

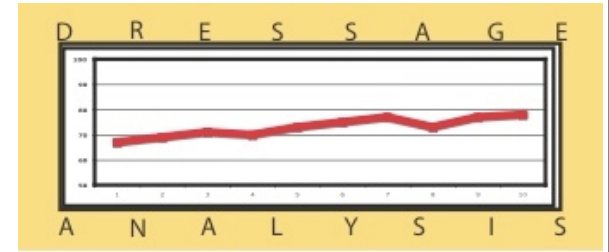
Judging, Myths, Facts and Solutions. EC 2009

Oct 27, 2009

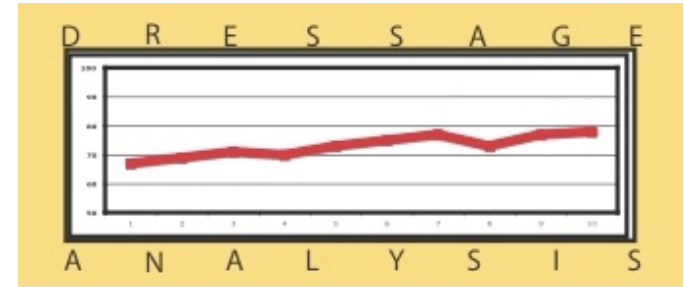
Dr. David Stickland

www.dressage-analysis.com

Outline

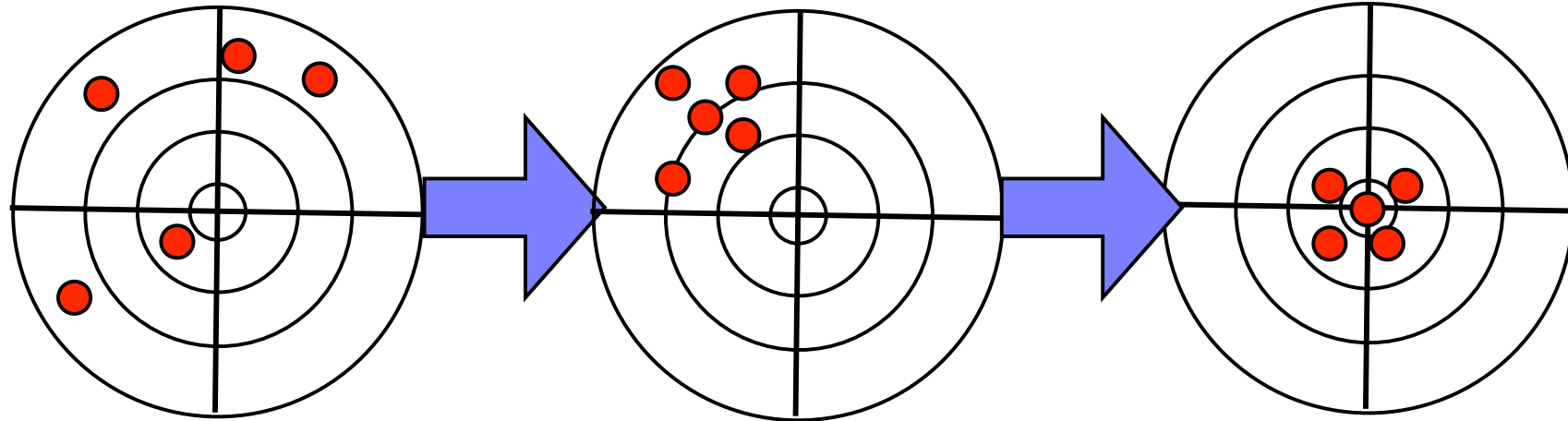
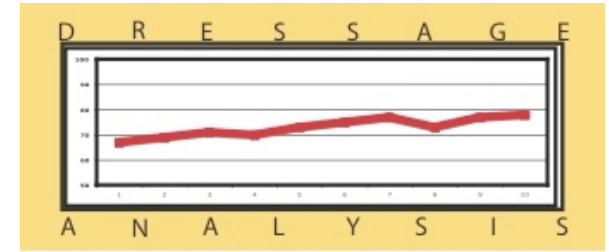


- Part I. Accuracy and Consistency
 - ➔ How Accurate and Consistent are we now?
- Part II. How was Windsor EC2009 Overall?
 - ➔ How good was the Judging?
- Part III. What are the problems of judging?
 - ➔ Windsor as the example
- Part IV. What can we do to improve judging?
Discussion



Part I: Accuracy and Consistency

Our Target:



Inconsistent
and Inaccurate
Judging

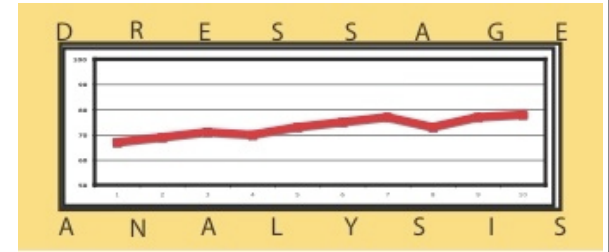
Consistent,
but still Inaccurate
Judging

Consistent
and Accurate
Judging

- How to achieve the Holy-Grail of Consistency and Accuracy ?

(Thanks to Ana E. Diaz for use of this pictorial analogy)

Consistency and Accuracy



- CONSISTENCY

- ➔ The ability of judges to agree on the same score for a given performance.

- Statistics can measure this

- ACCURACY

- ➔ The refinement of the judging system to better define what the correct score is for a given performance

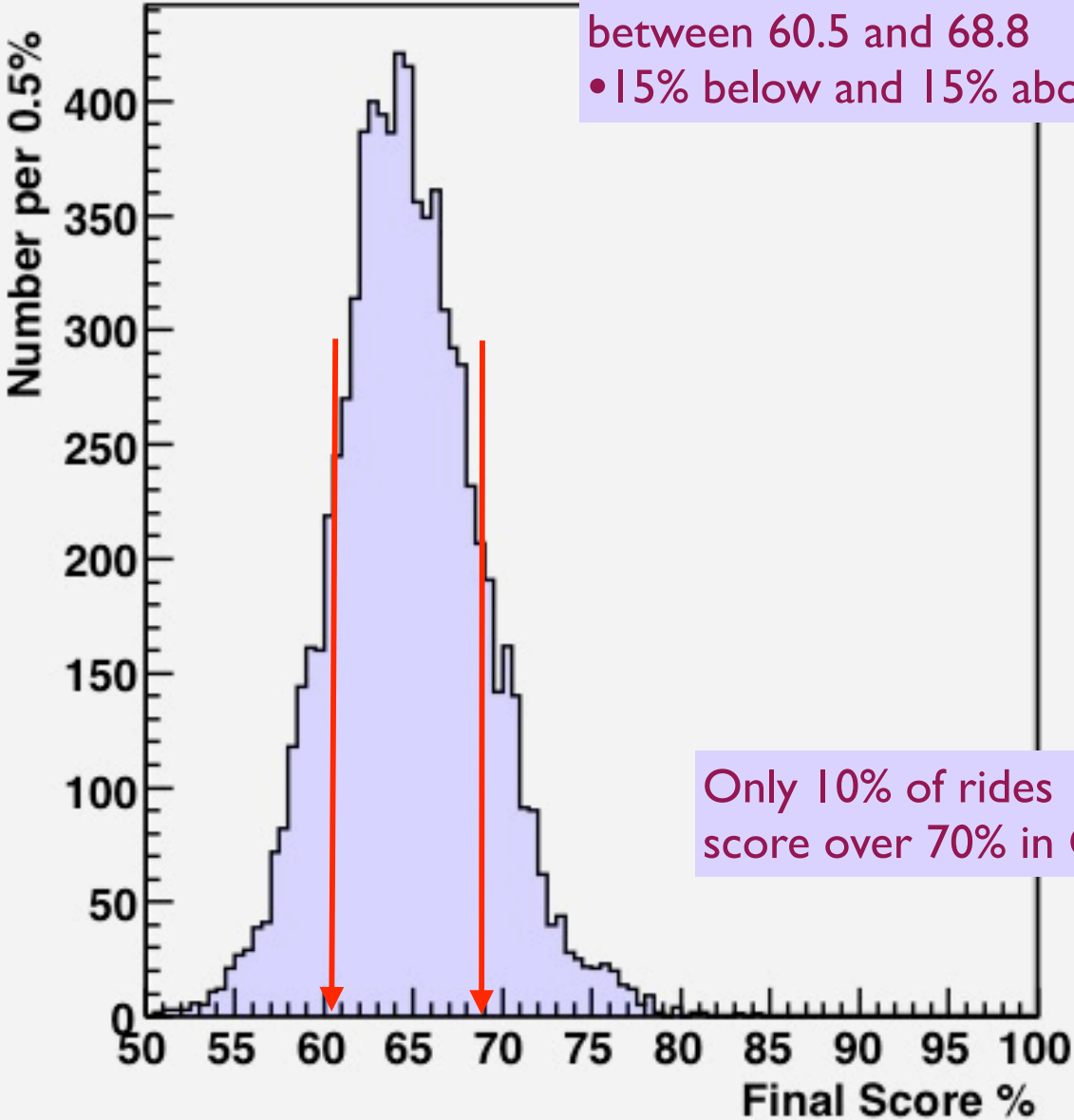
- Dressage experts need to do this

Five years CDIs, more than 13,000 tests

GP Scores

Average Score 64.68

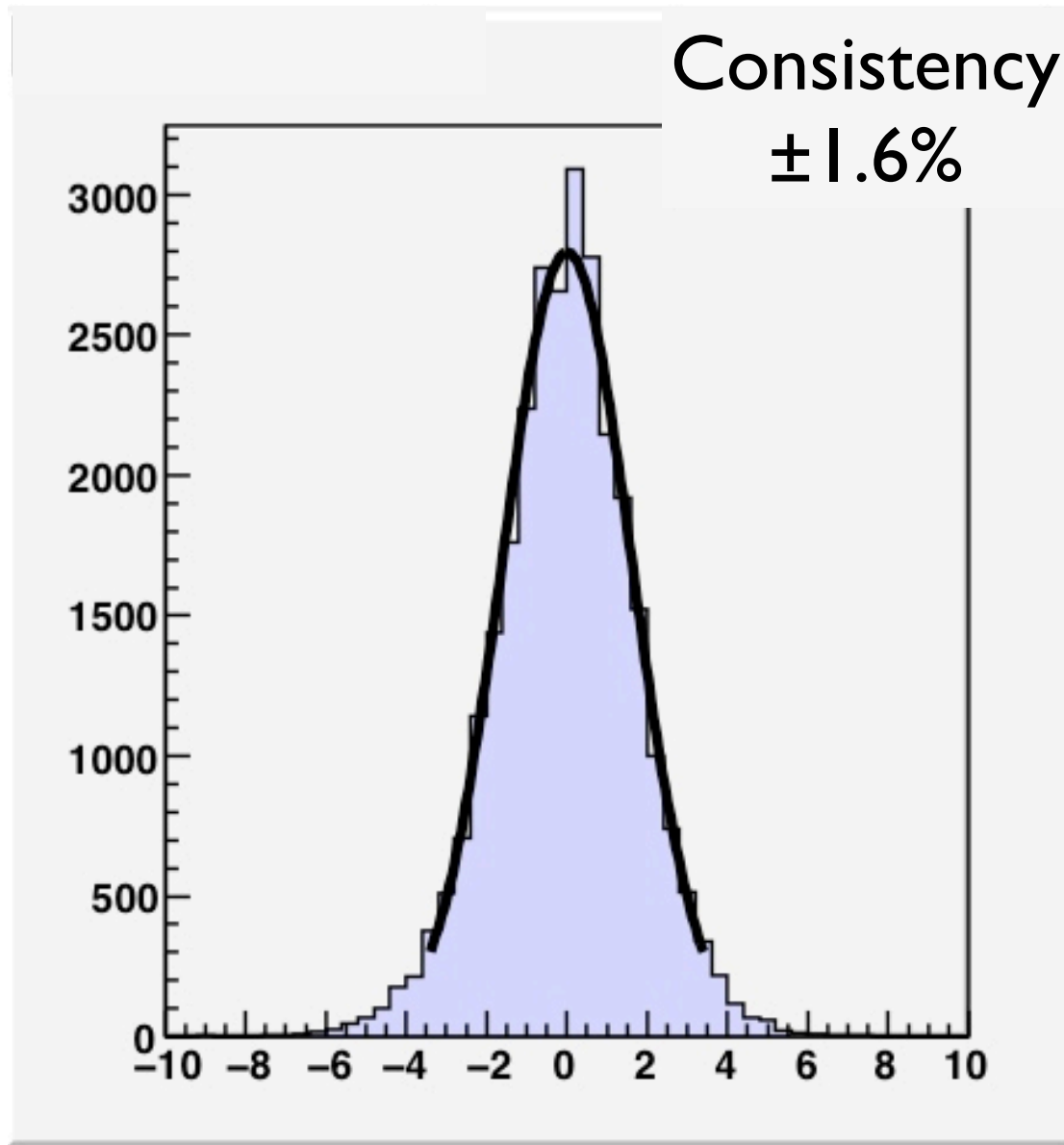
- 70% of all rides score between 60.5 and 68.8
- 15% below and 15% above



Only 10% of rides score over 70% in GP



Difference in final scores between each judge and the average of the other four judges.



Judging Consistency is also $\pm 1.6\%$ for GPS and $\pm 2\%$ for the Freestyle.

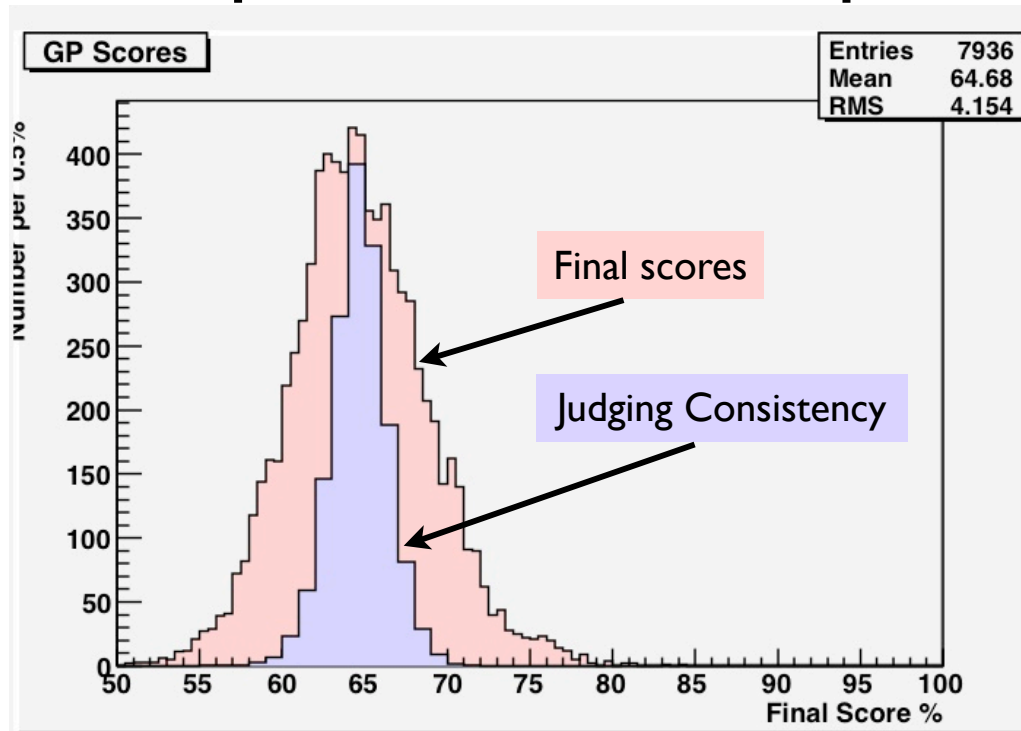
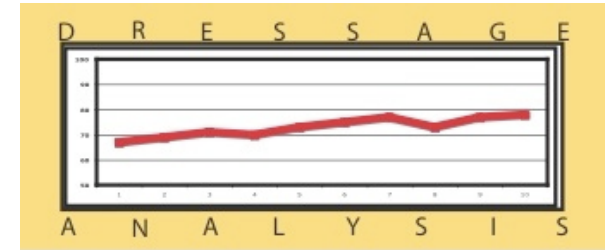
$\pm 1.6\%$ Consistency means that 2/3 of the time the judge is within 1.6% of the other judges.

But 1/3 of the time more than 1.6%,

1/20 times more than 3.2% different

and 1/100 times more than 4.8%....

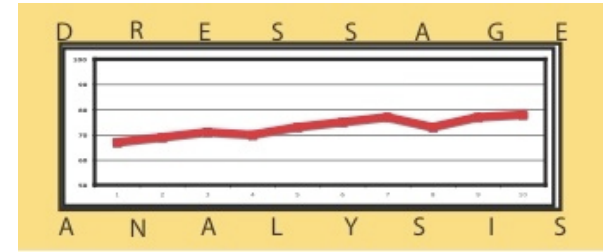
Is the system good enough to separate most competitors?



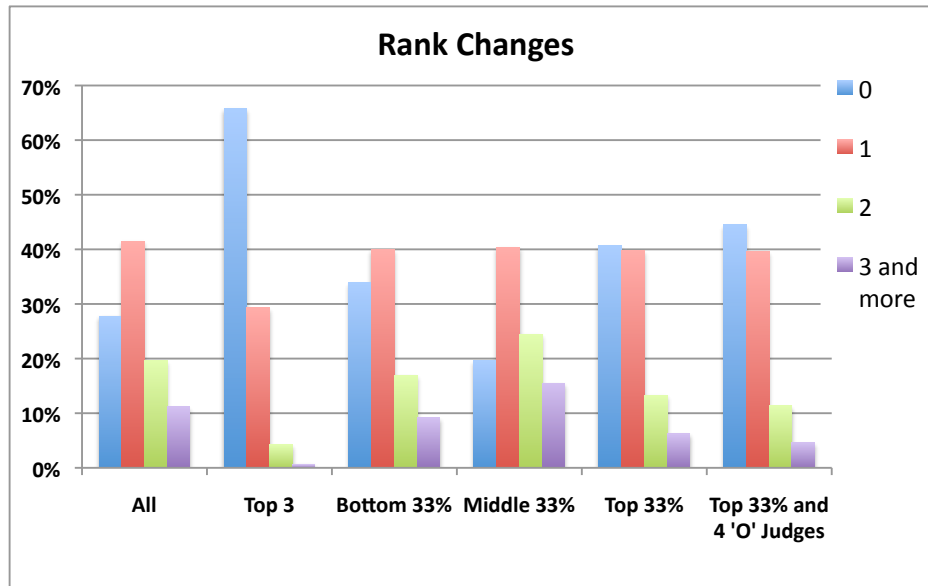
- The average score in a GP is $64.7 \pm 4\%$
- A judges consistency is about $\pm 1.6\%$

- ➔ A single judge's consistency covers much of the range of final score for most riders.
- ➔ With 5 good judges, the system achieves 0.7% precision in the final score.
- ➔ **With less than 5 judges the system is inadequate to rank most riders correctly**

“But anyway, the ranking is correct”....?



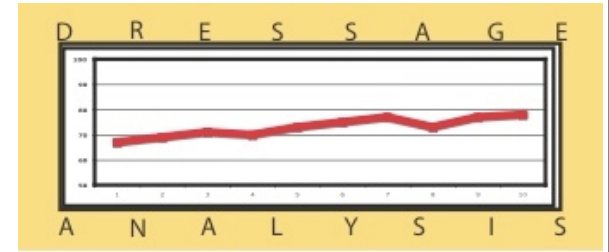
- Examine all CDI Grand Prix
 - ➔ How many ranks *actually* got changed?
 - Not just the different ranks given by different judges, but “how often does the final rank actually get changed”?



- ➔ Overall 72% of all ranks are changed
- ➔ 34% of Podium ranks are changed
- ➔ In the top 1/3 of ranks, 59% are changed
- ➔ In the mid 1/3 of ranks, 80% are changed
 - (16% by 3 or more places!)
- ➔ In the lower 1/3 of ranks, 66% are changed
- ➔ Even with 4 or more ‘O’ judges judging, 56% of places in the top 1/3 get changed

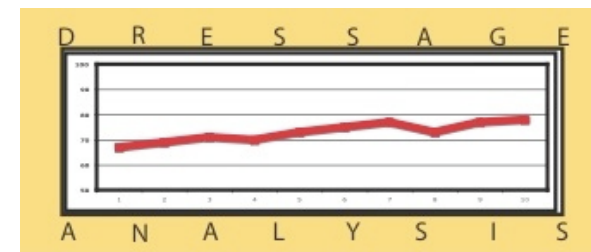
No, the ranking is not safe either

How good could perfect judges be with this system?



- Judges today can only give integer scores (5,6,7,8..)
- Even if a perfect judge knew the score should be 7.342, he could only give a 7
 - ➔ This introduces an ultimate precision beyond which no judge can improve, that turns out to be about $\pm 0.5\%$ for a test with 36 movements
 - ➔ $\pm 1.6\%$ is still quite a way from $\pm 0.5\%$, so what is happening?

“Randomized Tests”

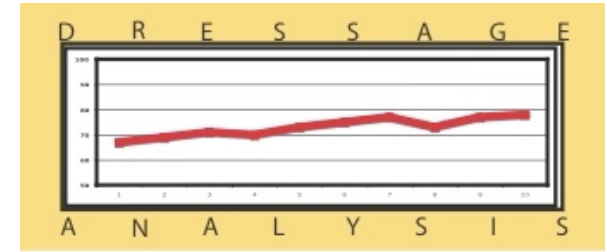


When we have the figure-by-figure scores we can perform an informative experiment

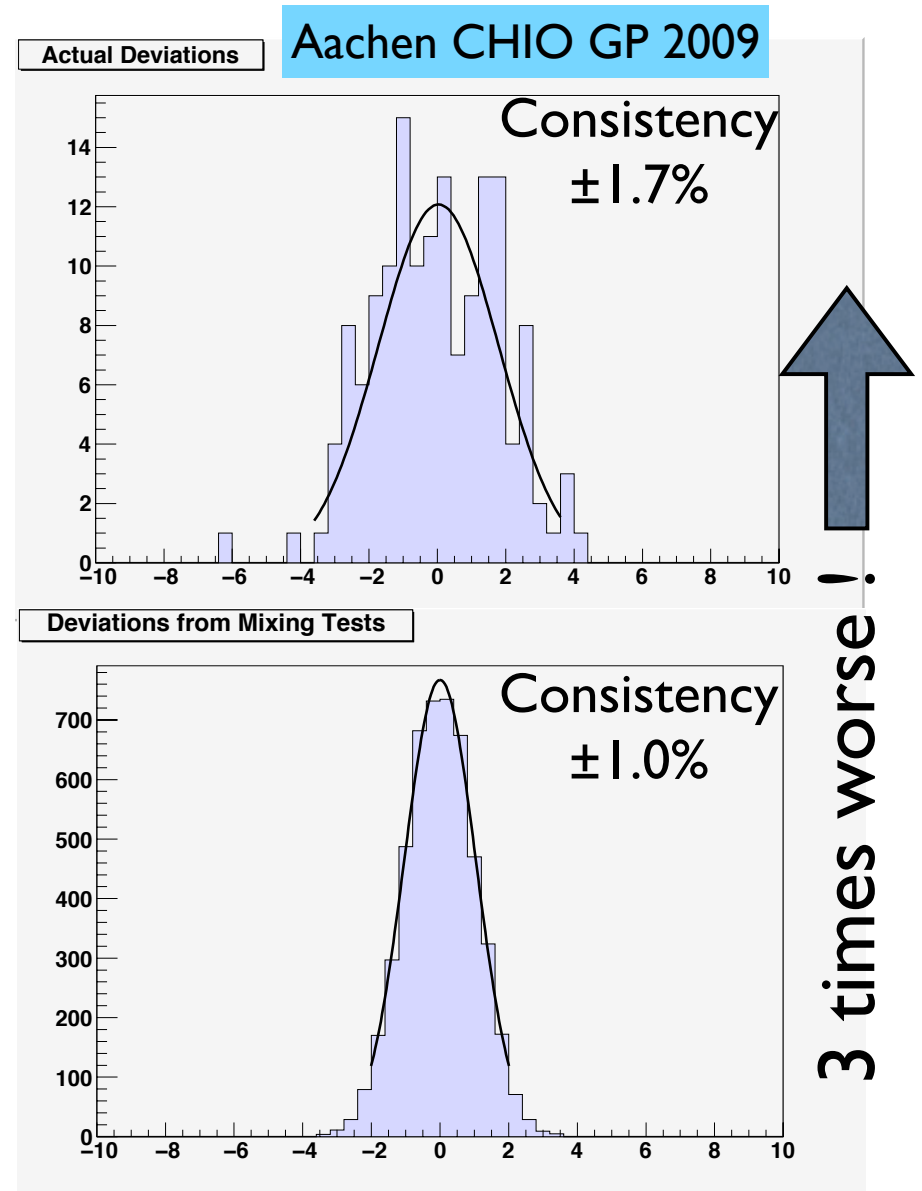
- ➔ Mix figures up randomly (ie take Figure 1 from Rider 6, Figure 2 from Rider 17, Figure 3....) to make fake tests, repeat thousands of time
- ➔ Study the consistency of these randomized tests

Figure	Rider	E	H	C	M	B
Halt-immobility-salute 1	Fiona Bigwood	7	8	6	6	8
Extended trot 2	Carl Hester	7	7	7	7	7
Half-pass right 3	Jeroen Devroe	10	14	14	12	12
Half-pass left 4	Tinne Vilhelmson-Silfven	14	14	12	14	14
Rein back 5 steps 5	Tinne Vilhelmson-Silfven	6	7	5	6	6
Extended trot 6	Stefan van Ingelgem	7	7	7	7	7
Passage 7	Stefan van Ingelgem	7	7	8	7	6
Piaffe 12 to 15 steps 8	Anky van Grunsven	8	8	9	9	8
Transitions passage-piaffe-passage 9	Tinne Vilhelmson-Silfven	7	8	8	6	7
Passage 10	Anna Merveldt	6	6	6	7	7
Extended walk 11	Matthias Alexander Rath	16	16	16	16	18
Collected walk 12	Hans Peter Minderhoud	14	16	14	14	16
Transition collected walk-passage 13	Stefan van Ingelgem	6	7	7	8	8
Passage 14	Ellen Schulten-Baumer	7	7	7	8	8
Piaffe 12 to 15 steps 15	Hayley Beresford	7	7	6	8	7
Transitions passage-piaffe-passage 16	Marcela Krinke Susmelj	7	7	7	8	8
Passage 17	Jeroen Devroe	6	7	7	7	6
Collected canter 18	Ellen Schulten-Baumer	7	8	7	6	7
9 flying changes every 2nd 19	Lyndal Oatley	6	7	6	8	7
Extended canter 20	Catherine Haddad	7	7	7	7	6
Flying change of leg 21	Heike Kemmer	7	8	7	7	7
5 half-passes 22	Tinne Vilhelmson-Silfven	14	16	14	16	16
15 flying changes every stride 23	Anky van Grunsven	14	16	18	16	14
Pirouette left 24	Mikala MÅnter Gundersen	8	12	6	6	10
Flying change of leg 25	Marcela Krinke Susmelj	7	7	6	7	7
Pirouette right 26	Victoria Max-Theurer	12	18	14	14	14
Collected trot 27	Steffen Peters	7	8	7	7	8
Extended trot 28	Marcela Krinke Susmelj	7	6	6	7	6
Passage 29	Stefan van Ingelgem	7	7	8	7	6
Piaffe 12 to 15 steps 30	Sander Marijnissen	6	6	7	6	5
Transitions passage-piaffe-passage 31	Ulla Salzgeber	7	7	7	7	7
Passage 32	Fiona Bigwood	7	7	6	8	7
Halt-immobility-salute 33	Andreas Helgstrand	7	6	7	6	7
Paces 34	Jeroen Devroe	7	7	7	6	7
Impulsion 35	Tinne Vilhelmson-Silfven	7	7	7	7	7
Submission 36	Anky van Grunsven	14	14	14	14	14
Rider's position and seat 37	Ellen Schulten-Baumer	14	14	14	16	16
	Final %	68.26	74.13	69.78	71.30	71.96

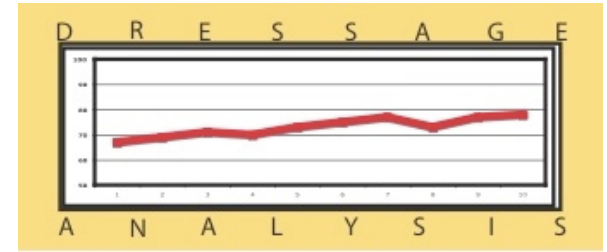
Randomized Combinations



- In the top plot the consistency achieved in a real event
- In the bottom plot the consistency achieved in the randomized combinations
 - ➔ Remarkably, the judges do better on the fake tests than the real ones
 - What is happening?

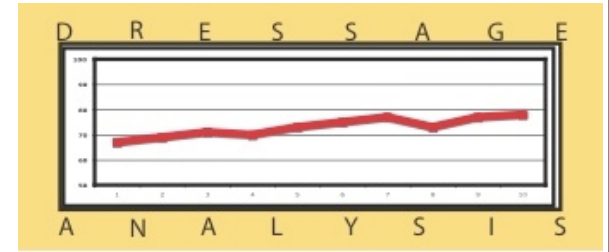


Anonymous tests tell us what is really happening



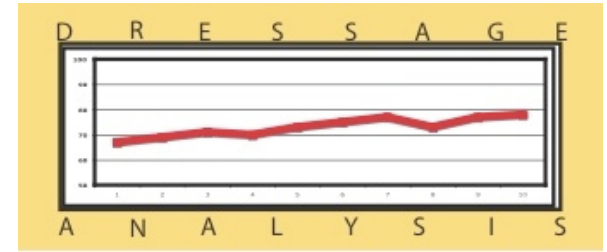
- The **only difference** between the fake tests and the real tests is that in the fake tests there are **no correlations** between the figures, they come from different riders.
 - ➔ Ergo, Judges are less consistent with real riders **because** of the correlations.
 - Combination Bias/Preference
 - Judges have a bias, up or down, for the whole test on particular riders/horses
 - This results in increasing their inconsistency from an intrinsic $\pm 1\%$ to an actual $\pm 1.7\%$ (for this event).
 - **This is an enormous effect.**

Judging as a Team Sport



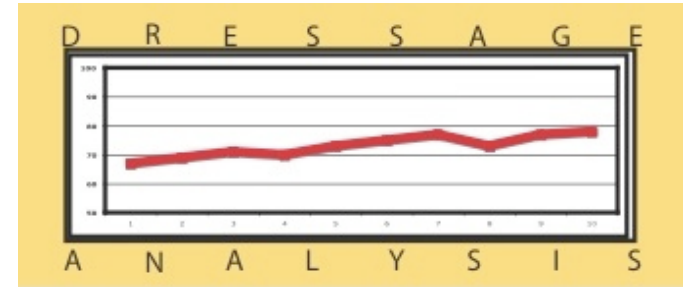
- The judging-team does limit the impact of mistakes
 - ➔ But this team needs every member to be at their peak
 - ➔ Every judge has an equal voice in the jury.
 - The best ones on the day get the same influence on the score as the weaker ones.
 - *(Unlike a real team sport, you can't avoid giving the ball to the weaker judge...)*
- And what about for the rest of us who do not have 5 judges, but 3, or 2 or only 1?
 - ➔ Every judge is entitled to the best feedback and training that we can give them so they can perform at their best, always

Summary I

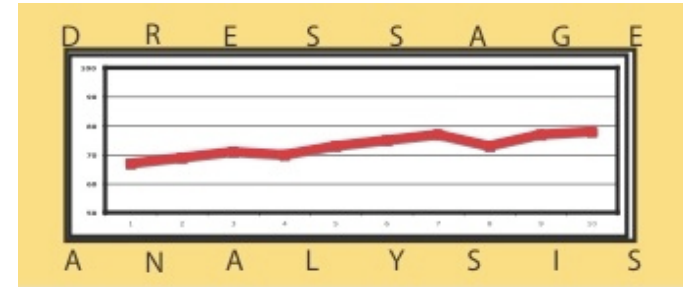


- Averaged over thousands of tests, judges show a precision per judge of about $\pm 1.6\%$
 - ➔ But, 70% of all riders get $65 \pm 4\%$, so with this precision it is hard to get the ranking right for most riders
- The ultimate precision possible with today's system is about $\pm 0.5\%$
- Randomized tests tell us that without the “combination-bias” judges could achieve close to $\pm 1\%$ agreement.
- This would be a fantastic improvement, and judges could do it if they would really be able to do what the manual says,

“Judge each figure on its merits”

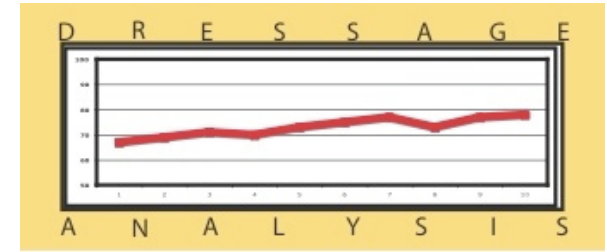


Pause. Let brains rest....



Part II: Windsor EC2009.

Windsor 2009 Consistency

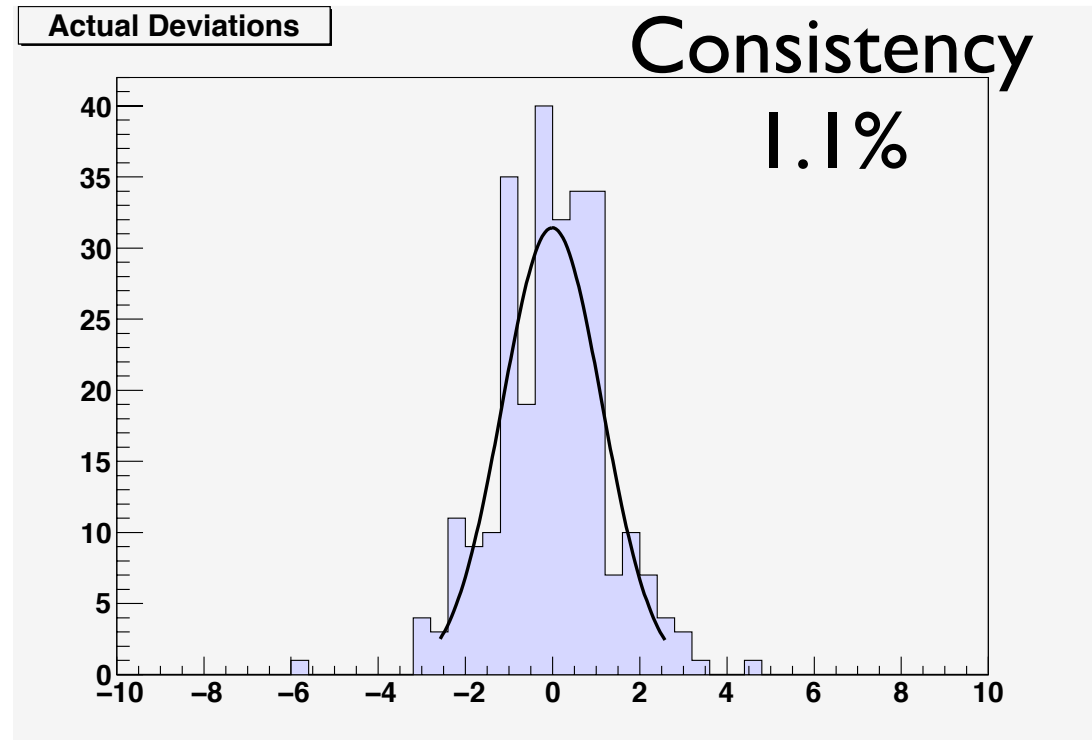


- The basic measure of judging consistency is about the best I have seen in a major event over 5 years.

➔ 1.1% for the GP

➔ 1.2% for the GPS

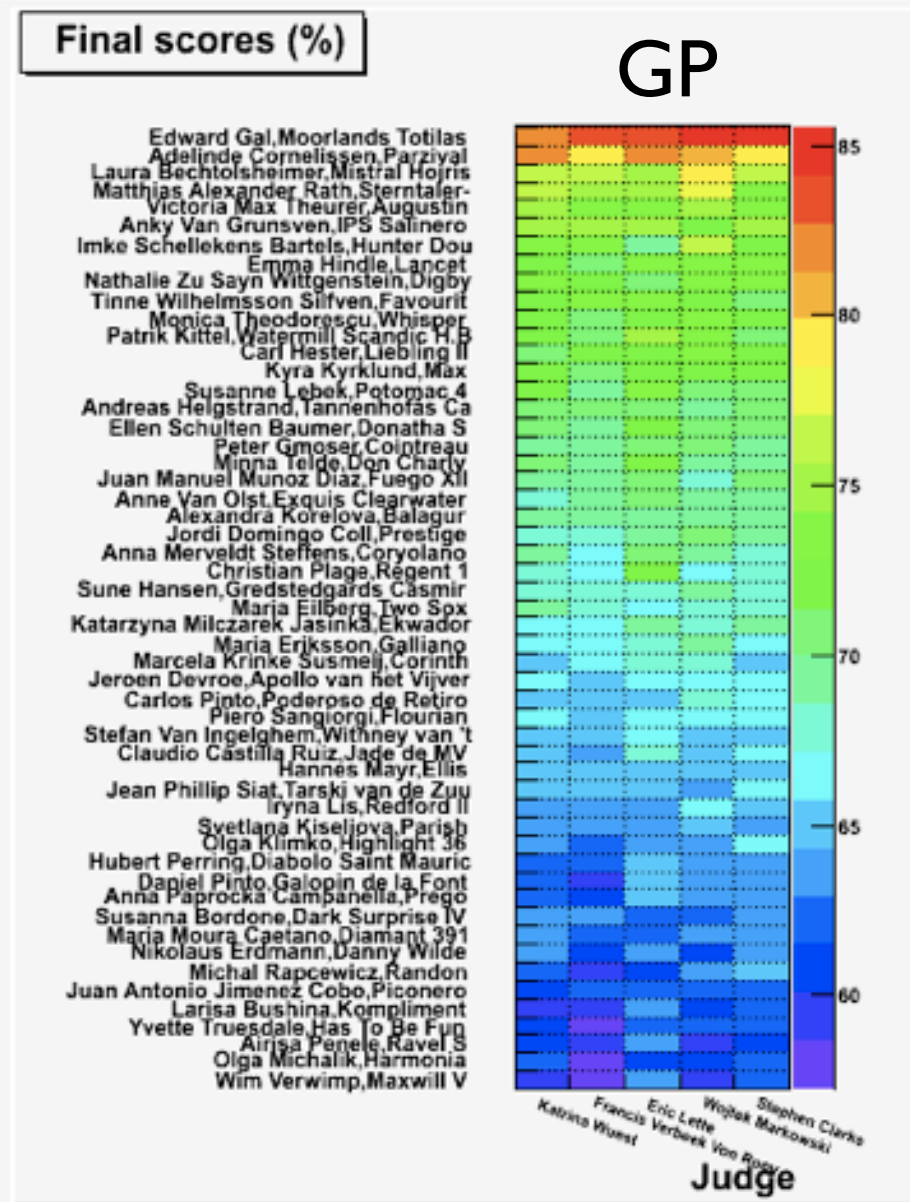
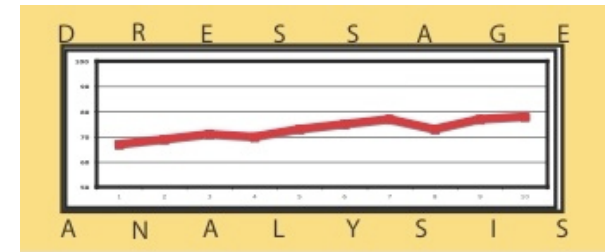
➔ 1.8% for the GPK
(Compared to typically 2%)



These judges achieved 1.1% for the GP and 1% for the “Randomized Tests”. They did the best I have ever seen at reducing “Combination-Bias”.

(ps: The same judges at Aachen achieved 1.7%...???)

Final Scores: No Glaring Problems



- ➔ In this figure I show graphically the final scores
- Red being highest
- Violet being lowest
- “Out of Place” colors can show you quickly if any large discrepancies are present
- GP, GPS, GPK all look good!

Overall Analysis GP

Judge Katrina Wuest at position E	Value	95% CL	99% CL	Consistency
Statistical Technical Diff to other 4 judges	0.8%	> 3.0%	> 4.5%	8/10
Average Technical Score Difference to Other Judges	-0.1%	>± 0.4%	>± 0.6%	8/10
Gave Highest Technical Score	5	> 18	> 26	10/10
Gave Lowest Technical Score	9	> 18	> 26	10/10
Technical Scores with Probability less than 1%	0	>1	>2	10/10
Rank Changes in top 5 Ranks	0	> 3	> 6	10/10

Judge Verbeek scored on average 1.1% less than the others.

Judge Lette 0.7% higher (This does not change rankings)

Judge Francis Verbeek Von Rooy at position H	Value	95% CL	99% CL	Consistency
Statistical Technical Diff to other 4 judges	1.2%	> 3.0%	> 4.5%	8/10
Average Technical Score Difference to Other Judges	-1.1%	>± 0.4%	>± 0.6%	2/10
Gave Highest Technical Score	4	> 18	> 26	8/10
Gave Lowest Technical Score	26	> 18	> 26	4/10
Technical Scores with Probability less than 1%	3	>1	>2	2/10
Rank Changes in top 5 Ranks	0	> 3	> 6	10/10

3 "low probability scores"

Judge Eric Lette at position C	Value	95% CL	99% CL	Consistency
Statistical Technical Diff to other 4 judges	1.7%	> 3.0%	> 4.5%	6/10
Average Technical Score Difference to Other Judges	0.7%	>± 0.4%	>± 0.6%	2/10
Gave Highest Technical Score	25	> 18	> 26	4/10
Gave Lowest Technical Score	8	> 18	> 26	10/10
Technical Scores with Probability less than 1%	2	>1	>2	4/10
Rank Changes in top 5 Ranks	1	> 3	> 6	8/10

2 "low probability scores"

Judge Wojtek Markowski at position M	Value	95% CL	99% CL	Consistency
Statistical Technical Diff to other 4 judges	1.5%	> 3.0%	> 4.5%	8/10
Average Technical Score Difference to Other Judges	0.3%	>± 0.4%	>± 0.6%	6/10
Gave Highest Technical Score	17	> 18	> 26	6/10
Gave Lowest Technical Score	7	> 18	> 26	10/10
Technical Scores with Probability less than 1%	1	>1	>2	8/10
Rank Changes in top 5 Ranks	1	> 3	> 6	8/10

1 "low probability scores"

Judge Stephen Clarke at position B	Value	95% CL	99% CL	Consistency
Statistical Technical Diff to other 4 judges	1.1%	> 3.0%	> 4.5%	8/10
Average Technical Score Difference to Other Judges	0.2%	>± 0.4%	>± 0.6%	8/10
Gave Highest Technical Score	8	> 18	> 26	10/10
Gave Lowest Technical Score	9	> 18	> 26	10/10
Technical Scores with Probability less than 1%	0	>1	>2	10/10
Rank Changes in top 5 Ranks	0	> 3	> 6	10/10

GP

Low Probability Scores.
 → Where one or more of the final scores does not fit in with the expectation of ~1.6% consistency

Francis Verbeek van Rooy
 3 large deviations
 all towards bottom of the ranking

Daniel Pinto - Galopin de la Font - POR		Test Judging Consistency 1.0 %: 6/10					
	Rk	Score	E	H	C	M	B
Final Scores	42	63.11	62.55	60.00	65.32	63.40	64.26
Score Diff.			-0.69	-3.88	2.77	0.37	1.45
This Judge's Rank			44	47	35	42	41
Final Rank Changes			1	-1	4	2	3
Final Score Consistency			10/10	4/10	7/10	10/10	9/10

Yvette Truesdale - Has To Be Fun - IRL		Test Judging Consistency 1.7 %: 6/10					
	Rk	Score	E	H	C	M	B
Final Scores	50	60.89	60.64	57.66	62.34	61.91	61.92
Score Diff.			-0.31	-4.04	1.81	1.28	1.29
This Judge's Rank			51	52	49	48	49
Final Rank Changes			0	-1	0	2	1
Final Score Consistency			10/10	4/10	9/10	9/10	9/10

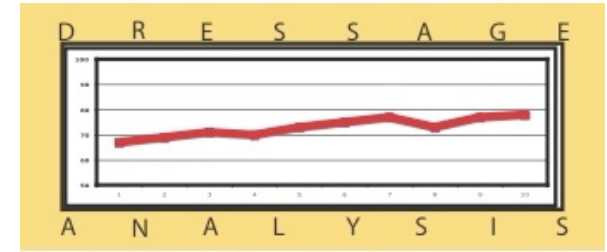
Wim Verwimp - Maxwill V - BEL		Test Judging Consistency 0.38 %: 2/10					
	Rk	Score	E	H	C	M	B
Final Scores	53	60.43	60.00	57.23	62.98	60.00	61.92
Score Diff.			-0.53	-3.99	3.20	-0.53	1.87
This Judge's Rank			52	53	46	52	49
Final Rank Changes			-1	-2	0	0	0
Final Score Consistency			10/10	4/10	6/10	10/10	9/10

Christian Plage - Regent 1 - SUI		Test Judging Consistency 0.74 %: 4/10					
	Rk	Score	E	H	C	M	B
Final Scores	25	68.09	68.09	67.02	71.49	65.96	67.87
Score Diff.			0.01	-1.33	4.26	-2.65	-0.27
This Judge's Rank			27	26	15	32	26
Final Rank Changes			0	0	2	-2	0
Final Score Consistency			10/10	9/10	2/10	7/10	10/10

Eric Lette
 2 large deviations
 Imke's result: *Eric Lette Ranks 25th*
 Wojciech Markowski 5th

Imke Schellekens Bartels - Hunter Douglas Sunrise - NED		Test Judging Consistency 0.007 %: 2/10					
	Rk	Score	E	H	C	M	B
Final Scores	7	73.15	73.62	73.19	69.15	77.02	72.77
Score Diff.			0.59	0.05	-5.00	4.84	-0.47
This Judge's Rank			7	7	25	5	9
Final Rank Changes			0	1	-2	3	0
Final Score Consistency			10/10	10/10	0/10	1/10	10/10

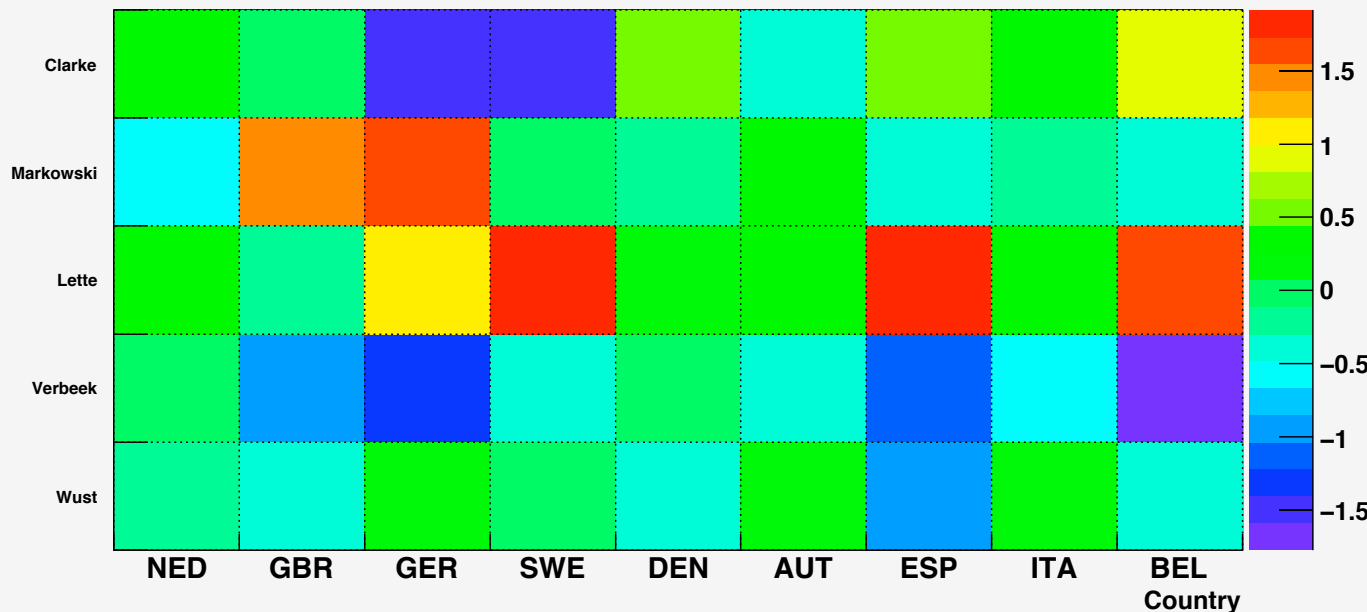
Nationalistic Judging?



Nat	Final	Score/Rank Without This Judge									
		Wust	Rk	Verbeek	Rk	Lette	Rk	Markowski	Rk	Clarke	Rk
NED	79.53	79.57	1	79.52	1	79.47	1	79.65	1	79.45	1
GBR	73.89	73.97	2	74.06	2	73.92	2	73.58	2	73.90	2
GER	73.08	73.03	3	73.35	3	72.87	3	72.77	3	73.37	3
SWE	71.52	71.52	4	71.61	4	71.13	4	71.51	4	71.81	4
DEN	71.04	71.12	5	71.05	5	71.01	5	71.06	5	70.94	5
AUT	69.70	69.68	6	69.79	6	69.65	6	69.63	6	69.77	6
ESP	68.04	68.24	7	68.28	7	67.66	7	68.10	7	67.93	7
ITA	64.01	63.99	9	64.13	9	63.95	8	64.06	8	63.94	8
BEL	63.97	64.04	8	64.33	8	63.65	9	64.06	8	63.78	9

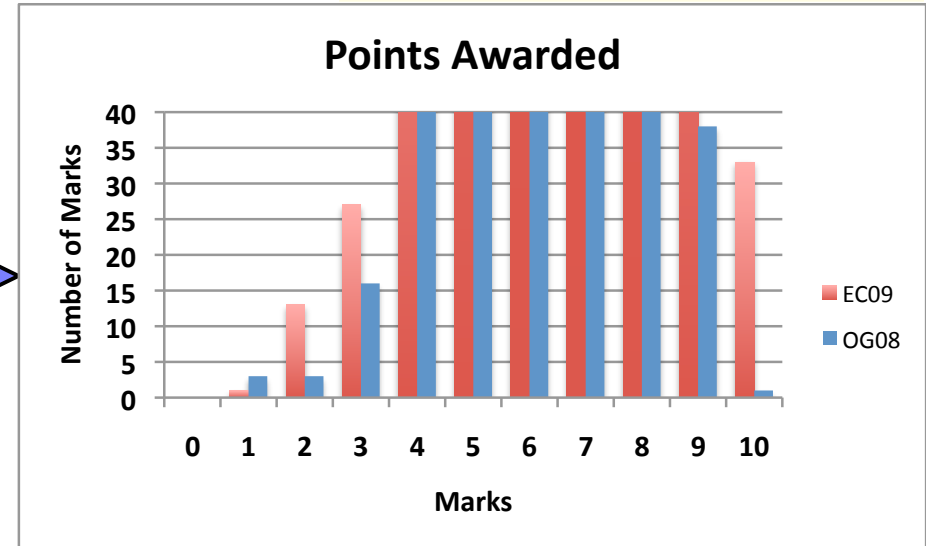
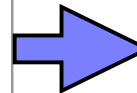
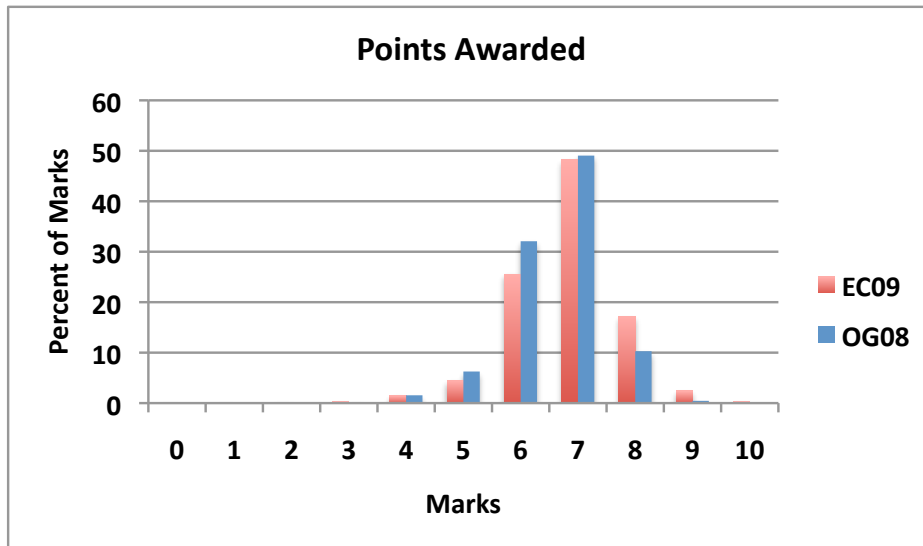
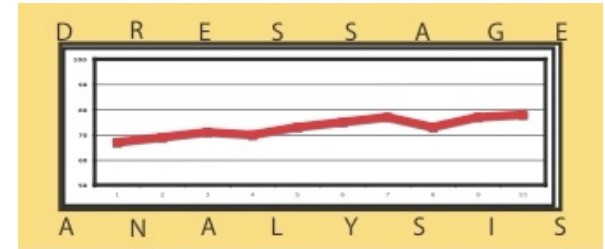
NO INDIVIDUAL JUDGE
CHANGED THE FINAL
NATIONAL RANKING

Countries Rank the Judges !

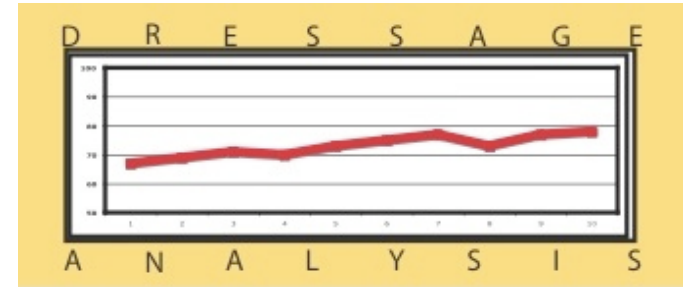


NO CLEAR
PATTERNS OF
NATIONALISTIC
JUDGING?

Use of the Points Scale

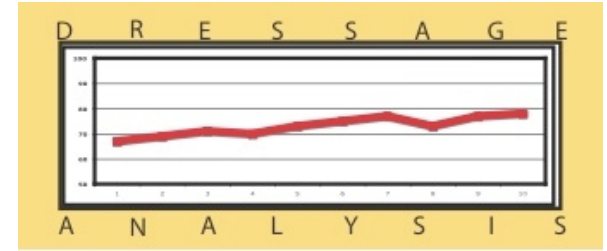


- Due to the presence of Moorlands Totilas (25) and Parzival (7) the population of the 10 bin is radically different this year!
Not to forget Peter Gmoser and Cointreau who also received a 10
- ➔ Even the “9” bin has five times more population this year than last year.
- Still, 50% of all points used are “7”,
 - ➔ 74% of marks are a 6 or a 7 (was 81% on OG)



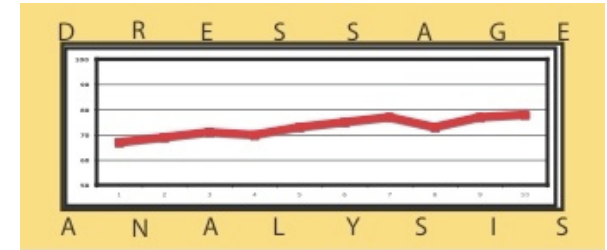
Part III: Types of judging problems

Causes of Inconsistency



- **“Mistake/Viewpoint”**,
 - ➔ Typically due to a poor view of a figure, lapse of attention, etc
 - The judge would probably change the score if shown the figure again or from a different angle
- **“Error”**,
 - ➔ Figure scores that are out of line with the “normal scores” for a particular performance.
 - Even after review, the judge may feel this is the “right” score.
- **“Bias”**,
 - ➔ Small deviations figure by figure that become a large inconsistency overall.
 - A general tendency of the judge to be high or low, for this test

Mistakes/Errors



- Select figures with more than 2 points difference between one judge and the average of the others

Rider	Figure	E	H	C	M	B	Effect
Gal, Edward Moorlands Totilas	Halt 33	10	10	7	9	8	-0.10
Cornelissen, Adelinde Parzival	Collected Canter EKAF 18	5	5	6	4	8	0.13
Cornelissen, Adelinde Parzival	Collected Canter KA 21	8	8	7	7	5	-0.11
Schellekens-Bartels, Imke Hunter Douglas Sunrise	Changes 23	5	5	4	8	5	0.28
Schulten-Baumer, Ellen Donatha S	Passage DFP 10	7	7	5	7	8	-0.10
Merveldt, Anna Coryolano	Halt Rein-Back 5	6	5	3	6	6	-0.12
Siat, Jean Philippe Tarski van de Zuuth	Flying Change 25	4	4	3	4	6	0.10
Klimko, Olga Highlight 36	Extended Walk 11	8	7	8	5	7	-0.21
Perring, Hubert Diabolo St Maurice	Transition 31	4	5	7	5	5	0.10
Caetano, Maria Diamant 391	Zig-Zag 22	4	6	4	3	4	0.19
Bushina, Larisa Kompliment	Collected Walk 12	5	4	7	5	5	0.19

- Judge Supervisory Panel “JSP” or even an automatic correction would be able to solve many of these problems

Mistakes & Errors?

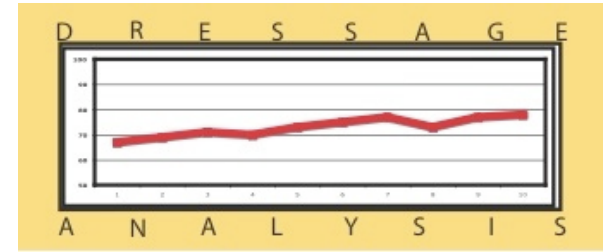


Figure Scores					
Figure	E	H	C	M	B
Transition 34	4	6	7	2	6
Piaffe 33	4	5	7	1	5
Passage 32	8	7	7	7	7

- Matthias Rath had not enough steps on his final piaffe (and traveling?) on the centerline.
 - ➔ Judges E,H,B quite consistent with 4,5,5, and 4,6,6 for the transition
 - ➔ Judge C could not see the traveling, and the steps are harder to see... so his 7 was a justifiable “mistake”? A JSP could presumably correct this.
 - ➔ Judge at M gave a 1 for the piaffe and a 2 for the transition.
 - **Something can't be right.** If M gave the correct combined score of $1+2=3$, then why did EHB give an average combined score of 10, or vice-versa?

E	H	C	M	B
73.62	73.19	69.15	77.02	72.77

“Bias”

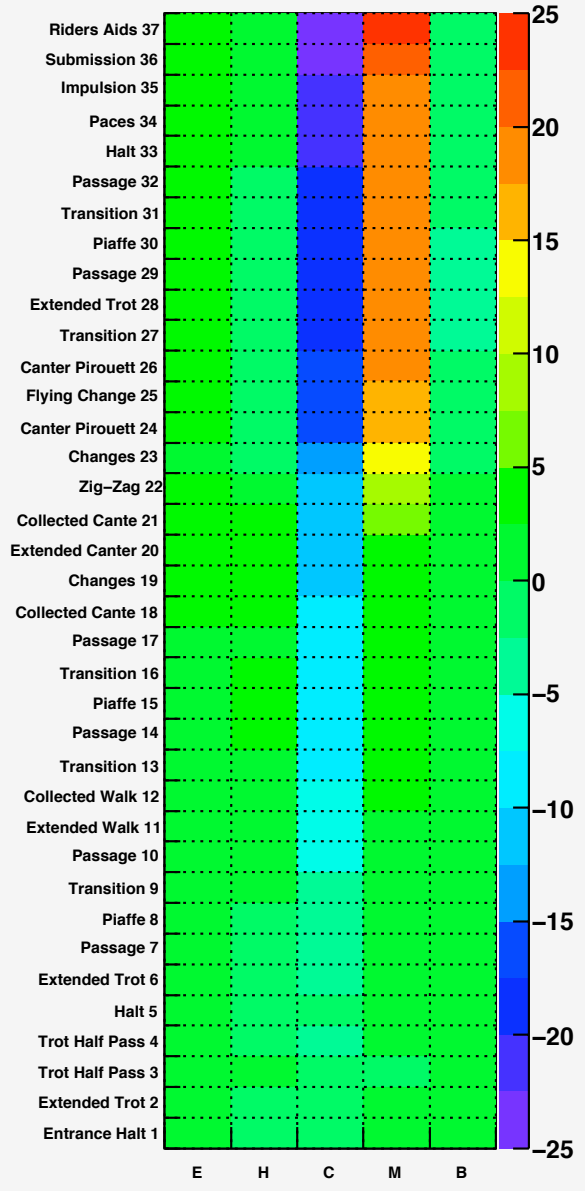
E	H	C	M	B
67.07	66.04	61.46	66.04	64.79

Schellekens-Bartels, Imke N - Hunter Douglas Sunr

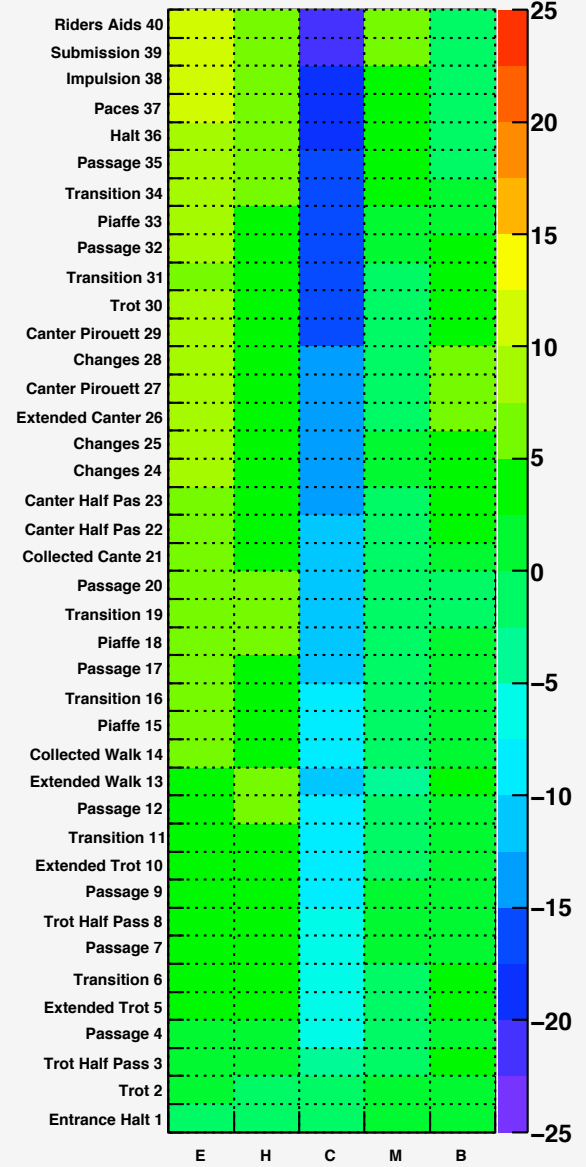
Helgstrand, Andreas D - Tannenhof's Carabas

GP

GPS



Cumulative Judge-Ave(Other4)



Cumulative Judge-Ave(Other4)

Detail Imke GP

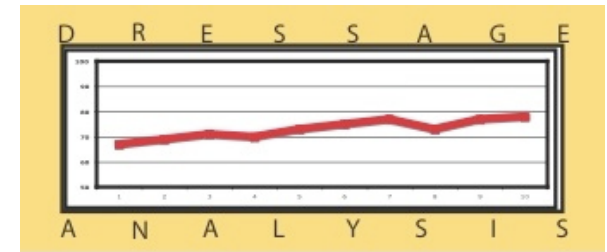
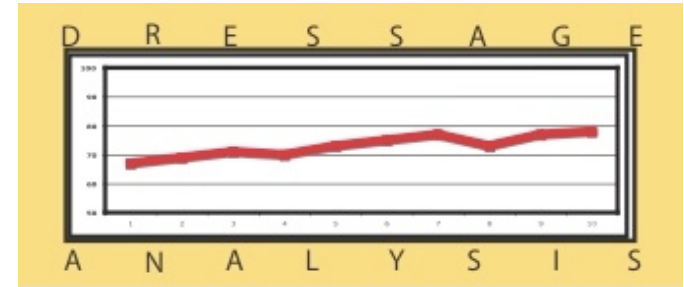


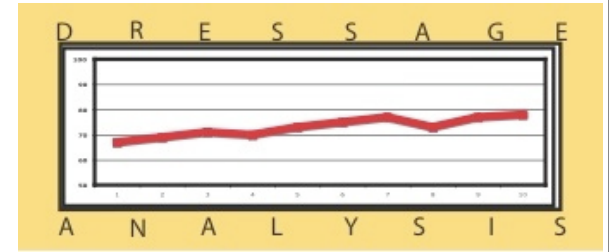
Figure	Figure Scores					Cumulative Scores				
	E	H	C	M	B	E	H	C	M	B
Riders Aids 37	8	8	8	9	8	346	344	325	362	342
Submission 36	7	7	7	8	7	330	328	309	344	326
Impulsion 35	8	8	7	8	8	316	314	295	328	312
Paces 34	7	8	7	7	7	308	306	288	320	304
Halt 33	7	8	6	8	8	301	298	281	313	297
Passage 32	7	7	7	7	7	294	290	275	305	289
Transition 31	7	8	7	7	8	287	283	268	298	282
Piaffe 30	7	7	6	7	7	280	275	261	291	274
Passage 29	7	7	7	7	7	273	268	255	284	267
Extended Trot 28	8	7	7	7	7	266	261	248	277	260
Transition 27	7	8	7	8	7	258	254	241	270	253
Canter Pirouett 26	8	7	7	8	7	251	246	234	262	246
Flying Change 25	7	8	7	8	8	235	232	220	246	232
Canter Pirouette 24	9	8	8	9	8	228	224	213	238	224
Changes 23	5	5	4	8	5	210	208	197	220	208
Zig-Zag 22	8	7	8	9	8	200	198	189	204	198
Collected Canter 21	7	7	8	8	7	184	184	173	186	182
Extended Canter 20	8	8	7	8	8	177	177	165	178	175
Changes 19	8	8	7	8	8	169	169	158	170	167
Collected Canter 18	8	8	7	8	7	161	161	151	162	159
Passage 17	8	7	7	8	7	153	153	144	154	152
Transition 16	7	7	7	7	8	145	146	137	146	145
Piaffe 15	7	7	8	7	7	138	139	130	139	137
Passage 14	7	8	6	7	7	131	132	122	132	130
Transition 13	8	8	7	8	7	124	124	116	125	123
Collected Walk 12	7	7	7	8	7	116	116	109	117	116
Extended Walk 11	7	7	6	6	7	102	102	95	101	102
Passage 10	8	8	7	8	8	88	88	83	89	88
Transition 9	7	8	7	7	7	80	80	76	81	80
Piaffe 8	7	8	7	8	7	73	72	69	74	73
Passage 7	7	7	7	7	7	66	64	62	66	66
Extended Trot 6	8	8	6	8	8	59	57	55	59	59
Halt 5	6	6	8	7	7	51	49	49	51	51
Trot Half Pass 4	8	7	7	8	7	45	43	41	44	44
Trot Half Pass 3	7	8	7	7	8	29	29	27	28	30
Extended Trot 2	8	7	7	7	7	15	13	13	14	14
Entrance Halt 1	7	6	6	7	7	7	6	6	7	7

- Blue Boxes are the lowest at each figure
- Red boxes are the highest
- Judges at E,H,B are extremely close throughout the test !
- The judge at C end up 20 points below E,H,B due to being the lowest score at almost every figure
 - ➔ None of his scores are extreme, but the combined effect is -5% and 18 places lower in the ranking
 - Maybe half-points will help?
- The judge at M could not see the mistake in the changes so scored 6 points higher,
 - ➔ But also he is highest in almost every canter figure



Part IV: Solutions, for Discussion

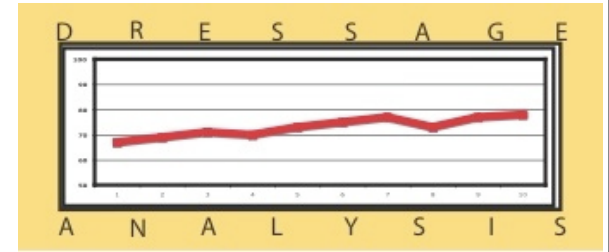
The rest of us



- Most of us do not perform at CDI, or even Grand Prix, level
- We seldom get judged by 3 judges, virtually never by 5
- We seldom get judged by FEI level judges
- In many countries a single judge is the norm for most competitors.
 - ➔ I don't *know* how accurate all those other judges are, but I suspect it is not as good as 1.6%...
 - ➔ So everything in this talk applies at national level, and probably more-so!
 - ➔ Including of course, the solutions....

Solutions I

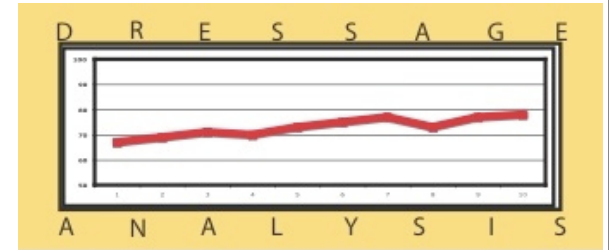
(Simple Changes to the system)



- Reviewing the Mistakes/Errors
 - ➔ Judge Supervisory Panels could fix the more obvious mistakes.
 - (But the proposal from the DTF only addresses a subset of mistakes/errors and only with one sign!)
 - Actually I suspect an Automatic System for large outlying scores would fix more problems than it introduced. It should be examined with a larger set of data from a range of CDI types.

Solutions I (contd)

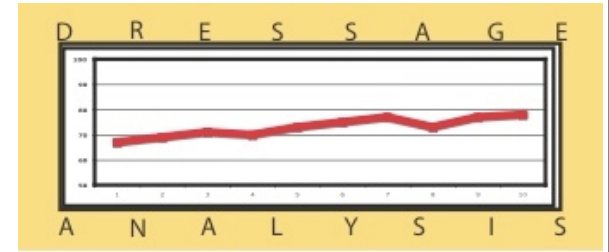
(Simple Changes to the system)



- Half-Points are an idea whose time has come!
 - ➔ Judges (some, not all) use half-points today
 - ➔ It is much better to give them the tools they need to use them correctly than to have them used by some judges and not by others.
 - ➔ Anything that encourages/allows judges NOT to consider the other figures when scoring the one in front of them would be a significant help

Solutions II.

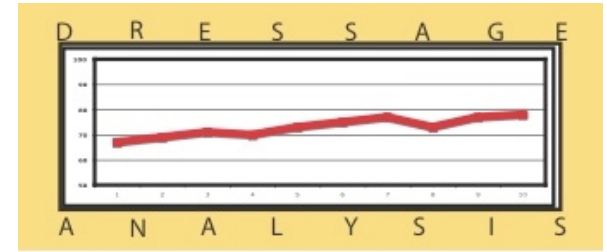
Major Changes to the System



- Radical changes could include decomposition of the movements into their “training-scale” components, and having (some) judges measuring only these components (straightness, impulsion etc)
- This is closer to the system used in gymnastics/skating.
 - ➔ It has the advantage of being very fine-grained so it is possible to develop a more exact code of points.
 - And the rider gets very explicit detailed indication of the faults to correct
 - Was tested at fairly successfully in Aachen, I think after full analysis it should be investigated in more studies, with better prepared judges - it is not any easy transition.

Solutions III

Judge Development



- **Feedback:**

- ➔ Judges currently get no formal feedback after an event. *Lets change that*

- **Definition:**

- ➔ The FEI Dressage handbook is a great start, but it is not a code of points as exists in other similar sports. *A Video handbook would be excellent*

- **Training:**

- ➔ Equal opportunity judge training worldwide does not exist. *e-learning*

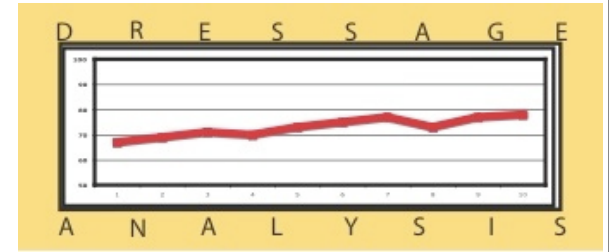
- ➔ Training seminars are important, but they are infrequent (in space and time)

- **Testing:**

- ➔ Testing should be an an integral part of initial appointment and of ongoing in-service training and skills refinement. *e-learning/testing*

Solutions IV

Openness



- Open scoring is the most powerful tool we have for the advancement of the sport.
- Analysis of the results and feedback to judges, riders and organizers can only help everyone involved.
- The Dressage public is in fact an educated and concerned population.
 - ➔ When you go to a soccer match you want to see how the winning team wins, who scored the goals, not just to be told by the referee who won. The same applies to Dressage...