Will ECF Elo grades move in line with linked FIDE ratings?

1. Overview

There are no plans for ECF grades to be synchronised with FIDE ratings. However the two methods are similar, albeit based on different sets of results, so they might be expected to move in similar ways.

This note started out answering a separate question from ecforum 4NCL online thread:

"Roger Lancaster post_id=243250 time=1586703171 user_id=8751]

I'm sure this topic must have been worked to death elsewhere but, when I've managed an OTB 4NCL team, it has almost always been the case that the conversions resulting from the {7.5 x ECF} + 700 formula result, at any rate for those below ELO2000, in higher [sometimes significantly higher] figures than the actual FIDE ratings.   In the case of juniors, my experience has been that the difference is even greater than for adults."

My interests were twofold:

a) To defend the ECF formula
b) To consider if the formula was still appropriate for the starting conversion for monthly grades from existing ECF grades.

It quickly became clear that Roger's assertion was correct and hence I had to concede point 1.

On point 2, I have no plans to change the starting formula. This is because it is quite clear that the grade/rating ratio will be unstable, and so the relationship is likely to change, making some initial "correct" conversion spurious and hostage to fortune.

The remainder of this note summarises my investigation.

2. History

The ECF grading system needs a conversion of FIDE grades where games played abroad against ungraded players are accepted for grading. This requirement is very narrow, but it has more widespread other uses where the formula may not be ideal. It has always been considered too complicated to increase the range of conversions for different uses, but officials are not bound to use this version in other situations. In 2009 the ECF grading team shrunk the dispersion of grades. A one-off adjustment to grades below 217 was made: new= .7752*old+49. This called in question the traditional conversion to FIDE of FIDE=8*ECF+600 . This conversion was looked at in 2013. Data was based on all players active in both lists who were aged over 18 (A fit including juniors was weaker).

The relationship is close to linear. Since there is no way to determine the dependent variable, both regressions were carried out:

FIDE = 7.15*ECF + 755

ECF = .127*FIDE -78 or representing by inversion FIDE = 7.9*ECF + 620

It was decided to go with FIDE = 7.5*ECF + 700.

This formula was reviewed in 2017 where it was recognised the relationship had changed, but not by enough to change it. There is some useful mathematical theory[i] which shows that the gradient should be a bit higher than 6.95.

The 2020 review instigated by the question above was based on the same criteria using the Jan 2020 ECF grades update and the April 2020 FIDE rating download. The two best fit regressions are:

FIDE = 7.64*ECF + 602

ECF= .121*FIDE -59 inverting to FIDE = 8.30*ECF + 491

The standard error of the gradient in the first formula is .045 suggesting that it is higher than 7.5, contradicted by the 2nd formula where the slope is unlikely to be less than eight!

3. Under the bonnet – ECF grades

The ECF method has been largely unchanged during the period. In 2015 we introduced a F-grade which gave a grade to players who had played 5 or more games, but less than 9 in the preceding 3 years. This grade had less standing than those for the more active players in categories A-F. The following table gives the dynamics of the whole list.

Table 3.1

	Old Mean	New Mean	Old SD	New SD	Old Count	New Count
Changes	124	118	44	51	4462	6423
Common	140	139	37	40	5646	5646
All	133	128	41	47	10108	12069

"Common" refers to players in both lists; "Changes" refers to those players in the Old-2013 list only and are only in the New-2020 list. Hence the 12,069 players in the new list are the 10,108 in the Old List less the 4,464 who have left plus the 6,423 who have joined. The new list has 871 F-grade players so the ECF has added more graded players, by the old definition, over the period.

Over the period most of the change in average grade can be put down to the introduction of weaker F-grade players. Supporting evidence is that the grade of those in both lists is almost unchanged. Newer

players tend to be weaker, but then improve. The overall average is also stabilised by the higher turnover in lower graded players. It is the stability of these opposing forces that has maintained a remarkably stable all listed players average over the last ten years. If these forces move out of equilibrium then the average would trend.

The middle columns give the variance standard deviations. The dispersion has increased as the shrinking made in 2009 has unwound to the natural level of dispersion.

4. Under the bonnet – FIDE ratings

The equivalent table for FIDE ratings is this:

Table 4.1

|	Old Mean	New Mean	Old SD	New SD	Old Count	New Count
Changes	1825	1493	266	304	39153	129465
Common	1906	1897	268	277	63464	63464
All	1875	1626	270	351	102617	192929

It is immediately clear that something different has happened here. FIDE have managed to grow the list of rated players by nearly 90%. I do not have a history of rating changes, but during the period there have been two big drivers:

   a. Players now have a rating if it comes out at above 1,000. Until quite recently the limit was 2,000.
   b. Players receive a rating on playing five rated players; this has come down from 9 results.

Those players in both lists have an average loss of rating of about 2 ECF points, I would regard this as stable. However, the average of the whole list has dropped significantly as the increase in population has come from the more recreational end of the player spectrum. This influx has also materially widened the dispersion of ratings.

The other aspect is that the new players are young. This graph below is the players only in the later list, by year of birth. ECF records of a measure of age are not as complete, but while there is a high proportion of new players at young ages it is not thought to be so high. Another point is that those only in the new list will include players returning to the game (in the ECF case there are "foreigners" who

intermittently return to big events).

New Entants by YoB
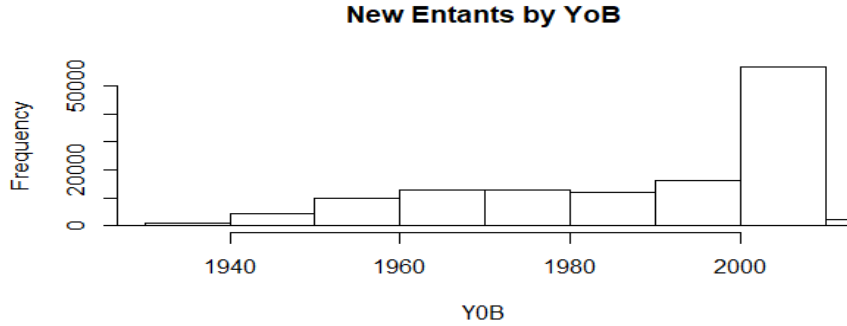
5. The sample set

The previous two sections identify the different drivers in the two full lists. Overtime these drivers will influence the dataset chosen to look at the differences between Monthly ECF grades and FIDE ratings. This dataset selection can be crucial. I have redone the analysis using three lists of players in both lists, these are:

- All players
- All players over the age of 18
- All players over the age of 18 with a FIDE nationality of ENG.

There are two censuses, 2013 and 2018 of all players appearing in each lists as active. My preference has been to use the middle definition, this gives preferable results when looking at a census as it passes over the improving junior issue and is a larger size than the third option. When looking at the changes over the period the difference becomes minimal and the same inferences can be made on all three datasets. The populations are 836, 758 and 539 respectively; only the results using the middle definition are used here.

Comparing the summary statistics of each census gives the following table:

Table 5.1

	Old Mean	New Mean	Old SD	New SD	Old#	New#
ECF	181	176	34	44	1539	2420
FIDE	2049	1945	259	348	1539	2420

Over the period a lot more ECF players have obtained a FIDE rating the sample sizes increase from 1,539 to 2,420. The influx has dropped the mean, but while the ECF grade has dropped by about 40 points in FIDE currency the FIDE mean is down over 100 points. The standard deviation has also increased disproportionately. This is a key indicator since a first approximation of the gradient in a linear regression would be the ratio of the standard deviations[ii].

If we analyse the data again but restrict it players who appear in both lists. The equivalent table is:

Table 5.2

	Old Mean	New Mean	Old SD	New SD	Old#	New#
ECF	183	183	33	36	758	758
FIDE	2065	2020	253	275	758	758

By design the number of players is identical. The change in average FIDE rating is of note given the ECF grade is stable. I have no explanation for this, given that both the full lists have a stable mean for players in both lists (see sections 3 and 4). The ECF dispersion has moved because of natural unwinding to the 2009 shrink; the FIDE dispersion is unexplained but of equivalent proportion.

6. Conclusion

It appears that the change in conversion equation derives from the change in populations over the period. The gradient of the linear regression is closely linked to the ratio of standard deviations and in Table 5.2 that ratio has remained around 7.6. This supports the view that the change is due to population change. Since the theoretical relationship is much nearer 7 than the latest best fit, then the forces are likely to flatten the gradient.

We can expect the forces with regard to juniors to move the relationship in the Elo numbers. There is clear evidence that juniors improve. In any system their grade will lag any change in strength. The way Elo works is to redistribute rating points. Without a countervailing force this will tend to deflate ratings of more mature players without a change in their strength. The FIDE approach is to vary k-factors with juniors using k=40 and most others k=20. The FIDE method will push points into the system when juniors out-perform their expected score based on ratings when playing adults. The ECF approach is to use that higher k for all juniors outperforming that expectation against all players. The suitability of approach depends on the mix of matches.

In future with each system based on the Elo the movement of ratings should be similar. However the forces moving both the populations rated in each system will be different. These forces will also affect English ratings under both lists in different ways. Any comparison will also be affected by the choice of and changes in the dataset chosen to make that comparison.

There seems little point in trying to align the grades to ratings at some point if they are going to move apart in unpredictable ways. In fact there is a case for breaking the link at inception to avoid any confusion that they must be linked. Nevertheless one test of the new monthly grading list will be how consistent it is against the FIDE benchmark.

I do not see a strong case to change the conversion factor.

Brian Valentine

Manager ECF grading

18th April 2020

---

[i] Ignoring the extreme differences, the Elo model has

$$We_{ij} = (1 + 10^{-\Delta/400})^{-1}$$

Where $\Delta = R_i - R_j$

Over a number of games, n, one gets

$$\Sigma We = \Sigma(1+10^{-\Delta/400})^{-1}$$

which is not very malleable, but by Taylor's series at point 0:

given $d(1+10^{-x})^{-1}/dx = 10^x ln(10)/(1+10^x)^2$

$$\approx \Sigma(\tfrac{1}{2}+ \tfrac{1}{4}ln(10)\Delta/400)$$

We $= n/2+\Sigma\Delta/695$

ECF has Expected score $n/2+ \Sigma\Delta/100$, $695/100=6.95$, 2nd order terms are +ve.

[ii] The linear model is $y = a*x + b +error$. The mean of y will be $a*(x-mean) +b$, but the standard deviation of y has no contribution from the constant b and is $a*x$. Hence $a \approx s.d.(y)/s.d.(x)$. Any difference to the fit of parameter a comes from differences in the shape of the underlying distributions.