A Summary of Judge Performance in the MFTHBA Show and Celebrations for the years 1998 to 2009 by Ken Kemp, Ph.D.

Introduction

Events such as the MFTHBA's annual Show and Celebrations are centered on various judged horse related events. It is the desire of nearly everyone involved to see all exhibitors be treated respectfully and fairly. One key element for this to happen is the employment of competent and fair judges. It could easily be argued that the single most important factor determining the success of any horse show depends on the hiring of competent and unbiased judges. Good horse judges are those who are well trained, well experienced, completely ethical, and have the mental skills needed to enable them to process a lot of information quickly and reach an appropriate decision as to which horses should be placed high and which should be placed lower in large classes of horses.

As a statistician and a member of the MFTHBA, I have been interested in evaluating the performances of judges at the MFTHBA S & C for the past several years. Judge evaluations are difficult because when their opinions are given it is very difficult to compare them to the available facts and then determine whether their placings are reasonable. There is generally no way to determine whether a horse show judge placed a class correctly except for timed events. Performance classes at the MFTHBA S & C are judged by 5 judges using the Olympic scoring method to determine the class placings. The judges seldom completely agree and it is not known, or knowable in most cases, which, if any, of the judges placed a class "correctly". Given that a judge's placings cannot be compared to the correct placings to determine how well he/she performed, some other criterion for determining the quality of a judge's performance must be used. That criterion must be something unbiased and completely objective.

Many organizations use judge evaluation forms to obtain information and feedback concerning their judges' performances. However, data from sources such as show management and exhibitors who show under a judge can easily be biased. They are especially prone to bias if they are not obtained by random sampling methods because self selected responders tend to be primarily from the more disgruntled exhibitors because they are highly motivated to evaluate a judge by their feeling that their horse was treated unfairly. This kind of feedback tends to be more negative than a judge deserves because of disproportionate response rates between the satisfied and disgruntled exhibitors. If the exhibitors are sampled by show management taking a randomly drawn sample, the response rate is likely to be low and the results again are likely to be biased because the disgruntled would still tend to respond at a higher rate than others and that again would bias the data for a judge's performance in a negative direction. We must remember that the happy exhibitors are those who were placed highly and they are a very small proportion of those who fill out judge performance survey forms. This is not to say that some useful information isn't to be obtained from such surveys but the drawback is that we never know how big the bias in the data is and whether it is larger for some judges than others which it probably is.

Ideally judges should judge horses based on the breed standard. A judge should not place horses based on his/her personal preferences and a judge should realize that his/her personal preferences are not synonymous with the breed standard. It is not whether a judge likes or dislikes a horse that should matter. He/she should always evaluate how horses compare to the breed standard along with how they are presented in the show ring and consider no more than that. Of course judges, like everyone, have their own preferences and biases when it comes to horses. However, given that the judges should be familiar with and judge according to the breed standard, good judges should be in agreement as to which horses come closest to the breed standard in each class. Of course agreement will not be complete, but competent, well trained, unbiased, ethical judges will select the same subset of horses to be toward the top of a class and another common subset of horses that will be placed in middle to lower positions. Based on this premise, I developed a method by which the degree to which a group of judges agree in placing the top 3 horses can be numerically quantified. Unlike other methods of judge performance evaluation, this method is completely objective. The top 3 placings are considered the most important placings and a lot of information regarding the degree to which judges agree can be gotten by measuring the disparity of their placings in the 3 top positions. I explain my method of calculation in detail in the judge tutorial wherein data from the 1999 Celebration are used as a source for examples. Understanding the results in the 12 year summary will be much easier if one is familiar with the information given in the tutorial. In a nutshell, if judges agree on the first 3 placings in a class, i.e. both place the same horse in 1st place, both agree on another horse as their 2nd choice, and likewise agree on another horse for their 3rd choice, the distance that their placings are apart is zero. To the extent that the top 3 choices of the judges grow more and more different, the distance between them increases. A distance can be calculated for every class that a judge judges. Computing the distance a judge's placement is away from a consensus placement by his 4 peer judges for each class he judges allows us to calculate an average distance for each judge over a set of classes. Judges with short average distances are in better agreement with their peers than are judges who have larger average distances. As is pointed out in the judge tutorial, a distance of less than 3 indicates close agreement, such as switching pairs among the top 3 positions or perhaps including two out of the consensus top 3 horses in their own top 3 along with a horse that is near the top such as a the consensus 4th or 5th place horse. A distance around 4 indicates more substantial disagreement as in picking horses that others placed in the consensus middle in a judge's own top 3. Distances larger than about 4.5 are associated with serious disagreements among the judges. If a judge is that far away from his peers on average, we have to question why. Is it because of his/her incompetence, lack of training, because he/she judges according to his own personal preferences rather than the breed standard, or perhaps such a judge is affected by factors that are not based on the horses in front of him as would be the case if he were bribed or placed his friends' horses high or his enemies' horses low. Of course a large average distance for a judge could also mean that he is the one who sees a class more correctly than his peers but if this is the case there is a bigger problem because it would mean several of his peers did not judge classes correctly in several of the classes that were averaged over. Large average distances mean that the judges do not agree which horses are the better horses in a class and that is a problem regardless of the cause.

There is a second measure of disagreement called theta that is used in the summaries that I will be referring to. It is similar to the average distance measure but is bounded between 0 and 1. Perfect agreement is 0 and good agreement is in the 0-0.3 range. When theta is between 0.3 and 0.4 it is similar to a distance in the region of 4 and when theta is in the range of 0.45 and higher, it means there is serious disagreement in class placings.

The data from 2005 were not usable because there were only 3 judges used to judge performance classes that year and a consensus of only two peer judges wouldn't be much of a consensus. All other years are included in the summary from 1998 through 2009. In all years an attempt was made to use as many of the WGC and champion open classes as possible. The open classes generally include the most experienced and proficient exhibitors and they are the ones we would expect to be the most consistent in presenting their horses to the judges. For most years there were between 8 and 10 classes that qualified for use. A class had to include at least 10 horses to be eligible for use as smaller classes do not challenge a judge's abilities nearly as much as larger classes do. In 2009 there were only 5 qualifying classes among the open division of classes. As a result some amateur classes were used as is shown in Table 3. Some previous work I have done with the MFTHBA performance data have shown that there is more variability among the judges placings in the amateur classes than in the open classes. I attribute that to amateurs not being as consistent in presenting their horses such that all judges see the same performance as their horse circles the show ring or perhaps most but not all judges see mistakes during the presentation. The inclusion of the amateur classes in 2009 means we should keep in mind that the 2009 judges would probably have been a little more consistent with each other, i.e. had slightly shorter average distances, had it not been necessary to include 5 amateur classes in order to get enough classes over which to summarize their performances. See MFTHBA 2009 discussion for a discussion of the 2009 judging results. The 2004 data should not be interpreted the same as the other years of data because of judge misconduct that year, as some may remember. The data in Table 10 depicts how differently a judge who was not suspected of wrong doing placed horses that year. See MFTHBA 2004 for a discussion of the 2004 judging results.

Before discussing the results over the 10 years of usable data, one more thing must be pointed out. Although we have calculated the distances and thetas for all years and it is tempting to compare the average distances of judges from one year to another, it is not a legitimate thing to do. Keep in mind that each year a particular judge judges with 4, 5, or 6 other judges. Because a judge is compared to the consensus of the judges that he judges classes with, his distances or theta values are related to how he agrees or differs with them and them only. A judge may agree with one set of peers resulting in a small average distance from their composite placings but when judging with another set of judges, he may not agree with them as often or as well and as a result he would have a larger average distance over a set of classes that he judged with them. Also keep in mind that the classes being judged from one year to the next include different horses, different riders, are of different sizes, and have different sets of circumstances and all of that will affect how well judges may agree over various classes. Thus, tempting as it is, average distances or average thetas should not be compared between judges of different years. What we can count on is that within a year we can measure the distances among the set of judges who judged together and identify which judges agreed with their peers that year and which ones tended not to. If we are willing to assume that a judge who agrees with his peers is doing a better job than one who doesn't, then we can identify the better judges within the various years. Some judges judged multiple years in the 12 year span of these data and it is very much of interest to identify judges who agree with different sets of peers over multiple years. There is no doubt in my mind that any judge who has closely agreed with his peers over multiple years is a good judge. I am happy to say that there are such judges.

Summary of Data

Figure 1 shows the distances for all judges over for all years between 1998 and 2009 except for 2005. The first thing to note is that the line connecting the average year distances for the various years is that the trend over time is basically flat. The lack of a downward trend shows that even though several efforts have been made over the years to improve judge training, and their subsequent performances, those efforts have not resulted in an improvement in judging consistency. If they had, the year average distances would have decreased over time, i.e. the mean distance of the 2009 judges would have been noticeably lower than the earlier years such as 1998 or 1999 which were prior to implementing the recent changes in judge training. However, as Table 2 shows, the average distances for the 1998, 1999, and 2009 were 4.00, 3.63, and 3.81, respectively. Only the 3.13 average distance for 2002 was noticeably smaller than the other years.

Figure 1 shows the average distances for the individual judges as data points above and below their respective year averages. Judge values above the line are above average for the respective years and are shown in red while values below average are shown in green. The judges in 1998 had individual average distances that were more similar than the other years while the 1999 judges were the most diverse regarding their respective average distance values. The average year distance is a measure of how closely the judges agreed and the spread of the individual judge values around the respective year averages show how consistent or uniform the judges were. Thus 2002 was the year where the judges tended to agree with one another the most. It has the smallest year average distance, 3.13, and there is no judge with an average distance value very far above the line, i.e. judge Clamp had the highest average distance for that year which was 3.85 but it was not far above the 2002 year average, the deviation being ***0.717*** = *3.85 – 3.13* as is shown in Table 1. The small year average distance and the close clustering of the judges' scores around the mean both show that the judges of 2002 were more consistent with each other than any other year. The 2002 data display what we would like to see for all years of data but regretfully it stands as the lone example.

At the other extreme were the judges of 2003 (see MFTHBA 2003 for details) and 2008. The average distances for those years were 4.63 and 4.60, respectively (Figure 1 & Table 2). Values as high as these indicate that the judges tended not to agree with each other in most of the classes they judged (Table 3). The high average distances for both of these years are the result of having one or more judges that did not judge according the same criteria as the other judges were using or there were judges who were either not properly trained or were simply incapable of judging classes as big and as competitive as those they were asked to deal with. I call these judges "maverick" judges. Clearly the selections of some of the judges in these, and other, years were mistakes that had major negative ramifications in the quality of the judges' performances as a group. Note how the presence of a "maverick" judge affected the average distance of the groups of judges in these two years. Either judge Rogers in 2003 and judge Hicks in 2008 were not judging according to the breed standard or they were the only ones who were. Table 9 shows how differently judges Rogers, Bailey, and Thompson placed classes with all 3 of them frequently placing their peer judges' consensus 8th, 9th, or 10th picks as one of their top 3 horses. The large variances for judges Rogers and Thompson indicate that they were in close agreement with their peers in some classes and in nearly complete disagreement in other classes while the small variance and large average distance for judge Bailey indicates that she almost never agreed with her peer judges, i.e. her distance from her peers was high in nearly all the classes she judged so she almost always disagreed. Table 13 gives similar information for the 2008 judges. Judges Hicks, Hammer, and Garland all had large average distances and large variances indicating that they tended to disagree with their peers but in some classes the amount of disagreement was very extreme while in others not as extreme.

The most useful information to come from this summarization of the judge distances is that it allows us to objectively identify which judges were in most agreement with their peers for each year. Because the presence of "maverick" judges cause the average year distance to in increase for all judges in the group that year, as mentioned before, comparing judges' average distances across years is not a fair method of comparison. A better method of comparing judges is to track which judges were more consistent with their peers in the respective years in which they judged. The best measure of that is the deviation of a judge's average distance from the respective year's average distance. These values are given in Table 1. A positive deviation is an indication that the judge was in more disagreement with his peers than was average for that year and negative deviations indicate that a judge agreed with his peers more than the average judge for a specific year. Table 1 shows that the better judges in 1998 were Free and Mizer, in 1999 they were Harris, Thompson, and Roark. Figure 2 shows a plot of the deviations from the year average values for the respective years. Judges below the line are those who tended to agree with their peers and their average deviation values are shown in green in Figure 2. Simply limiting judge selection in future shows to those who are below the average year distance for the year they judged, below the line in Figure2 or negative in Table 1, would go a long way in improving judge performance at future Celebrations or other future shows.

Using the summary data as a guideline for future judge selection is particularly appealing when considering the performances of judges who judged more than one Celebration in the 12 year period. There were 14 judges who judged multiple shows. Eleven of them had negative average deviations for the shows they judged (Table 15). My opinion is that judges who have judged multiple shows and have maintained a negative average deviation over those shows should be highly sought after. Consider Roark who judged 4 shows and Wilkerson who judged 3, both had average deviations less than -.6. This shows they were consistently in close agreement with the judges they judged with and that included 4 and 3 different groups of judges, respectively. Both have outstanding performance records in my opinion. Jameson's average deviation for the two years he judged was -1.10. He was the judge in most agreement with his peers both years that he judged. That is an impressive record. The other 8 judges with negative average deviations over multiple shows are also very good judges by the average deviation criterion. Some of the judges just barely into the positive range are no doubt capable judges but it would be difficult to differentiate those who are from those who aren't without more data.

Where to go from Here?

After nearly every Celebration exhibitors complain about the quality of judges and inconsistencies in their placings. Of course some of this goes with the territory because in large only the winners are truly happy. But much of the complaining is warranted as the data in Tables 1 and 2, Tables 4 through 14 and Figures 1 and 2 clearly show. Using the recent judging performances in the foregoing summary can definitely be helpful in making better future choices in judge selection. However, historic data can't be used when new judges need to be brought into the mix and of course that will happen sooner than later. The methods used in the past for judge selection have been badly flawed as is evidenced by some of the judges who have been selected and are toward the top of the graphs in Figures 1 & 2 or what happened in 2004. The methods used in the past have been very subjective and riddled with conflicts of interest. It is always a bad idea to let anyone who is going to exhibit their horse in a show be involved in selecting the judges they will show under because the criterion for judge selection rather than being objective and based on an appraisal of how well a judge can judge becomes very subjective and based how the persons involved in making the selections think a particular judge may judge their horses. Having members of the BOD and show committee who show in a Celebration be involved in or have influence on which judges are selected is fraught with conflicts of interest, lack of objectivity, and gives unfair advantage to those involved in the selection process. This has not gone unnoticed by the membership of the organization nor exhibitors not fortunate enough to have influence on which judges are hired. Clearly an impartial method of judge selection should be used to assure fairness for all. I have proposed a method (Judge Training & Testing Procedure) that would use videoed classes as both a training and testing tool for judges. Classes that have been placed by expert judges such as those mentioned above who have outstanding judging performance records would place a set of classes by consensus among them and these classes could then serve to test the ability of candidate judges by having them place some of the classes. The candidates' placings would be compared to the "official" placings by computing how far the average distance for each candidate is from the respective "official" placings of the same classes and the best performing judges could be easily identified and then become candidates for judging future shows. This is an objective method that would assure the better judges are selected for shows and their selection would be fair to all.

Although there are requirements that judges in the MFTHBA must meet, they are not rigorous enough to assure that those who may be selected to judge are competent. Currently judge candidates must attend a clinic and then pass an exam and also have some mentoring before they become a qualified judge but there is very little performance testing involved. It is one thing to pass a quiz and quite another to judge a major horse show. Knowing key facts in the rulebook is important but it is far from sufficient to assure that someone will be a good judge. Certain mental skills and abilities are necessary for one to be able to take in and remember and then process all the information that comes at a judge when 20 or more horses circle a show ring. Only a certain few, in my opinion, have what it takes to apply the rules under the conditions mentioned for multiple classes over a string of several days.

The TWH Association is much more particular about who is allowed to judge their Celebration. To avoid conflicts of interest they have S.H.O.W. Inc. supply their judges and given that the two organizations are independent of each other, there are fewer conflicts of interest regarding which judges are selected. The eligibility requirements are very stringent for a judge to be qualified to judge a TWH Celebration. A judge must be AAA rated and the requirements for being AAA rated are that they must have attended a two day clinic, taken a written test, judged at least 12 shows, and have been a judge for at least 5 years before they qualify. The results of the more stringent requirements are reflected in the performances of their judges. Data from the 2003 TWH Celebration shows how much better they performed. The average distance for 19 TWH WGC classes was 2.56 and the average theta was 0.29. Both values are indicative of a generally high level of agreement among the judges. The MFTHBA average distances range from 3.13 to 4.60 and average 3.86 (Table 2) and the only year with an average distance less than 3.5 was the 3.13 from 2002 which was the best year among the 12 years of data. There is a much greater disagreement among judges who average 3.86 and those who average 2.56.   In addition, the 2003 TWH open WGC class had an average distance of 1.4 and a theta value of 0.16. Contrast that to the 2009 MFTHBA WGC class where the average distance was 3.31 and theta was 0.37. In the MFTHBA WGC class one judge placed a 7th place horse 3rd and another placed an 8th place horse 3rd (see MFTHBA 2009 for details). While some of the TWH judges may have switched an adjacent pair among the top 3 positions, some of the MFTHBA judges placed 7th and 8th place horses in 3rd when picking the breed champion!

The advantages of outside judges are that they would generally not be acquainted with MFTHBA exhibitors, would not be familiar with who is who within the breed, would not have been selected by members of a committee or the BoD, and thus would be much less likely to have conflicts of interest that may affect their judging and would therefore be in a better position to judge objectively than judges from within the breed would be. Having an independent organization in charge of DQPs and rules enforcement would likely work better than what was done at this year's Celebration where show management was inappropriately involved with the DQPs and the judges apparently placed a horse that was bleeding. Having an independent agency enforce rules would be fairer to all members who show because no one would get favorable treatment because of their status in the MFTHBA. Some organizations worthy of investigation as a source of judges are: Natural Walking Horse Association (NWHA), Friends of Sound Horses (FOSH), and Horse Protection Commission (HPC). The requirements for being licensed by these organizations and the amount of experience the judges who judge for such organizations have, should both be considerably higher than what we find among the MFTHBA judges, and these two things should translate to more competent judges being available for the MFTHBA Celebration in addition to them being more objective.

Conclusions

Major findings from the 1998-2009 MFTHBA Show & Celebration Summary

1. During the 12 year period despite all efforts to improve judge performance there has been no progress in getting judges to agree more in their placings. Continuing to use the same methods as have been used in the past will not somehow become more effective at some point in the future.

2. We currently have objective information regarding which judges were the most consistent in placing horses with their peers over the recent past which may be useful for judge selection in the near future

3. New methods for training and selecting or hiring judges must be adopted

    a. Use video aided training and testing to first train and then to identify judge candidates best suited for future shows based on objective performance criteria, or

    b. Use outside judges for access to better trained, more experienced, and well tested judges, who will not have conflicts of interest, and who have proven they can judge gaited horses in a consistent manner with their peer judges.

Table 1. Judges Ordered by Decreasing Average Distance within Years

Year	Judge	Distance	Deviation
1998	Mackie	4.54	0.536
1998	Gilbert	4.50	0.496
1998	PHollingsworth	4.23	0.226
1998	Free	3.82	-0.184
1998	Mizer	2.93	-1.074
1999	Jlaughlin	6.74	3.108
1999	Casper	4.12	0.488
1999	Baker	3.88	0.248
1999	Harris	3.08	-0.551
1999	RobThompson	2.17	-1.461
1999	Roark	1.80	-1.831
2000	Berry	4.90	1.342
2000	Trussel	4.07	0.512
2000	Cox	3.68	0.122
2000	Stringer	2.78	-0.778
2000	Jameson	2.36	-1.198
2001	Evans	4.58	0.956
2001	BLaughlin	4.41	0.786
2001	Spiceland	4.35	0.726
2001	Wilkerson	3.62	-0.003
2001	Hays	3.05	-0.573
2001	DHollingsworth	1.73	-1.893
2002	Clamp	3.85	0.717
2002	Barton	3.78	0.647
2002	Roark	3.48	0.347
2002	JanThompson	3.25	0.117
2002	Burks	2.53	-0.603
2002	Wilkerson	1.91	-1.223
2003	Rogers	6.15	1.520
2003	Bailey	5.36	0.730
2003	RrdThompson	4.72	0.090
2003	PHollingsworth	4.07	-0.560
2003	Day	3.86	-0.770
2003	Jameson	3.62	-1.010
2004	Spiceland	6.16	2.658
2004	Cox	3.26	-0.242
2004	McBride	3.16	-0.342
2004	Eagleburger	3.06	-0.442
2004	Wilkerson	2.83	-0.672
2004	Clamp	2.54	-0.962
2006	Dehner	5.46	1.316
2006	RrdThompson	5.21	1.066
2006	Spiceland	4.35	0.206
2006	Porter	4.14	-0.004
2006	Baysinger	3.97	-0.174
2006	Harris	3.07	-1.074
2006	Burks	2.81	-1.334
2007	RobThompson	6.40	2.599
2007	Moore	4.87	1.069
2007	Roark	4.25	0.449
2007	Arthur	3.35	-0.451
2007	Ball	2.77	-1.031
2007	Gettys	2.76	-1.041
2007	Nichols	2.21	-1.591

(Table 1. Cont)

Year	Judge	Distance	Deviation
2008	Hicks	6.63	2.033
2008	Garland	5.16	0.563
2008	Hammer	4.78	0.183
2008	Eagleburger	4.23	-0.367
2008	Maxey	3.60	-0.997
2008	Porter	3.18	-1.417
2009	Marion	5.21	1.400
2009	Gates	4.52	0.710
2009	Hart	3.93	0.120
2009	Brewster	3.85	0.040
2009	Hays	3.30	-0.510
2009	Roark	2.05	-1.760

Table 2. Average Distance and Theta by Year

MFTHBA

		Average	Average
Year	# of Judges	Distance	Theta
1998	5	4.00	0.40
1999	6	3.63	0.38
2000	5	3.56	0.38
2001	6	3.62	0.36
2002	6	3.13	0.33
2003	6	4.63	0.46
2004	6	3.50	0.36
2006	7	4.14	0.40
2007	7	3.80	0.39
2008	6	4.60	0.43
2009	6	3.81	0.41
Grand Mean		**3.86**	**0.39**

THW 2003

19 WGC classes avg.	2.56	0.29
Open WGC class	1.40	0.16

Table 3. Classes used for each year from 1998 through 2009 at MFTHBA Show & Celebrations

Year	WGC 5 & older	WGC 4	WGC 3	WGC 2	5 Year S & G	5 Year Mares	4 Year S & G	4 Year Mares	3 Year Mares	4 Year & older	3 Year S & G	2 Year S & G	2 Year Mares	WGC Am 5+	WGC Am 4 year old	WGC Am 3 year old	Number of Classes
1998	X		X	X		X			X		X	X	X				8
1999	X		X	X	X	X	X		X		X	X	X				10
2000	X		X	X	X				X		X	X	X				8
2001	X		X	X		X			X		X	X	X				8
2002	X		X	X		X		X	X	X	X	X	X				10
2003	X	X	X	X	X	X			X				X				8
2004	X	X	X	X		X			X		X	X	X				9
2006	X		X	X		X			X			X		X	X		8
2007	X		X	X	X	X			X		X						7
2008	X		X			X			X		X			X	X	X	8
2009	X		X		X	X			X						X		10*

* Because of limited qualifying Open classes in 2009 the following classes were also included: Open Am 4 Yr & Older S&G, Open Am 4 Year Old Mares, Open Am 5 Yr & Older Mares, and Open Am 3 Year Old Mares

Table 4. Grand averages of 8 classes for 1998

Judge	Placed an 8,9,10 as 1, 2, or 3	Placed a 1,2,3 as 8, 9, or 10	Was only one to place a particular horse
Free	0.12	0.00	0.38
Gilbert	0.38	0.38	0.50
PHollingsworth	0.12	0.12	0.12
Mackie	0.38	0.12	0.38
Mizer	0.12	0.00	0.38

Average theta: 0.40

	1-10	1-5	1-3	Variance	# classes
Judge-Free:	9.22	5.41	3.82	7.869	8
Judge-Gilbert:	11.06	6.83	4.50	5.852	8
Judge-PHollingsworth:	9.24	6.17	4.23	15.854	8
Judge-Mackie:	9.24	6.00	4.54	7.624	8
Judge-Mizer:	8.76	5.76	2.93	15.302	8
Averages	9.50	6.03	4.00		

Table 5. Grand averages of 11 classes for 1999

	Placed an 8,9,10	Placed a 1,2,3	Was only one to place
Judge	as 1, 2, or 3	as 8, 9, or 10	a particular horse
Casper	0.43	0.43	0.29
RobThompson	0.00	0.20	0.20
Harris	0.18	0.00	0.36
Jlaughlin	0.67	0.00	0.44
Roark	0.00	0.00	0.25
Baker	0.20	0.00	0.30

Average theta: 0.38

	1-10	1-5	1-3	Variance	# classes
Judge- Casper:	9.90	6.82	4.12	18.846	7
Judge- RobThompson:	9.01	5.75	2.17	1.457	10
Judge- Harris:	8.82	6.16	3.08	5.856	11
Judge- Jlaughlin:	10.37	8.29	6.74	12.994	9
Judge- Roark:	8.84	5.72	1.80	2.426	8
Judge- Baker:	7.73	4.97	3.88	11.164	10
Averages	9.11	6.29	3.63		

Table 6. Grand averages of 8 classes for 2000

	Placed an 8,9,10	Placed a 1,2,3	Was only one to place
Judge	as 1, 2, or 3	as 8, 9, or 10	a particular horse
Cox	0.12	0.25	0.12
Trussel	0.25	0.25	0.50
Stringer	0.00	0.12	0.25
Jameson	0.00	0.12	0.00
Berry	0.25	0.25	0.25

Average theta: 0.38

	1-10	1-5	1-3	Variance	# classes
Judge-Cox:	9.66	5.82	3.68	5.115	8
Judge-Trussel:	9.56	5.98	4.07	6.195	8
Judge-Stringer:	8.29	4.86	2.78	0.620	8
Judge-Jameson:	8.57	4.77	2.36	1.925	8
Judge-Berry:	9.70	6.52	4.90	4.244	8
Averages	9.16	5.59	3.56		

Table 7. Grand averages of 9 classes for 2001

Judge	Placed an 8,9,10 as 1, 2, or 3	Placed a 1,2,3 as 8, 9, or 10	Was only one to place a particular horse
Hays	0.00	0.00	0.14
BLaughlin	0.25	0.50	0.75
Spiceland	0.40	0.40	0.40
Evans	0.29	0.14	0.14
DHollingsworth	0.00	0.17	0.17
Wilkerson	0.14	0.00	0.29

Average theta: 0.36

	1-10	1-5	1-3	Variance	# classes
Judge-Hays:	8.80	5.26	3.05	5.511	7
Judge-BLaughlin:	11.53	7.44	4.41	7.214	8
Judge-Spiceland:	9.65	5.98	4.35	16.559	5
Judge-Evans:	9.98	6.32	4.58	2.827	7
Judge-DHollingsworth:	7.57	4.38	1.73	4.405	6
Judge-Wilkerson:	9.66	6.23	3.62	7.724	7
Averages	9.53	5.93	3.62		

Table 8. Grand averages of 10 classes for 2002

	Placed an 8,9,10	Placed a 1,2,3	Was only one to place
Judge	as 1, 2, or 3	as 8, 9, or 10	a particular horse
Clamp	0.12	0.00	0.38
Roark	0.00	0.00	0.00
JanThompson	0.10	0.20	0.30
Wilkerson	0.00	0.00	0.00
Barton	0.25	0.00	0.38
Burks	0.00	0.11	0.44

Average theta: 0.33

	1-10	1-5	1-3	Variance	# classes
Judge-Clamp:	8.90	5.76	3.85	13.232	8
Judge-Roark:	7.63	4.41	3.48	1.145	5
Judge-JanThompson:	8.31	5.42	3.25	3.157	10
Judge-Wilkerson:	7.03	3.18	1.91	3.181	10
Judge-Barton:	8.77	6.38	3.78	11.773	8
Judge-Burks:	7.80	3.74	2.53	3.409	9
Averages	8.07	4.81	3.13		

Table 9. Grand averages of 8 classes for 2003

```
           Placed an 8,9,10   Placed a 1,2,3   Was only one to place
 Judge     as 1, 2, or 3      as 8, 9, or 10   a particular horse
RrdThompson    0.50                0.00                0.17
Bailey         0.43                0.29                0.00
Rogers         0.50                0.38                0.38
Day            0.14                0.14                0.43
Jameson        0.00                0.33                0.17
PHollingsworth 0.00                0.00                0.50

 Average theta: 0.46

                      1-10    1-5    1-3    Variance    # classes(10)
Judge-RrdThompson:    7.37   6.19   4.72    24.041           6
Judge-Bailey:         9.65   6.03   5.36     2.855           7
Judge-Rogers:        10.99   7.76   6.15    16.159           8
Judge-Day:            9.46   5.64   3.86     4.818           7
Judge-Jameson:        8.89   4.94   3.62     2.452           6
Judge-PHollingsworth:11.18   6.14   4.07     4.746           6
Averages              9.59   6.11   4.63
```

Table 10. Grand averages of 9 classes for 2004

Judge	Placed an 8,9,10 as 1, 2, or 3	Placed a 1,2,3 as 8, 9, or 10	Was only one to place a particular horse
Clamp	0.00	0.25	0.25
Spiceland	0.50	0.17	0.33
McBride	0.14	0.14	0.29
Eagleburger	0.00	0.22	0.22
Cox	0.14	0.14	0.14
Wilkerson	0.12	0.12	0.12

Average theta: 0.36

	1-10	1-5	1-3	Variance	# classes
Judge-Clamp:	7.26	3.88	2.54	1.029	8
Judge-Spiceland:	10.57	7.75	6.16	24.711	6
Judge-McBride:	7.23	4.55	3.16	4.161	7
Judge-Eagleburger:	7.51	4.71	3.06	4.114	9
Judge-Cox:	8.72	5.64	3.26	4.235	7
Judge-Wilkerson:	6.99	4.30	2.83	6.019	8
Averages	8.05	5.14	3.50		

Table 11. Grand averages of 10 classes for 2006

Judge	Placed an 8,9,10 as 1, 2, or 3	Placed a 1,2,3 as 8, 9, or 10	Was only one to place a particular horse
RrdThompson	0.33	0.33	0.33
Porter	0.14	0.00	0.14
Burks	0.00	0.00	0.50
Spiceland	0.33	0.33	0.67
Dehner	0.60	0.40	0.00
Baysinger	0.17	0.00	0.33
Harris	0.17	0.67	0.00

Average theta: 0.40

	1-10	1-5	1-3	Variance	# classes
Judge-RrdThompson:	10.39	6.57	5.21	13.236	6
Judge-Porter:	8.63	5.83	4.14	3.805	7
Judge-Burks:	7.67	3.88	2.81	1.819	4
Judge-Spiceland:	11.52	8.35	4.35	6.932	6
Judge-Dehner:	10.07	7.02	5.46	15.439	5
Judge-Baysinger:	9.63	5.68	3.97	5.269	6
Judge-Harris:	9.32	6.13	3.07	4.874	6
Averages	9.60	6.21	4.14		

Table 12. Grand averages of 8 classes for 2007

Judge	Placed an 8,9,10 as 1, 2, or 3	Placed a 1,2,3 as 8, 9, or 10	Was only one to place a particular horse
Arthur	0.17	0.00	0.00
Ball	0.00	0.29	0.00
Gettys	0.00	0.14	0.14
Moore	0.33	0.17	0.50
Nichols	0.00	0.29	0.29
Roark	0.25	0.25	0.00
RobThompson	0.67	0.33	0.33

Average theta: 0.39

	1-10	1-5	1-3	Variance	# classes
Judge-Arthur:	7.18	5.19	3.35	2.143	6
Judge-Ball:	7.08	4.77	2.77	2.875	7
Judge-Gettys:	6.72	3.94	2.76	2.258	7
Judge-Moore:	7.24	5.32	4.87	11.995	6
Judge-Nichols:	6.21	3.57	2.21	0.947	7
Judge-Roark:	8.78	6.10	4.25	7.531	4
Judge-RobThompson:	11.35	9.00	6.40	8.648	3
Averages	7.79	5.41	3.80		

Table 13. Grand averages of 8 classes for 2008

Judge	Placed an 8,9,10 as 1, 2, or 3	Placed a 1,2,3 as 8, 9, or 10	Was only one to place a particular horse
Eagleburger	0.14	0.00	0.29
Garland	0.20	0.00	0.00
Hammer	0.29	0.14	0.43
Hicks	0.57	0.29	0.43
Maxey	0.14	0.29	0.29
Porter	0.17	0.00	0.17

Average theta: 0.43

	1-10	1-5	1-3	Variance	# classes
Judge-Eagleburger:	8.31	6.09	4.23	21.248	7
Judge-Garland:	8.58	6.72	5.16	20.706	5
Judge-Hammer:	9.03	6.65	4.78	22.711	7
Judge-Hicks:	11.30	7.96	6.63	20.369	7
Judge-Maxey:	9.10	6.15	3.60	8.361	7
Judge-Porter:	6.58	4.09	3.18	2.670	6
Averages	8.82	6.27	4.60		

Table 14. Grand averages of 10 classes for 2009

Judge	Placed an 8,9,10 as 1, 2, or 3	Placed a 1,2,3 as 8, 9, or 10	Was only one to place a particular horse
Brewster	0.29	0.29	0.29
Gates	0.33	0.00	0.00
Marion	0.67	0.11	0.56
Roark	0.00	0.11	0.11
Hays	0.12	0.12	0.12
Hart	0.25	0.25	0.00

Average theta: 0.41

	1-10	1-5	1-3	Variance	# classes
Judge-Brewster:	7.59	5.27	3.85	7.403	7
Judge-Gates:	7.74	5.46	4.52	6.056	9
Judge-Marion:	9.04	6.39	5.21	7.910	9
Judge-Roark:	6.35	4.04	2.05	2.769	9
Judge-Hays:	7.28	4.95	3.30	6.730	8
Judge-Hart:	7.29	5.11	3.93	3.497	8
Averages	7.55	5.21	3.8		

Table 15. Judges ordered from more to less in agreement with their peers

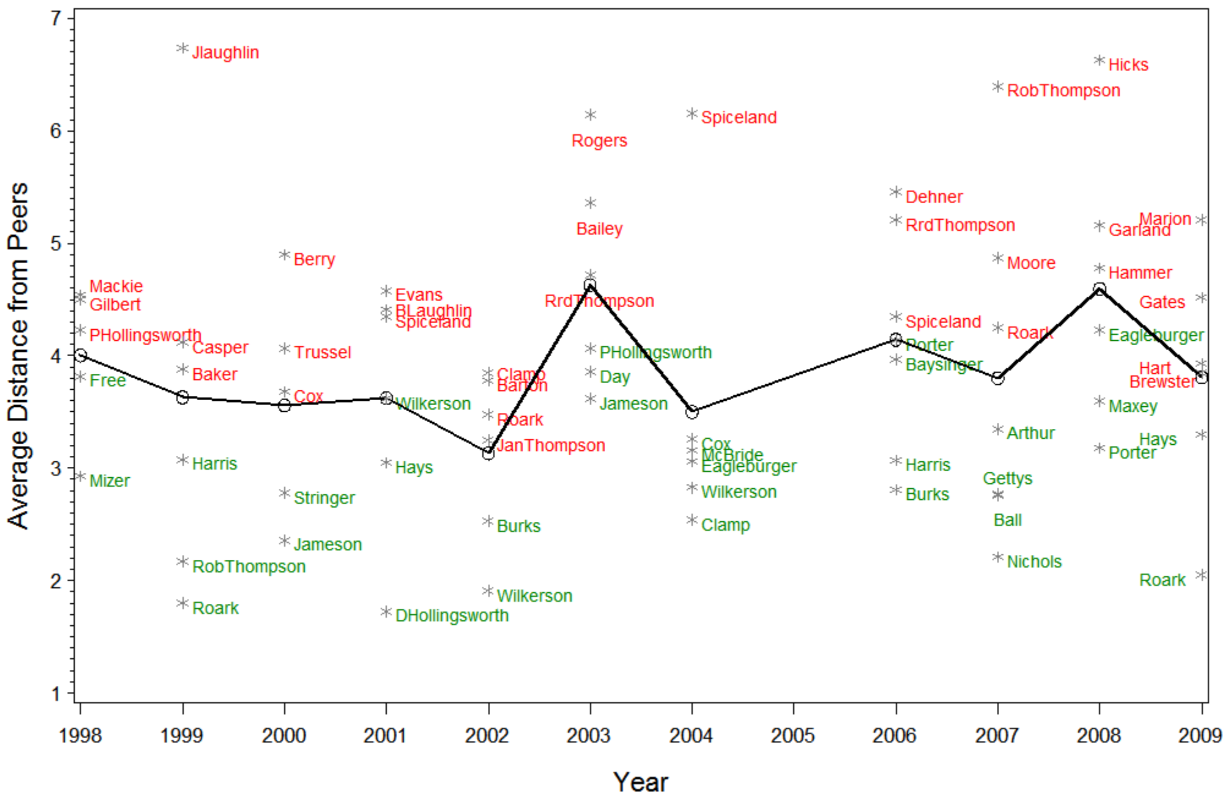
Name	# of Years	Average Deviation	
DHollingsworth	1	-1.89333	
Nichols	1	-1.59143	
Jameson	**2**	***-1.10400***	
Mizer	1	-1.07400	
Gettys	1	-1.04143	
Ball	1	-1.03143	
Maxey	1	-0.99667	
Burks	**2**	***-0.96881***	
Harris	**2**	***-0.81298***	
Stringer	1	-0.77800	
Day	1	-0.77000	
Porter	**2**	***-0.71048***	
Roark	**4**	***-0.69911***	
Wilkerson	**3**	***-0.63278***	
Hays	**2**	***-0.54167***	
Arthur	1	-0.45143	
Eagleburger	**2**	***-0.40417***	
McBride	1	-0.34167	
Free	1	-0.18400	
Baysinger	1	-0.17429	
PHollingsworth	**2**	***-0.16700***	
Clamp	**2**	***-0.12250***	
Cox	**2**	***-0.05983***	
Brewster	1	0.04000	
JanThompson	1	0.11667	
Hart	1	0.12000	
Hammer	1	0.18333	
Baker	1	0.24833	
Spiceland*	2	0.46600	*(w/o 2004 data)
Casper	1	0.48833	
Gilbert	1	0.49600	
Trussel	1	0.51200	
Mackie	1	0.53600	
Garland	1	0.56333	
RobThompson	2	0.56845	
RrdThompson	2	0.57786	
Barton	1	0.64667	
Gates	1	0.71000	
Bailey	1	0.73000	
BLaughlin	1	0.78667	
Evans	1	0.95667	
Moore	1	1.06857	
Spiceland*	3	1.19690	(with 2004 data)
Dehner	1	1.31571	

```
Berry                1         1.34200
Marion               1         1.40000
Rogers               1         1.52000
Hicks                1         2.03333
Jlaughlin            1         3.10833
```

Figure 1. S & C Judge Distances for Time Period 1998-2009

Figure 1. MFTHBA Show and Celebration judge distances for time period: 1998-2009

Note: Lack of a downward trend shows there was no improvement in consistency over the 12 years



Points above the line indicate a judge was less consistent than peers

Figure 2. S & C Judge Deviations for Time Period 1998-2009

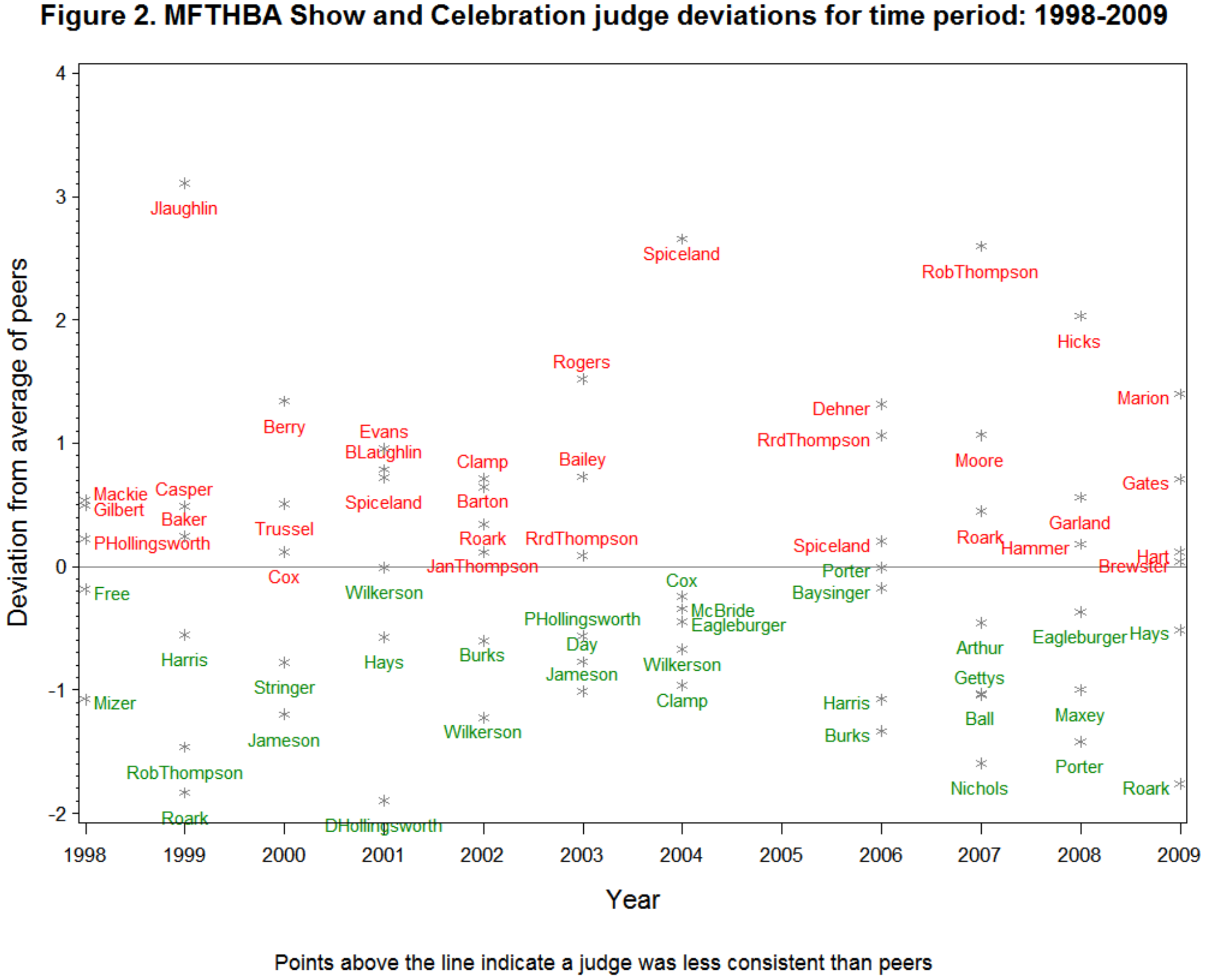


Figure 2. MFTHBA Show and Celebration judge deviations for time period: 1998-2009
Deviation from average of peers
Year
Jlaughlin
Spiceland
RobThompson
Hicks
Rogers
Marion
Dehner
RrdThompson
Berry
Evans
BLaughlin
Clamp
Bailey
Moore
Mackie
Gilbert
Casper
Baker
Gates
Trussel
Spiceland
Barton
Garland
PHollingsworth
Roark
RrdThompson
Spiceland
Roark
Hammer
Cox
JanThompson
Hart
Brewster
Free
Wilkerson
Cox
Porter
Baysinger
McBride
Eagleburger
PHollingsworth
Day
Arthur
Eagleburger
Hays
Harris
Hays
Burks
Jameson
Wilkerson
Stringer
Clamp
Gettys
Mizer
Harris
Maxey
Jameson
Wilkerson
Ball
Burks
RobThompson
Porter
Nichols
Roark
Roark
DHollingsworth
Points above the line indicate a judge was less consistent than peers