# Report of the
# USCF Ratings Committee

## August 1995

**Mark E. Glickman, Ph.D.**
**Chairman, USCF Ratings Committee**

# Contents

# Ratings Committee Members

**Harry Cohen** Ph.D. in Operations Research, Massachusetts Institute of Technology (1975). Member of the USCF Ratings Committee 1988–present. Currently a principal in a transportation planning and management services consulting firm.

**Mark Glickman** Ph.D. in Statistics, Harvard University (1993). B.A. in Statistics, *Summa Cum Laude*, Princeton University (1986). Member of the USCF Ratings Committee 1985–present; chairman 1992–present. Begins an Assistant Professorship in Mathematics at Boston University, Boston, Massachusetts, in September 1995.

**Bill Goichberg** B.A. in Political Science, New York University (1963). USCF Policy Board member, 1975–1978, and 1989–1992. Professional chess organizer and director.

**Albyn Jones** Ph.D. in Statistics, Yale University (1986). B.A. in Mathematics, UCLA (1978). Currently holds an Associate Professorship in Statistics at Reed College, Portland, Oregon.

**Alan Losoff** B.S. in Mathematics, Illinois Institute of Technology (1969). Proofreader for "The Ratings of Chessplayers" by Arpad Elo. Member of the USCF Ratings Committee 1991–present. Computer programmer working on mathematical modeling for a financial derivatives firm.

**Kenneth Sloan** Ph.D. in Computer and Information Science, University of Pennsylvania (1977). Sc.B. in Applied Mathematics, Brown University (1970). Currently holds an Associate Professorship at the University of Alabama in Birmingham.

## Policy Board Liaison:

**Frank Camaratta** M.S. in Applied Mathematics, Applied Mechanics, and Aerospace Engineering, Drexel Institute of Technology (1968). Treasurer and former vice president of USCF. Chairman of the Computer Rating Agency, 1985–present. Ratings Committee chairman, 1986–1990.

# Committee Motions

(Note: in the text, "the Committee report" refers to this document)

1. The Policy Board recognizes and authorizes the use of the FIDE-to-USCF conversion and the CFC-to-USCF conversion as described in Section 1 of the Committee report.

2. The Policy Board authorizes the use of the FIDE-to-USCF and CFC-to-USCF conversions for the purpose of assigning ratings to unrated players with either CFC or FIDE ratings, as described in Section 1. The converted rating would be treated as a provisional USCF rating based on 10 games when updating ratings from an event.

3. The Policy Board authorizes the following temporary measure to accelerate rating changes for scholastic players: For players under 12 years old who have ratings less than 800, the value of $K$ in the rating formula will be multiplied by 4. For games in which a player competes against an opponent less than 12 years old and with a rating less than 800, the value of $K$ to be used is 1/4 of its normal value for that player.

4. The Policy Board authorizes the hiring of an independent computer programmer to rewrite the program that calculates USCF ratings and titles as described in Section 4 of the Committee report. The specifications for the program, which will be determined by the Ratings Committee chair and the Liaison to the Policy Board from appropriate documents, will serve as the basis for the ratings program implementation. The selection of a programmer must be agreed upon by both the Ratings Committee chair and the Liaison to the Policy Board.

5. The Policy Board recognizes that any proposed change to the USCF rating system or USCF title system must be submitted to the Ratings Committee for study and recommendation.

6. For tournament players with USCF prison memberships, the Policy Board authorizes to move the rating floor down 200 points so that the floor is at least 300 points below the highest rating attained. Under this motion, a player in prison who has had a highest rating of 1750 would have a rating floor at 1400, whereas currently the player would have a rating floor at 1600.

# 1 1995 Rating Conversions

According to last year's accepted Ratings Committee proposals, if an unrated player has a USCF-unrated opponent with a FIDE rating, the FIDE-rating would be converted to a USCF rating given by a conversion table. The converted rating would then be used as if the player had an established USCF rating. This conversion is only used when an unrated player's opponent is unrated. This year we present an updated table of adjustments to convert FIDE ratings to USCF ratings, and also propose a conversion from CFC (Canadian Chess Federation) ratings to USCF ratings.

The Committee has determined a 1995 conversion of FIDE ratings to USCF ratings, as shown in Table 1. The 1995 conversion of CFC ratings to USCF ratings is shown in Table 2. If an unrated opponent has both a CFC and FIDE rating, then we propose that the CFC rating have precedence. Details of the conversions are described in Appendix A.

Table 1: USCF rating conversion from FIDE rating

| FIDE | USCF | FIDE | USCF | FIDE | USCF | FIDE | USCF | FIDE | USCF |
|---|---|---|---|---|---|---|---|---|---|
| 2200 | 2217 | 2300 | 2362 | 2400 | 2477 | 2500 | 2584 | 2600 | 2659 |
| 2205 | 2225 | 2305 | 2368 | 2405 | 2483 | 2505 | 2589 | 2605 | 2662 |
| 2210 | 2233 | 2310 | 2375 | 2410 | 2489 | 2510 | 2593 | 2610 | 2664 |
| 2215 | 2241 | 2315 | 2381 | 2415 | 2495 | 2515 | 2598 | 2615 | 2666 |
| 2220 | 2249 | 2320 | 2387 | 2420 | 2501 | 2520 | 2602 | 2620 | 2669 |
| 2225 | 2257 | 2325 | 2392 | 2425 | 2507 | 2525 | 2606 | 2625 | 2671 |
| 2230 | 2265 | 2330 | 2398 | 2430 | 2513 | 2530 | 2611 | 2630 | 2673 |
| 2235 | 2272 | 2335 | 2404 | 2435 | 2520 | 2535 | 2615 | 2635 | 2674 |
| 2240 | 2280 | 2340 | 2409 | 2440 | 2525 | 2540 | 2619 | 2640 | 2676 |
| 2245 | 2287 | 2345 | 2414 | 2445 | 2531 | 2545 | 2623 | 2645 | 2678 |
| 2250 | 2295 | 2350 | 2420 | 2450 | 2536 | 2550 | 2627 | 2650 | 2679 |
| 2255 | 2302 | 2355 | 2425 | 2455 | 2542 | 2555 | 2631 | 2655 | 2681 |
| 2260 | 2309 | 2360 | 2431 | 2460 | 2547 | 2560 | 2635 | 2660 | 2682 |
| 2265 | 2316 | 2365 | 2436 | 2465 | 2552 | 2565 | 2638 | 2665 | 2683 |
| 2270 | 2323 | 2370 | 2442 | 2470 | 2557 | 2570 | 2642 | 2670 | 2685 |
| 2275 | 2329 | 2375 | 2448 | 2475 | 2561 | 2575 | 2645 | 2675 | 2686 |
| 2280 | 2336 | 2380 | 2454 | 2480 | 2566 | 2580 | 2648 | 2680 | 2687 |
| 2285 | 2343 | 2385 | 2460 | 2485 | 2571 | 2585 | 2651 | 2685 | 2688 |
| 2290 | 2349 | 2390 | 2465 | 2490 | 2575 | 2590 | 2654 | 2690 | 2689 |
| 2295 | 2356 | 2395 | 2471 | 2495 | 2580 | 2595 | 2657 | 2695 | 2690 |

Table 2: USCF rating conversion from CFC rating

| CFC | USCF | CFC | USCF | CFC | USCF | CFC | USCF | CFC | USCF |
|---|---|---|---|---|---|---|---|---|---|
| 1510 | 1483 | 1710 | 1675 | 1910 | 1897 | 2110 | 2060 | 2310 | 2283 |
| 1520 | 1492 | 1720 | 1686 | 1920 | 1907 | 2120 | 2069 | 2320 | 2296 |
| 1530 | 1501 | 1730 | 1697 | 1930 | 1918 | 2130 | 2079 | 2330 | 2310 |
| 1540 | 1510 | 1740 | 1707 | 1940 | 1928 | 2140 | 2088 | 2340 | 2323 |
| 1550 | 1519 | 1750 | 1718 | 1950 | 1938 | 2150 | 2098 | 2350 | 2337 |
| 1560 | 1528 | 1760 | 1729 | 1960 | 1948 | 2160 | 2108 | 2360 | 2351 |
| 1570 | 1537 | 1770 | 1740 | 1970 | 1958 | 2170 | 2118 | 2370 | 2365 |
| 1580 | 1547 | 1780 | 1751 | 1980 | 1967 | 2180 | 2128 | 2380 | 2379 |
| 1590 | 1556 | 1790 | 1762 | 1990 | 1976 | 2190 | 2138 | 2390 | 2394 |
| 1600 | 1566 | 1800 | 1774 | 2000 | 1985 | 2200 | 2149 | 2400 | 2408 |
| 1610 | 1575 | 1810 | 1786 | 2010 | 1992 | 2210 | 2160 | 2410 | 2423 |
| 1620 | 1585 | 1820 | 1798 | 2020 | 1999 | 2220 | 2172 | 2420 | 2438 |
| 1630 | 1595 | 1830 | 1809 | 2030 | 2006 | 2230 | 2183 | 2430 | 2453 |
| 1640 | 1604 | 1840 | 1820 | 2040 | 2012 | 2240 | 2195 | 2440 | 2469 |
| 1650 | 1614 | 1850 | 1831 | 2050 | 2018 | 2250 | 2207 | 2450 | 2484 |
| 1660 | 1624 | 1860 | 1842 | 2060 | 2024 | 2260 | 2219 | 2460 | 2500 |
| 1670 | 1634 | 1870 | 1853 | 2070 | 2030 | 2270 | 2232 | 2470 | 2516 |
| 1680 | 1644 | 1880 | 1864 | 2080 | 2037 | 2280 | 2244 | 2480 | 2532 |
| 1690 | 1655 | 1890 | 1875 | 2090 | 2044 | 2290 | 2257 | 2490 | 2548 |
| 1700 | 1665 | 1900 | 1886 | 2100 | 2052 | 2300 | 2270 | 2500 | 2565 |

It is worth noting that the calculated conversions are not constant for all FIDE or CFC values. For example, to obtain the estimated USCF rating for a 2250 FIDE-rated player, 45 points need to be added; for a 2400 FIDE-rated player, 77 points need to be added; and for a 2670 FIDE-rated player, 15 points need to be added. Similarly, to obtain the estimated USCF rating for a CFC-rated player of 1550, 31 points need to be subtracted; for a 2000 CFC-rated player, 15 points need to be subtracted; and for a 2400 CFC-rated player, 8 points need to be added.

## Additional use of converted ratings in rating system

The Ratings Committee recommends the use of the FIDE-to-USCF or CFC-to-USCF conversions for the purpose of updating ratings from an event. We propose that prior to any other computations, USCF unrated players in an event with either CFC or FIDE established ratings have their ratings converted to USCF ratings by Tables 1 or 2. These ratings are then treated as provisional based on 10 games. Again the CFC conversion takes precedence if a USCF-unrated player has ratings in both the CFC and FIDE systems.

Two examples:

1. Suppose an established USCF player competes against four opponents, one of whom is unrated but with a CFC rating of 1970, the other three with established USCF ratings. Under the proposed change, the CFC-rated player would be assigned a (provisional) rating of 1958. Updating would then proceed as usual (see the 1994 Ratings Committee Report for more details).

2. Suppose a 2440 FIDE-rated player without a USCF rating competes against four opponents with established USCF ratings. Prior to applying the rating formulas, the 2440 FIDE rating is converted to a USCF rating of 2525 based on 10 games. Updating would then proceed as usual.

# 2 A method to accelerate rating changes in the scholastic pool

Rating deflation is the tendency for ratings to decrease over time while players' abilities generally remain unchanged. This is a problem that is suspected to occur at the lower end of the rating scale. One plausible mechanism for deflation arises from pools of young players only competing amongst themselves in scholastic events – on average, their ratings remain roughly the same even though they improve enormously. When they begin competing against non-scholastic opponents, they are found to be grossly underrated and easily defeat similarly rated opponents. The net effect is that the more established non-scholastic players' ratings move downward.

Recognizing the difficulties with possible rating deflation, the Ratings Committee proposes a temporary solution which accelerates rating changes for scholastic players. We propose that for players under 12 years old who have ratings less than 800, the value of $K$ in the rating formula will be multiplied by 4. Also, for games in which a player competes against an opponent less than 12 years old and with a rating less than 800, the value of $K$ to be used is 1/4 of its normal value for that player. When two players compete who are both under 12 years old and both have ratings less than 800, the two conditions above "cancel out" and ratings are updated normally.

Two examples:

1. Suppose a 10-year old with an established rating of 700 competes against opponents rated 600, 750, 900, and 1000, defeating the first three and losing to the last. The winning expectancies for a 700-rated player against these opponents is 0.640, 0.429, 0.240, and 0.151. Assume the player with the 750 rating is under 12, and the remaining three are 12 or older.

   To update the rating, we note that the value of $K$ for each game, respectively, is 128, 32, 128 and 128. The value 128 is obtained by multiplying the "normal" value of 32 by 4. The value of $K = 32$ for the second game is obtained by magnifying $K$ by 4, and then dividing by 4 because the opponent has a rating less than 800 and is under 12 years old. The calculation to update the 10-year old's rating is:

   $$R_{new} = 700 + 128(1 - 0.640) + 32(1 - 0.429) + 128(1 - 0.240) + 128(0 - 0.151) = 842$$

   Under the current system, the updated rating would be 749. Using the proposed system moves the player's rating upward more quickly.

2. Suppose an adult with a 1200 rating competes against opponents rated 750, 1100, and 1150, losing to the first, drawing the second, and defeating the third. The winning expectancies for a 1200-rated player against these opponents are 0.930, 0.640, and 0.571. Assume the player with the 750 rating is under 12, and the other two are 12 or older.

   The value of $K$ for each of these games is 8, 32 and 32, respectively. We compute $K = 8$ for the first game because the usual value of $K = 32$ is divided by 4 when the opponent is under 12 years old and

has a rating less than 800. The calculation to update the adult's rating is:

$$R_{new} = 1200 + 8(0 - 0.930) + 32(0.5 - 0.64) + 32(1 - .571) = 1202.$$

Under the current system, the updated rating would be 1179. The proposed system "protects" the adult's rating when the scholastic opponent may be underrated.

The Ratings Committee is currently exploring more principled methods for controlling deflation that systematically bias upwards the rating changes of scholastic players. This approach recognizes that the current rating system framework is unable to detect overall improvement in scholastic players' abilities when they primarily compete amongst themselves. This area requires further examination.

# 3 A Major Extension to the Elo Rating System

Last year, we described the basis of an extension to the Elo system that accounts for differing amounts of uncertainty in players' ratings. This system has been simplified, and the resulting product is called the "Glicko" system. In the Glicko system, each player not only receives a rating, but also a measure of trust in the rating. This second measure is called a "rating deviation" (RD). Large RD's correspond to ratings that are not reliable, and low RD's indicate ratings that are precise measures of ability. When a player has a large RD, his or her rating is susceptible to large changes. Provisionally rated players and players who have not competed for a long time typically have high RD's. When a player's *opponent* has a large RD, the player's rating change is likely to be small because there is little information conveyed in the result of the game. The Glicko system has been implemented on both the Internet Chess Club (ICC) and the Free Internet Chess Server (FICS). On FICS, the Glicko system has replaced the Elo system in its entirety. Appendix B includes a description of the Glicko system implemented on FICS.

The advantage of the Glicko system over Elo-based systems, such as those used currently by the USCF and FIDE, is that ratings in the Glicko system are better predictors of performance. The disadvantage of the Glicko system is that it is subject to manipulation, which we feel is of greater concern to the USCF than for the administrators of the internet chess servers. The problem of rating manipulation stems from the practice that ratings are used to section events and award prizes. As we suggested in our report two years ago, and currently maintain, sectioning events and awarding prizes according to USCF titles rather than rating ranges would likely reduce the problems of rating manipulation.

The Ratings Committee continues to explore potential use of this system for rating USCF events.

# 4 Rewriting the ratings program

The Ratings Committee examined the USCF ratings program and determined that the implemented algorithm does not match the description on the USCF Rating System sheet. Apparently several changes have been made to the system that have not been clearly and conveniently documented. Rather than take the approach of modifying the current rating code, we propose to have the ratings program rewritten by a computer programmer under the supervision of the Ratings Committee chairman and the Committee's Policy Board Liaison.

The proposal to rewrite the ratings program involves several parts. These are listed as follows:

- Write the specifications for the ratings program and title program.
- Create and maintain a file that stores current ratings-related information about each USCF member.
- According to the specifications, write well-commented code in a common computer language (e.g., C) which
  - reads tournament rating reports either from a computer file or from manual data entry,
  - calculates rating updates from tournament and match results,
  - adjusts title information,
  - stores rating and title information in a file that can be read by the USCF database program,
  - produces summary information to separate files (e.g., crosstables).
- Thoroughly test program before implementing.

This task will require a small amount of time of the USCF programmer to provide assistance with descriptions of file formats, and to assist with the implementation of the new program on USCF computers. We expect the entire procedure will take no more than 10 months of time to complete, and no more than 100 hours of programmer time (most of which would be used for testing the code).

There are several important reasons for adopting our proposal. The most important is that the ratings program would become straightforward to maintain. Currently, the USCF is at the mercy of a program that is very difficult to read and understand. No one on the Ratings Committee, including competent computer programmers, has been able to read the current code and describe with certainty the implemented ratings algorithm. Because the current rating code is not well understood, and the specifications for the implemented code are unavailable, rewriting the ratings program and its specifications makes potential changes to the system easy to test and implement. Another important benefit of our proposal is that it relieves the USCF programmer of time that would otherwise be devoted to the ratings program. From conversations with the USCF Deputy Director, the USCF programmer's time is very limited, and it is difficult to schedule programmer time to make modifications to the rating system. The rating proposals that have been approved by the Policy Board in August 1994 are only now in the process of being implemented, and the proposals

passed in February 1995 have not yet been given programmer time to be implemented. With a new system in place, modifications to the code could be performed by programmers outside the USCF office.

# A FIDE and CFC to USCF conversion

The Committee has determined a conversion to predict a USCF rating from a FIDE rating, and a USCF rating from a CFC rating, for purposes of pairing FIDE-rated USCF-unrated players into USCF-rated events. This is accomplished by identifying players common to both the active USCF and FIDE or CFC pool of players, and fitting a local regression model ("loess") to the data.

Among the 778 players who competed in both USCF and FIDE events in 1994, only players with established USCF ratings, FIDE ratings of at least 2200, and 10 or more FIDE-rated games in 1994, were included. This resulted in a total of 207 players used in the FIDE analysis. For the CFC analysis, 127 players were identified as being active in 1994 in both the US and Canada based on the annual USCF and CFC lists. We included only players who had established USCF and established CFC ratings which resulted in a total of 56 players used in the analysis.

The loess fits were performed as robust procedures (not adversely affected by outliers) using a smoothness criterion based on 75% samples of the data at each point. The results of the fits appear on Table 1 and Table 2 in Section 1. The loess fits revealed non-linearity in the relationships between FIDE and USCF ratings, and between CFC and USCF ratings.

# B FICS description of the Glicko System

The following is the "help file" for the Glicko system on the Free Internet Chess Server (FICS) with a few minor edits. It was written by Kevin Leeds of Georgia Tech.

An explanation of the Glicko system

As many have noticed, each FICS player now has a rating and an RD.

RD stands for "ratings deviation."

Why a new system?

The new system with the RD improves upon the binary categorization that was used before on fics and elsewhere, where players with fewer than 20 games were labeled "provisional" and others were labeled "established." Instead of two separate ratings formulas for the two categories, there is now a single formula incorporating the two ratings and the two RD's to find the ratings changes for you and your opponent after a game.

What RD represents

The Ratings Deviation is used to measure how much a player's current rating should be trusted. A high RD indicates that the player may not be competing frequently or that the player has not played very many games yet at the current rating level. A low RD indicates that the player's rating is fairly well established. This is described in more detail below under "RD Interpretation."

How RD Affects Ratings Changes

In general, if your RD is high, then your rating will change a lot each time you play. As it gets smaller, the ratings change per game will go down. However, your opponent's RD will have the opposite effect, to a smaller extent: if his RD is high, then your ratings change will be somewhat smaller than it would be otherwise.

A further use of RD's:

Kevin asked Mark Glickman the following:

> Given player one with rating $r_1$, error $s_1$, and player two with $r_2$ and $s_2$, do you have a formula for the probability that player 1's "true" rating is greater than player 2's ?

Mark said:

> Yes - it's:
>
> $$\frac{1}{1+10^{-(r_1-r_2)f(\sqrt{s_1^2+s_2^2})/400}}.$$
>
> where f(s) is the function applied to RD in Step 2 below.

How RD is Updated

In this system, the RD will decrease somewhat each time you play a game, because when you play more games there is a stronger basis for concluding what your rating should be. However, if you go for a long time without playing any games, your RD will increase to reflect the increased uncertainty in your rating due to the passage of time. Also, your RD will decrease more if your opponent's rating is similar to yours, and decrease less if your opponent's rating is much different.

Why Ratings Changes Aren't Balanced

In the other system, except for provisional games, the ratings changes for the two players in a game would balance each other out – if A wins 16 points, B loses 16 points. That is not the case with this system. Here is the explanation I received from Mark Glickman:

> The system does not conserve rating points - and with good reason! Suppose two players both have ratings of 1700, except one has not played in awhile and the other playing constantly. In the former case, the player's rating is not a reliable measure while in the latter case the rating is a fairly reliable measure. Let's say the player with the uncertain rating defeats the player with the precisely measured rating. Then I would claim that the player with the imprecisely measured rating should have his rating increase a fair amount (because we have learned something informative from defeating a player with a precisely measured ability) and the player with the precise rating should have his rating decrease by a very small amount (because losing to a player with an imprecise rating contains little information). That's the intuitive gist of my extension to the Elo system.
>
> On average, the system will stay roughly constant (by the law of large numbers). In other words, the above scenario in the long run should occur just as often with the imprecisely rated player losing.

Mathematical Interpretation of RD

Direct from Mark Glickman:

> Each player can be characterized as having a true (but unknown) rating that may be thought of as the player's average ability. We never get to know that value, partly because we only observe

a finite number of games, but also because that true rating changes over time as a player's ability changes. But we can *estimate* the unknown rating. Rather than restrict oneself to a single estimate of the true rating, we can describe our estimate as an *interval* of plausible values. The interval is wider if we are less sure about the player's unknown true rating, and the interval is narrower if we are more sure about the unknown rating. The RD quantifies the uncertainty in terms of probability:

The interval formed by Current rating ±RD contains your true rating with probability of about 0.67.

The interval formed by Current rating ±2RD contains your true rating with probability of about 0.95.

The interval formed by Current rating ±3RD contains your true rating with probability of about 0.997.

The Formulas

Algorithm to calculate ratings change for a game against a given opponent:

**Step 1.** Before a game, calculate initial rating and RD for each player.

**(a)** If no games yet, initial rating assumed to be 1720. Otherwise, use existing rating. (The 1720 is used only for FICS, and is not intended to carry over to the USCF system).

**(b)** If no RD yet, initial RD assumed to be 350 if you have no games. Otherwise, calculate new RD, based on the RD that was obtained after the most recent game played, and on the amount of time $t$ that has passed since that game, as follows:

$$\mathrm{RD}' = \sqrt{\mathrm{RD}^2 + c\ln(1+t)}$$

where $c$ is a numerical constant chosen so that predictions made according to the ratings from this system will be approximately optimal.

**Step 2.** Calculate the "attenuating factor" due to your OPPONENT's RD, for use in later steps.

$$f(RD) = 1/\sqrt{1 + p(\mathrm{RD})^2}.$$

Here $p$ is the mathematical constant

$$\frac{3(\ln 10)^2}{(400\pi)^2}.$$

Note that this is between 0 and 1 – if RD is very big, then $f(\mathrm{RD})$ will be closer to 0.

**Step 3.** Let

$$\begin{aligned} r_1 &= \text{your rating} \\ r_2 &= \text{opponent's rating,} \end{aligned}$$

$$E = \frac{1}{1 + 10^{-(r_1 - r_2) f(\mathrm{RD})/400}}.$$

The quantity $E$ is the winning expectancy accounting for the uncertainty in the opponent's rating.

**Step 4.**

$$K = \frac{q f(\mathrm{RD})}{1/\mathrm{RD}^2 + q^2 f(\mathrm{RD})^2 E(1-E)}$$

where $q = (\ln 10)/400$.

**Step 5.** This is the $K$ factor for the game, so

$$\text{Your new rating} = \text{pre-game rating} + K(w - E)$$

where $w$ is 1 for a win, .5 for a draw, and 0 for a loss.

**Step 6.** Your new RD is calculated as

$$\text{RD}' = \frac{1}{\sqrt{1/\text{RD}^2 + q^2 f(\text{RD})^2 E(1-E)}}.$$

The same steps are done for your opponent.