Report of the USCF Ratings Committee

August 1994

Mark E. Glickman, Ph.D.

Chairman, USCF Ratings Committee

Contents

Preface – General Philosophy of Current Ratings Committee Work

Ratings Committee work currently is dominated by two goals:

- To produce a rating system that *predicts* performances as accurately as possible, and
- To produce separate measurement systems that may be based on the rating system which serve other functions, such as rewarding players for meritorious performances, or determining sectioning strategies for tournaments.

The current ratings system serves many functions, some of which conflict with each other. For example, while the current rating system tries to predict performances accurately, it also serves as a promotional tool where a player may be more encouraged to continue tournament participation if his or her rating increases, and conversely discouraged if a player's rating decreases. To relieve this and other burdens from the rating system, the rating system is being viewed solely a predictor of performances. Other measures (such as titles), whose computation may rely on ratings, are intended to enhance and encourage tournament participation. This year's proposal takes steps towards implementing this philosophy.

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. Currently a post-doctoral fellow at the Department of Health Care Policy, Harvard Medical School.

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.

Ratings Committee Special Advisor:

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.

Committee Motions

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

1. The Policy Board authorizes the modification of the "Lifetime Title System" by adopting the *Delta* (Δ) schedule described in Section 2.2 of the Committee report, and the adoption of the Norm and Title rules described in Section 2.3 of the Committee report.

2. The Policy Board authorizes the restructuring of titles to correspond to every 200 rating points, and the conversion from players' current norms and titles to the proposed norms and titles as described in item 2 of Section 2.6 of the Committee report.

3. In connection with the previous motion, the Policy Board authorizes the following title designations, as described in item 2 of Section 2.6 of the Committee report: Advanced Senior Master (2600), Senior Master (2400), Master (2200), Expert (2000), Category I (1800), Category II (1600), Category III (1400), Category IV (1200), Category V (1000), Category VI (800), Category VII (600) and Category VIII (400).

4. The Policy Board authorizes the reinstatement of the "Life Master" title, which requires earning 20 norms at the 2200 level. The Policy Board also authorizes the adoption of the "Life Senior Master" title, which requires earning 20 norms at the 2400 level.

5. The Policy Board authorizes the renaming of the current Lifetime Titles system, "The USCF Title System" as described in item 3 of Section 2.6 of the Committee report.

6. The Policy Board authorizes the USCF to mail computer-printed postcards, or other forms of notification, when a player earns a title from an event. The Policy Board also authorizes the USCF to issue membership cards indicating a player's current title. The Policy Board further authorizes the USCF to issue a certificate when a player earns the title of Expert, Master, Senior Master, Advanced Senior Master, Life Master, or Life Senior Master.

7. The Policy Board authorizes the replacement of the currently implemented provisional rating system with the one described in Section 1 of the Committee report. In summary, the provisional rating formula is replaced by either the currently implemented established rating formulas using a large value of K that decreases from 600 to 32 over a player's first 20 rated games, or an event performance rating calculated on the cumulative results through the first 20 rated games. The Committee will present a specific recommendation to the Policy Board at the August meetings. In either case, unrated players obtain ratings utilizing procedures described in Section 1.4 of the Committee report.

1 Provisional Rating System Modifications

1.1 Introduction

The currently implemented provisional rating system suffers from several problems, including an ad hoc algorithm to rate tournament sections that involve all or mostly unrated players. The Committee has determined an algorithm based on statistical principles that substantially improves on the current algorithm. Our proposed modifications, which were described in last year's proposal, involve modifying the provisional rating formulas so that only the established rating formula is used with a K-factor that depends on the number of tournament games in which a player has competed. The key element to this system is that K would be large for players who have played only a small number of rated games. This approach provides continuity in the rating calculations, and allows for a smooth transition from provisional ratings to established ratings.

1.2 The Current Provisional Rating System

Committee members Ken Sloan and Al Losoff have examined the USCF rating program to understand the implemented algorithm. This was prompted by two letters to the USCF from Illinois tournament organizer Paul Mills. The USCF Rating System sheet does not provide a complete description of the actual implemented algorithm. Below we discuss our understanding of the implemented algorithm, which has been validated from the replication of actual tournament crosstables.

- A tournament is rated by sequentially performing the following steps:
  1. Calculate the ratings of unrated players conditional on the opponents pre-tournament ratings, imputing a rating of 1000 for each unrated opponent.
  2. Update the ratings of provisionally rated players conditional on the resulting ratings of the previous step.
  3. Rerate the unrated players conditional on the ratings from the last two steps.
  4. Update the established players' ratings conditional on the last three steps.
- If an unrated player defeats an opponent with a rating of 600 or less, then the opponent's rating is treated as 601 in the rating calculation for the unrated player (so, in effect, the player does not lose rating points for the win).

- If provisionally rated player with a pre-tournament rating of R defeats an opponent with a rating of less than or equal to $R-400$, then the opponent's rating is treated as $R-399$ in the rating calculation for the provisionally rated player.
- If a player's provisional rating is based on 14 or more games, then if he or she plays in an event of 6 or more rounds, the established rating formula is in effect.

The main problem with the current provisional rating algorithm is that it incorporates arbitrary decisions that have unclear consequences for estimating a player's ability. For example, the current provisional rating system computes a rating by taking an average of opponents' ratings (plus or minus 400). This process of averaging ratings to produce a provisional rating has no theoretical justification consistent with the underlying Elo model – it was intended as an ad hoc computation that could be performed with paper and pencil. Furthermore, the process of assigning all unrated opponents a rating of 1000 seems too restrictive. The current implementation allows a player's provisional rating to increase as the result of a loss to a much higher rated player, though we recognize that this problem can be fixed under the current framework.

1.3 Outline of the Proposed Rating Procedure

Consider rating a tournament section consisting of an arbitrary number of unrated players, provisionally rated players, and established players. The following steps are to be performed sequentially:

1. Compute the ratings of each unrated player using the event performance rating (described in Section 4) if all opponents are rated. If some opponents are unrated, then impute either an assigned rating based on a player's age or a converted rating from a foreign rating system, and then calculate the event performance rating.
2. Update the ratings of each provisionally rated player using the appropriate formula.
3. Update the ratings of established players in the usual manner.

Ideally, we would prefer to ignore the distinction between unrated, provisional, and established ratings in these computations and perform all calculations simultaneously. The problem, however, is that when rating an established player's results against an unrated or provisional player, the pre-tournament rating (or initial assigned rating, in the case of an unrated opponent) may be very inaccurate.

1.4 Rating an Unrated Player

The currently used algorithm for computing an initial rating is based on a crude approximation of a player's event performance rating. An event performance rating is the rating at which a player's actual total score in the event is equal to his or her expected total score. Until recently, only an iterative algorithm was available to compute an accurate event performance rating; however, the Committee has derived a simple non-iterative algorithm that calculates an event performance rating with a substantially greater degree of precision than the currently used formula. The algorithm is presented in Section 4. The Committee recommends the adoption of this algorithm as the basis for computing a player's initial rating.

The Committee identifies two types of situations for computing an initial rating for an unrated player. The first case involves rating a player when all of his or her opponents have established or provisional ratings. The second case involves rating a player when at least one of his or her opponents is unrated. We discuss these two cases below.

1.4.1 All Opponents are Rated

When an unrated player's opponents all have either established or provisional ratings, then the performance rating computation of Section 4 may be directly applied. An exception to this rule occurs when a player achieves either all wins or all losses, which is discussed in Section 1.4.3.

1.4.2 Not all Opponents are rated

When not all opponents of an unrated player are rated, the Committee recommends, for each unrated opponent, imputing ratings based on either an opponent's FIDE rating (if it exists) or based on a player's age. These two imputation methods are discussed in the following sections. Once ratings are imputed for unrated opponents, the rating computation of Section 4 may be directly applied. Again, an exception occurs when a player achieves all wins or all losses. This is discussed in Section 1.4.3.

Rating Assignment for FIDE-rated opponents

During the rating computation for an unrated player, if an unrated opponent has a FIDE rating, then the initial USCF rating assigned for that player would be computed using a conversion derived by the Committee. The Committee has determined a 1994 conversion of FIDE ratings to USCF ratings, as shown in Table 1. Figure 1 plots the conversion along with the data that produced the conversion. Details of this conversion are described in Appendix A.

For players with FIDE ratings less than 2200, we recommend adding 15 to the FIDE rating to produce the converted USCF rating, and for players with FIDE ratings 2700 and higher, we recommend adding 30 to the FIDE rating to produce the converted USCF rating.

Future work involves constructing conversions from other rating systems such as the CFC (Canadian) system.

Table 1: USCF rating conversion from FIDE rating

FIDE	USCF	FIDE	USCF	FIDE	USCF	FIDE	USCF	FIDE	USCF
2200	2214	2300	2354	2400	2474	2500	2581	2600	2665
2205	2222	2305	2361	2405	2480	2505	2586	2605	2669
2210	2229	2310	2367	2410	2486	2510	2591	2610	2672
2215	2237	2315	2373	2415	2492	2515	2595	2615	2676
2220	2245	2320	2379	2420	2498	2520	2600	2620	2679
2225	2252	2325	2385	2425	2504	2525	2605	2625	2682
2230	2259	2330	2390	2430	2509	2530	2609	2630	2686
2235	2267	2335	2396	2435	2515	2535	2614	2635	2689
2240	2274	2340	2402	2440	2520	2540	2618	2640	2692
2245	2281	2345	2408	2445	2526	2545	2622	2645	2695
2250	2288	2350	2414	2450	2531	2550	2627	2650	2698
2255	2295	2355	2419	2455	2536	2555	2631	2655	2701
2260	2302	2360	2425	2460	2541	2560	2635	2660	2703
2265	2309	2365	2431	2465	2546	2565	2639	2665	2706
2270	2315	2370	2437	2470	2552	2570	2643	2670	2709
2275	2322	2375	2443	2475	2557	2575	2647	2675	2712
2280	2329	2380	2449	2480	2562	2580	2651	2680	2715
2285	2335	2385	2455	2485	2567	2585	2654	2685	2718
2290	2342	2390	2462	2490	2571	2590	2658	2690	2721
2295	2348	2395	2468	2495	2576	2595	2662	2695	2724

Rating Assignment based on age

If an unrated opponent has no FIDE rating, the Committee proposes to assign ratings to such opponents according to age. From examining the relationship between provisionally rated players' ages and their ratings, we determined the initial age-based initial rating assignment in Table 2. Figure 2 shows the conversion along with the ratings of all 12366 provisionally rated players from the 1994 January rating list. Details of the analysis that produced these values are described in Appendix B.

Each rating on Table 2 may be interpreted as the rating for an average player of the given age upon first entering tournament play.

Table 2: Initial rating assignment given a player's age

Age (years)	Initial Rating	Age (years)	Initial Rating	Age (years)	Initial Rating	Age (years)	Initial Rating
5	543	24	1218	43	1419	62	1386
6	584	25	1240	44	1422	63	1379
7	624	26	1260	45	1424	64	1372
8	664	27	1278	46	1426	65	1365
9	703	28	1295	47	1427	66	1357
10	743	29	1311	48	1428	67	1349
11	782	30	1325	49	1428	68	1340
12	822	31	1338	50	1428	69	1331
13	862	32	1350	51	1427	70	1321
14	900	33	1360	52	1425	71	1311
15	937	34	1370	53	1424	72	1301
16	973	35	1378	54	1421	73	1290
17	1009	36	1385	55	1419	74	1279
18	1043	37	1392	56	1415	75	1267
19	1077	38	1398	57	1411	76	1255
20	1109	39	1403	58	1407	77	1243
21	1139	40	1408	59	1403	78	1230
22	1168	41	1412	60	1397	79	1217
23	1194	42	1416	61	1392	80	1203

1.4.3 Two Exceptions

Regardless of the rating status of an unrated player's opponents, we recommend a player receives an initial rating only if he or she plays *at least four games* in a USCF event, as is the practice in the currently implemented system. Otherwise, the player does not receive a post-event rating, and the outcomes of the games are saved on file. The outcomes of these games are treated as if they occurred during the player's subsequent event. However, the ratings of the opponents are updated based on the results against the unrated player by imputing either a converted FIDE rating (if it exists) or an age-based rating for the unrated player as described in Sections 1.4.

When an unrated player achieves either all wins or all losses, a slightly different procedure is applied. The event performance rating formula is still used to compute a (nominal) post-event provisional rating for the player, which is then used as the rating upon which his or her opponents' provisional or established ratings are updated. However, this player's game outcomes and opponents' pre-event ratings (possibly some of which are imputed) are saved on file. At this player's next event, the nominal provisional rating is used for updating unrated and provisionally rated opponents, but the player's rating is updated as if the outcomes from the previous event and the current event all occurred in one event, and are rated simultaneously using the event performance rating formula. If, after the second event, the player still has either all wins or all losses, the same procedure is applied for the third and subsequent events until the player no longer has all wins or all losses.

1.5 Provisional rating algorithm

We describe here the method for computing provisional ratings for players who have played in at least one event of at least 4 games, not having achieved all wins or all losses. The calculation mimics the established rating formulas, except with a value of K that is initially large, but then decreases with the number of games played (indicating the decrease in uncertainty in one's rating).

1.5.1 Framework for provisional rating calculations

The current established rating formula can be used as a framework for provisional ratings. The update calculation is given by

$$R_{post} = R_{pre} + K(W - We),$$

where R_{post} is the player's post-event rating, R_{pre} is the player's pre-event rating, W is the player's total score in the event, We is the player's expected total score given by the "winning expectancy" formula, and K is an attenuation factor that determines the amount of change in one's rating. For established rating calculations, $K = 32$ when a player's rating is less than 2100. The values of K in the rating update formula are proposed to be larger for a provisionally rated player than for an established player.

1.5.2 Formula for computing K

If a player has already played a total of N rated games ($N > 3$), and then competes in an m-round event, we propose that the value of K should be

$$K = \begin{cases} 600/(N + m - 1) & \text{if } 4 \leq N + m \leq 20 \\ 32 & \text{if } N + m > 20 \end{cases}$$

After an event in which $N + m$ exceeds 20, the player's rating is governed by the established rating formulas along with the appropriate values of K.

As an example, suppose a player who has already played 10 rated games competes in a 5-round event. The value of K is calculated to be

$$K = 600/(10 + 5 - 1) = 43.$$

1.5.3 Recognizing propensity to improve

A possible addition to the provisional rating system involves the idea that players abilities generally improve over their first several tournaments. One possible implementation is to compute an event performance rating based on all game outcomes through the current event, and then set the provisional rating to be the greater of this performance rating and the rating calculated from the proposed provisional rating formulas with variable K. This approach ensures that if a provisional player's performance in recent events was considerably worse than in his or her initial events, the rating would not decrease substantially.

1.5.4 Alternative computation for Provisional Ratings

Instead of using the established rating formula with large K, the Committee is also considering the exclusive use of the event performance rating calculation to determine provisional ratings. Under this system, game outcomes and opponents' ratings (possibly imputed ratings) are saved on file. After each event while a player's rating is provisional, the event performance rating is calculated based on the players' accumulated game results. This procedure is performed until a player competes in an event that brings his or her total number of rated games over 20, after which the established rating formula would govern the rating computation. As in the system using the established rating formula, unrated opponents have their ratings imputed during the rating calculation.

1.6 Differences between Proposed System and Current System

The proposed provisional rating system introduces several changes to the system. We describe these changes below.

- The current rating system operates under two separate sets of formulas; one for rating unrated and provisionally rated players, and another for rating established players. The proposed modifications unify the rating system by prescribing rating updates from only one formula with a parameter, K, that depends on the number of games played.
- When rating an unrated player after a first event, the current rating system assumes that all unrated opponents are initially rated 1000 in the rating computations. The proposed system more accurately reflects the abilities of unrated players by incorporating either a rating based on the age of the unrated opponent, or the opponent's FIDE rating if it exists, into the computation, thereby using more information to estimate an unrated opponent's pre-tournament rating.
- The current rating system uses a crude computation to estimate a player's initial rating. The proposed system determines with a high degree of precision the rating at which a player's actual total score equals his or her expected total score.

1.7 Example calculations

The following analyses show some resulting rating calculations under the current system and our proposed system.

1. An unrated player competes against players rated 700, 850, 950, 1200, and 1500. For all possible total scores, the following table summarizes the resulting rating calculation.

	CurrSys	NewSys
0.0	640	256
0.5	720	531
1.0	800	689
1.5	880	806
2.0	960	912
2.5	1040	1018
3.0	1120	1130
3.5	1200	1255
4.0	1280	1403
4.5	1360	1592
5.0	1440	2043

2. The same unrated player competes against 5 opponents, all of whom are rated 1040. The following table summarizes the resulting rating calculations.

	CurrSys	NewSys
0.0	640	428
0.5	720	658
1.0	800	799
1.5	880	893
2.0	960	970
2.5	1040	1040
3.0	1120	1110
3.5	1200	1187
4.0	1280	1281
4.5	1360	1422
5.0	1440	1652

These first two examples involve opponents' whose average rating is 1040. Notice that, for this example, the proposed system is affected by the particular distribution of opponents' ratings given that the average is 1040, whereas the current system ignores the opponents' specific ratings. The proposed system recognizes that greater variability in opponents' ratings provide more potential information about a player's performance, and this is reflected in a wider spectrum of possible post-event ratings in the first example.

3. A provisionally rated 700 player based on 10 games competes against opponents rated 900, 1100, 1350, 1400, and 1450.

	CurrSys	NewSys
0.0	747	684
0.5	773	705
1.0	800	726
1.5	827	748
2.0	853	769
2.5	880	791
3.0	907	812
3.5	933	834
4.0	960	855
4.5	987	876
5.0	1013	898

One aspect of the currently implemented system is that all losses can still result in an increase in one's rating. The proposed system prevents this from occurring in a methodologically sound manner. Also, in this example, a low score does not substantially increase one's rating after competing against relatively strong opposition.

2 Title System Modifications

2.1 Introduction

The "Lifetime Title" system has been recently introduced to US chess competition so that players who demonstrate repeated exceptional performances are awarded with a title that they keep for life. Titles correspond to different levels of ability ranging from "Certified Class E" (corresponding to a rating of 1000) to "7-star life master" (corresponding to a rating of 2900). The currently implemented system awards a player multiple "norm points" from the results of an event, and upon the acquisition of 10 norm points, the player is awarded the relevant title. The current terminology defines a "norm" as worth 2 norm points; thus receiving 5 norms earns a player a title.

We feel there are several modification that could greatly improve the coherency and interpretability of the currently implemented lifetime title system. We discuss these changes below.

2.2 Proposed Delta schedule

The current framework for the lifetime title system requires a player's total score in an event to exceed the expected total score of a player rated Y by a specified value, Δ (Delta), to earn norm points for the Y-rated title. The value of Δ depends on the number of games played. We continue to advocate this framework, though we find problems with the current choices of Δ's. Specifically, we suggest the following framework to revise the values of Δ:

- If a player competes in a tournament and attains a total score of w points, then to earn a norm corresponding to a rating of Y, w must be greater than the highest total score a Y-rated player would attain simply by chance $P\%$ of the time.

To demonstrate the ability of a Y-rated player, the norm candidate must provide evidence that he or she can compete at the level of a player rated Y, so the rule enforces that a player below Y-strength will earn a norm for the Y rating by chance (e.g., via a fluke performance) less than $P\%$ of the time. When a player demonstrates this level of performance five times, then he or she should be awarded the title. For multiple norms in a single event, the Committee recommends a player would need to obtain a score that a Y-rated player would exceed by chance with probability $(P/100)^2$ for two norms, $(P/100)^3$ for three norms, and so on. This approach to awarding multiple norms equates the difficulty of earning two norms from one event with the difficulty of earning two norms singly from two consecutive events (and analogously for three or more norms).

The implications of the choice of P are that lower values correspond to greater the difficulty earning a title. For example, if $P = 15$, then a player who truly possesses an ability of a rating Y will be expected to earn the Y-rated title in 10 tournaments with a probability of about 0.02; if $P = 40$, then the probability increases dramatically to about 0.99. To have about a 50% chance at earning a title from 10 tournaments, P must be about 27. This is the choice we recommend.

It can also be shown that a player having an ability less than the rating of Y will earn a norm with small frequency. For example, a player with an ability corresponding to a rating of $Y - 100$ will earn a norm for the Y rating with a probability of about 11%, which corresponds to earning a title in 10 tournaments with less than a 5 in a thousand chance. A player with an ability corresponding to a rating of $Y - 200$ will earn a norm for the Y rating with a probability of about 3%, which converts to a 3 in a million chance of earning the title in 10 tournaments. These calculations assume a 4-round tournament, and that all four opponents have a rating of Y.

Using conservative estimates of the variability of game results given players' ratings, and using the normal distribution as a conservative approximation to the distribution of a player's total score in an event, the calculation of Δ's corresponding to the above rule are straightforward. From a single event in which a competitor plays n games, the following norm schedule is proposed consistent with $P = 27$:

- To earn 1 norm, $\Delta = 0.306\sqrt{n}$.
- To earn 2 norms, $\Delta = 0.727\sqrt{n}$.
- To earn 3 norms, $\Delta = 1.030\sqrt{n}$.
- To earn 4 norms, $\Delta = 1.277\sqrt{n}$.
- To earn the title, $\Delta = 1.491\sqrt{n}$.

Note that the Δ's become larger for increased number of games, which is in contrast to the current system. Under this system, the values of Δ to earn 1 norm from an event are 0.612, 0.684, 0.750, 0.810, 0.865, 0.918, and 0.968 for a 4-round tournament through a 10-round tournament, respectively. We recommend that this schedule be used only when $n \geq 4$ games in an event. Note that this formulation of a Δ schedule precludes the necessity for norm points, and makes many of the norm rules irrelevant (e.g., rules b, c, d, e, f, and g), thus simplifying the system.

2.3 Proposed Norm and Title rules

1. Norms can only be earned in events of 4 rounds or more.
2. A norm is earned, or multiple norms are earned, towards a Y-rated title when a player's total score in an event exceeds the expected total score of a Y-rated player by the value of Δ given in Section 2.2.
3. A player's results from an event apply simultaneously to every norm for titles not already earned. Thus, a player may be working on several titles at once.

2.4 Differences between Proposed System and Current System

A number of notable differences exist between the currently implemented Lifetime Titles system and the proposed modifications. We outline the changes below.

- The current system calculates norm points to earn norms and titles. The proposed modifications to the system use norms rather than norm points as the unit on which titles are based.
- The proposed Δ-schedule makes it slightly easier to earn a single norm in a 4-round event than in the current system, but more difficult to earn single norms in events of 5 rounds or longer (and especially so in long tournaments). This revision, however, ensures that it is equally difficult to earn norms, independent of the number of rounds in the event. Under the current system, it is substantially easier to earn norms in events with more rounds.
- Multiple norms in the proposed system are awarded under the assumption that earning n norms in a single event should be just as difficult as earning single norms in n consecutive events. Multiple norms are determined using the Δ-schedule, whereas the current system uses a sequence of ad hoc rules. Earning multiple norms are more difficult in the proposed modifications than in the current system.
- The current system only allows a player to be working on two titles at any time because lower titles are awarded automatically. The proposed system removes this restriction.
- The current system awards a title if a player's rating is substantially higher than the rating corresponding to the unearned title. In the spirit of separating the meaning of ratings and titles, this restriction is removed in the proposed system.
- The current system only allows players with established ratings to earn norms or titles. The proposed system removes this restriction, which, again, is consistent with the separation of the meaning of ratings and titles.
- Under the current system, a player must score at least two points in an event to earn a norm or title. The proposed system removes this restriction.

2.5 An Example

Suppose an established 1750-rated player competes in a 4-round event against players rated 1800, 1900, 2000 and 2100 and scores 3.5 points. Further suppose this player has no norms or titles prior to becoming rated 1750. The expected total score against such opponents with particular ratings is given in the table below. A player with an ability of 1750 is expected to attain 1.034 points against such opposition.

Expected total score against opponents
rated 1800, 1900, 2000 and 2100

Rating	Expected Score
1500	0.326
1600	0.535
1700	0.842
1800	1.251
1900	1.740
2000	2.260
2100	2.749
2200	3.158
2300	3.465

Under the current system:

Before competing in the 4-round event, the player is automatically awarded the 1600-title due to being rated above 1700. The value of Δ for this event is 0.7 under the current system, implying that with a total score of 3.5 points, the player is awarded

- the 2100-norm (because $2.749 + 0.7 = 3.449 < 3.5$),
- 5 norm points towards the 2000-title, and
- the 1900-title.

Under the proposed system:

The values of Δ for a four round event are 0.612, 1.454, 2.060, 2.554 and 2.982 for 1 norm through 5 norms, respectively. With a total score of 3.5 points, the player is awarded

- the 1500-title (because $0.326 + 2.982 = 3.308 < 3.5$),
- 4 norms towards the 1600-title (because $0.535 + 2.554 = 3.089 < 3.5$),
- 4 norms towards the 1700-title (because $0.842 + 2.554 = 3.396 < 3.5$),
- 3 norms towards the 1800-title (because $1.251 + 2.060 = 3.311 < 3.5$),
- 2 norms towards the 1900-title (because $1.740 + 1.454 = 3.194 < 3.5$),
- 1 norm towards the 2000-title (because $2.260 + 0.612 = 2.872 < 3.5$), and
- 1 norm towards the 2100-title (because $2.749 + 0.612 = 3.361 < 3.5$).

It is worth noting that the proposed system awards higher level norms and titles in long events than short events when a player has achieved the same overall percentage score. This reflects the idea that it is more difficult to sustain an exceptional level of performance for a greater number of games. To show this, suppose the player in the previous example played each opponent twice rather than just once, and attained a total score of 7 out of 8. In this case, the player has scored 87.5%, which is identical to the percentage score of 3.5 out of 4. However, because the player was able to sustain this level of performance for a greater number of rounds, higher level norms and titles are awarded. For a score of 7 out of 8, it can be shown from a similar computation to the preceding example that the player earns

- the 1800-title,
- 3 norms towards the 1900-title,
- 2 norms towards the 2000-title, and
- 1 norm towards the 2100-title.

These awards for an 87.5% score in an 8-round event are more impressive than the corresponding awards in a 4-round event, especially at the lower title levels.

2.6 Miscellaneous Title System Modifications

Here we enumerate other suggestions for modifying the title system.

1. Reinstate the "life master" title, and create the "life senior master" title. The requirements for being awarded the life master title and life senior master title are to obtain 20 norms at the 2200 and 2400 levels, respectively. The choice of 20 norms to earn a life master or life senior master title has some analogy to the old rule of keeping one's rating above 2200 for 300 games. With $P = 27$, a player who possesses a 2200 ability can be expected to earn 20 2200-norms in 74 tournaments, on average (because $0.27(74) \approx 20$). If every tournament is 4 rounds, then a player of 2200 strength can be expected to earn a life master title in roughly $4 \times 74 = 296$ games, or close to 300 games.

2. Rather than have titles corresponding to every 100 rating points, have titles for every 200 rating points. To convert current norm and title status for all players to the proposed delineation of titles and norms for every 200 points, we suggest

   - a player with a title in the current system that does not exist in the proposed system is awarded 3 norms towards the next highest title in the proposed system, and
   - a player with x norms in the current system for a title-level not existing in the proposed system is awarded $x/2$ norms (rounded upward) for the next higher title.
   - after the two preceding steps are performed, a player whose highest norm level is "L" (or that the player has the "L" title and no norms at higher levels) is awarded with that norm status for all levels below "L" for which the player does not already have the title. For example, if after performing the first two steps, a player has 3 norms for the 1600-title, and has the 1200-title, then the player is awarded 3 norms for the 1400-title.

   A possible delineation of titles could be

Rating Level	Title	Norms required for Title
400	Category VIII	5
600	Category VII	5
800	Category VI	5
1000	Category V	5
1200	Category IV	5
1400	Category III	5
1600	Category II	5
1800	Category I	5
2000	Expert	5
2200	Master	5
2400	Senior Master	5
2600	Advanced Senior Master	5
2200	Life Master	20
2400	Life Senior Master	20

3. Rename the system, "The USCF Title System." The Committee has heard many complaints that earning titles is too effortless under the currently implemented system, so describing the titles as "Lifetime Titles" mistakenly connotes lifetime performance rather than a title that is kept for life.

4. Whenever a player earns a title from a tournament, the USCF should mail a computer-printed postcard, or some other form of notification, indicating the player's improved title status. Rather than printing a player's norm/title status on Chess Life mailing labels, we recommend printing the highest attained title achieved.

5. A player's USCF membership card should indicate the highest non-life title achieved. Certificates should be awarded for the titles of Expert, Master, Senior Master, Advanced Senior Master, Life Master, and Life Senior Master.

3 A Major Extension to the Elo Rating System

One of the most problematic features of the Elo rating model is that, except in very specific instances, it assumes that players' ratings are perfectly estimated. This assumption is used, for example, when an established player's rating is updated from the results of a tournament, as the winning expectancy formula treats the opponents' ratings as fixed and known. Published ratings are subject to some amount of uncertainty, and in some cases, this amount of uncertainty is substantial (e.g., for players who are provisionally rated, and for established players who have not competed in a long time). At first, this may not seem like a major issue. But consider the following situations which are not properly handled by the current rating system:

- An established player who hasn't competed in a long time will undergo the same magnitude of rating change as someone who competes regularly. Intuitively, for the player who hasn't competed in a long time, a recent tournament performance should have a substantial impact on his rating relative to the player who competes regularly. The current USCF system makes no distinction between these two situations.
- An established player who competes against a provisionally rated opponent (or any opponent with a poorly estimated rating) will undergo the same magnitude of rating change against an opponent whose rating is well-estimated. It can be argued that an established player's rating should not be substantially affected by the outcome of a game against a player whose rating is poorly estimated. Whereas a player's rating may change by 16 points against an established opponent of a similar rating under the current system, his rating should only change by a fraction of this amount when competing against a player with a poorly estimated rating to reflect this lack of information.
- Occasionally a tournament organizer will be late in submitting a tournament report to the USCF for rating the event. In some cases, not only have players involved in such an event competed more recently, but they may have had their ratings already updated from more recent events. Such a rating update would modify a rating *backward* in time, so that the information from the less recent tournament would have greater impact on current rating than the more recent tournament. The current system makes no distinction in the time order of events. The results of a tournament that occurred *before* a recently occurring event should have only a small impact on a player's rating.

The methodology that extends the Elo model using sound statistical principles to allow the uncertainty in players' ratings to be incorporated formally into the rating system has been worked out. The key component to the extension is that every tournament players' strength can be described by a probability distribution (parametrized by two values) rather than just by a single number, as is currently the case. The flexibility in assuming a probability distribution rather than a single value is that one can distinguish between players' whose abilities are precisely estimated and those whose abilities are not well-estimated. This framework removes the distinction between labeling players as unrated, provisional or established.

In addition to providing a formal remedy to problems associated with the uncertainty in players' ratings, the new methodology supplies solutions to the problems of

- the general tendency for ratings to deflate over time, and
- the differing rates at which younger and older players improve over time.

All of these issues are directly addressed in this extension of Elo's model.

The Committee is working with the Policy Board and the USCF office staff to test the proposed system. Results and recommendations will be presented in next year's report. The documentation of the proposed system can be obtained directly from Mark Glickman (Home address: 20 Belton St., Arlington, MA 02174; e-mail: glickman@figaro.med.harvard.edu).

4 Computation of Performance Ratings

A tournament or event performance rating is usually defined as the rating at which a player's total expected score equals his or her attained score. For example, if a player scores 3.5 points against opponents rated 1600, 1700, 1750, 1800, and 1850, then his or her performance rating is 1896, because the sum of winning expectancies for a player rated 1896 against these opponents would result in 3.5. A commonly used approximation to calculate a performance rating is to average the opponents' ratings, and then add 400 times the number of wins minus losses, divided by the number of games. This is a very crude approximation, and there is no guaranteed accuracy of the resulting computation. Computing a performance rating cannot be accomplished in closed form, but it is fairly straightforward to implement an iterative procedure (e.g., the "Newton-Raphson" algorithm) that quickly converges to the performance rating. In some situations, iterative computation is undesirable – for example, computing performance ratings by hand using an iterative procedure could be time-consuming. In response to such a concern, the following algorithm has been devised which accurately computes performance ratings without any iteration. It should be noted that performance ratings do not exist for total scores comprised of all wins or all losses (they correspond to ratings of positive infinity and negative infinity, respectively), but this algorithm computes a rating which is substantially lower than the result of all losses and a single draw, and is substantially higher than the result of all wins and a single draw.

Suppose a player competes against m opponents with ratings $r_1, \ldots, r_m$, and scores a total of W points out of m possible. First, calculate an initial guess, r_g, of the performance rating by letting

$$r_g = \frac{1}{m}\sum_{i=1}^{m} r_i + \frac{400(2W-m)}{m}.$$

Now let $h_i = 10^{(r_i/400)}$ for each opponent i, and let $h_g = 10^{(r_g/400)}$.

Also let

$$p_i = \frac{h_g}{h_g + h_i}$$

for each i. Compute the following four quantities:

$$\begin{aligned}
a &= \sum_{i=1}^{m} p_i \\
b &= \sum_{i=1}^{m} p_i(1-p_i) \\
c &= \sum_{i=1}^{m} p_i(1-p_i)\frac{h_i - h_g}{h_i + h_g} \\
D &= \begin{cases} \sqrt{b^2 + 2c(W-a)} & \text{if } b^2 > 2c(W-a) \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

Finally, letting $k = 400/\log(10) \approx 173.4$, the approximate performance rating, r_p, is

$$r_p = r_g + k\frac{D - b}{c}.$$

This algorithm can be shown empirically to obtain accuracy comparable to iterative methods.

The Committee plans to use this computation in the calculation of an initial provisional rating, as described in Section 1.4. The Committee has also asked the Policy Board to mandate that players' event performance ratings appear on printed tournament crosstables along with pre-event and post-event ratings.

5 Rating Deflation

A recent concern among tournament organizers in areas where a growing number of tournament participants are scholastic players is the issue of rating deflation. Deflation is the tendency for a rating to become lower over time while a player's ability does not change. A primary cause for rating deflation in areas where scholastic tournaments are prominent is that, typically, the scholastic players compete almost exclusively amongst themselves for awhile before competing against the adult population. The rating system can measure relative strengths among these scholastic players, but as they dramatically improve, the rating system does not simultaneously move all these players' rating upwards to reflect this increase in ability. Instead, the scholastic players as a pool retain roughly the same average rating, so that in comparison to their adult counterparts, they may become drastically underrated. Once these underrated scholastic players begin to play in open tournaments, their ratings begin to increase at the expense of the adults' ratings decreasing. The outcome of such a phenomenon is an overall decline in the average rating for all players.

5.1 Determining the extent of rating deflation

The only technical way to determine the extent of deflation caused by isolated scholastic pools is to examine the outcomes of games played between scholastic players and adults in areas where deflation is suspected to occur. A data analysis would proceed by examining whether younger players are outscoring their adult opponents more often then the ratings predict. It is not sufficient merely to examine the distribution of ratings by age in a particular region and compare the distribution to the rest of the nation because it is impossible to distinguish whether deflation exists or whether players in a region are better, on average, then the rest of the country.

The ratings committee is taking measures to collect data on the outcomes of all rated games from the USCF office as tournaments are rated. One of our tasks will be to determine the extent of rating deflation. We are presently unable to provide a technical assessment of the extent of deflation.

5.2 Preventing deflation

Several approaches can be taken to prevent deflation. We outline two methods below.

Rating floors: Establishing rating floors prevents ratings from declining too far. Unfortunately, this method of preventing deflation has no basis in theory, and may even cause ratings to become inaccurate predictors of performance. While rating floors are simple to implement, they are an ad hoc solution to the problem of rating deflation.

Rating adjustments due to aging: A more formal approach to attacking rating deflation involves adjusting players' ratings due to the passage of time. When players are young and low-rated, we expect their abilities to improve at a relatively faster rate than similarly rated adults. This idea can be incorporated into the rating system by adjusting players ratings according to the amount of time passing between events. For example, if a young player has not competed in 2 years, the current system assumes that the player's ability has not changed. It is much more realistic to believe that the young player, in all likelihood, has improved somewhat, so the player's rating should be adjusted to represent this belief. Committee chairman Mark Glickman has begun the analysis of ratings changes over time to measure the expected rating changes due to age and initial rating level. Not only does this approach make ratings more accurate predictors of performance, but addresses deflationary tendencies of the rating system head on. This approach is described in greater detail in the documentation for the proposed major extensions to the Elo system, available directly from Mark Glickman (see Section 3).

6 Adjusting USCF Ratings from FIDE events

This section describes a procedure adopted by the Policy Board in February 1994 to update players' USCF ratings from the results of FIDE-rated events not normally rated by the USCF. This procedure is now currently in use.

6.1 Background

Every six months, the USCF receives a printout from FIDE detailing rating adjustments to players that are US residents. The USCF has an interest in updating these players' USCF ratings from the outcomes of FIDE rated games to better reflect their current ability. For each US resident competing in FIDE events within the last six months, the computer printout from FIDE lists each FIDE event in which the player competed. For each event, the following useful information is provided:

- nation in which the event occurred,
- pre-tournament FIDE rating
- total score in the event
- total number of games played in the event, and
- difference between the total score and the expected total score.

The algorithm previously used to compute USCF rating updates contained flaws which carried uncertain consequences. The following algorithm was adopted to update USCF ratings from FIDE events.

6.2 The Algorithm

Identify the FIDE-rated events that were not USCF-rated (these should be the ones played outside the US). Perform the following computations for the non-USCF events:

1. Calculate the total number of games competed in the last six months. Call this N.
2. Calculate the sum of the total scores. Call this value W.

3. Calculate the sum of the differences between the total score and expected score (sum of "Chg"). Call this value d.

4. Compute the sum of expected scores, E, by subtraction:

$$E = W - d$$

5. Let A be the amount that needs to be added to the player's FIDE rating to place it on the USCF scale. This value depends on the player's FIDE rating, and is obtained from Table 1 of this report.

6. Let R_f be the pre-tournament FIDE rating, and let R_u be the player's current USCF rating. Now compute the "adjusted" expected score, E_{adj} using the formula:

$$E_{adj} = \frac{N}{1 + 10^{(R_f - R_u + A)/400}(\frac{N}{E} - 1)}.$$

7. Now compute the player's updated rating using the formula

$$R_{new.u} = R_u + K(W - E_{adj})$$

   where

$$K = 8$$

6.3 An Example

FIDE supplied the USCF office with the following information on Gata Kamsky's performances in FIDE events from the first half of 1993:

Start	Fed	City	Ro	Rc	W	N	Chg	K	Kchg
22.02.93	ESP	Linares	2655	2678	5.5	13	-0.61	10	-6.10
20.03.93	USA	New York	2655	2372	2.5	3	-0.02	10	-0.20
10.04.93	GER	Dortmund	2655	2634	3.5	7	-0.21	10	-2.10
18.04.93	ARG	Bns Aires	2655	2563	7.0	11	0.07	10	0.70

The algorithm to update Kamsky's USCF rating proceeds as follows (Kamsky's USCF rating was 2795):

1. Calculate the total number of games competed in the last six months:

$$N = 13 + 7 + 11 = 31$$

2. Calculate the sum of the total scores:

$$W = 5.5 + 3.5 + 7.0 = 16$$

3. Calculate the sum of the differences between the total score and expected score:

$$d = -0.61 + (-0.21) + 0.07 = -0.75$$

4. Compute the sum of expected scores:

$$E = W - d = 16 - (-0.75) = 16.75$$

5. For a player with a FIDE rating of 2655, the corresponding estimated USCF rating from Table 2 is 2701. The value of A is therefore $2701 - 2655 = 46$.

6. Compute the adjusted sum of expected scores:

$$E_{adj} = \frac{31}{1 + 10^{(2655-2795+46)/400}(\frac{31}{16.75} - 1)} = \frac{31}{1 + (0.5821)(0.8507)} = \frac{31}{1.50} = 20.7$$

7. Compute the new USCF rating:

$$R_{new.u} = 2795 + 8 \times (16 - 20.7) = 2795 - 37.6 \approx 2757$$

6.4 Rationale

Ideally, if we knew each of the opponents' FIDE ratings, we could perform a procedure in which we

- convert a player's USCF rating to a FIDE rating from a USCF to FIDE conversion,
- calculate and sum the winning expectancies for each game based on this converted rating (call this E_{new}),
- obtain the updated USCF rating by calculating $K(W - E_{new})$ and adding this to the pre-tournament USCF rating.

This procedure cannot be implemented because the USCF does not receive such detailed information on individual opponents. Instead, we assume that all opponents' ratings are identical subject to the constraint that they produce the statistics on the FIDE printout. Carrying out the mathematics the results in the algorithm described above.

The value of $K = 8$ in the updating computation is half of what is used for USCF rating updates of players rated at least 2400. This smaller K is used because the FIDE results should have less effect on a USCF rating than USCF results would produce. A smaller K reflects that the FIDE events may have occurred at least six months ago, and that the conversion between the USCF and FIDE scales is imprecise.

A FIDE to USCF conversion

The Committee determines a conversion on an annual basis to predict a USCF rating from a FIDE 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 pool of players, and fitting a local regression model ("loess") to the data. Among the 489 players who competed in both USCF and FIDE events in 1993, only players with established USCF ratings, FIDE ratings of at least 2200, and 10 or more FIDE-rated games in 1993, were included. This resulted in a total of 194 players used in the analysis.

The loess fit was performed as a robust procedure (not adversely affected by outliers), using a smoothness criterion based on 75% samples of the data at each point. From diagnostic plots, this variability appears independent of FIDE rating. The results of the fit appear on Table 1 in Section 1.4.2, and are displayed in Figure 1. The loess fit revealed non-linearity in the relationship between FIDE and USCF ratings.

B Initial Age-Based Ratings

An analysis was performed to obtain initial USCF rating assignments for players without a FIDE rating. Players with provisional ratings on the 1993 annual rating list were identified, and only those who were at least 5 years old were included in the analysis (some ratings for players under 5 years did not seem plausible). Also, attention was restricted to players whose provisional ratings were based on at least 6 games because ratings based on fewer than 6 games may not be reliable. A loess fit (see Appendix A) of provisional rating was performed as a function of age and the number of games on which the provisional rating was based. The ratings predicted from the model when the number of games was equal to 6 was used as the initial assignment for different ages. These values are displayed in Table 2 in Section 1.4.2 and shown in Figure 2, where the rating entry on the table corresponds to the predicted value from the model at 6 months after a player's birthday. It is worth noting that the fitted values are slightly lower than the average provisional rating by age. This may be apparent from observing Figure 2. The reason is that the fitted values correspond to provisional ratings based on only 6 games, which tend to be less than the average provisional rating.