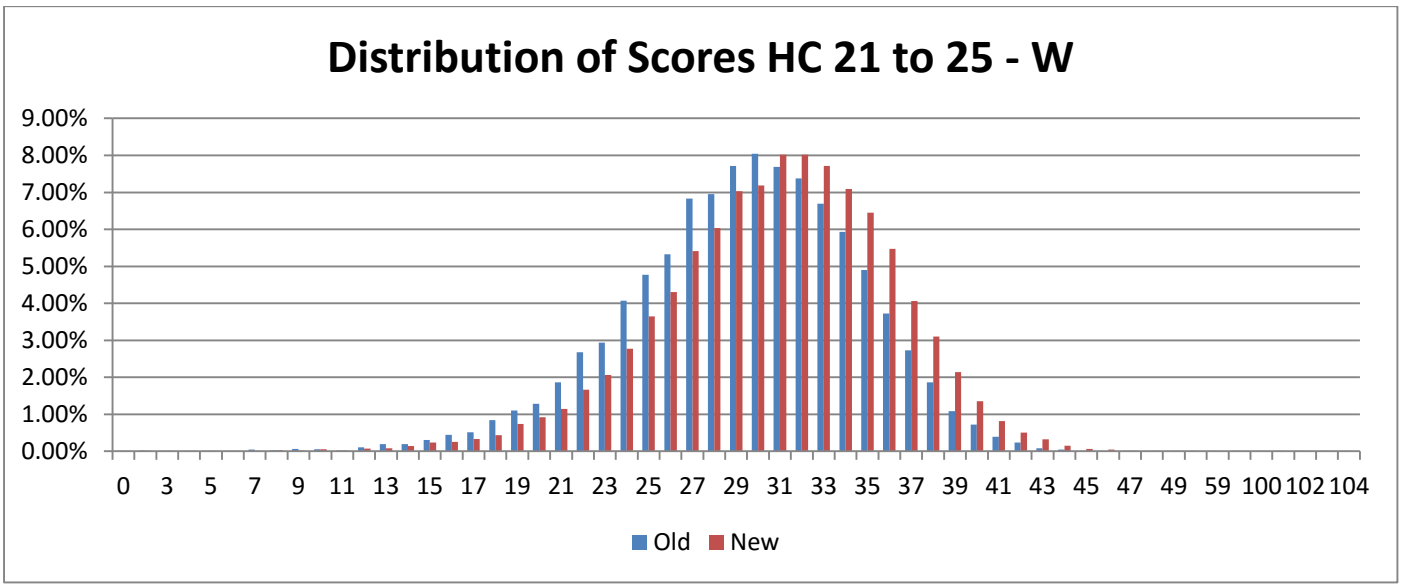


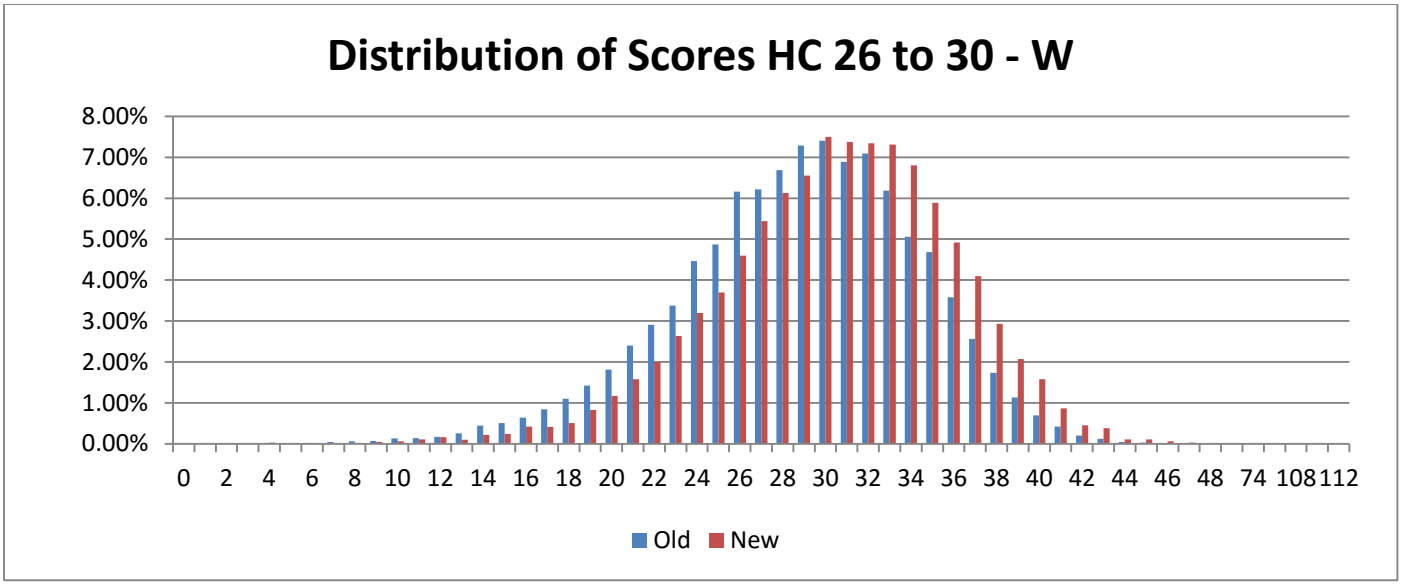


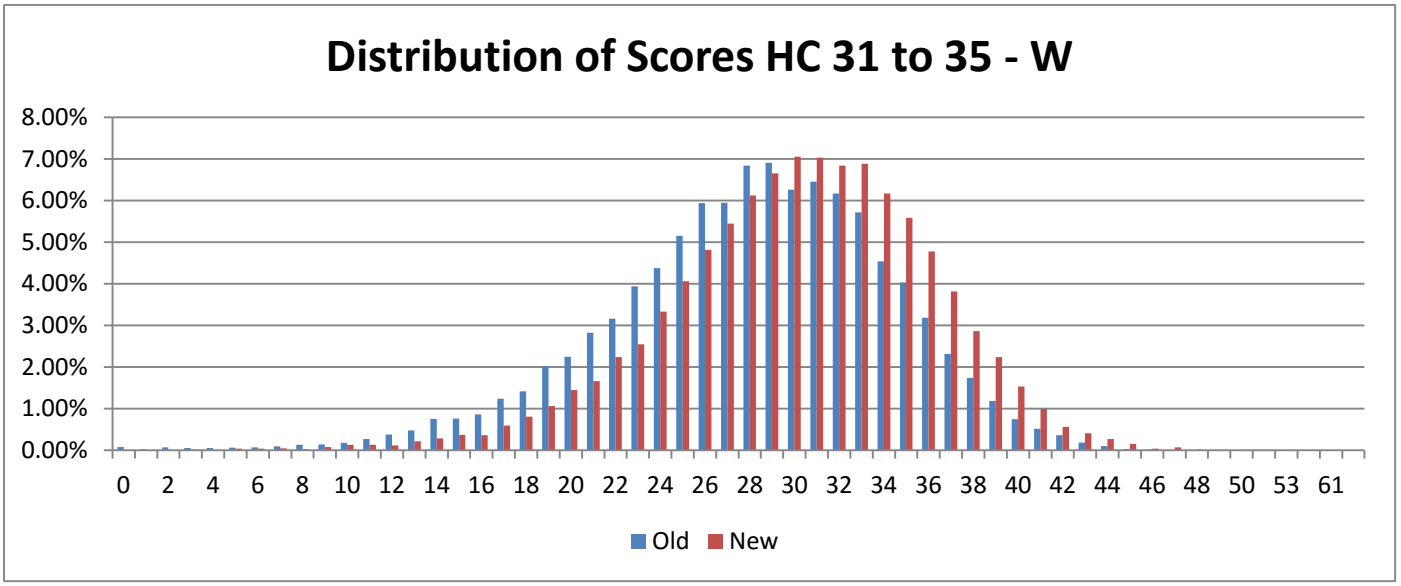


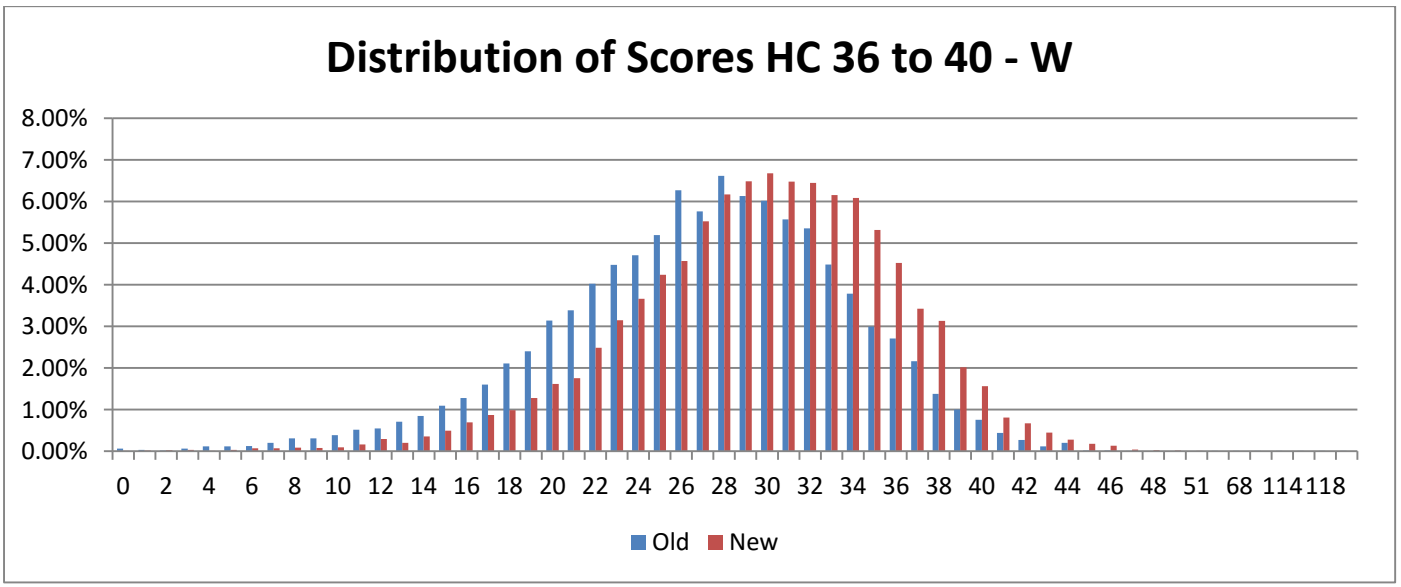

Distribution of Scores HC 11 to 15 - W
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
8 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 47 87 89
Old
New

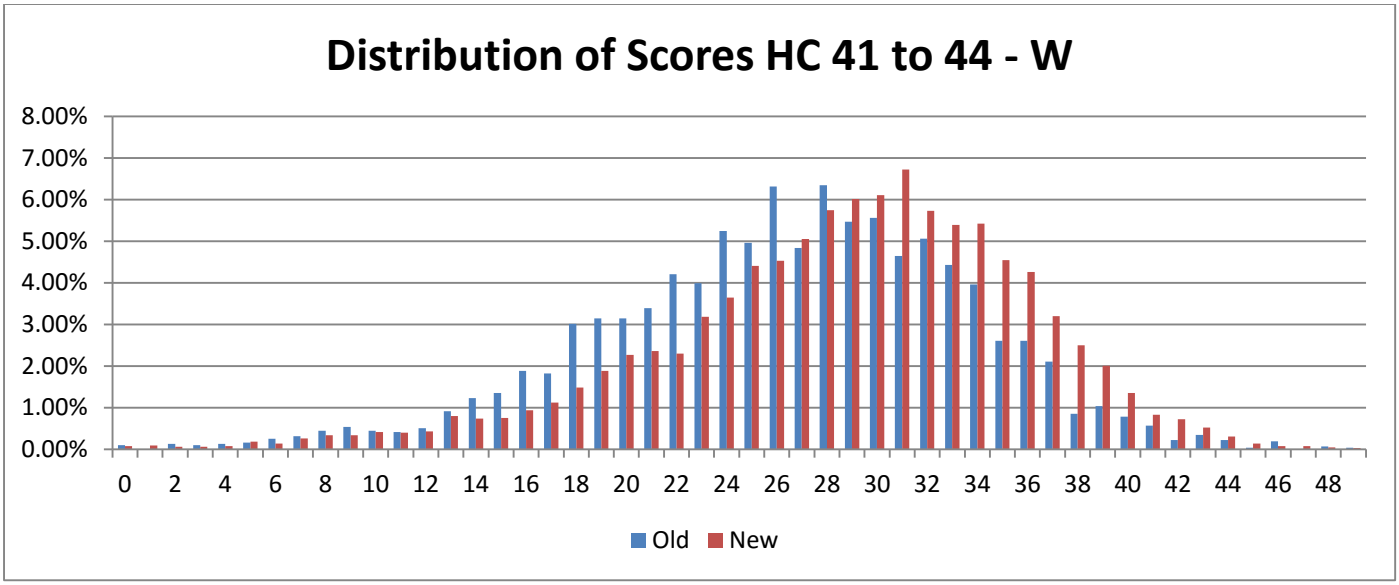

Distribution of Scores HC 16 to 20 - W
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
0 5 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 53 93 98 106
Old
New

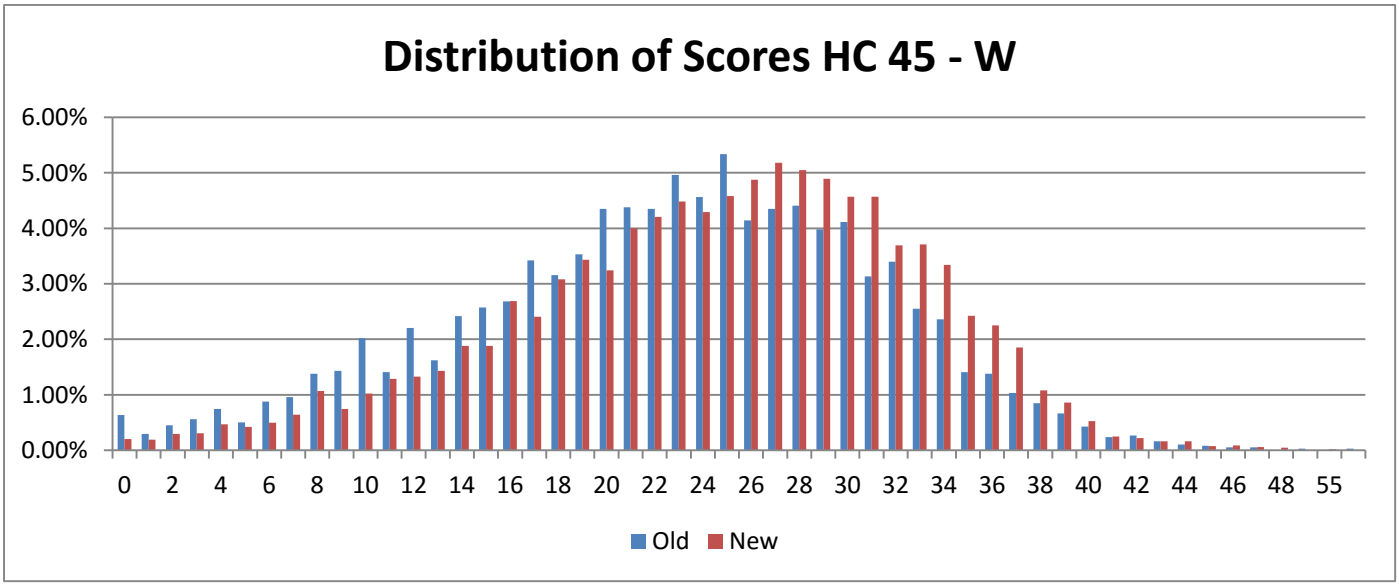

Distribution of Scores HC 21 to 25 - W
9.00%
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
0 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 59 100 102 104
Old
New


Distribution of Scores HC 26 to 30 - W
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 74 108 112
Old
New


Distribution of Scores HC 31 to 35 - W
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 53 61
Old
New

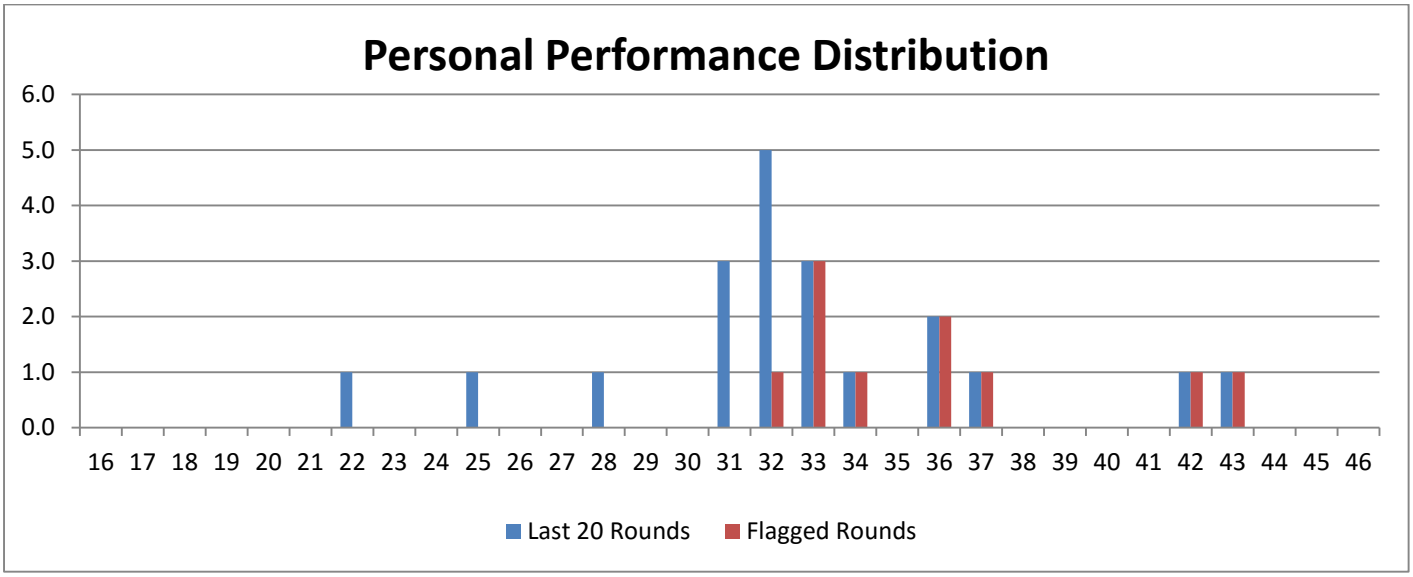

Distribution of Scores HC 36 to 40 - W
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 51 68 114 118
Old
New


Distribution of Scores HC 41 to 44 - W
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48
Old
New

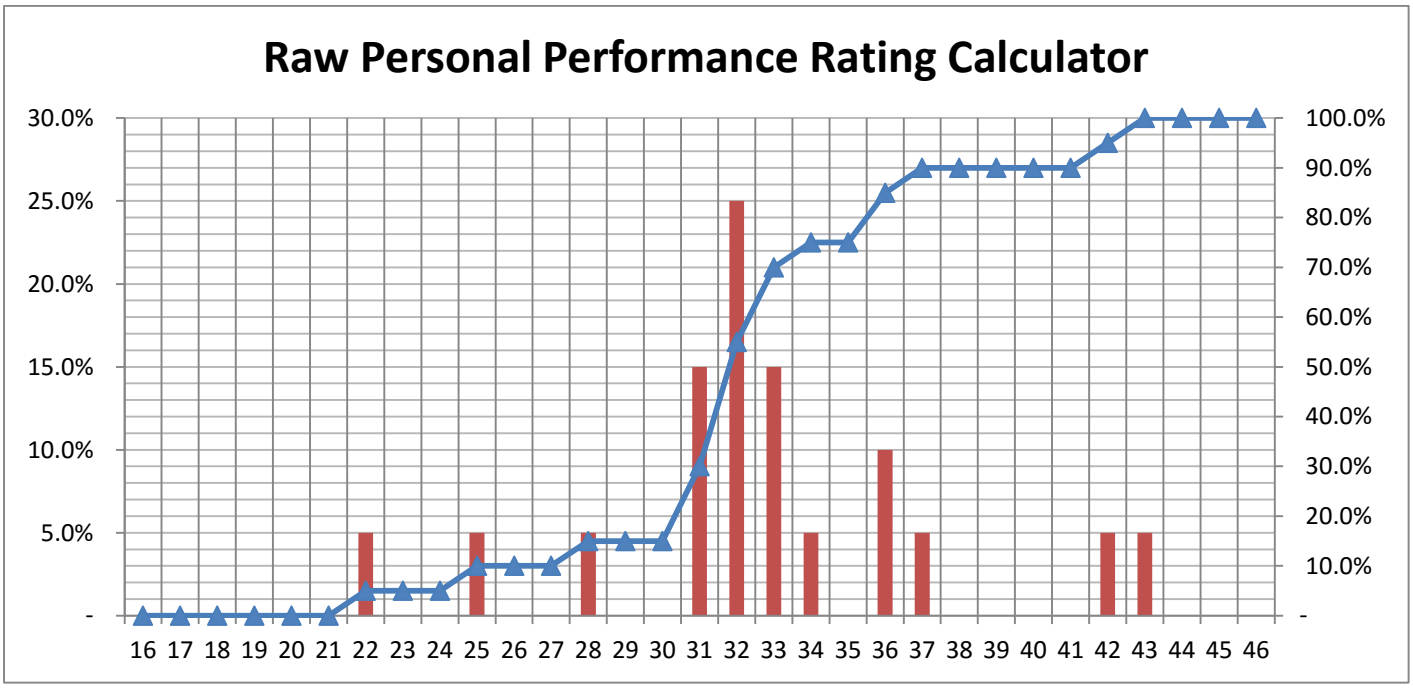

Distribution of Scores HC 45 - W
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
0.00%
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 55
Old
New

Appendix 4 – An "Equitable" Performance System

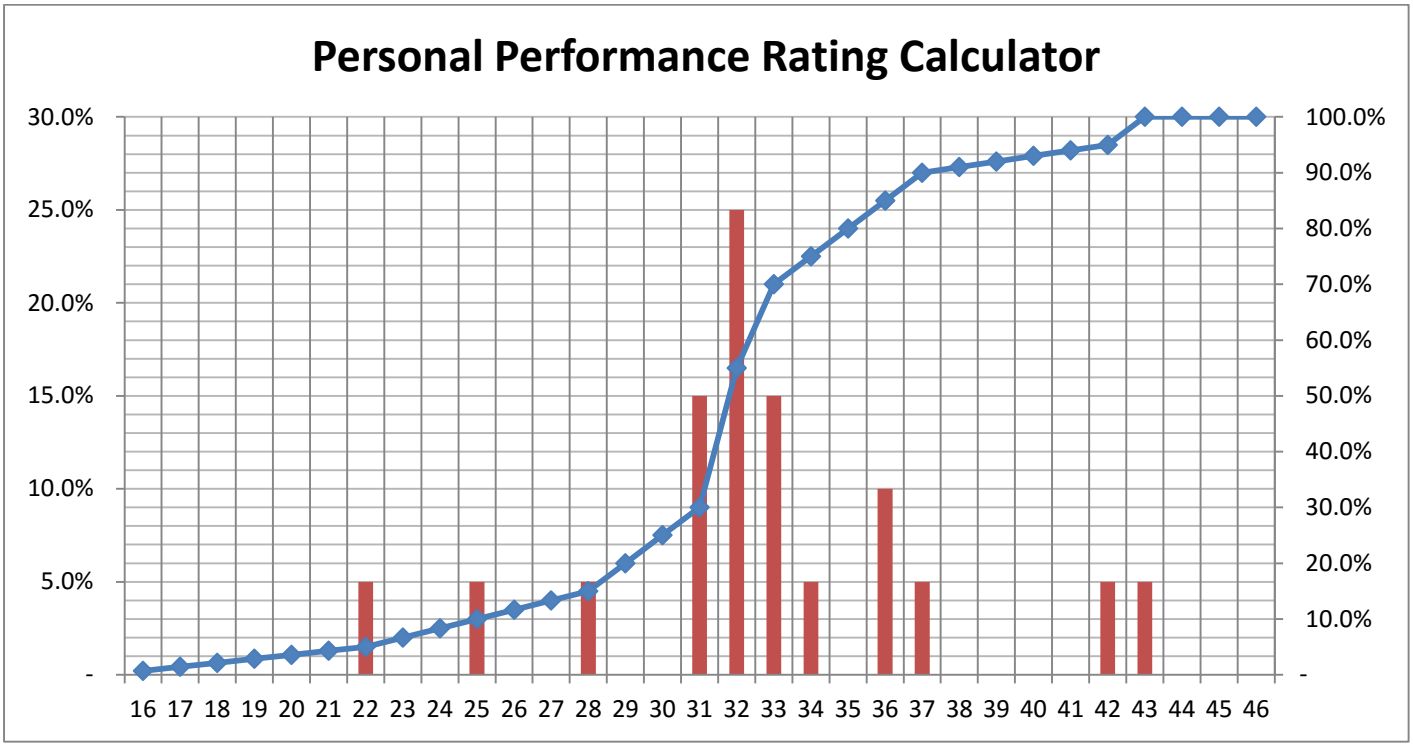
Suppose all players have available a record of their last 20 rounds of golf. This record could be used to establish a Personal Performance Distribution (PPD) of their performance. The best 10 scores would still be used to establish the Playing Handicap, with a 0.96 or the like Bonus for Excellence. This tends to normalise more recent form compared to less recent form, up to 20 rounds ago. It might look like this:



This can be converted to include a cumulative distribution which forms the basis of a Personal Performance Rating Calculator:



And finally the "jumps", eg between 37 and 42 can be smoothed:



Can we assume that in an Equitable Performance System all players have a close to equal chance of winning the competition, and that a player who performs at the 99th percentile of his PPD has performed better than one who performs at, say the 90th percentile? Suppose we call the percentile of performance a player's Personal Performance Rating (PPR).

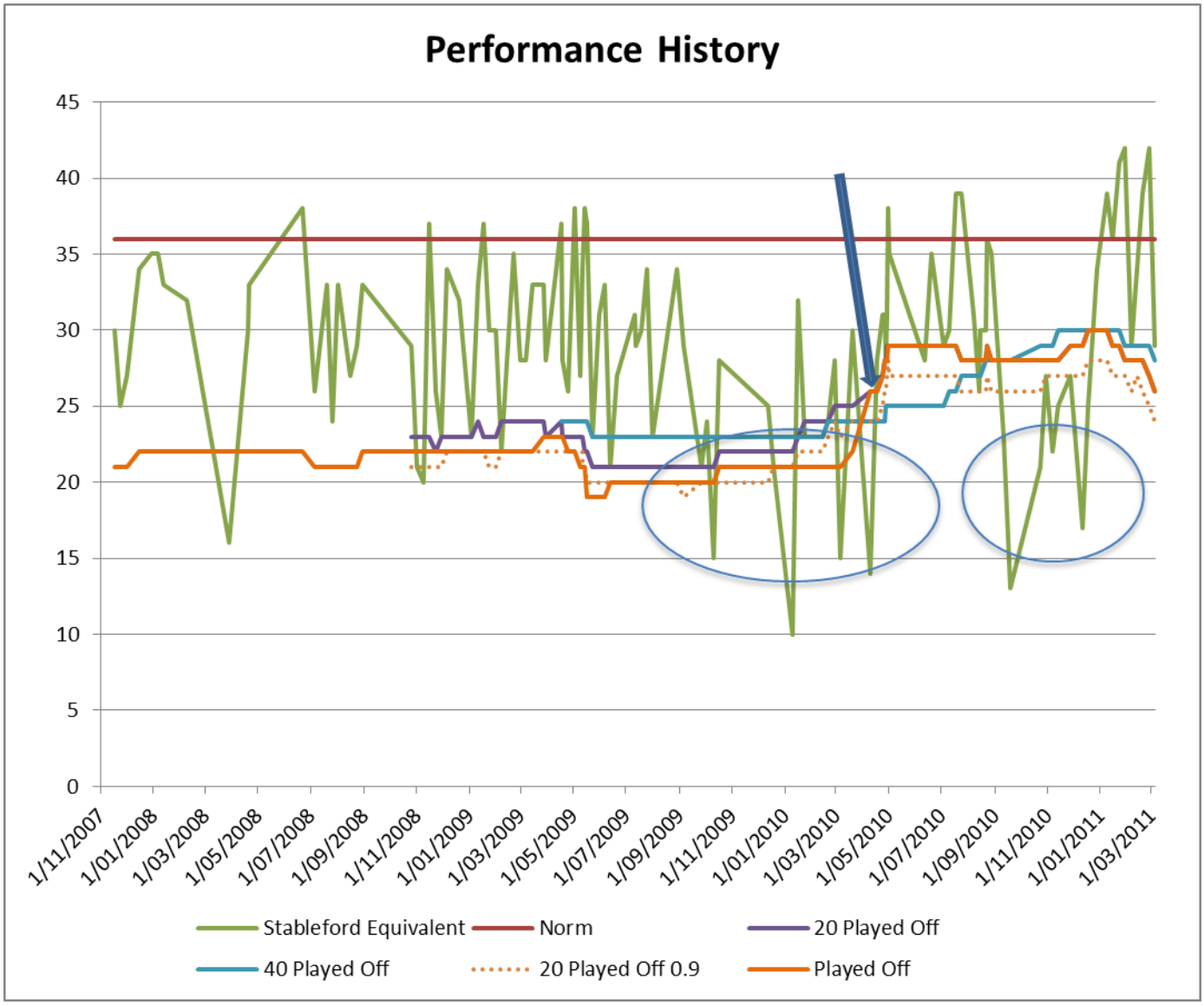
Using the PPR Calculator above, if this player scores 30 he or she would have a PPR of 25, if they score 36 they would have a PPR of 85, and 43 and over would be 100. One can see that 100 is the equivalent of a Personal Best.

The PPR rather than the score would then be used to rank the field to determine the winner and the prize-winners.

Appendix 5 – One Golfer's Tale

Consider a golfer who would consider himself to realistically be a 22-24 handicapper but believes he went through a horror patch and is now starting to come out the other side:

Here is his performance since 2007:



No wonder he considers himself to be a 22-24 handicapper – he played in this range from November 2007 until the new handicap system was introduced in 2010. In fact he went through two "horror patches": the first from September 2009 through April 2010, and the second shorter one from September 2010 through mid-December 2010.

During the first horror patch, he received little handicap relief (the Orange line is what he actually played off) until one magic day, 10/4/2010, the New handicap system recognised the recent performance and he jumped four strokes to 26. After a few more rounds he was off 29. At this level he had two rounds recording 39 points. His previous best scores had been 38 in 2008 and 2009.

The second horror patch did not change his handicap much, only one stroke, to 30, at which point he scored another 39. His handicap dropped one stroke and he scored 41. It dropped another stroke to 28 and he scored 42. Another 39, another stroke lost, and then another 42. Last round was 29 off 27.

There is no doubt that he has recently scored more rounds in the 40s than ever before. And no doubt that if the Old system had remained in place he would still be playing off close to 22, and not getting scores of 41 or 42.

So what if the New system had been in place all along? This is the purple line above. You can see that it would have responded to the first horror patch sooner, and taking him up through 24, 25 and 26 sooner.

What if we had 40 rounds in the calculation, not 20? This is the cyan line. A 40 sample moves more slowly than a twenty sample, and so the cyan line lags in responding to the horror patches compared to the purple twenty sample. The result is that if we were on a 40 sample, he would have been playing of a higher handicap (29 v 28) when he got the first 42 turning it into a 43, and playing off two strokes higher (29 v 27) when he got the second 42, turning it into a 44.

What if we were using the twenty sample, but with a 0.9 bonus for excellence rather than 0.96? This is the dotted orange line. It is one to two strokes below the actual Played to after the New system was introduce, and one to two strokes below the purple line. Obviously it would cause the high scores of 41 and 42 to drop by one to two strokes.

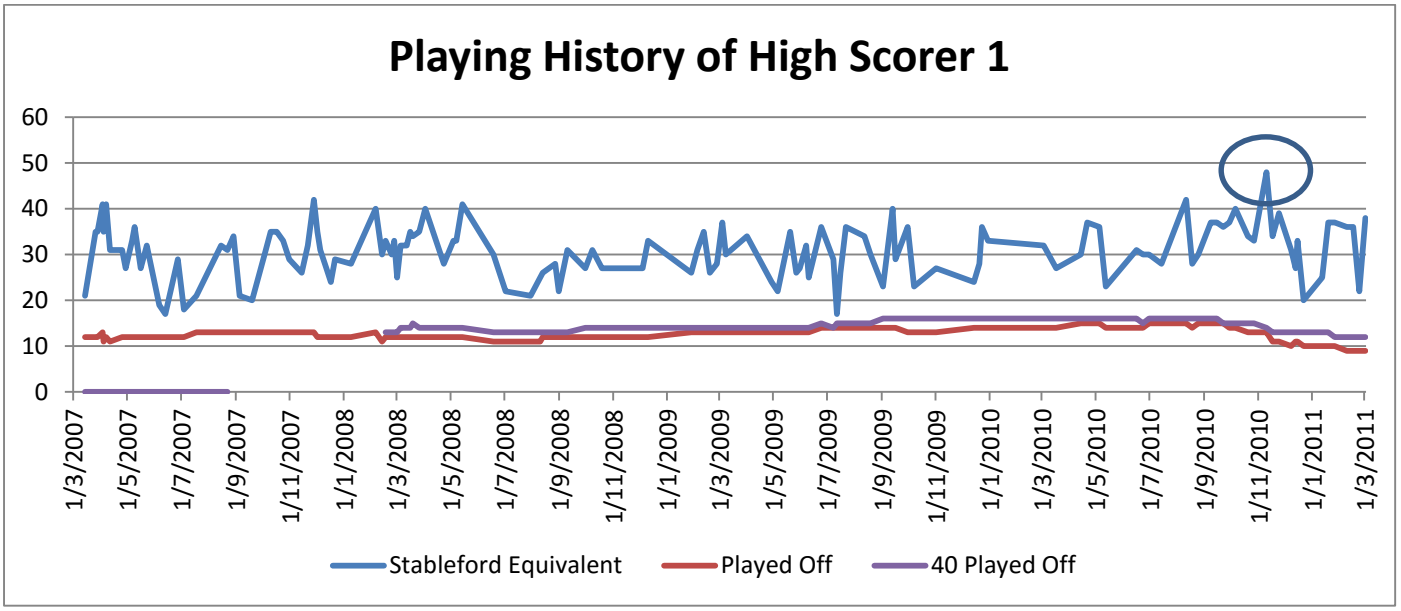
What do we conclude from this? The reason that the player concludes that he is a 20 to 22 player, is that the Old system was not very responsive to deterioration of form. Had the new system been in force all along, he would no doubt have considered himself a 22 to 24 player, all the way through January 2010. At this time, the deterioration of form, which really started some six months prior, would have been recognised in the handicap and it would have eased out to 29. Playing off this level through December 2010, he was more or less competitive – no more or less than in the past.

The fabulous return to form in January and February 2011 has been recognised by the system, and he is back to the 27 which is a reflection of his current average. An average which is clearly less consistent than it was in the past.

Appendix 6 – Three of the best

There were only five rounds of golf in the sample where the stableford equivalent score was 50 or more. Recall that the stableford equivalent score is already adjusted for Course Rating. Eg if a player has 50 points on a course with a par of 72, but a course rating of only 66, then the stableford equivalent score would be 44 points.

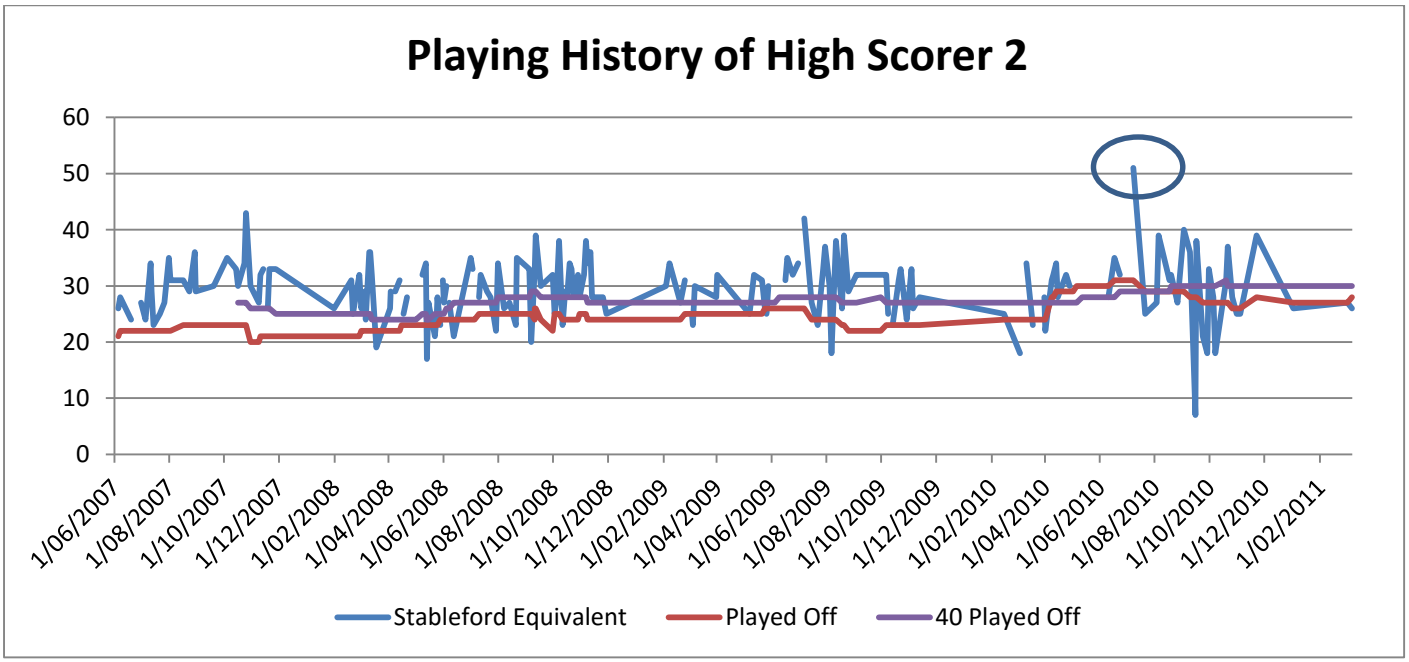
These five rounds were selected for further analysis. Firstly, two of them were achieved under the Old system, so they were clearly not "enabled" by the New system. Of the other three, their performance history is as follows:



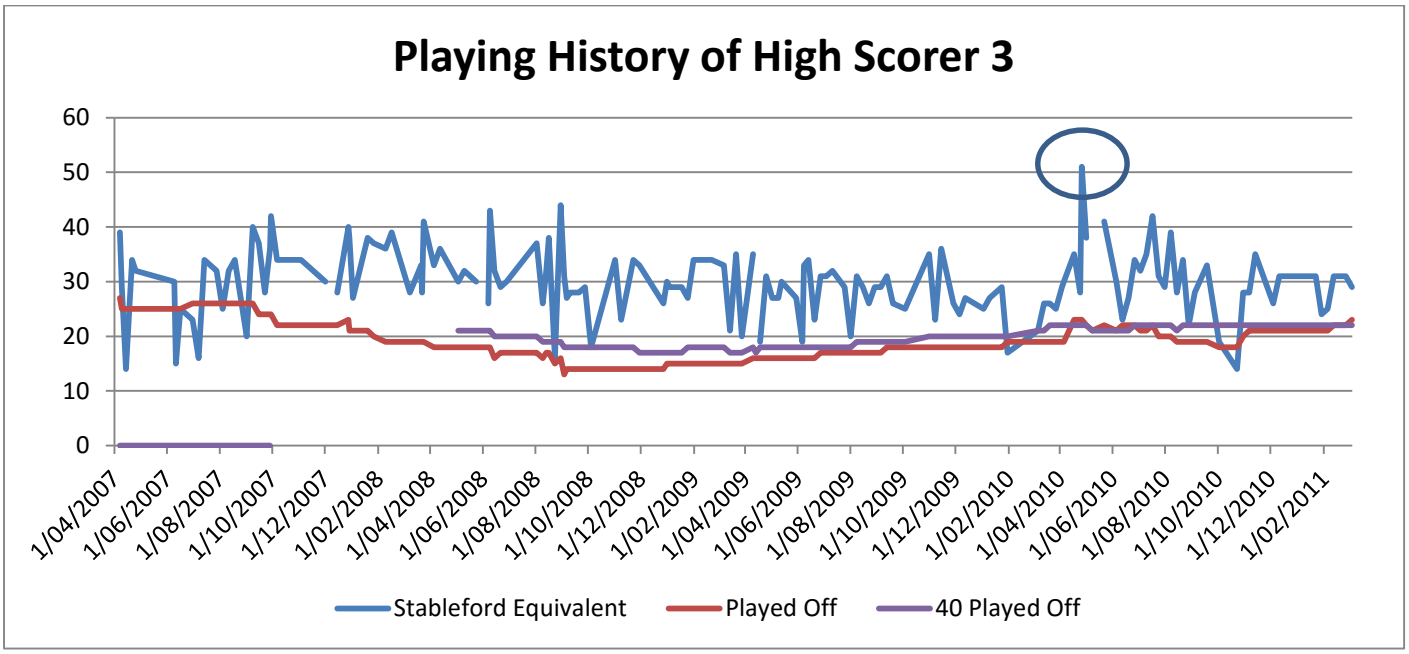
Here is a male player with a fairly consistent set of scores, and a fairly consistent handicap – both under the Old system, and also under the new system from 9th April 2010. On 10th November he achieved 50 points. On this day, his handicap had fallen back the 13 he had been playing off under the Old system.

Clearly the fabulous result has little to do with the handicap system, and a lot to do with everything going right on the day.

If the handicap system was based on the last forty rounds rather than twenty, his handicap would have been up to three strokes higher since that round. This is shown by the purple line.



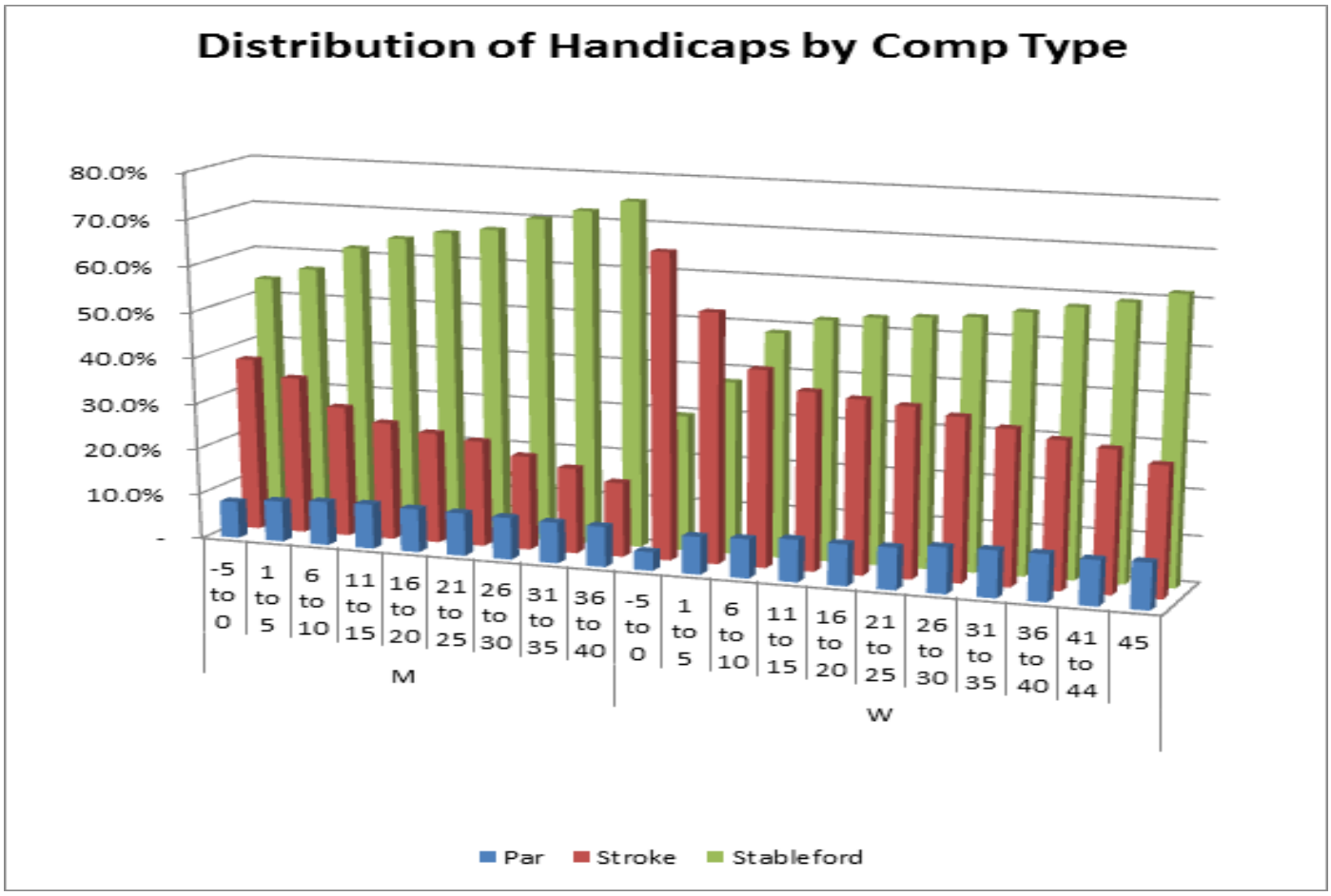
While it is true that this lady player's handicap rose after the New system was introduced, the history has a number of "No Score Not Approved" and No Score Approved" entries – the gaps in the chart above. The player's history is more volatile than the prior player, and the 51 appears not to be related to a slump in form (there were recent scores of 32 and 35) as a remarkable round. Again, the purple line shows what she would have played off if a forty sample rather than twenty was used – in this case she would have had a handicap two stroke less on the day she scored 51, but after that score the player would have a higher handicap still over the following four months. In fact the score of 51 only reduced the exact by 0.3 strokes, as this Played To of 16 replaced a Played To of 22, 6 points different, but only an impact of 0.3 in an average of 20 best rounds.



The high score here was achieved on 26/4/2010, shortly after the New system was introduced. This man had a four stroke increase in handicap with the New system. Had a sample size of forty instead of twenty been used, this would have been reduced by one stroke. But most subsequent rounds would have been one stroke or more higher in handicap with the forty sample than with the twenty. It seems that this player did have a magic round, and that the impact of handicap was minimal.

Interestingly, the three players, while not low markers, were far from being the highest markers. The two men were off 13 and 22 respectively, the lady was off 31.

Appendix 7 – Competition Type by Handicap



Golfers% Handicap Group	Par	Stroke	Stableford
M	**9.5%**	**24.5%**	**66.0%**
-5 to 0	8.0%	37.9%	54.1%
1 to 5	8.9%	34.4%	56.7%
6 to 10	9.5%	28.5%	61.9%
11 to 15	9.8%	25.7%	64.5%
16 to 20	9.5%	24.2%	66.3%
21 to 25	9.4%	23.1%	67.5%
26 to 30	9.2%	20.6%	70.3%
31 to 35	8.9%	18.6%	72.5%
36 to 40	8.8%	16.2%	75.0%
W	**9.7%**	**35.1%**	**55.2%**
-5 to 0	4.1%	66.3%	29.6%
1 to 5	8.2%	54.2%	37.6%
6 to 10	8.5%	42.7%	48.8%
11 to 15	9.2%	38.7%	52.1%
16 to 20	9.0%	37.7%	53.2%
21 to 25	9.2%	37.0%	53.9%
26 to 30	10.0%	35.5%	54.5%
31 to 35	10.2%	33.6%	56.2%
36 to 40	10.2%	32.0%	57.8%
41 to 44	9.8%	30.8%	59.4%
45	10.1%	28.2%	61.7%
Grand Total	**9.5%**	**26.6%**	**63.9%**

Appendix 8 – About the Author

GA commissioned a statistician with significant experience in developing practical solutions to conduct the data analysis. Michael Maher also has long-term peak-level management and board experience with major Australian and international companies. Michael comes into this with a fresh outlook and without any industry baggage. Michael plays regular club competition golf in Sydney. His current professional focus is FX trading using automated algorithms.