CIVA Judging Seminar

The CIVA FairPlay system as used in the ACRO software

This is extracted from the explanation in the “Help” section of the ACRO program, and describes what ACRO’s built-in FPS coding does.

The FairPlay System (FPS) - why do we need a "system" at all?

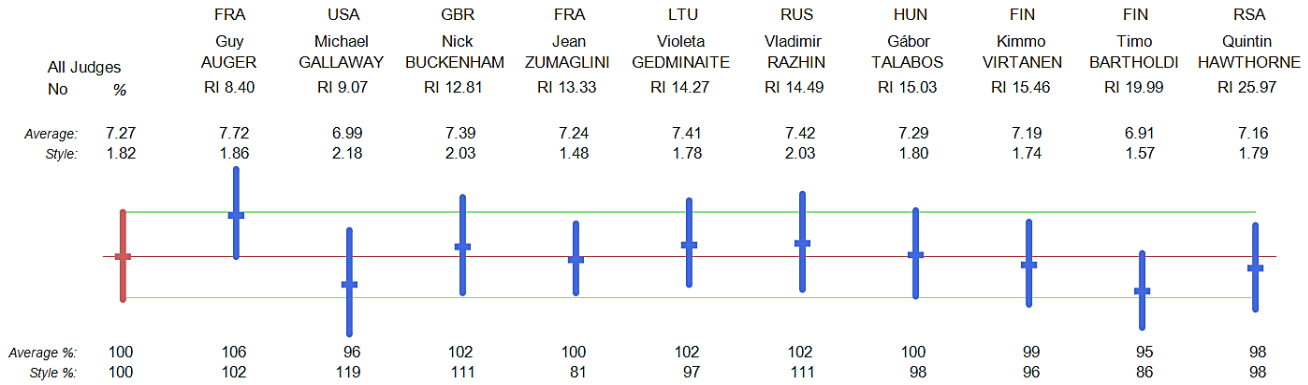
Aerobatic sequences are usually judged by 5-8 judges, and it is unlikely that each judge will see the same 'faults' and assess them in precisely the same way. Also each judge’s experience and perhaps their prior time as a competitor will influence their personal style to favour or disadvantage some pilot, aeroplane and flight characteristics. Because in our marking system we subtract the faults we see from a bank of ten marks, the ‘kind’ or inexperienced judge will tend to give higher marks and so be more influential than his harsher colleagues – the reverse of what we would like to see!

These unavoidable human characteristics create marks variations that can be significant and will put the fairness of the result in doubt. While minor anomalies can be casually left to 'average out' between the judges, in instances where one or more judges marks clearly do not fit the overall panel view or even their own style of marking it would be unreasonable to ignore them. For these 'unusual' marks a carefully engineered detection and resolution system is essential. This can also provide the foundation for a thorough analysis of the performance of each judge in comparison to his or her peers, a vital tool in judge assessment and longer term training.

What does FairPlay do?

1. <u>It divides the marks into suitable “Groups” for analysis</u>:
First the system uses figure type (Aresti family and SuperFamily) and figure complexity (K factor) to divide all the Judges' marks into tabular groups of data, so that within each group the pilots have all executed identical or very similar figures. This ensures that the judging expertise applied to each figure was confined to a relatively narrow range, and the marks should be very similar. In this way the system strives always to compare like with like.

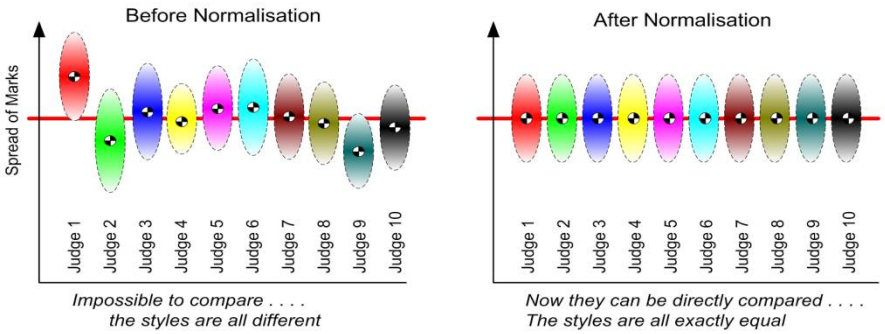
2. <u>It “normalises” the marks in each Group</u>:
In every sequence the judging ‘styles’ will always be different, even though all judges see the same things and all follow the same downgrading rules to arrive at their marks. For example, here is the Chief Judges Raw Grades graphic for all figures in the 2nd Free Unknown at WAAC 2012:



You can see that their average marks vary from 6.91 to 7.72 (+/- 0.406) and the spread of marks (here shown by 2x the Standard Deviation for each judge) ranges from 1.48 to 2.18 (+/- 0.350).

Within the figure data groups FairPlay 'normalises' each Judges complete set of marks, to level or balance them by comparison with the other Judges. To do this, FPS moves each judges' whole set of marks up or down so that the average becomes the same as the all-judges average, and at the same time it increases or reduces the vertical spread of marks so they all become equal to the all-judges average spread. For each judge this doesn't change the *relative* marking of each pilot, but as all judges have now been brought to the same 'style' it is possible to make direct comparisons between the judges' marks for each figure.

After normalisation each Judge will have equal status within the group, the effects of experience and style are effectively eliminated, and the marks can be assessed figure-by-figure / judge-by-judge on a fair and equivalent basis.



This 'normalisation' method is commonly used as a first stage in many areas of numeric analysis, to ensure that apparently similar streams of data are free of underlying style and bias differences and may thereafter be reliably compared with each other.

3. <u>It looks for "odd" or "unusual" grades and resolves them</u>:
Now FairPlay seeks out unusual or ill-fitting marks within each group on a statistical basis, by testing each against a 'uncertainty' value that has been calculated for each judge and each figure. Any marks that fail this uncertainty test are considered to be unsuitable and are discarded, in each case being proportionally replaced by a Fitted Value that is calculated after all the anomalies have been removed – i.e. it is uninfluenced by the discarded "bad" marks.

4. <u>It calculates pilots points totals per judge, then looks for "unusual" scores and resolves them</u>:
The adjusted marks provided by the above stages can now combined with their K-factors to

produce a points total for each judge / each figure flown, and these are totalled to provide an overall score per judge for each pilot. The last stage of FairPlay is to use the normalisation process again, this time on the judges' scores for each pilot, to ensure that any remaining overall bias is detected and eliminated. The scores are now completely free of any detectable anomalies, and can reliably be used create the table of results.

5. It subtracts the pilots Penalties:
Finally any penalties that have been given are subtracted from the marks of the pilots concerned, and the final table of results can be published.

Using FairPlay as a training tool

ACRO provides several Judging Analysis formats to complement the Ranking Index required by CIVA, both printable and uploaded to the web that displays the contest results. These look for patterns in the judging that differ from the majority view, and can be used to provide advice or more extended training to address the issues seen.

a) The Judges Individual Sequence Analysis:
This is a printed and/or online personal report showing every "raw" mark given for every pilot, together with the FPS handling of each mark and the sequence marks, with boxes added to show data that the system has changed. Pilots are ranked by the panel FPS mark before penalties are applied, and a comparison made with the judges personal ranking after the minimum processing has been applied to resolve requests for "AV"s; note that for RI purposes averages are replaced by raw grades calculated to the nearest half mark. The changed data is summarised and a histogram given that shows the judges use of each possible mark (10.0 to 0.5, SZ and HZ) compared to the FPS panel average – this very clearly reveals for instance the uneven use of marks of whole marks and half marks. The judges' CIVA indexes are given too.

To see a Judge's Individual Sequence Analysis online – visit a single sequence results page (i.e. not the overall all-sequences one) and click a Judge's name at the foot of the page.

b) The Overall Judging Analysis:
This printed report is for the Chief Judge only, and can be produced either for one single sequence or a collection of sequences. It collates the data from all the Judges in the many categories assessed by FPS, as shown in each judges' individual sequence analysis sheet. The report allows the Chief Judge to review and compare the performances of each judge within each FPS area, and if necessary to discuss with a judge his/her handling of elements that he feels would benefit from additional attention. This is the data that is collated by CIVA and used as the basis for ongoing international judge selections: the Ranking Index for each judge / for all sequences at each event is averaged and added to the CIVA Judging Database to give a rolling three-year average RI for every judge, providing the main criterion used at the initial selection stage each year.

Judges Single Sequence Processed Marks Analysis for J2 - Some Judge

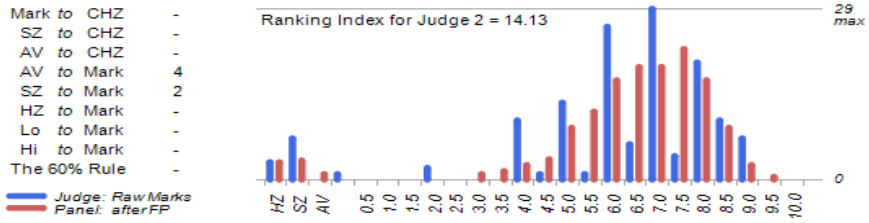
Icicle & Newbold Trophies, Sandtoft, 10th April

Level: Advanced - Power, Sequence: 1st Unknown Sequence

FP Rank before Penalties	Fig-1	Fig-2	Fig-3	Fig-4	Fig-5	Fig-6	Fig-7	Fig-8	Fig-9	Fig-10	Posi	No.	From RI marks x K's: RI / FP	J-rank	Diff
1: 007 Alan Cassidy	*7.0*	*7.0*	*7.0*	*6.0*	*8.5*	*9.0*	*8.0*	*8.5*	*7.0*	*6.5*	*8.0*	*0*	*2125.7*	*1*	
Pitts S-1T N-666BM	7.35	7.35	7.09	6.10	8.04	8.00	7.67	8.30	6.82	6.05	8.05		2113.9		
2: 010 Nick Wakefield	*7.0*	*8.0*	*8.5*	*7.0*	*6.5*	*8.5*	*6.0*	*7.0*	*7.0*	*7.5*	*7.5*	*0*	*2103.8*	*2*	
Pitts S-1T G-WILD	7.35	8.17	8.35	7.41	6.53	7.69	6.39	7.23	6.82	6.89	7.42		2092.3		
3: 024 Paul Tomlinson	*6.5*	*8.0*	*8.5*	*7.0*	*8.0*	*8.0*	*7.0*	*9.0*	*2.0*	*8.0*	*8.5*	*0*	*2078.0*	*3*	
Edge-360 G-ZVKO	7.04	8.17	8.35	7.41	7.66	7.37	7.03	8.66	2.96	7.31	8.68		2068.1		
4: 013 Gary Ferriman	*4.0*	*5.0*	*9.0*	*6.0*	*6.5*	*Lo SZ*	*8.0*	*6.0*	*7.0*	*6.5*	*8.5*	*1*	*1824.3*	*7*	*-3*
Extra-230 G-ROMP	5.48	5.72	8.77	6.10	6.53	7.44	7.67	6.52	6.82	6.05	8.68		2007.5		
5: 027 Marco Kalweit	*9.0*	*8.0*	*7.0*	*6.0*	*8.5*	*8.0*	*4.0*	*4.0*	*4.0*	*5.5*	*8.0*	*0*	*1931.0*	*5*	
Pitts S-2C G-FDPS	8.59	8.17	7.09	6.10	8.04	7.37	5.11	5.09	4.51	5.21	8.05		1924.1		
6: 006 Eddie Goggins	*8.0*	*9.0*	*9.0*	*7.5*	*8.5*	*HZ*	*8.5*	*7.0*	*7.0*	*6.0*	*7.0*	*0*	*Hi 1999.1*	*4*	*+2*
Extra-300L G-IJMI	7.97	8.98	8.77	8.07	8.04	HZ	7.99	7.23	6.82	5.63	6.79		1779.8		
7: 016 Ron Allan	*7.0*	*AV*	*7.0*	*7.0*	*4.0*	*5.0*	*8.0*	*6.0*	*SZ*	*4.0*	*7.0*	*1*	*1664.0*	*8*	*-1*
Pitts S-2B G-IIII	7.35	8.19	7.09	7.41	4.64	5.47	7.67	6.52	0.00	3.94	6.79		1668.6		
8: 011 Richard Buchan	*6.0*	*6.0*	*7.0*	*6.0*	*6.0*	*4.0*	*5.0*	*6.0*	*SZ*	*5.0*	*8.0*	*0*	*1608.6*	*10*	*-2*
CAP 231 G-OZZO	6.73	6.53	7.09	6.10	6.15	4.84	5.75	6.52	0.00	4.78	8.05		1618.0		
9: 026 Simon Cattlin	*8.0*	*7.0*	*7.5*	*6.0*	*7.0*	*AV*	*HZ*	*AV*	*Lo SZ*	*AV*	*6.0*	*4*	*1457.4*	*12*	*-3*
Yak-55M G-NOIZ	7.97	7.35	7.51	6.10	6.91	5.64	HZ	7.74	5.45	4.74	5.53		1671.4		
10: 023 Brian McCartney	*7.0*	*7.0*	*8.0*	*6.0*	*6.0*	*6.0*	*4.0*	*7.0*	*5.0*	*HZ*	*6.0*	*0*	*1656.7*	*9*	*+1*
Pitts S-2A G-TIII	7.35	7.35	7.93	6.10	6.15	6.10	5.11	7.23	5.28	HZ	5.53		1662.6		
11: 017 Tony Maxwell	*7.0*	*7.0*	*8.0*	*5.0*	*5.0*	*5.0*	*8.0*	*9.0*	*4.5*	*8.0*	*7.0*	*0*	*Hi 1882.5*	*6*	*+5*
Pitts S-2B G-IIII	7.35	7.35	7.93	4.79	5.40	5.47	7.67	8.66	4.89	7.31	6.79		1663.4		
12: 019 Cas Smith	*8.0*	*7.0*	*6.0*	*6.0*	*4.0*	*2.0*	*6.0*	*5.0*	*SZ*	*6.0*	*6.5*	*0*	*1528.2*	*11*	*+1*
Pitts S-2A G-ICAS	7.97	7.35	6.25	6.10	4.64	3.57	6.39	5.80	0.00	5.63	6.16		1538.0		
13: 028 Andrew Holman-West	*4.0*	*5.0*	*5.0*	*5.0*	*6.0*	*6.0*	*6.0*	*6.0*	*SZ*	*SZ*	*5.0*	*0*	*1266.5*	*13*	
Yak-50 G-YKSO	5.48	5.72	5.41	4.79	6.15	6.10	6.39	6.52	0.00	0.00	4.27		1290.8		
	0	*1*	*0*	*0*	*0*	*2*	*0*	*1*	*1*	*1*	*0*	*6*	*2*		*18*

Use of Marks:

Mark *to* CHZ	-
SZ *to* CHZ	-
AV *to* CHZ	-
AV *to* Mark	4
SZ *to* Mark	2
HZ *to* Mark	-
Lo *to* Mark	-
Hi *to* Mark	-
The 60% Rule	-



Aerobatic Contest Results Organiser, Version 3.0 Build 28-01-11
Calculations by: FairPlay (scoring CJ + CHZ Summary) method
This report created at 11:55 on 26 February 2011

Judges Anomaly Review for J2 - Some Judge

Icicle & Newbold Trophies, Sandtoft, 10th April

Level: Advanced - Power, Sequence: 1st Unknown Sequence

Marking anomalies:		Ch.Judge	**Judge 2**	Judge 3	Judge 4	Judge 5	CHZ's
013 Gary Ferriman	*Fig-6*	*8.0*	**Lo SZ**	*7.5*	*7.0*	*7.0*	*OK*
Extra-230 G-ROMP		7.92	**7.44**	8.03	7.09	6.72	
016 Ron Allan	*Fig-2*	*7.0*	**AVGE**	*8.5*	*9.5*	*8.0*	*OK*
Pitts S-2B G-IIII		7.83	**8.19**	8.15	8.84	7.93	
026 Simon Cattlin	*Fig-6*	*5.0*	**AVGE**	*6.0*	*6.0*	*6.0*	*OK*
Yak-55M G-NOIZ		5.58	**5.64**	6.25	5.56	5.16	
026 Simon Cattlin	*Fig-8*	*6.0*	**AVGE**	*8.0*	*8.5*	*8.0*	*OK*
Yak-55M G-NOIZ		7.08	**7.74**	7.61	8.05	8.21	
026 Simon Cattlin	*Fig-9*	*5.0*	**Lo SZ**	*7.5*	*5.5*	*6.5*	*OK*
Yak-55M G-NOIZ		6.58	**5.45**	7.31	5.24	5.62	
026 Simon Cattlin	*Fig-10*	*5.0*	**AVGE**	*AVGE*	*3.0*	*6.0*	*OK*
Yak-55M G-NOIZ		5.12	**4.74**	4.74	4.03	5.76	
Score anomalies:							
006 Eddie Goggins		1767.0	**Hi 1999.1**	1852.2	1776.5	Lo 1591.1	
Extra-300L G-IJMI			**1779.8**			1850.3	
017 Tony Maxwell		1714.8	**Hi 1882.5**	1560.1	1616.9	1703.7	
Pitts S-2B G-IIII			**1663.4**				

Aerobatic Contest Results Organiser, Version 3.0 Build 28-01-11
Calculations by: FairPlay (scoring CJ + CHZ Summary) method
This report created at 11:55 on 26 February 2011

Analysis of Judges Combined Anomalies

Sequence: Programme 2: FreeUnknown Compulsary (ADV)

9th FAI WAAC
Radom
05. - 15. Aug. 2010

Use of Marks:	All Judges No	All Judges %	RUS Kotelnikov, Vladimir RI 9.92	POL Bialek, Maciej RI 15.38	FRA Courtois, Bernard RI 15.59	UKR Zelenina, Lyudmila RI 16.37	LTU Gedminaite, Violeta RI 16.97	FIN Virtanen, Kimmo RI 17.76	DEN Maxen, Jan RI 18.03	RSA Hawthorne, Quintin RI 19.17	GER Borowik, Isabella RI 27.52	USA Adams, Tom RI 29.73
HZ - Hard Zeros	116	*1.3*	10 *1.2*	9 *1.0*	11 *1.3*	10 *1.2*	11 *1.3*	11 *1.3*	13 *1.5*	12 *1.4*	14 *1.6*	15 *1.7*
SZ - Soft Zeros	31	*0.4*	4 *0.5*	1 *0.1*	3 *0.3*	2 *0.2*	0 *0.0*	0 *0.0*	3 *0.3*	9 *1.0*	7 *0.8*	2 *0.2*
Marks from 0.5 to 6.5	1864	*21.5*	247 *28.5*	222 *25.6*	232 *26.8*	167 *19.3*	107 *12.3*	157 *18.1*	101 *11.6*	300 *34.6*	185 *21.3*	146 *16.8*
Marks from 7.0 to 10.0	6650	*76.7*	606 *69.9*	635 *73.2*	620 *71.5*	688 *79.4*	747 *86.2*	699 *80.6*	748 *86.3*	546 *63.0*	660 *76.1*	701 *80.9*
AV - averages	9	*0.1*	0 *0.0*	0 *0.0*	1 *0.1*	0 *0.0*	2 *0.2*	0 *0.0*	2 *0.2*	0 *0.0*	1 *0.1*	3 *0.3*
Total marks ... (Pilots/Judge)	8670		867 ... (58)	867 ... (58)	867 ... (58)	867 ... (58)	867 ... (58)	867 ... (58)	867 ... (58)	867 ... (58)	867 ... (58)	867 ... (58)
Style Comparison:	*Average:*	7.32	6.99	7.24	7.17	7.24	7.41	7.30	7.65	7.08	7.39	7.70
	Style:	1.63	1.74	1.90	1.53	1.33	1.64	1.39	1.46	1.84	1.56	1.92
Raw Marks Factors:	*Average %:*	100	95	99	98	99	101	100	105	97	101	105
	Style %:	100	107	117	94	82	101	85	89	113	95	118
Figure anomalies												
HZ to fitted value	35	*0.4*	1	2	2	1	3	2	6	3	6	9
Mark to confirmed HZ	10	*0.1*	-	2	-	-	1	1	2	-	1	3
SZ to confirmed HZ	0	*0.0*	-	-	-	-	-	-	-	-	-	-
SZ to fitted value	22	*0.3*	2	-	2	1	-	-	2	8	5	2
AV to confirmed HZ	0	*0.0*	-	-	-	-	-	-	-	-	-	-
AV to fitted value	9	*0.1*	-	-	1	-	2	-	2	-	1	3
Lo to fitted value	114	*1.3*	9	9	11	10	15	12	9	13	13	13
Hi to fitted value	53	*0.6*	7	5	4	5	8	7	5	1	5	6
The 60% Rule	20	*0.2*	2	2	2	2	2	2	2	2	2	2
Total figure anomalies	263		21	20	22	19	31	24	28	27	33	38

Sequence anomalies

CZE	Czech Republic	1 Lo 0 Hi	- -	- -	- -	- -	- -	- -	- -	1 Lo -	- -	- -
FIN	Finland	0 Lo 1 Hi	- -	- -	- -	- -	- -	- 1 Hi	- -	- -	- -	- -
FRA	France	1 Lo 2 Hi	- -	- -	- 2 Hi	1 Lo -	- -	- -	- -	- -	- -	- -
GER	Germany	0 Lo 2 Hi	- -	- -	- -	- -	- -	- 1 Hi	- -	- -	- -	- 1 Hi
GBR	Great Britain	2 Lo 2 Hi	- -	1 Lo -	- -	- -	- -	- 1 Hi	1 Lo -	- -	- 1 Hi	- -
HUN	Hungary	3 Lo 3 Hi	- -	- -	1 Lo -	1 Lo -	- -	- 1 Hi	- 1 Hi	1 Lo -	- -	- 1 Hi
RUS	Russia	1 Lo 2 Hi	- 1 Hi	- -	- -	- 1 Hi	- -	- -	- -	- -	1 Lo -	- -
SLO	Slovenia	0 Lo 1 Hi	- -	- -	- -	- -	- -	- -	- -	- -	- 1 Hi	- -
RSA	South Africa	2 Lo 3 Hi	1 Lo -	- -	- -	- -	1 Lo 1 Hi	- -	- -	- 2 Hi	- -	- -
USA	United States	3 Lo 3 Hi	- -	- 1 Hi	- -	- -	- -	1 Lo -	- -	1 Lo -	1 Lo -	- 2 Hi
POL	Poland	1 Lo 0 Hi	1 Lo -	- -	- -	- -	- -	- -	- -	- -	- -	- -
BLR	Belarus	1 Lo 0 Hi	- -	- -	- -	- -	- -	1 Lo -	- -	- -	- -	- -
UKR	Ukraine	4 Lo 2 Hi	1 Lo -	- -	1 Lo -	- -	- -	- 1 Hi	1 Lo -	- 1 Hi	1 Lo -	- -
LTU	Lithuania	2 Lo 0 Hi	1 Lo -	- -	- -	- -	- -	- -	1 Lo -	- -	- -	- -
Total sequence anomalies		21 Lo 21 Hi	4 Lo 1 Hi	1 Lo 1 Hi	2 Lo 2 Hi	2 Lo 1 Hi	1 Lo 1 Hi	2 Lo 5 Hi	3 Lo 1 Hi	3 Lo 3 Hi	3 Lo 2 Hi	- 4 Hi

Aerobatic Contest Results Organiser, Version 3.0 Build 28-01-11
Calculations by: FairPlay (non-scoring CJ + CHZ Summary) method
This report created at 11:59 on 26 February 2011