GRADING MATTERS

by Fred Harte

A LOOK AT THE GRADING SYSTEM

The Irish Chess Union uses a grading system based upon the ELO system since it was first devised by Professor Arpad Elo for the United States Federation, and later adopted by FIDE.

The Elo system distributes players of different strengths along a numerical scale in such a way that players' relative strengths can be assessed. In other words, if a player of 1300 is pitted against another rated 1475, the outcome of a match (of a sufficient number of games) between them may be predicted. A superiority of 175 points translates to a probability of 0.73 of winning. Therefore, it would be expected that the stronger player would win three games in every four.

The numbers used for the scale are not significant in themselves; they are merely historical in origin. The scale has no reproducible fixed points as has a temperature scale, e.g. the freezing, and boiling points of water at 0, and 100 degrees C respectively. This means that the management of a grading system must involve the control of the tendency of ratings to drift over time. Without this control, our 1300 and 1475 players, say, who played each other in 1984 may differ significantly in strength from another two players similarly graded in ten years time even though the predicted outcome of the match would be the same.

GRADING DISTRIBUTIONS

If the distribution of the grades of players in the ICU grading list is graphed (number of players in a given rating band) it will be seen that most grades are heaped around the 1450 – 1500 mark falling off quickly to the left, i.e. lower than 1450, and tailing off much more slowly to the right, i.e. higher than 1500. Although they exist, no grade below 700 gets on to the ICU list. The average rating is about 1480. Approximately one player in nine is rated 1800 or more. 2000 is reckoned to be expert level. A 2200/2250 grading is needed for Irish National Master qualification. The highest rating attained by an Irish player was 2438 – Bernard Kernan. This is International Master territory. The more humble of Grandmasters begin to flourish about the 2500 mark, while those on the top exist in the rarefied air of the 2700's. The highest rating ever was that of Fischer who achieved 2780.

JOINING THE SCALE

After this brief look at the scale that contains all players from beginner to World Champion, let us see how a player first gets on to it.

At first, some estimate of the beginner's strength must be made so that he may be introduced on to the scale at the point corresponding to his standard. This can only be done by reference to his results against players already graded. For example, a player enters his first tournament, and achieves a score of 2½ points against opposition rated 1000, 800, 950, 1170, 1050, 980.

The sum of the opposition's grades = 5950. Awarding 400 points for a win, and -400 for a loss, 2½/6 gives -400. Then we have (5950-400)/6 = 925 which is known as the performance rating. The player then has an estimated rating of 925. Had another ungraded player been one of his opponents the calculations would have involved simultaneous equations. In practice, there can be considerable interplay amongst ungraded players in a large tournament such as the City of Dublin, and the grader must resort to a tedious iterative process in order to get a set of self-consistent ratings for players hitherto ungraded.

Suppose our example, now estimated at 925, plays in another tournament versus, say, 1200, 1100, 1048, 880, 920, 998, and scores 2/6.

As in the previous example the performance rating is calculated which in this case is 891. This result is merged with his previous one of 925 to give a new estimated rating of (6*925+6*891)/12 = 908.

ACCELERATION PRINCIPLE

Had the score instead been sufficient to bring the performance rating over 1200 [say 5/6 giving a performance rating of 1291] then an acceleration principle would have been used. The old performance rating of 925 would have been ignored, and replaced by 1200 as a basis for further calculations, e.g. (6*1200+6*1291)/12 = 1246. If the new performance rating had been less than 1200, but more than the old performance rating, the new estimated rating would have equalled the new performance. The reason for the acceleration process is to take account of players who improve rapidly at this stage, and to avoid penalising their opponents when it comes to their turn to be rated.

FULL GRADING CALCULATIONS

Grades are estimated in this way until such time as the player has played 12 games. Then the grade is no longer regarded as provisional. Strictly speaking, this should be a higher figure – around 18 (sic) according to theory – but there are arguments for keeping the lower figure. When more tournaments are entered the calculations are different.

Now suppose our example enters yet another event. He will be treated as a fully graded player of 908. Let us say he scores 1½/5 versus 950, 1150, 1225, 1280, 1000.

We apply the standard formula for a graded player less than 2000.

- award 16 for a win, and -16 for a loss.
- where opponent's ratings differ by more than 350 points, treat those ratings as being different by exactly 350 points, e.g. the 1280 above will be counted as 1258.
- award 4% of the adjusted rating difference to the lower rated player.

Thus we have his new grade as
(950+1150+1225+1258+1000-5*908)*4% = 41.72 --> 42
From 42 is deducted 32 [arrived at under note a)] giving a +10 points gain. Our player now becomes 918.

However, had he done a little better, he might have qualified for the application of rules for acceleration since his grade was less than 1200 when entering the tournament. These rules are similar to the ones that apply to provisionally graded players whose estimated grades are also less than 1200.

In order to qualify for acceleration a player must score at least 40% in not less than 5 games. If he does qualify, his performance rating is calculated, and then one of three situations will arise.

MORE ACCELERATION

To illustrate all three situations we shall conjure up a fourth tournament for our example who enters graded 918.

1. He plays against opponents rated 797, 810, 1014, 925, and 1169, and scores 2/5. His performance rating is (4715-400)/5 = 863 which is lower than his tournament grading so he will not qualify for acceleration, but will be regraded according to the standard formula. He will emerge from this tournament with a new grading of 907, i.e. old grade minus 11 points. [(797+810+1014+925+1169-5*918)*4%-16 = -11]

2. But, supposing he had scored instead, say, 3½/5. His performance rating would have been then 1103. This is less than 1200, but more than his old grade so 1103 is taken to be his new grade.

3. This time we suppose a score of 5/5 where the performance rating would be 1343. In this instance his old pre-tournament rating is struck out, and replaced by 1200. He is then regraded by the standard formula when he would gain a further 32 points [(850+850+1014+925+1169-5*1200)*4%+80)] bringing his grade to 1232.

ORDER OF CALCULATION

The order in which grading is carried out is important. First, we find estimated grades for ungraded players. Then provisionally graded players are regraded. Next, a point not mentioned until now, ungraded, and provisionally graded players are regraded a <u>second</u> time if a provisionally graded player was one of their opponents. In practice this happens a lot, and can mean quite a sizeable adjustment to a player's grade especially if one or more provisional opponents did very well, and their grades were accelerated. This is because the new grades are the ones used in the second cycle of calculations. To go back to tournament number one for a moment. Here, our player, then ungraded, scored 2½/6 against 1000, 800, 950, 1170, 1050, 980 and came away with an estimate of 925.

But, suppose that the opponent whose pre-tournament rating was 950 was only provisionally graded, and in later calculations became 1250. This is a gain of 300 so in the second run through our example player's estimate would be adjusted accordingly, i.e. 925+(300/6) = 975

After the second cycle of calculations just described we do the fully graded players whose ratings are less than 1200. A second cycle may operate here too. We then come to the main body of players: those fully graded players whose ratings are in excess of 1200. However, included in this group may be

i. those players less than 1200 who did not qualify for acceleration [5 games, and 40% rule].
ii. those whose performance rating was less than pre-tournament.
iii. those whose pre-tournament grade was accelerated to 1200.

THE STANDARD FORMULA

The formula [see above] is really a linear approximation to the percentage expectancy curve, itself derived from the normal distribution curve. To apply a non-linear system would be too burdensome for one that is manually operated. In any case the distortion of ratings that occurs through the application of the linear approximation formula – especially where the difference in ratings between players is large – is not usually severe.

The United States Chess Federation itself, where the Elo system was born, uses a linear approximation. As there is an ongoing search for the distribution that best describes the distribution of chess players' performances no doubt, one day, our system will need an overhaul so that it may be closer to "that elusive reality" as Elo himself put it. Even then the grader will not be free from the task of monitoring the integrity of his rating pool - more about this later.

The standard formula contains a coefficient K. The value of K can be varied in the course of the management of the rating system. The Irish Chess Union uses a value of 32 for players under 2000 and half this figure for players 2000+. The higher figure gives more weight to more recent performances while the lower one puts more emphasis on past performances. This is logical enough. The system recognises the likelihood of players rapidly improving at first, and facilitates the faster upward movement of their grades while providing some stability to the grades of established players. Two examples:

1. a 1485 player scores 3½/6 v 1600, 1085, 1860, 1485, 1550, 1705. He is regraded to 1517.

2. a 2100 player scores 3½/4 v 1825, 1750, 1650, 1875. His grade remains unchanged. Had he achieved a score of 4/4 he would have gained just 8 points.

BONUS POINTS

Let us suppose that the 1485 player had scored 4/6 instead. His normal gain would have passed over a certain threshold that indicated a statistically exceptional performance, and that he might qualify for bonus points. His normal gain would work out to be +48. The threshold for 6 games is 38. A gain in excess of a threshold is doubled. Therefore, a normal gain of 48 would be increased to 38+2(48-38) = 58

Bonus points are awarded so that the grades, especially of new or young players keep pace with fast improving levels of play. In this way it operates somewhat like a high K.

FEEDBACK

It may be argued that a bonus earner was undergraded when entering a tournament. By using his pre-tournament grade in rating his opponents, it could be claimed that each opponent's post-tournament grade was somewhat less than it should have been. Feedback points are therefore awarded to the opponents of bonus earners. This is done, in effect, by recalculating the grades of those who played others who turned out to be bonus earners, using the post-tournament grade in the new calculation.

This is just one more aspect of the management of a grading system, and deflation/inflation control. Even when using all the weapons in the grader's arsenal – low, high, multiple K's; bonuses; feedback; additions for those aged under 21 [not used at the moment by the ICU], etc., it may still be necessary to make a global adjustment from time to time to compensate for a general drift.

DRIFT CONTROL

As mentioned earlier, the ICU uses the Elo grading system. This system is the one also used by FIDE, the International Chess Federation. The main difference between the management of the two systems stems from the obvious differences between the natures of the two rating pools. On the one hand we have a national pool made up of every conceivable kind of player: long term, transient, local, beginner, old, young, those whose final rating will not be much more than their initial estimated grade, and those whose ratings are destined to increase by 1000 points or so.
On the other hand the pool of players with which FIDE deals is composed largely of players whose initial burst of improvement is over, and who have attained a level of competence that enables them to engage in international play. Furthermore, FIDE only deals with the interplay between such players. The FIDE pool is therefore much more stable, and is the one to which local pools should be aligned so that a given rating on any one list would reflect the same standard of play as the same rating on another.

REGULAR SCRUTINIES

The task, then, of a national grader is to scrutinize the performances of local players when they participate in international tournaments, and meet with a large number of players who have FIDE ratings. Some, or all, of the local players may have played in previous international events, and may have FIDE ratings themselves. For example, in an international tournament let us say three players have results as follows:
A. 2½/6; B. 4/10; C. 4½/8. When we extract only those results pertaining to FIDE rated opponents we might get:

	ICU grade			OPPONENTS FIDE GRADES					
A	2350	2½/6	2200	2300	2400	2500	2550	2450	
B	2290	2/7	2530	2480	2260	2380	2410	2600	2440

C	2365	3/6	2375	2420	2390	2370	2340	2345	

We can now calculate a performance rating R(p) by the usual procedure for each player that was achieved against FIDE rated opponents:

Rp(A) = 2333.3; Rp(B) = 2271.4; Rp(C) = 2373.3

The average weighted Rp is:
(2333.3 *6)+(2271.4*7)+(2373.3*6) = 44140/19 = 2323.2

Now we work out the average weighted ICU rating:
(2350*6)+(2290*7)+(2365*6) = 44320/19 = 2332.6

Comparing the two ratings we see that it is indicated that the ICU ratings in the example are inflated by about 10 Elo points which might warrant a general deduction of 10 points from everyone in the ICU rating pool. Such a global adjustment has not been found necessary in the case of Irish ratings since 1979.

FIDE INTEGRITY

Does the FIDE pool itself drift? Can FIDE ratings suffer from inflation or deflation too? Certainly, any drift will be small, and a strong attempt is made to hammer it in place with statistical nails. Professor Elo claims that it is possible to maintain the integrity of the ratings even from era to era so that by executing the rating procedures retrospectively it is possible to calculate the ratings of players long since dead. Professor Elo discusses this in some detail in his book "The Rating of Chessplayers, Past and Present" [Batsford, 1978]. Readers might like to reflect on the highest ratings achieved by some famous players which have been extracted from a diagram in the book. The figure on the left shows the year in which the maximum was attained.

1856 Anderssen 2595	1921 Capablanca 2730	1955 Keres 2670
1880 Steinitz 2640	1926 Nimzovitch 2630	1960 Tal 2710
1888 Chigorin 2600	1934 Alekhin 2690	1964 Petrosian 2680
1900 Tarrasch 2610	1936 Euwe 2645	1972 Spassky 2690
1908 Lasker 2725	1946 Botvinnik 2730	1972 Fischer 2780
1915 Rubinstein 2660	1953 Smyslov 2700	1978 Karpov 2725

Karpov's grade on 1st January 1985 had declined to 2705 while Kasparov's stood higher at 2715. It is interesting to note that Kasparov was not 22 years old until last April [1985]. Fischer was 22 before he broke the 2700 mark. The average advance from this age to a peak rating at age 36 is about 100 points. So watch out for Kasparov. If he can keep motivated for the next 14 years we may see him break 2800.

UPSETS EXCEPTED

Could Muhammad Ali have "whupped" Jack Dempsey? We shall never know. However, the continuous grading of chess players enables us to predict the outcome of a proposed chess match with much more certainty than can our pugilistic friends a boxing bout. It is obvious that if the Lasker of 1908 played Kasparov now, we could easily state that Lasker would have the edge. But the tables of the normal curve, which is the basic tool of the Elo rating system, allow us to be more specific. We could expect Lasker to win by 12½ to 11½. Needless to say, it is not ordained that this must be the result. The underdog might upset the odds, and come out on top. This happened in the case of Alekhin vs. Euwe in 1935. Alekhin, who led Euwe by 50 Elo points, would have been expected to score 17 from the 30 games played. Nonetheless, he only scored 14½, and lost his world crown. In the return match two years later Alekhin won easily, thereby realising his expected score over the two matches quite closely.

This aspect of the grading system is the most significant. Of course, any system that gains the confidence and acceptance of players has another use too. It provides an enormous incentive for players to improve their standard of play even if this interferes with the pleasure of playing for its own sake. If a win is achieved against someone rated 250 points less, it may be felt that the game was a waste of time. “Players can become too rating conscious” it has been said. This attitude has been criticised, but those chessplayers who are not in line for trophies, or prize money can by way of compensation get much satisfaction from registering a new personal best grade. No doubt the rating of chessplayers will continue to stimulate interest in the game through a heightened desire to improve one’s degree of determination, persistence, and understanding.

SOME STATISTICS

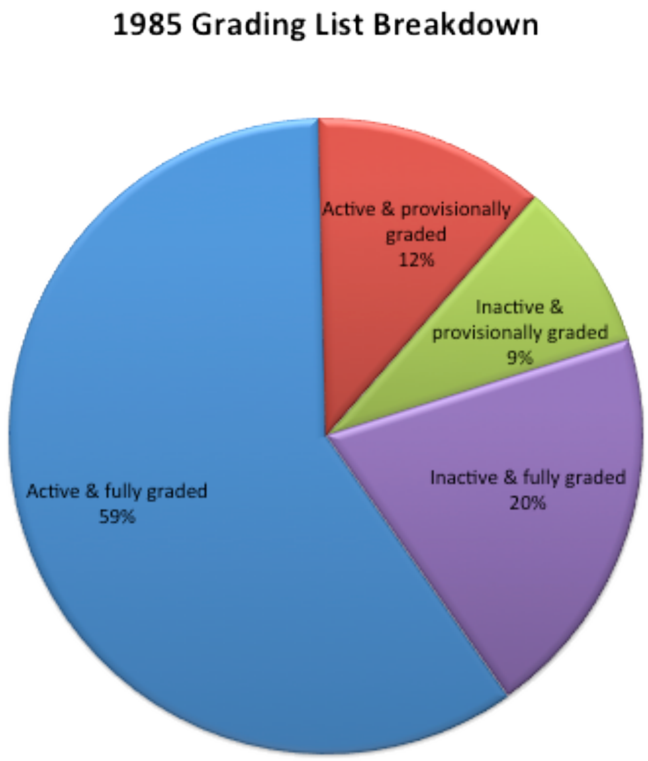
The production of statistics on grades is not undertaken to satisfy an idle curiosity, or merely to while away the time. It is a bit too laborious for that. It is necessary to analyse the rating pool in order to spot either the drift of grades or changes in the character of the pool that may presage a drift. One example is a breakdown of 2217 entries in the 1985 rating list. The figures for 1984 are for comparison.

	Active85	Active84	Inactive85	Inactive84	Total85	Total84
Graded	1316	1278	447	404	1763	1682
Provisional	262	258	192	176	454	434
TOTAL	1578	1536	639	580	2217	2116

Expressed in percentage terms, the figures are presented as a pie chart (see below).

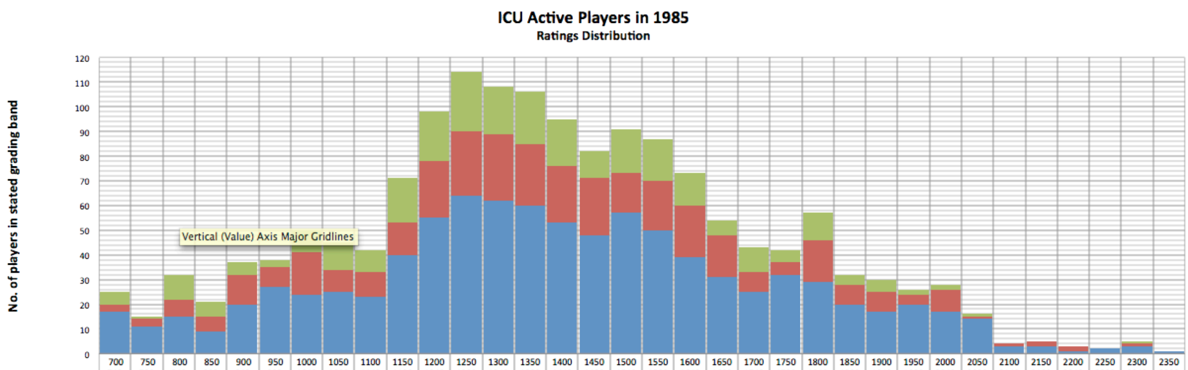
Of course, when we talk about inactive players we only refer to those who have recently become inactive, and whose names are retained in the register. There is a large body of erstwhile players now dropped from the current rating list who seem destined to live out the rest of their existence in obscurity. Cases of resurrection from the group border on the miraculous as the following figures indicate. The data was taken from the 1985 rating list.

	1 yr inactive	Returned	2 yrs inactive	Returned
Connacht	8	0	16	0
Leinster	158	30	157	8
Munster	61	13	37	10
Ulster	68	14	75	7
TOTAL	295	57	285	25



The table shows 295 who were inactive for one year and 285 inactive for two years, a total of 580 players. The number from each group who returned to play in the 1984/85 season is given in the column headed "Returned". Of those who had been inactive for just one year, 57 [19%] came back; of those inactive for two years, only 25 [<9%]. And from the anonymous masses, a mere trickle made it.

The distribution of ratings is an important pointer to the character of the rating pool. Included are all active players, both provisionally and fully rated. The average grade is 1412. If you are over 1600 you are in the top quarter; over 1850 and you are well within the top 10%. See the histogram below.



	700 749	750 799	800 849	850 899	900 949	950 999	1000 1049	1050 1099	1100 1149	1150 1199	1200 1249	1250 1299	1300 1349	1350 1399	1400 1449	1450 1499	1500 1549
Munster	5	1	10	6	5	3	9	11	9	18	20	24	19	21	19	11	18
Ulster	3	3	7	6	12	8	17	9	10	13	23	26	27	25	23	23	16
Leinster/Connacht	17	11	15	9	20	27	24	25	23	40	55	64	62	60	53	48	57

	1550 1599	1600 1649	1650 1699	1700 1749	1750 1799	1800 1849	1850 1899	1900 1949	1950 1999	2000 2049	2050 2099	2100 2149	2150 2199	2200 2249	2250 2299	2300 2349	2350 2399
Munster	17	13	6	10	5	11	4	5	2	2	1	0	0	0	0	1	0
Ulster	20	21	17	8	5	17	8	8	4	9	1	1	2	2	0	1	0
Leinster/Connacht	50	39	31	25	32	29	20	17	20	17	14	3	3	1	2	3	1

IMPORTANT ANNOUNCEMENT – CHANGE IN GRADING CALCULATION

At present [April 1986] there are only two rates of change [K factors] applicable, i.e. K=32 for those under 2000 Elo, and K=16 for those who are 2000+. In other words if we regard the change for those under 2000 as “full rate”, then 2000+ are changed at half rate. This is in recognition of the likelihood that the standard of play of those who are 2000+ has more or less stabilised.

At the other end of the scale the standard of play of those under 21 years of age is anything but stable, and is likely to show dramatic improvement over a relatively short period. In order to give more weight to the more recent performances, a K factor of 40 will apply to the under 21’s (i.e. “full rate” * 1.25).

This will mean that the grades of this group will tend to surge upwards more rapidly than under the present system, and besides having a direct positive effect on the grades within the U21 group will have a positive indirect effect on the grades of their opponents. It is, therefore, in every player’s interest to ensure that those who are under 21 are noted as such in the grading list.

END

The above article was published as a series in four issues [September & December 1984, July 1985, April 1986] of the Irish Chess Union bulletin “Fiacla Fichille” ISSN 0332 4664.

Fred Harte, December 2012
ICU Rating Officer (1981-1989)