

Propuesta: Modificación a la Normativa de Evaluación de la FCE

En vista de los problemas que hemos detectado con la actual normativa de evaluación de la FCE queremos presentar la siguiente propuesta de modificación del rating. Esta propuesta no incorpora la redacción reglamentaria, puesto que algunos de sus soluciones son abiertas y sometidas a la voluntad de la asamblea, y hasta que no se haya decidido cuál se escoge no se puede hacer la redacción definitiva.

Las cuestiones más importantes que se quieren solventar son:

- Los factores K utilizados.
- La evaluación de las partidas contra jugadores no federados a la FCE.
- La evaluación de las partidas de los torneos sub-12.

Como criterio principal se ha escogido el mantener el sistema lo más simple posible (el principio KISS, https://es.wikipedia.org/wiki/Principio_KISS) y que cualquier complicación al sistema tenga que estar bien fundamentada y con una justificación importante.

Al principio del documento está la propuesta puramente dicha, muy corta. Se añaden además unos anexos para justificar esa elección.

PROPUESTA

1. Se mantiene como sistema de rating el ideado por Elo, descartando el uso de Glicko o similares, y seguiremos usando las tablas del art 13 del actual reglamento.
2. Factor K. Se utilizará con carácter general un factor $K = 20$.
3. Se plantea a la asamblea la posibilidad de utilizar un factor K aumentado para menores y jugadores recién incorporados al sistema. Se propone:
 1. K40 a todos los jugadores hasta que hayan terminado eventos que sumen 30 o más partidas evaluadas.
 2. K40 a los jugadores menores de 18 años mientras su rating se mantenga por debajo de 2300.
4. Rating de entrada al sistema. Se propone:
 1. Usar el mismo sistema que utiliza la FIDE: calcular el rating inicial según la performance del primer (o los primeros) torneos del jugador, con un mínimo de cinco partidas. Hasta entonces el jugador se considerará unrated.
 2. En el caso de jugadores que ya tengan rating FIDE, su rating inicial será su rating FIDE + el diferencial publicado por la FCE (ver punto 8).
5. Evaluación de partidas contra jugadores no federados en la FCE. Se contemplan dos casos:
 1. Si el jugador tiene rating FIDE, se evaluará como si la partida se hubiera jugado contra un jugador de rating FCE igual al FIDE + el diferencial publicado por la FCE (ver punto 8).
 2. Si el jugador tampoco tiene rating FIDE la partida no se evaluará.
6. Evaluación de partidas sub-12. Una vez que los menores ya no entren al sistema con un elo de 1700 no tienen sentido mantener la K reducida, así que se evaluarían igual que el resto de partidas.
7. Las partidas perdidas o ganadas por incomparecencia se seguirán evaluando como hasta ahora.
8. Con el propósito de ser utilizado en los cálculos de los puntos 4.2 y 5.1, anualmente la FCE publicará un diferencial entre rating FIDE y rating de la FCE. Dicho diferencial se basará en la diferencia media entre ambos sistemas de rating con los datos de que disponga la FCE, y podrá dividirse en tramos si así lo estima conveniente.

En resumen el rating propuesto es básicamente el utilizado por la FIDE, con las siguientes diferencias:

1. Los jugadores con rating FIDE no entrarán al sistema como unrated, sino con su rating FIDE más un diferencial.
2. Las partidas contra jugadores sin rating FCE (pero sí con rating FIDE) no se quedarán sin evaluar, sino que se evaluarán como si fuera contra un jugador de rating FCE igual al rating FIDE más un diferencial.
3. Las partidas perdidas o ganadas por incomparecencia SÍ se evaluarán (a diferencia de lo que ocurre con el sistema de la FIDE)

Este último punto es por una razón práctica: tal y como explicó en su día Toni Ayza la introducción de esa medida ha servido para, de una forma sencilla, reducir en gran medida las incomparecencias en las primeras rondas de los torneos y, además, tampoco es descabellado incluir en el rating la tendencia de los jugadores a no presentarse a las partidas, puesto que también nos da información sobre sus resultados futuros.

ANEXOS

A. ¿POR QUÉ EL SISTEMA ELO? ¿NO ESTÁ ANTICUADO? ¿Y POR QUÉ NO EL GLICKO? ¿NO LO PODEMOS MEJORAR TOMANDO COSAS DE UNO U OTRO?

Cuando Arpad Elo diseñó su sistema lo hizo estableciendo una escala arbitraria según la cuál habría 200 puntos de diferencia entre cada categoría de jugador (en la USCF), asumiendo una distribución normal de los resultados según la diferencia de rating, y buscando un método sencillo acorde a la tecnología de la época. Hoy en día con la proliferación de los computadores se han diseñado otros sistemas más precisos siendo quizá Glicko <http://www.glicko.net/glicko.html> el más conocido de ellos y utilizado en muchos sitios de ajedrez online como Lichess, FICS, chess.com y en la Federación Australiana. El actual sistema de la FCE dice basarse también en él.

En 2011 la FIDE organizó una competición de sistemas de evaluación esponsorizada por Deloitte <https://www.kaggle.com/c/ChessRatings2> con un premio de \$10,000 buscando un sistema que pudiera sustituir al Elo y que se presentaría durante una reunión de expertos en ratings de la FIDE. Se presentaron 181 sistemas distintos de los cuales los 10 finalistas y un reserva fueron <https://www.kaggle.com/c/ChessRatings2/forums/t/501/fide-prize-final-standings>:

Rank: Private score (Public score, Submission date): Team name

#1: 0.256683 (0.256237, 04/25/2011 03:50): Reversi

#2: 0.257354 (0.257094, 05/04/2011 14:17): Uri Blass

#3: 0.257435 (0.257001, 05/04/2011 07:31): uqwn

#4: 0.257608 (0.257411, 04/22/2011 20:42): JAQ

#5: 0.257622 (0.257287, 05/04/2011 01:17): Real Glicko

#6: 0.257723 (0.257482, 05/04/2011 13:11): TrueGrit

--- Glicko Benchmark (using c=15.8) scored 0.257834 ---

#7: 0.258554 (0.258238, 04/11/2011 00:10): Stalemate

#8: 0.258950 (0.258358, 04/19/2011 11:36): chessnuts

--- Actual FIDE Ratings Benchmark scored 0.259751 ---

#9: 0.259901 (0.259560, 04/07/2011 05:24): Nirtak

#10: 0.259947 (0.259794, 05/03/2011 12:29): AFC

#11: 0.260296 (0.260350, 05/03/2011 14:45): Dave Poet

Como se puede observar el sistema de Elo se habría clasificado entre los diez finalistas de haberse presentado. La diferencia de precisión que aportan los mejores sistemas propuestos es tan pequeña que no tiene casi significación. Siendo el caso, ¿para qué vamos a complicar el sistema?

En cualquier caso es importante señalar que tomar elementos de varios sistemas distintos y mezclarlos sin más no mejora para nada el sistema sino que ocurre al contrario. Hay varios sistemas fiables y probados, podemos usar cualquiera de ellos pero tomar un poco de cada uno no tiene sentido.

B. EL LÍO DE LAS K Y LAS DELTAS

Mientras las tablas de resultado esperado vs diferencia de Elo han permanecido inalteradas desde que Arpad Elo presentó su sistema hasta hoy, determinar el valor de la K es, junto al rating de entrada en el sistema, lo que más polémica ha generado.

Cuando la FIDE pasó a publicar las listas con más frecuencia se hizo necesario modificar el valor de la K, ya que la mayor frecuencia de actualización del rating hacía que la efectividad de la K bajara (ie, hace falta una K más elevada para que se de el mismo efecto). Hubo entonces un interesante debate sobre el valor que debía tener la K que están recogidos en <http://en.chessbase.com> entre diversos expertos en el sistema Elo. Quizá la más clarificadora fue la aportación del GM Bartłomiej Macieja <http://en.chessbase.com/post/rating-debate-6-here-comes-the-proof-> (se puede ver el resto del debate en la lista de páginas relacionadas).

En su post explica que el objetivo del sistema de evaluación es muy simple: **que dos jugadores que tengan exactamente los mismos resultados deben acabar teniendo el mismo rating**. Cuánto más pequeño sea el valor de K más tiempo se necesitará para que eso ocurra; el tiempo que se tardará dependerá también de la diferencia inicial de rating entre ambos jugadores y del número de actualizaciones anuales del rating (de hecho la K tuvo que actualizarse al pasar a tener listados mensuales).

Como valor de compromiso Bartłomiej toma que dos jugadores con una diferencia inicial de 100 puntos de Elo consiguiendo exactamente los mismos resultados tendrían que acabar teniendo el mismo Elo al cabo de un año, incluso proporciona una fórmula para estimar una K que logre converger en un tiempo dado con un margen del 1%. Haciendo el mismo ejercicio podemos calcular cuánto tardaría un jugador que tuviera un desempeño de 100 puntos por debajo o por encima de su rating en ver reflejado ese desempeño en su evaluación. Cada columna es el tiempo que tarda en quedarse a 50 puntos de su desempeño (la mitad del recorrido), a 25, a 10, a 5 y a 1 (el objetivo marcado por Bartłomiej). Si contamos que cada torneo se juegan 9 partidas tenemos:

	'+50	'+25	'+10	'+5	'+1	Final
K40	1	2	4	5	6	2500
K30	2	3	5	7	9	2501
K20	3	5	7	9	14	2501
K10	6	10	15	22	-	2503
K5	11	20	34	-	-	2507

El gráfico que sigue muestra la evolución del rating de un jugador que tuviera un rating inicial de 2600, pero una performance continua y estable de 2500. Al cabo del tiempo su rating debería acomodarse a la fuerza de juego que muestra.

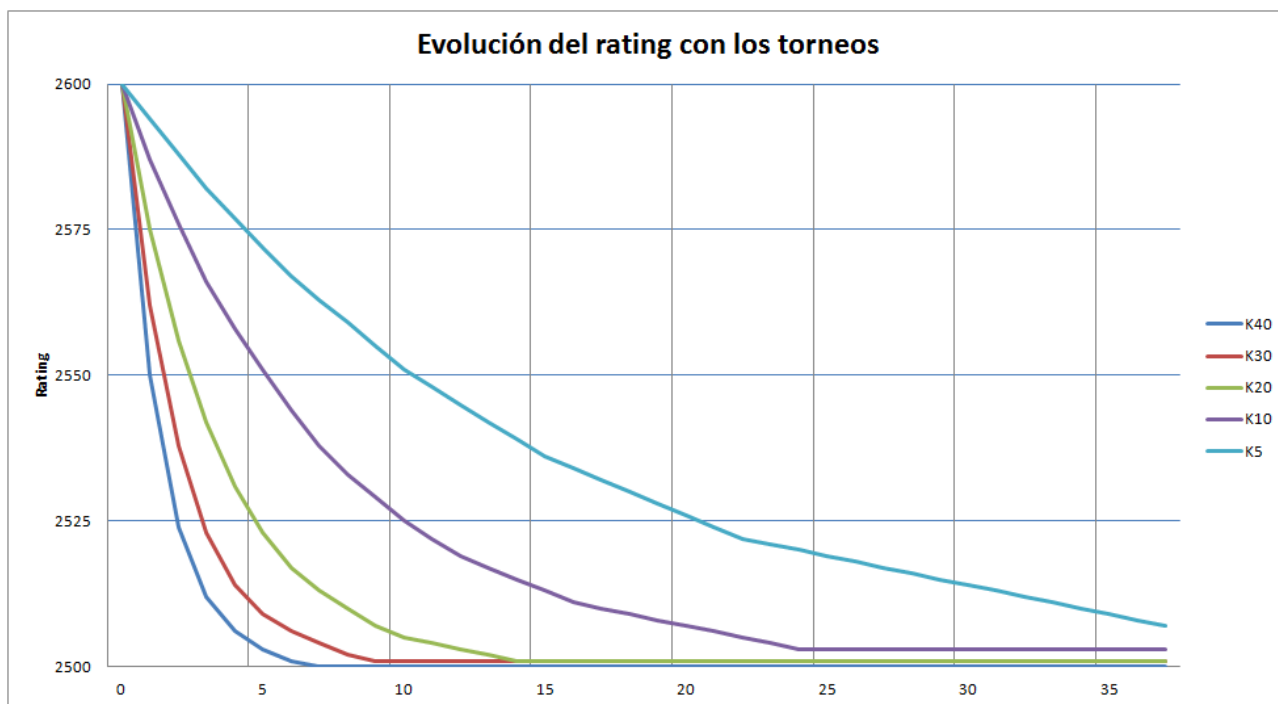


Ilustración 1: Variación del rating de un jugador con rating inicial 2600 con una performance estable de 2500

Si tomamos las estimaciones de la propia FCE en cuanto a torneos jugados anualmente (las que da para el nuevo cálculo de las delta-K) tenemos que el número normal de torneos sería de 4-6 al año. La tabla muestra que **una K20 es un valor bastante razonable**, que hace que en 9 torneos (unos dos años de media) el rating converja con un 5% de error. Citando a Bartlomiej si un jugador lleva unas 100 partidas jugando a un nivel determinado no es descabellado pensar que ese sea su nivel.

K40 consigue que el rating converja en un año (5 torneos), pero a costa de fuertes variaciones cada torneo (el 50% de los puntos llegan en el primer torneo). Una K10 haría que tardáramos unos cuatro años de media (22 torneos, unas 200 partidas) en alcanzar un error del 5%. La FIDE marca este valor de K para jugadores con elo superior a 2400, por entender que estos jugadores de mayor nivel juegan también un mayor número de partidas al año con lo que tardarían bastante menos de esos cuatro años en converger.

Una K5 como las que se usan en la FCE no llega a converger nunca ni al 1% ni al 5%, necesitaría 7 años para alcanzar un error del 10% (34 torneos, más de 300 partidas) y eso si la fuerza del jugador se mantiene estable durante todo este tiempo; y es totalmente inútil para seguir la evolución de ningún jugador. Las K3 que también se han usado y no contemplamos en la tabla tampoco llegan a converger al 10% y necesitan casi 450 partidas para un error por debajo del 15%.

C. NUEVO REDACTADO DEL “CAPÍTOL III. NORMATIVA D'AVALUACIÓ DE LA FCE”

Article 1 Introducció

El sistema d'avaluació de la Federació Catalana d'Escacs té com a objectiu la valoració de les actuacions dels jugadors afiliats mitjançant procediments estadístics inspirat en els sistemes ELO de la FIDE i la USCF.

Article 2 Àmbit

2.1 L'àmbit d'avaluació estarà integrat per les competicions oficials i les competicions que prèviament a la seva celebració, hagin estat registrades i homologades per la Federació Catalana d'Escacs. Aquest àmbit d'avaluació afectarà a tots els jugadors afiliats a la FCE.

2.2 La Federació Catalana d'Escacs es reserva el dret d'avaluar altres torneigs quan ho cregui convenient, tot i que no compleixin les condicions establertes.

Article 3 Avaluació

3.1 El sistema d'avaluació de la Federació Catalana d'Escacs comptabilitza, de forma diferenciada, dos tipus d'avaluació: l'avaluació de ritme estàndard i l'avaluació de ritmes ràpids.

Article 4 Reglamentació aplicable

4.1 Les partides es jugaran d'acord a les Lleis dels Escacs de la FIDE.

4.2 A totes les proves serà d'aplicació el Reglament General i de Competicions de la Federació Catalana d'Escacs.

Article 5 Ritme de joc

5.1 Perquè les partides siguin avaluades per a ritme estàndard cada jugador haurà de disposar d'un temps mínim per completar la partida assumint que es faran 60 jugades. El temps mínim serà de 60 minuts com a resultat de sumar el temps fix més el temps afegit per jugada multiplicat per 60 (ritme Standard).

5.2 Perquè les partides siguin avaluades per a ritmes ràpids cada jugador haurà de disposar d'un temps que pot ser fix, o que la suma del temps assignat més l'increment multiplicat per 60 sigui inferior a 60 minuts.

Article 6 Duració del torneig i calendari

6.1 El nombre de rondes diàries no serà superior a 3 per torneigs computats a ritme Standard.

6.2 La durada de la prova no pot ésser superior a 90 dies, excepte en les competicions per equips.

Article 7 Partides avaluades

7.1 Per calcular la primera performance d'un jugador sense avaluació aquest haurà haver puntuat mínim 1/2 punt.

7.2 Pels jugadors amb avaluació només les partides contra jugadors amb avaluació es tindran en

compte.

7.3 Partides contra jugadors sense avaluació catalana però amb avaluació FIDE

7.3.1 Una partida jugada contra un jugador sense avaluació catalana però amb avaluació FIDE equivaldran a una partida jugada amb un jugador que tingui una avaluació igual al seu elo FIDE més el diferencial publicat per la Federació a principi d'any segons l'article 9.5, i es tindrà en compte tant per l'avaluació de jugadors amb avaluació catalana com pel càlcul de la primera performance dels jugadors no avaluats.

7.3.2 La avaluació FIDE a utilitzar en el càlcul anterior serà corresponent al ritme de joc de les partides avaluades.

7.3.3 En el cas de partides a ritme ràpid, si el jugador no té avaluació FIDE corresponent al ritme de joc de les partides avaluades, s'utilitzarà la seva avaluació FIDE d'escacs estàndard.

Article 8 Incompareixences a les partides

8.1 Les partides guanyades per incompareixença no seran vàlides per l'avaluació catalana.

8.2 Les partides perdudes per incompareixença seran vàlides per l'avaluació catalana. Una partida perduda per incompareixença equivaldrà a una partida jugada amb un jugador del mateix ELO que el propi jugador, i la meitat del factor K del jugador.

8.3 Les partides no jugades per incompareixença de qualsevol dels jugadors no es tindran en compte per calcular la primera performance d'un jugador.

Article 9 Llistes oficials Procediments d'avaluació

9.1 La Federació publicarà anualment dotze (12) llistes de l'avaluació catalana per a ritme estàndard i dotze (12) llistes per a ritmes ràpids. Les dates de publicació seran el primer dia de cada mes. Aquesta relació incorporarà els resultats de les proves avaluades des de l'anterior publicació i serà vigent mentre no es publiqui una de nova.

9.2 Les llistes d'avaluació catalana incorporaran tots els jugadors que tinguin tramitada la llicència per la federació catalana d'escacs a l'actual temporada.

9.3 Els jugadors podran presentar reclamacions del seu ELO dintre dels vint-i-cinc (25) primers dies, comptant a partir de la data de publicació de cada llista oficial.

9.4 Cada Llista oficial d'ELO inclourà les següents dades:

Nom del jugador
Codi d'identificació
Titulació (WCM, CM, MC, MN, WMF, MF, WMI, MI, WGM i GM)
Club
Avaluació
Factor K
Nombre de partides avaluades en el corresponent període
Altra informació rellevant

9.5 La Federació publicarà anualment un diferencial entre avaluació FIDE i catalana que s'aplicarà a

9.5.1 Avaluació de les partides jugades contra jugadors sense avaluació catalana però amb avaluació FIDE, com s'estableix a l'article 7.3.

9.5.2 Primera avaluació catalana dels jugadors que tenen avaluació FIDE, com s'estableix a l'article 10.1.

Article 10 Avaluació inicial dels jugadors per la llista d'avaluació d'escacs estàndard

10.1 Als jugadors avaluats per primera vegada amb avaluació FIDE se'ls assignarà una avaluació inicial igual a la seva avaluació FIDE més el diferencial publicat per la Federació.

10.2 A la resta de jugadors avaluats per primera vegada:

10.2.1 Als jugadors avaluats per primera vegada se'ls assignarà una avaluació inicial quan hagin jugat un mínim de 5 partides contra jugadors avaluats, inclosos els rivals sense avaluació catalana però amb avaluació FIDE que s'avaluaran conforme l'article 7.3.

10.2.2 Les 5 partides que demana l'article anterior no s'han de completar necessàriament en un únic torneig. Si es necessiten dos o més tornejos per completar 5 o més partides, s'avaluaran totes les partides com si s'haguessin jugat al mateix torneig.

10.2.3 Si el primer torneig d'un jugador aquest no suma cap punt, les partides d'aquest torneig es descartaran.

Article 11 Avaluació inicial dels jugadors per a la llista d'avaluació d'escacs ràpids

11.1 Als jugadors avaluats per primera vegada que tinguin avaluació d'escacs estàndard, se'ls assignarà una avaluació inicial d'escacs ràpids igual a la seva avaluació d'escacs estàndard.

11.2 Als jugadors avaluats per primera vegada que tinguin avaluació FIDE, se'ls assignarà una avaluació inicial d'escacs ràpid igual a la seva avaluació FIDE més el diferencial publicat per la Federació. La avaluació FIDE a utilitzar serà, per aquest ordre:

11.2.1 Avaluació FIDE corresponent a la modalitat de les partides a avaluar, si la té (ràpid o blitz, segons la normativa FIDE).

11.2.2 Avaluació FIDE d'escacs estàndards, si el jugador no té avaluació per la modalitat de les partides a avaluar però en té avaluació estàndard

11.3 Als jugadors avaluats per primera vegada que no tinguin avaluació d'escacs estàndard, ni avaluació FIDE, se'ls assignarà una avaluació inicial seguint el procediment de l'article 10.2

Article 12 Avaluació dels jugadors que aconseguen la seva avaluació inicial abans de finalitzar un torneig

12.1 Si un jugador obté la seva avaluació inicial abans de que un torneig determinat sigui avaluat, seguirà comptant com si no tingués avaluació pels seus rivals (i la partida no es tindrà en compte) però el serà avaluat com un jugador amb avaluació, pel que fa a la seva pròpia avaluació.

Article 13 Descripció del sistema ELO

13.1 El sistema ELO és un procediment numèric en el qual els percentatges de puntuació es converteixen en diferències d'avaluació i viceversa. La seva funció es produir una estimació estadística el més acurada possible de la força de joc d'un jugador.

13.2 La Taula 1 mostra la conversió de diferència d'avaluació D en percentatge de puntuació PD per el jugador més valorat M i el menys valorat m.

D	M	m	D	M	m	D	M	M
0 – 3	.50	.50	122 – 129	.67	.33	279 – 290	.84	.16
4 – 10	.51	.49	130 – 137	.68	.32	291 – 302	.85	.15
11 – 17	.52	.48	138 – 145	.69	.31	303 – 315	.86	.14
18 – 25	.53	.47	146 – 153	.70	.30	316 – 328	.87	.13
26 – 32	.54	.46	154 – 162	.71	.29	329 – 344	.88	.12
33 – 39	.55	.45	163 – 170	.72	.28	345 – 357	.89	.11
40 – 46	.56	.44	171 – 179	.73	.27	358 – 374	.90	.10
47 – 53	.57	.43	180 – 188	.74	.26	375 – 391	.91	.09
54 – 61	.58	.42	189 – 197	.75	.25	392 – 411	.92	.08
62 – 68	.59	.41	198 – 206	.76	.24	412 – 432	.93	.07
69 – 76	.60	.40	207 – 215	.77	.23	433 – 456	.94	.06
77 – 83	.61	.39	216 – 225	.78	.22	457 – 484	.95	.05
84 – 91	.62	.38	226 – 235	.79	.21	485 – 517	.96	.04
92 – 98	.63	.37	236 – 245	.80	.20	518 – 559	.97	.03
99 – 106	.64	.36	246 – 256	.81	.19	560 – 619	.98	.02
107 – 113	.65	.35	257 – 267	.82	.18	620 – 735	.99	.01
114 – 121	.66	.34	268 – 278	.83	.17	> 735	.99	.01

Taula 1. Conversió de diferència d'avaluació D en percentatge PD. M = Jugador més avaluat. m = Jugador menys avaluat.

13.3 La Taula 2, necessàriament mirall de la taula anterior, mostra la probabilitat de puntuació en funció de les diferències d'avaluació dels dos jugadors.

D	M	D	M	D	M	D	M	D	M	D	M
		.83	273	.66	117	.49	-7	.32	-133	.15	-296
.99	677	.82	262	.65	110	.48	-14	.31	-141	.14	-309
.98	589	.81	251	.64	102	.47	-21	.30	-149	.13	-322
.97	538	.80	240	.63	95	.46	-29	.29	-158	.12	-336
.96	501	.79	230	.62	87	.45	-36	.28	-166	.11	-351
.95	470	.78	220	.61	80	.44	-43	.27	-175	.10	-366
.94	444	.77	211	.60	72	.43	-50	.26	-184	.09	-383
.93	422	.76	202	.59	65	.42	-57	.25	-193	.08	-401
.92	401	.75	193	.58	57	.41	-65	.24	-202	.07	-422
.91	383	.74	184	.57	50	.40	-72	.23	-211	.06	-444
.90	366	.73	175	.56	43	.39	-80	.22	-220	.05	-470
.89	351	.72	166	.55	36	.38	-87	.21	-230	.04	-501
.88	336	.71	158	.54	29	.37	-95	.20	-240	.03	-538
.87	322	.70	149	.53	21	.36	-102	.19	-251	.02	-589
.86	309	.69	141	.52	14	.35	-110	.18	-262	.01	-677
.85	296	.68	133	.51	7	.34	-117	.17	-273		
.84	284	.67	125	.50	0	.33	-125	.16	-284		

Taula 2. Conversió de diferència d'avaluació p en probabilitat de puntuació dp

Article 14 Càlcul de la avaluació inicial d'un jugador sense avaluació

14.1 Es calcula la mitjana d'ELO dels seus oponents avaluats durant el torneig (R_a).

14.1.1 Als jugadors sense avaluació catalana però amb avaluació FIDE se'ls compararà la seva avaluació FIDE més un diferencial, seguint el procediment de l'article 7.3.

14.2 Es calcula el percentatge de punts realitzat a les partides avaluades al punt anterior.

14.3 Es calcula amb la Taula 2 (dp) corresponent al percentatge de punts realitzat.

14.3 Si fa el 50% dels punts, la seva avaluació inicial serà $R_i = R_a$.

14.4 Si fa més del 50% dels punts, la seva avaluació inicial serà $R_i = R_a + 20$ per cada mig punt per sobre del 50% dels punts.

14.5 Si fa menys del 50% dels punts, $R_i = R_a + dp$

Article 15 Variació d'ELO en una partida

15.1 Utilitzant la Taula 1, es calcula el resultat esperat PD (M o m) per cada jugador.

15.2 Es calcula la variació d'ELO R d'aquella partida com: $R = K (P - P_d)$

A on:

P = Resultat de la partida

P_D = Resultat esperat (M o m)

K = Factor d'estabilització

Article 16 Factor K

16.1 El factor K és un coeficient d'estabilització del sistema ELO utilitzat per compensar la incertesa en la estimació de la valoració d'un jugador.

16.2 El factor K de cada jugador serà el següent:

16.2.1 Per l'avaluació d'escacs estàndard:

K = 40 pels nous jugadors, fins que completin tornejos per 30 o més partides.

K = 40 pels jugadors menys de 18 anys i menys de 2200 d'avaluació.

K = 20 a la resta de casos

16.2.2 Per l'avaluació d'escacs a ritme ràpid K és sempre 20.

Article 17 Variació d'ELO en una avaluació - R(n)

17.1 La variació d'ELO en una avaluació R(n) es calcula com la suma de les variacions d'avaluació de totes les partides computades dins del termini d'avaluació.

17.2 Una vegada feta la suma, R(n) s'arrodoneix cap als valors més pròxims a la unitat.

17.3 Si en algun moment la variació d'ELO d'un jugador fa que la seva valoració sigui inferior a 1000 punts, la seva valoració serà 1000.