Table of Contents

2011 MFTHBA Show and Celebration Judges and Judge Evaluation Methods Compared *by Ken Kemp, Ph.D.*

Judge Evaluation Methods

Following are summaries of judge performances over 17 classes from the 2011 Show and Celebration that are the basis of a discussion of the six participating judges and two judge evaluation methods. One judge evaluation method focuses on placements 1 through 10 while the other focuses on the top 3 placements only. I have been using the top 3 placements method to determine which judges are the better judges for many years. As is outlined in the judge tutorial (http://www.kenkemp.net/judging/tutorial.htm), the amount of disagreement between a judge and his peers who judge the same classes together with him is quantified as a distance measure based on the difference in the ranking a judge provides for a class and the consensus ranking provided by the other 4 judges who judged the same class with him. As is pointed out in the judge tutorial a distance of less than 3 is associated with placements where the two rankings are in close agreement, namely that there is very good agreement as to which horses belong at the top of a class. The larger the distance, the more disagreement there is.

There are five reasons for concentrating on the top 3 positions:

1. The placement of the top of a class is more important than the placing of the rest of the class as it is these placements that receive the most attention and acclamation, they receive the more coveted ribbons, and when trophies or payback monies are awarded, they almost always go only to the top few places.

2. Judges will generally spend more time, which is especially limited in large classes, deciding how to rank the top end of a class rather than the middle or bottom primarily because there is not enough time to evaluate the subtle differences among all the horses. Thus, more judge effort and concentration goes into deciding both which horses deserve to be at the top of a class and the order in which they should be in, than the other positions.

3. Until there are judges who can consistently place the top rankings of classes similarly to one another, it makes little or no sense to be looking for judges who can place the top 10 positions consistently with their peers or a group that establishes what may be considered correct placements for classes. You have to walk before you can run. The 12 year summary results show that judges who agree closely with their peers on the top 3 places are not very common in all those years of data. Given that this is the case, there is no chance that there are judges who are consistent with each other in their top 10 choices and it would therefore be futile to try to find them. In fact any effort to do so actually makes identifying those who agree on the top 3 places a nearly impossible task as the results below show.

4. My experience has been that judges who do the best job of picking the top of a class are also generally the ones who do better at placing the rest of the class as well. If they do fail to get the bottom of a class

consistent with their peers, it is not nearly as serious a problem as not getting the top of a class correct, i.e. consistent with their peers, would be.

5. The major drawback to considering all 10 placements is that a judge who blows the placements at the top of the class has too many chances to redeem himself by getting some of the poorer performers near the bottom of the class or some of the mediocre performers near the middle of a class. This tends to make all of the judges' scores more similar and when comparisons are made between their scores it is not possible to tell if a judge's score is due to him placing the top of the class well, the middle of the class well, the bottom of the class well, or none of the class was placed well but he was right for a few intermingled choices. In my opinion, it is important to identify the judges who place the top of a class well and makes little sense to reward a judge who has trouble identifying the better horses in class for his ability, or perhaps just his good luck, in his placement of the mediocre or poorer horses in a class.

The second method of evaluation to be considered is the one adopted by the MFTHBA in 2010. It is a method that converts differences between two rankings into a percentage of error and then subtracts that from 100% to get the "percentage of agreement" as the score for quantifying a judge's performance. If a judge's placement is exactly the same as a placement to which his is compared, there would be no disagreement or error and the score in such a case would be 100%. The calculation of the score can be done as follows: Score = 100% - sum{ absval(judge rank(i) – i)} summed over i for i = 1 to 10 where rank(i) is the ranking of the ith position given by a judge. Below his placement is compared to the correct placing which would be 1, 2, 3,…,10. For example, if a judge placed a class as **5, 2, 6, 1, 7, 3, 4, 8, 9, 10** when the correct placing was 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, his score would be:

$$\text{Score} = 100 - \{ \ |\mathbf{5}\text{-}1| + |\mathbf{2}\text{-}2| + |\mathbf{6}\text{-}3| + |\mathbf{1}\text{-}4| + |\mathbf{7}\text{-}5| + |\mathbf{3}\text{-}6| + |\mathbf{4}\text{-}7| + |\mathbf{8}\text{-}8| + |\mathbf{9}\text{-}9| + |\mathbf{10}\text{-}10| \ \}$$
$$= 100\% - \{ \ 4 \ + \ 0 \ + \ 3 \ + \ 3 \ + \ 2 \ + \ 3 \ + \ 3 \ + \ 0 \ + \ 0 \ + \ 0 \ \}$$
$$= 100\% - 18\% = \mathbf{82} \ \%, \textit{where the vertical bars mean absolute values of the pairwise differences.}$$

Note that this example represents a very poor placement of a class where neither the top nor middle of the class was placed well. However, this score would be enough to rate a MFTHBA judge as a AAA judge, the highest rating they give, and such a judge would be considered qualified to judge the annual Show and Celebration as well as any other sanctioned shows. This example also shows what including all ten places in the calculation of a score can do to obliterate the fact that this example judge did not come close to finding the top of the class. A judge placing the top of a class correctly and failing to do well placing the rest of the class can also score 82, or less, by not placing the bottom of the class well, in which case there may be two judges both with scores of 82 where one ranked the top of the class well and the other did not. This insensitivity to such an important difference in performances is the main weakness of this method of scoring. Note in the above calculation it makes no difference what order the numbers in the summation are in. Any ordering where there are four 0's, four 3's, a 2, and a 4, or any other combination of differences that add to 18 for that matter, will produce a score of 82. This scoring method is based on the premise that it is just as important to get the middle and last place horses in the correct order as it is to get the first 3 places in the correct order. It is doubtful that many serious competitors would agree with this premise, in my opinion. I definitely disagree with it.

Another problem with the above method of scoring judge performance is that if the top 10 horses are included in any possible order in a judge's ranking, the minimum possible score is 50%, not zero. If a judge places a class completely backwards with the top horse last, the last horse first and all other horses similarly misplaced in between, his score will be 50%. Thus the actual scale for this metric is 50% to 100%, not 0 to 100% as we might expect a "percentage of agreement" measurement to be. Also, as I have shown via simulation results elsewhere (http://www.kenkemp.net/judging/MFTHBA_Judgeratings_Kemp.pdf), that the expected score for someone who places a class w/o even seeing the horses in doing so is about 66%, i.e. half of those picking

horses at random would score 66% or more. Given this is the case, it clearly is not too difficult to score 14 percentage points higher if one knows anything about Missouri Foxtrotting Horses. A score of 70% may not seem too low unless one realizes that the range of the MFT metric is just 50-100 and that half the scores of someone just guessing at placements would likely be 66 or above. If you understand that, it makes a score of 70% take on a different interpretation.

Results from the six 2011 MFTHBA Show and Celebration judges are presented below. These data allow for both an assessment of each judge's performance and for a comparison of the two judge evaluation methods based on actual judging results. The classes included are primarily championship and WGC performance classes. Only classes that had at least 10 horses are discussed and included in calculating the averages of the judges' scores. Small classes are considerably easier to judge than larger classes and tend to not discriminate between judges' judging abilities so some of the small classes are shown as a matter of possible interest to the reader but they were not included in the summarization statistics. The consensus placings of peer judges are the basis of comparisons for both evaluation methods. For a judge to score well, he must agree with the choices of his peer judges who judge the same class. This criterion is based on the assumption that competent judges who judge by the MFTHBA breed standard, who judge classes of well trained horses that are shown by competent riders, should closely agree with the way they place each class.

Discussion of Judge Performances for the 2011 MFTHBA Show and Celebration

Judge Gengelbach

The table below shows that Judge Gengelbach had a major problem with class 132 placing his peers' 15th place horse as his 3rd place choice. He had problems with two other classes also but placed 5 classes quite well. Please note in the two charts below how the 1-3 distance measure, based on the first 3 places, shows larger differences among his performances than do the scores from the MFTHBA scoring method. Note that according to the MFTHBA metric his performance in classes 134 and 148 indicate similar performances, scores of 75 and 74, respectively, even though he picked a 10th place horse 3rd in class 134, while the 1-3 distance indicates that the performance in class 134 was considerably worse than that in class 148, with distances of 7.7 and 4.5, respectively. Neither class was placed well. Looking at the last 4 classes, note that the scores are almost all the same according to the MFT metric while the 1-3 distance picked up on the fact that he did better at picking the top 3 horses in some of those classes than others, especially in class 64. A 1-3 distance value less than 3 indicates close agreement with the peer judges' ranking of the top of a class. Note that classes with green data points on the 1-3 distance chart shows 5 classes where this judge agreed with his peer judges as to which entrees belonged at the top of the respective classes. These were classes 150, 153, 137, 64, and 62B. The red data points indicate the classes where he was not in harmony, to put it nicely, with the choices made by his peers. Note class 150 where he scored 91 and class 137 where he scored 96 for placing those entire classes very well. Keep these examples in mind as you peruse the results for the other judges that follow because there are few classes where a score of 90 or higher is attained. The results that follow provide support for my assertion that close agreement between class placements is quite unusual when all 10 positions are considered.

Class	132	134	139	148	150	151	152	153	137	109	108	64	62b	62	Avg
	1	4	1	5	1	4	1	1	1	2	2	2	3	2	
	4	1	3	4	2	2	2	3	2	4	1	1	1	4	
	15	10	2	3	3	1	4	5	3	5	3	3	4	1	
	3	2	6	11	4	8	5	4	4	1	5	5	2	5	
	2	7	5	1	11	7	3	6	5	7	6	6	6	3	
	6	6	4	2	5	3	7	7	6	3	7	4	7	8	
	16	3	8	9	8	5	6	2	7	8	9	13	5	7	
	5	5	7	8	7	9	8	9	9	6	8	8	9	6	
	8	11	9	6	9	6	9	10	10	9	4	11	10	10	
	11	9		10	10	10		8	8	10		10	8	9	
MFT Score	68	75		74	91	80		86	96	84		86	86	86	82.9
1-3 Dist.	12.2	7.7		4.5	0.0	3.6		2.2	0.0	3.0		1.4	2.5	3.0	3.64
Class Count	1	2		3	4	5		6	7	8		9	10	11	

Judge Gengelbach(G)
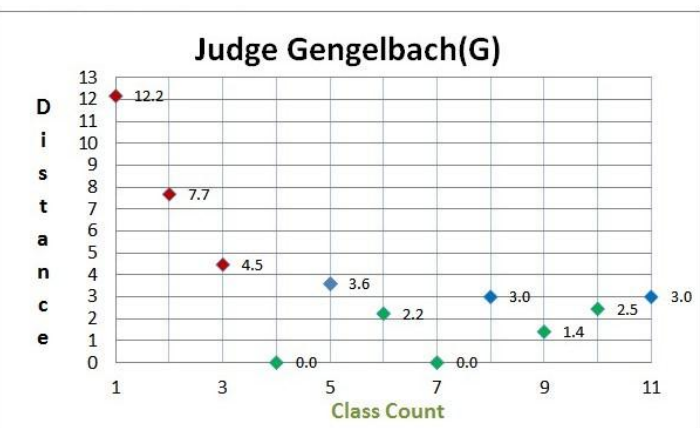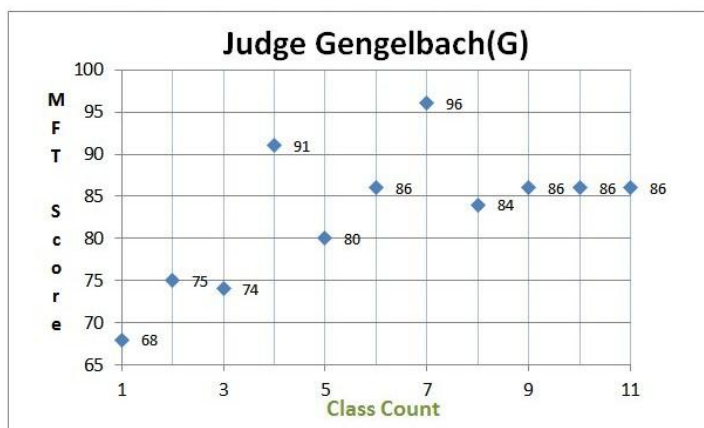
Judge Gengelbach(G)

Summary Table Highlighting Gengelbach

Judge	Placed an 8,9,10 as 1, 2, or 3	Placed a 1,2,3 as 8, 9, or 10	Was the only one to place particular horse	Average of 1st two
D	0.00	0.00	0.00	0.00
E	0.08	0.08	0.08	0.08
G	0.18	0.00	0.09	0.09
O	0.00	0.10	0.00	0.05
R	0.23	0.15	0.23	0.19
T	0.00	0.15	0.38	0.08

	1-10	1-5	1-3	Variance	# classes With 10	MFTHBA Avg Score
Judge-D:	6.37	3.93	2.46	2.538	11	84.5
Judge-E:	7.94	5.65	3.39	4.744	12	80.8
Judge-G:	7.38	5.27	3.64	12.537	11	82.9
Judge-O:	7.88	4.54	3.41	2.092	10	80.9
Judge-R:	8.22	5.37	3.90	5.719	13	80.1
Judge-T:	8.25	5.81	3.57	3.138	13	79.8
Averages	7.67	5.09	3.39			81.5

Judge Elmer

Judge Elmer did well in half of the classes he judged according to the 1-3 distance measure. He completely blew classes 104, 134, and 109 according to the same performance criterion. In 104 and 109 he placed his peers' 7th place horses first and in class 109 he placed their 8th place horse 2nd and their 6th place horse 3rd. Note that the MFT score for class 134 completely missed the fact that the class was poorly judged as his score was 84 which is 3 points higher than his average MFT score (80.8). The reason this happened was because in this class he placed the 4th, 6th, 8th, and 10th place horses correctly. These intermingled matching placements illustrate the point made earlier about how major mistakes in placing the top of a class are often compensated for by getting other placements correct, or nearly correct, further down in a class. Note also that his lowest MFT score came from a class where he did a fair job at finding the top of class 132 (4, 1, 3) but he was penalized because he did a poor job of finding the horses at the lower positions. He got a lower score for doing that than he did in class 109 where he placed the 8th horse 2nd, the 6th horse 3rd, and the 1st horse 6th, and the 3rd horse 8th. This happened because the lower end placements in class 109 were closer to being correct than was the case in class 132. In class 109 the poor choices for the top of the class were compensated for by reasonably close choices at the bottom of the class. We have to ask if that makes sense when we are trying to identify judges with outstanding ability.

Class No.	104	135	132	134	139	149	150	151	152	153	109	108	64	62b	62	Avg
	7	1	4	7	2	1	2	1	1	5	2	2	3	3	1	
	3	2	1	3	1	3	1	3	2	1	8	1	2	1	4	
	1	3	3	1	3	5	4	4	3	4	6	6	6	4	2	
	9	6	15	4	6	4	8	8	7	2	7	3	1	2	3	
	5	4	6	2	5	2	3	2	4	7	4	7	10	6	5	
	6	7	2	6	4	6	9	9	5	6	1	8	5	9	7	
	2	9	16	5	7	7	5	6	8	3	5	9	13	5	8	
	8	5	11	8	8	8	6	7	6	9	9	5	11	8	6	
	11	10	12	11	9	9	10	5	9	8	3	4	4	10	9	
	4	8	8	10		10	11	10		10	10		9	7	10	
MFT Score	73	88	63	84		94	82	82		84	72		71	84	92	80.8
1-3 Dist.	6.4	0.0	3.2	6.4		2.2	1.7	1.4		4.2	6.8		3.6	2.4	2.2	3.4
Class count	1	2	3	4		5	6	7		8	9		10	11	12	

Judge Elmer — MFT Score vs Class Count

Data points: 73, 88, 94, 84, 82, 82, 84, 84, 72, 71, 63, 92

Judge Elmer — 1-3 Distance vs Class Count

Data points: 6.4, 6.4, 6.8, 4.2, 3.6, 3.2, 2.2, 2.4, 2.2, 1.7, 1.4, 0.0

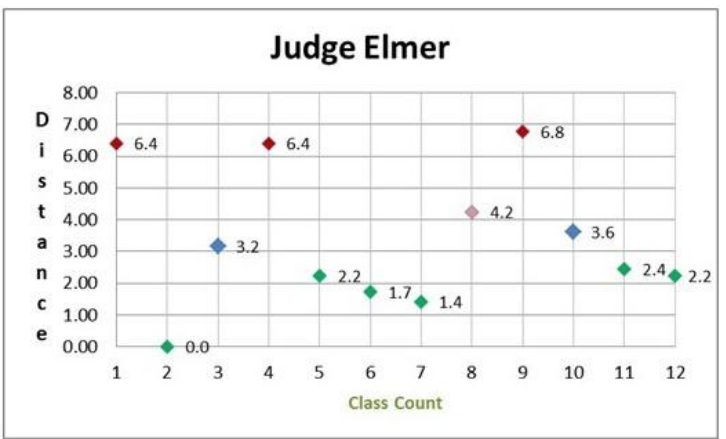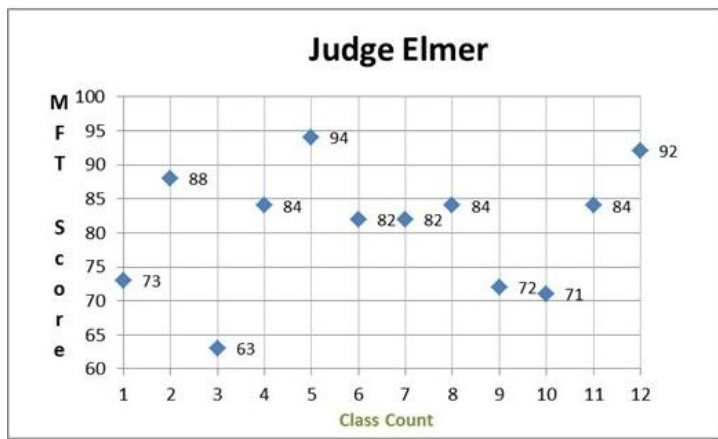| Judge	Placed an 8,9,10 as 1, 2, or 3	Placed a 1,2,3 as 8, 9, or 10	Was the only one to place particular horse	Average of 1st two
D	0.00	0.00	0.00	0.00
E	0.08	0.08	0.08	0.08
G	0.18	0.00	0.09	0.09
O	0.00	0.10	0.00	0.05
R	0.23	0.15	0.23	0.19
T	0.00	0.15	0.38	0.08

	1-10	1-5	1-3	Variance	# classes With 10	MFTHBA Avg Score
Judge-D:	6.37	3.93	2.46	2.538	11	84.5
Judge-E:	**7.94**	**5.65**	**3.39**	**4.744**	**12**	**80.8**
Judge-G:	7.38	5.27	3.64	12.537	11	82.9
Judge-O:	7.88	4.54	3.41	2.092	10	80.9
Judge-R:	8.22	5.37	3.90	5.719	13	80.1
Judge-T:	8.25	5.81	3.57	3.138	13	79.8
Averages	**7.67**	**5.09**	**3.39**			**81.5**

Judge Green

Judge Green seemed to have the most problems. The 1-3 distance chart below shows he did well placing 5 out of 13 classes. The MFT chart shows he scored lower than the AAA cut off value in 6 of the 13 classes he judged. Note that class 132 with a 1, 5, 3 top placement had an MFT score considerably lower than class 109 which had a 7, 1, 8 top 3 placement, 70 versus 78, respectively, because of the way the bottom of the two classes were placed. It would have taken very little to change his score of 78 in class 109 to a score of 80, one slight change was all that was needed, and it would have been good enough for it to be rated as a AAA judge performance for this class. This is a scary thought for anyone who shows, in my opinion. He placed an 8th or lower placed horse in his top 3 in 3 classes and placed a 3rd or higher placed horse 8th or lower in 2 classes having done both in class 109. He also had the highest average 1-3 distance, 3.90, and the 2nd lowest average MFT score, 80.1.

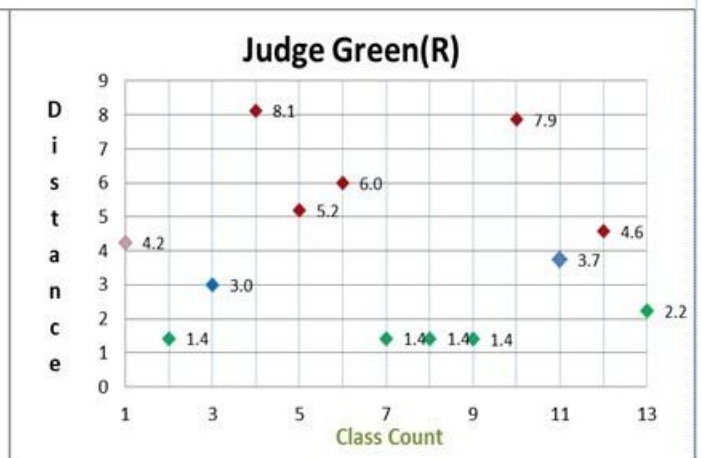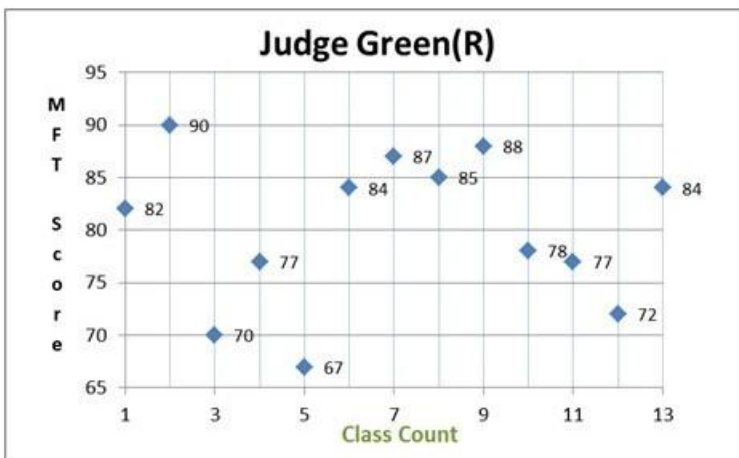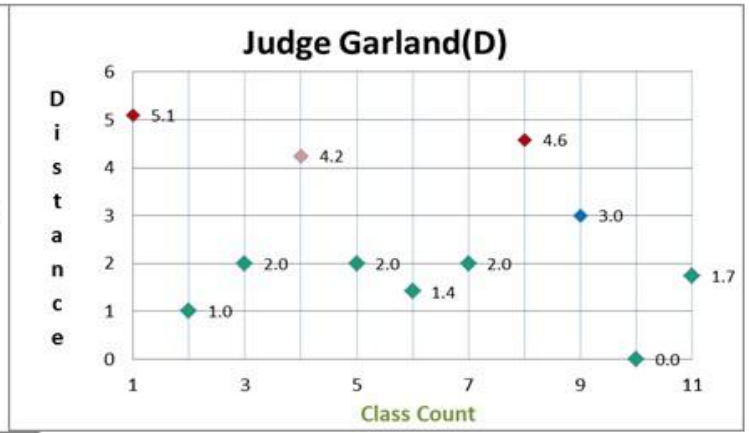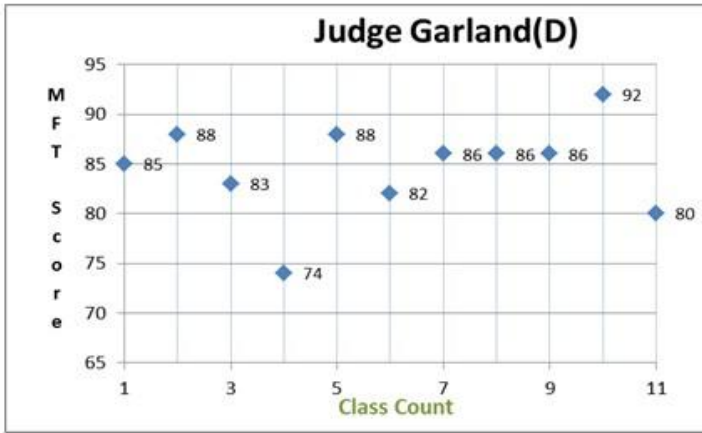| Class No.	104	135	132	134	148	149	150	152	153	137	109	64	62b	62	Avg
	1	2	1	2	4	1	2	1	1	1	7	3	3	1	
	5	1	5	1	5	8	1	2	3	3	1	1	6	4	
	6	3	3	11	6	3	3	3	4	2	8	6	4	2	
	2	6	4	7	11	6	4	5	7	7	2	2	1	3	
	7	5	2	3	1	4	5	8	2	6	5	5	7	10	
	4	4	16	4	2	5	11	7	5	5	4	4	9	5	
	3	7	6	6	3	2	7	6	9	4	6	13	5	8	
	9	10	12	8	9	7	6	4	6	8	3	11	10	7	
	8	8	15	9	8	9	8	9	10	10	9	10	2	6	
	10	11	7	5	7	10	13		11	9	10	7	8	9	
MFT Score	82	90	70	77	67	84	87		85	88	78	77	72	84	**80.1**
1-3 Dist.	4.2	1.4	3.0	8.1	5.2	6.0	1.4		4.2	1.4	6.8	3.6	2.4	2.2	**3.90**
Class Count	1	2	3	4	5	6	7		8	9	10	11	12	13	

Judge	Placed an 8,9,10 as 1, 2, or 3	Placed a 1,2,3 as 8, 9, or 10	Was the only one to place particular horse	Average of 1st two
D	0.00	0.00	0.00	0.00
E	0.08	0.08	0.08	0.08
G	0.18	0.00	0.09	0.09
O	0.00	0.10	0.00	0.05
R	0.23	0.15	0.23	0.19
T	0.00	0.15	0.38	0.08

	1-10	1-5	1-3	Variance	# classes With 10	MFTHBA Avg Score
Judge-D:	6.37	3.93	2.46	2.538	11	84.5
Judge-E:	7.94	5.65	3.39	4.744	12	80.8
Judge-G:	7.38	5.27	3.64	12.537	11	82.9
Judge-O:	7.88	4.54	3.41	2.092	10	80.9
Judge-R:	8.22	5.37	3.90	5.719	13	80.1
Judge-T:	8.25	5.81	3.57	3.138	13	79.8
Averages	7.67	5.09	3.39			81.5

Judge Garland

Judge Garland did an outstanding job of agreeing with his peers. His 1-3 distance was below 3 in 7 of the 11 classes he judged with 10 or more entrees. He had the smallest average distance and the highest MFT score. Note that he had a problem with the top of class 104 where he placed a 7th place horse as his 2nd choice. The 1-3 distance was 5.1, a large value, while the mft score was 85 indicating a well-placed class. According to the MFT criterion he placed the class well but according the 1-3 distance it was his worst performance. Overall he did an outstanding job as is reflected in the charts and tables given below.

Class No.	104	135	134	139	148	149	150	151	152	137	109	108	62b	62	Avg
	1	1	1	1	4	1	1	1	2	1	2	2	1	2	
	7	2	2	2	5	4	3	4	1	5	6	1	2	3	
	2	4	5	3	3	3	4	3	3	3	1	8	3	4	
	5	6	3	6	1	5	2	2	5	2	4	3	5	1	
	3	3	6	5	2	6	10	7	4	7	5	6	6	8	
	4	5	7	7	9	8	5	6	7	4	8	5	7	10	
	6	8	11	4	10	2	7	9	6	6	3	9	4	7	
	9	10	10	8	11	7	11	8	8	9	7	4	9	6	
	8	9	4	9	6	9	6	10	9	10	9	7	8	5	
	11	7	9		8	10	8	5		8	10		10	9	
MFT Score	85	88	83		74	88	82	86		86	86		92	80	84.5
1-3 Dist.	5.1	1.0	2.0		4.2	2.0	1.4	2.0		3.0	4.6		0.0	1.7	2.46
Class Count	1	2	3		4	5	6	7		8	9		10	11	

Judge Garland(D)

MFT Score vs Class Count

Class Count	MFT Score
1	85
2	88
3	83
4	74
5	88
6	82
7	86
8	86
9	86
10	92
11	80

Judge Garland(D)

Distance vs Class Count

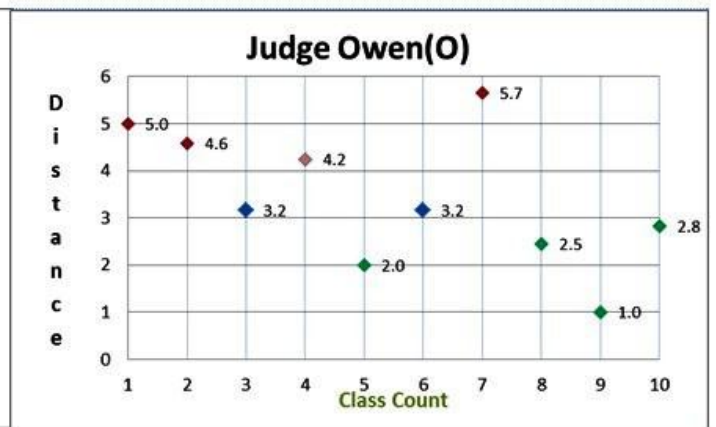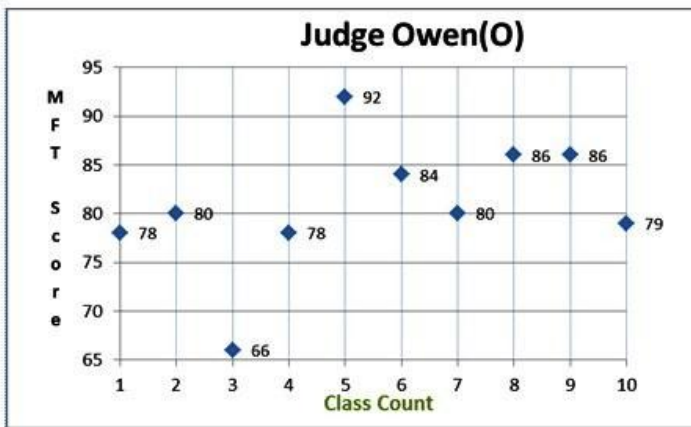| Class Count	Distance
1	5.1
2	1.0
3	2.0
4	4.2
5	2.0
6	1.4
7	2.0
8	4.6
9	3.0
10	0.0
11	1.7

Judge	Placed an 8,9,10 as 1, 2, or 3	Placed a 1,2,3 as 8, 9, or 10	Was the only one to place particular horse	Average of 1st two
D	0.00	0.00	0.00	0.00
E	0.08	0.08	0.08	0.08
G	0.18	0.00	0.09	0.09
O	0.00	0.10	0.00	0.05
R	0.23	0.15	0.23	0.19
T	0.00	0.15	0.38	0.08

	1-10	1-5	1-3	Variance	# classes With 10	MFTHBA Avg Score
Judge-D:	6.37	3.93	2.46	2.538	11	84.5
Judge-E:	7.94	5.65	3.39	4.744	12	80.8
Judge-G:	7.38	5.27	3.64	12.537	11	82.9
Judge-O:	7.88	4.54	3.41	2.092	10	80.9
Judge-R:	8.22	5.37	3.90	5.719	13	80.1
Judge-T:	8.25	5.81	3.57	3.138	13	79.8
Averages	7.67	5.09	3.39			81.5

Judge Owen

Judge Owen had a similar situation with class 132 as did judge Elmer where he did a fair job at placing the top of the class but was penalized by the MFT scoring method for not placing the bottom of the class well, producing his lowest MFT score. Note that he did a better job of placing the top of class 109 than class 137 but the MFT scores are the same for the two classes due to a better placement of the bottom of class 137. Also note that for class 153 the top 3 placings are 1, 6, 7, yet it meets the qualifying criterion for being a AAA MFTHBA judge performance. This happened primarily because the 8th and 10th horses were included among the bottom 3 placements. The 1-3 distance identifies class 153 as the worst placement he turned in among his classes because placing the 6th and 7th place horses in the top 3 is an indication he was lost in this class. It is hard to see how the placement of this class could be construed as evidence of outstanding judging ability.

Class No.	104	135	132	139	148	149	151	152	153	137	109	108	64	Avg
	1	3	4	3	5	1	4	1	1	3	1	1	3	
	5	6	1	6	3	2	1	2	6	1	2	2	2	
	7	2	3	1	2	5	3	3	7	2	4	7	1	
	2	1	5	2	4	4	7	8	2	7	5	6	6	
	4	5	2	7	10	7	6	5	4	4	7	5	4	
	6	4	15	5	1	6	2	4	3	8	8	8	5	
	9	10	11	4	9	3	8	7	9	5	3	9	13	
	8	9	16	8	8	8	5	6	5	6	9	3	11	
	12	7	6	9	6	9	9	9	8	9	6	4	10	
	3	8	8		11	10	10		10	10	10		7	
MFT Score	78	80	66		78	92	84		80	86	86		79	80.9
1-3 Dist.	5.0	4.6	3.2		4.2	2.0	3.2		5.7	2.5	1.0		2.8	3.41
Class Count	1	2	3		4	5	6		7	8	9		10	

Judge Owen(O) — MFT Score vs Class Count
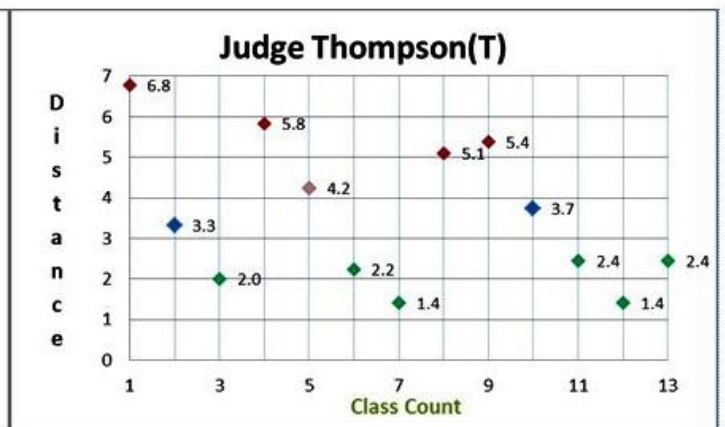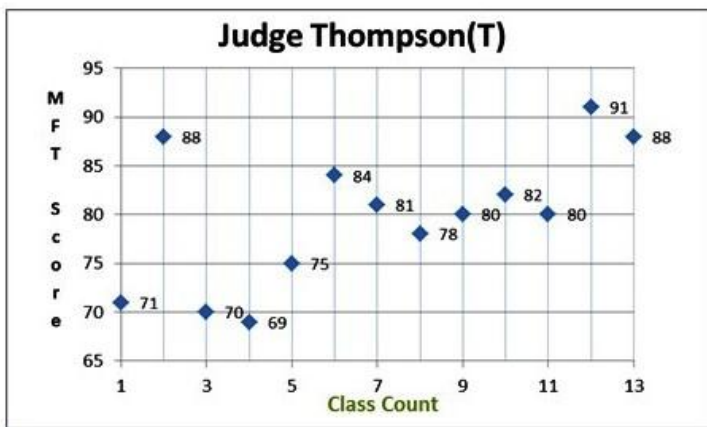

Judge Owen(O) — 1-3 Distance vs Class Count

Judge	Placed an 8,9,10 as 1, 2, or 3	Placed a 1,2,3 as 8, 9, or 10	Was the only one to place particular horse	Average of 1st two
D	0.00	0.00	0.00	0.00
E	0.08	0.08	0.08	0.08
G	0.18	0.00	0.09	0.09
O	**0.00**	**0.10**	**0.00**	**0.05**
R	0.23	0.15	0.23	0.19
T	0.00	0.15	0.38	0.08

	1-10	1-5	1-3	Variance	# classes With 10	MFTHBA Avg Score
Judge-D:	6.37	3.93	2.46	2.538	11	84.5
Judge-E:	7.94	5.65	3.39	4.744	12	80.8
Judge-G:	7.38	5.27	3.64	12.537	11	82.9
Judge-O:	**7.88**	**4.54**	**3.41**	**2.092**	**10**	**80.9**
Judge-R:	8.22	5.37	3.90	5.719	13	80.1
Judge-T:	8.25	5.81	3.57	3.138	13	79.8
Averages	**7.67**	**5.09**	**3.39**			**81.5**

Judge Thompson

Judge Thompson performed well in 6 of 13 classes according to the 1-3 distance chart but he disagreed with his peers on which horses were among the best in classes 104, 134, 148, 151, and 153. Placements 2,1,6 and 1, 7, 5 as the top horses were both good enough to meet the MFTHBA AAA judge performance cut-off in classes 135 and 153, respectively, while the ranking of 3,1,2 in class 64 barely made it because the bottom of the class was not placed well. Judge Thompson had the lowest MFT average score (79.8) and the second highest 1-3 average distance (3.57).

Class No.	104	135	132	134	139	148	149	150	151	153	137	108	64	62b	62	Avg
	7	2	1	1	1	4	1	1	4	1	3	7	3	1	2	
	5	1	4	7	6	5	4	3	1	7	1	4	1	3	3	
	2	6	3	6	3	3	2	2	7	5	6	6	2	2	1	
	1	3	16	2	5	7	5	11	2	3	7	2	5	4	8	
	6	5	2	11	7	1	8	5	6	6	2	9	6	5	4	
	9	7	13	5	4	2	3	4	8	2	4	5	4	9	5	
	3	4	5	9	8	11	6	8	5	9	8	3	13	6	6	
	4	8	9	8	2	9	9	13	3	8	5	1	12	7	7	
	11	10	10	3	9	10	10	7	10	4	9	8	11	8	9	
	12	9	8	4		8	7	10	9	10	10		10	11	10	
MFT Score	71	88	70	69		75	84	81	78	80	82		80	91	88	79.8
1-3 Dist.	6.8	3.3	2.0	5.8		4.2	2.2	1.4	5.1	5.4	3.74		2.4	1.4	2.4	3.57
Class Count	1	2	3	4		5	6	7	8	9	10		11	12	13	

```
            Placed an 8,9,10   Placed a 1,2,3   Was the only one to    Average of
             as 1, 2, or 3     as 8, 9, or 10   place particular horse  1ˢᵗ two
Judge
   D            0.00              0.00               0.00                 0.00
   E            0.08              0.08               0.08                 0.08
   G            0.18              0.00               0.09                 0.09
   O            0.00              0.10               0.00                 0.05
   R            0.23              0.15               0.23                 0.19
   T            0.00              0.15               0.38                 0.08


             1-10     1-5     1-3    Variance    # classes       MFTHBA
                                                 With 10        Avg Score
   Judge-D:  6.37    3.93    2.46     2.538         11             84.5
   Judge-E:  7.94    5.65    3.39     4.744         12             80.8
   Judge-G:  7.38    5.27    3.64    12.537         11             82.9
   Judge-O:  7.88    4.54    3.41     2.092         10             80.9
   Judge-R:  8.22    5.37    3.90     5.719         13             80.1
   Judge-T:  8.25    5.81    3.57     3.138         13             79.8
   Averages  7.67    5.09    3.39                                  81.5
```

Summary

Judges

The 1-3 distance measure shows only judge Garland having an average 1-3 distance less than 3. He clearly did the best job of being in agreement with his peers. Judge Garland also had the highest MFT average score of 84.5. Judges Elmer and Owen had similar 1-3 distance average values but judge Owen had a smaller 1-5 average distance, 4.54, which indicates he did a better job at placing the top 5 places than did judge Elmer. The other three judges had fairly large average 1-3 distances indicating that they frequently differed with the consensus placings of their peer judges.

Methods

The MFT average scores show that judge Garland did the best job placing all 10 places (84.5) and that agrees with his 1-10 average distance result which shows the same thing, that value being 6.37 is considerably smaller than the average of all judges which is 7.76. The 1-10 distance measure is comparable to the MFT score in that it is based on all 10 placements. However, note that there is relatively more separation in the 1-10 distance values ranging from 6.37 to 8.35 with an average of 7.67 than there is in the average MFT scores that range from 79.8 to 84.5 and average 81.5. Of course both methods suffer from the short comings associated with scoring all 10 places. If you review the results of all 6 judges you will see that most of the MFT scores were in the range of the low

to mid 80s over a wide range of performances from really bad to pretty good. Four judges had one class each that scored 90 or higher and two judges each had two. Those are the only classes where based on the MFT score we can be sure the judge placed a class well without first looking at the placement itself. There were only 8 such scores among all the scores computed. Scores below 80 are indicative of poor placements but scores in the 80s convey little information about how well a class was placed.

The primary goal of a judge evaluation procedure should be to discriminate between good judges and poor judges. The foregoing results show that the 1-3 distance measure is very helpful in identifying which classes were judged well according to the degree of agreement between a judge and his peer judges' consensus placement. In cases where the 1-3 distance is less than 3, the agreement at the top of a class is good and for larger distances it is not. There is no such interpretation to be made using the MFT scoring method. The method is too insensitive to differences in placements to be useful in identifying judges who place classes well from those who place classes poorly. The similarity of 4 of the 6 MFT judges' average scores, 79.8 to 80.9, indicates how little discrimination among judges resulted from scoring them over several classes.

The MFT scoring method is currently used to qualify MFTHBA AAA judges. It was evident from the beginning (see MFTHBA Rating) that the procedure was not well thought out before it was implemented. The qualifying score of 80% was chosen arbitrarily rather than having been based on preliminary test data with a specific goal in mind. It was also obvious from early on that the value of 80% is too low of a threshold. I believe it has been retained so all of those who want to judge will be able to so. The results from the 2011 S & C show that completely botched placements routinely generate scores of 80 or more. If such a low standard is used to qualify judges, there is no purpose served by qualifying them. A list of qualified judges based on such a low standard is worse than no list at all. At least without a list everyone knows they don't know whether the judges selected are capable or not. That is the case also using current qualified judge list but not everyone knows it and that makes it a deceptive practice, in my opinion. The MFTHBA should decide what they want a AAA rating to mean and then set a qualifying standard that will help them meet their goal. Possibilities include allowing only the upper 10 or 15 percent of all judges to qualify as a AAA judge. If there aren't enough judges that can qualify at that level, the standard may have to be lowered. Whatever the standard is, it should be known by all and it should apply to all. The current standard is so low that there really isn't one.

The S & C judge selection procedure has always been subjective and ripe for corruption. The procedure has been to have people on the Judge committee submit names of people who they think are good candidates and those names are taken to the board for approval. Many involved in the process of judge selection go on show under the judges selected. The board then runs the show and has authority over the judges during the S & C and many of them either show or have horses they own shown for them. They also have family members who show. There are major conflicts of interest in the process that make it unacceptable and unfair to those who show and have no influence on the selection of who they will be judged by. If the selection procedure is not be changed, it really doesn't matter very much who the judges are. People will think that placements in some classes are based on politics, prior agreements, prior relationships, etc. and they will most likely be correct.

The only fair way to select judges is to put the names of every qualified judge who is willing to judge an upcoming S & C in a sack and have an uninvolved party draw names at an open meeting. This is the only way to keep corruption and the appearance of corruption out of the process. To be able to do this, the MFTHBA must first have a list of truly qualified judges and it must be interested in having fairness in the process of judge selection rather than allowing those in charge to have an advantage over others.