

1. Background

The ECF director of home chess has asked for opinions on publishing a grading list more often than annually. There are various administration issues. This note looks at how this fragile system is working on an annual basis, since this might shed light on a more regular update.

The grading system has been running for many decades, but only recently has data been retained at the game level and even more recently has the list been made available electronically. The grading committee had been reviewing the system with this data.

The ECF grading system was re-calibrated in 2009 to meet a situation where stronger players were underperforming compared to their expected score and weaker players overperforming. The calculation of grades for juniors and new entrants were altered at the same time, this is recognised as part of the new environment, but not investigated here. Serious investigation requires data at the game level, whereas I only have the grading lists for 2008 (as re-calibrated), 2009 and 2010. The re-calibration broadly fitted the following formula:

$$\text{Adult New Grade} = 0.81 \times \text{Old Grade} + 43$$

(source: ECF grading database Help).

Previously I have deposited papers here: <http://www.ecforum.org.uk/viewtopic.php?f=4&t=1038>. This work has convinced me that, if the system was used in a crowded lighthouse with players playing broadly the same number of games then the average grade would be maintained and the standard deviation of such grades (a measure of dispersion or stretch) would be stable - the system would be self-repairing. The issues I have been looking at are the mild deflation and increasing dispersion seen in the lists.

The crude statistics are as follows:

2008: 10,245 players with an average grade of	133.9
2009: 10,151	133.7
2010: 10,074	133.3

Standard deviations of grades:

2007 old grades	42.3
2008	44.7
2008 new grades	34.5

2009	37.1
2010	37.5

The change between 2007 and 2008 suggests the cause of the increase is within the system. The re-calibration over-cooked the fix (amongst others, see my first entry on this page:

<http://www.ecforum.org.uk/viewtopic.php?f=4&t=22&start=210>). My rough estimate is that the 2008 new grades should have had a standard deviation of 35.8, but the trend has gone past this point.

The other evidence is to look at the "Keith and Mark" effect, where it was predicted that one effect of the overall inflation in the change may lead to mature active players grades increase themselves over time. There is some evidence of this. For instance, here are all the strong players who have been category A or B in all three lists (hence active and their grade is largely based on games played in that season) who were graded 225 or more in 2008.

Name	grade08	grade09	Grade10	change
Howell, David WL	247	247	258	+11
Gordon, Stephen J	239	238	242	+3
Pert, Nicholas	239	235	237	-2
Jones, Gawain CB	236	240	242	+6
Wells, Peter K	235	238	232	-3
Hebden, Mark L	233	232	242	+9
Conquest, Stuart C	232	243	229	-3
Lalic, Bogdan	231	233	234	+3
Arkell, Keith C	231	237	231	0
Emms, John M	231	229	229	-2
Davies, Nigel R	230	233	232	+2
Gormally, Daniel W	230	229	231	+1
Pert, Richard G	229	220	224	-5
Hawkins, Jonathan	226	231	235	+9
Cherniaev, Alexander	226	226	233	+7
Greet, Andrew N	226	219	232	+6
McNab, Colin A	226	232	226	0
Williams, Simon K	225	234	237	+12
D'Costa, Lorin AR	225	229	220	-5
average				+2.6

Clearly this table is not conclusive, but when seen alongside general deflation suggests some underlying systemic issue. One place to look is in the dynamics in the data.

2. Player level analysis

The next table summarises the movements for 2009

	with grade	with games	with cat	points	average
Players in 2008 list	10245	10479	10257	1373012	133.9
deduct leavers			-1595	-202128	126.7
			8662	1170884	135.2
deduct remaining juniors			737	83533	113.3
Rated by prior year grades			7925	1087351	137.2
2009 list	10149	10369	10151	1357190	133.7
deduct new entrants (including juniors)			-1489	-183387	123.2
deduct existing juniors			-737	-87717	119.0
			7925	1086086	137.0
change				-1265	-0.2

Notes:

1. Some entries have played games but have no grade; my count is based on having a category. The difference between have a grade (the way it is counted) is the number of players with zero grade.
2. Points are the sum of the grades.
3. Remaining/existing juniors are those that appear in both lists as juniors and are isolated because their prior grades are disregarded each year.
4. Included in the 2009 points are 5,170 points that are not based on performance, but are junior additions
5. The change shows that the grade for 7,925 players that form the reference set for the 2009 calculations fell by 0.2 points during that calculation.

This table has been updated for the 2010 list.

	with grade	with games	with cat	points	average
Players in 2009 list	10149	10369	10151	1357190	133.7
deduct leavers			-1543	-200676	130.1
			8608	1156514	134.4
deduct remaining juniors			695	79195	113.9
Rated by prior year grades			7913	1077319	136.1
2010 list	10072	10302	10074	1343104	133.3
deduct new entrants (including juniors)			-1466	-183597	125.2
deduct existing juniors			-695	-84659	121.8
			7913	1074848	135.8
change				-2471	-0.3

The junior additions for 2010 total 5,775. If these are ignored then the average for new entrants would be 123.8 and that of existing juniors 116.5.

3. Observations at the player level

If leavers have higher grades than joiners, then the crude average of all grades will decline. However even

the grades of the reference set are falling slowly. The reference set standard deviation also rose from 34.2 to 34.7 between 2009 and 2010.

This latter feature is known; here is quote from the ECF database reporting on the grading changes:

"There may be several separate reasons for grades to inflate, deflate or stretch out, but the main one is very simple: if we don't play lots of games against opponents who also play lots of games, the list is bound to become gradually stretched. This has been demonstrated by mathematical modelling, but the results cannot easily be summarised in a meaningful way. We are looking at ways of making this information available to those who are interested."

The modelling has not been made available. The maths I have published in the link given above, suggest that in the proverbial lighthouse (well manned by active chess players) the system is self repairing. Nevertheless the evidence away from the lighthouse, supports the graders.

4. Game level analysis

This year it is possible to investigate further. This is first year that three lists have been available on the new grades. This is important because the system uses games from up to three seasons to calculate grades where there are insufficient games in the current year. The above tables are at the player level, but inflation, deflation and stretch are probably driven at the game level. I have had a crack at reconstructing the data at the game level. It is clear that there is insufficient data and that at the margins I do not understand the exact methodology. However I think these are blemishes that can be ignored at the global level.

The idea is that in some sense the system cannot be inflationary or deflationary, since for most games the sum of the grading points in must be the sum of the grading points coming out. The exceptions are where juniors and new entrants are involved. In these cases a weighted average of existing reference grades are used as input. I use match points to describe grade inputs and outputs and "games" are half games since each game has two players providing data.

Analysis using grade points		MatchPtsIn	GamesIn	MatchPtsOut	
Rated by prior year grades	7,913	22,899,298	161,514	22,926,507	141.9
Existing Juniors	695		17,876	2,425,678	135.7
New Entrants (including new juniors)	1,466		18,833	2,375,572	126.1
			<u>198,223</u>	<u>27,727,757</u>	<u>139.9</u>
Misc		925	843	-	
		<u>22,900,223</u>	<u>199,066</u>		
Games from 2009			65,016	8,546,061	131.4
Games from 2008			24,752	3,015,091	121.8
	Total ignoring misc		<u>287,991</u>	<u>39,288,909</u>	<u>136.4</u>

Notes:

1. MatchpointsIn are the product of the games played in 2010 and the 2009 grade; MatchpointsOut are the product substituting the 2010 grades.
2. Output is after deducting the junior addition implicit in the final grades
3. The final column is the game weighted grades for the respective groups
4. Misc is for reconciliation and includes some games included by those not graded in 2010.
5. Games from 2009 and 2008 are those used from prior years to lift the games of each player to the category minimum.
6. Only the reference set players have in grades; new entrants and juniors have in grade imputed from the reference set grades through the iteration process.

The weakness of this analysis is that the 2010 grades have some 90,000 games out of 288,000 based on performance in previous years. So there is a requirement to back these out of the 2010 only performance. The graders appear to have taken performance (points obtained in that season) from previous years, whereas the only information I have is the gradings from previous years. These grades could be on the performance up to 5 seasons previous. I have used the older grades to derive a 2010 performance. This provides an alternative view of the previous table

Analysis using grade points		MatchPtsIn	GamesIn	PerformancePoints	
Rated by prior year grades	7,913	22,899,298	161,514	22,872,307	141.6
Existing Juniors	695		17,876	2,425,678	135.7
New Entrants (including new juniors)	1,466		18,833	2,386,704	126.7
			<u>198,223</u>	<u>27,684,689</u>	<u>139.7</u>
Misc		925	843		
		<u>22,900,223</u>	<u>199,066</u>		
Games from 2009			65,016		
Games from 2008			24,752		
	Total ignoring misc		<u>287,991</u>		

Whichever of these tables is used the following issues emerge as candidates to explain deflation and stretch:

- The difference between player average grades and game weighted averages
- The importance of the iteration process in creating grades for input
- The influence of prior season games
- The influence of the ongoing juniors (which may be one example of other important subsets) since is

diverse group make up 7% of the players, but 9% of the games.

As an alternative analysis the above statistics have been split by category:

Cat09		MatchPtsIn	GamesIn	MatchPtsOut		PerformancePointGWP		Ave Grade 09	
A	1972	12,200,282	81,757	12,280,182	150.2	12,279,503	150.2	147.2	
B	1642	4,992,281	35,887	5,009,175	139.6	5,000,683	139.3	138.4	
C	2405	4,278,707	32,284	4,272,336	132.3	4,236,219	131.2	132.7	
D	2348	2,674,801	20,716	2,696,895	130.2	2,692,455	130.0	128.6	
E	1784	1,075,954	8,754	1,093,597	124.9	1,089,125	124.4	122.6	
	10151	25,222,025	179,398	25,352,185	141.3	25,297,985	141.0	133.7	

Cat10								Ave Grade10	AvePerf
A	1940	13,071,522	91,162	13,556,508	148.7	13,556,508	148.7	147.4	147.4
B	1626	5,203,206	39,023	5,347,900	137.0	5,344,873	137.0	137.1	136.9
C	2455	4,665,400	35,753	4,741,281	132.6	4,725,972	132.2	132.9	132.3
D	2342	1,966,938	22,780	2,904,602	127.5	2,891,702	126.9	127.4	126.3
E	1711	314,034	9,505	1,177,466	123.9	1,165,634	122.6	122.4	118.6
	10074	25,221,100	198,223	27,727,757	139.9	27,684,689	139.7	133.3	131.7

Notes:

1. All data refers to the 2009-10 season, except the category and average grades in the top table, which refer to the previous list.
2. MatchPtsIn includes those imputed for new entrants and juniors
3. MatchPointsOut and Performance points in the top table exclude those for new entrants.

5. Observations

- Effects of deflation and stretch are delayed by the influence of old grades on the current grading list. The situation may be worse than the crude statistics suggest.
- While dispersion has increased it is unclear why. If 2010 adult grades are plotted against 2009 grades then a linear fit is the best available using statistical information criteria, but the derived parameters give a gradient of slightly less than 1, suggesting the grades should be converging. However the regression of performance gives statistically significant parameters with a gradient of more than 1 and a negative constant: implying stretch and deflation.
- One candidate feature is that the relatively stronger players play more games. Some jottings suggest that it depends on whom they play and the data to look at this is not in the public domain.
- Deflation in average grade depends on:

- o Leavers being stronger than new entrants and the relationship of both these group's grades to the reference set
- o Changes in the proportion of juniors in the system. Since:
 - § Their grade addition increases the global average
 - § They generally have lower grades
- o Changes in the proportion of games played by stronger players. This concept lacks a suitable measure at present.

And finally what might this suggest about more regular updates. I shall make the assumptions that administration issues are overcome and the annual list remains under the same format. The updates will be on the basis of "if the season ended now what would grades be".

Remarks would be:

- About half the players have more games counting from previous years than the current year. For these regular updates will hardly change their rating
- Juniors and new entrants could be done under the current system (especially if the calculation could be moved to its matrix equivalent), but would probably be misleading given the limited linkage to the reference set (which even applies over the whole year to many)
- Which rapid play grade to use for new entrants if there is a choice, could be difficult
- Step changes in grades can be expected as players pass through categories (changing the data for the calculation). This will occur whether categories are estimated on games so far on just follow existing rules as though it was a complete season
- Intermediate publication would not affect the systemic problems with the annual system