

Propuesta de Sonas: reparar el sistema de clasificación Elo estándar de la FIDE

-- por Jeff Sonas (20 de julio de 2023)

Introducción y resumen ejecutivo

Para decirlo sin rodeos, la última década ha traído una deflación extrema de calificación al sistema de calificación Elo estándar de la FIDE. Las calificaciones de los jugadores están mucho más separadas de lo que deberían y empeoran cada año. La FIDE debería tomar medidas firmes para revertir los efectos de esta deflación y evitar que vuelva a suceder. Propongo dos medidas correctivas para lograr estos objetivos conjuntos.

La primera medida correctiva, a la que llamo Compresión, traerá una mejora inmediata y poderosa a la precisión de las calificaciones. Todos los jugadores con una calificación de Elo inferior a 2000, que representa el 85% inferior de los jugadores calificados, deberían recibir un aumento de calificación único, en la lista de calificación de enero de 2024, usando una fórmula simple como se resalta en amarillo en el gráfico adjunto.

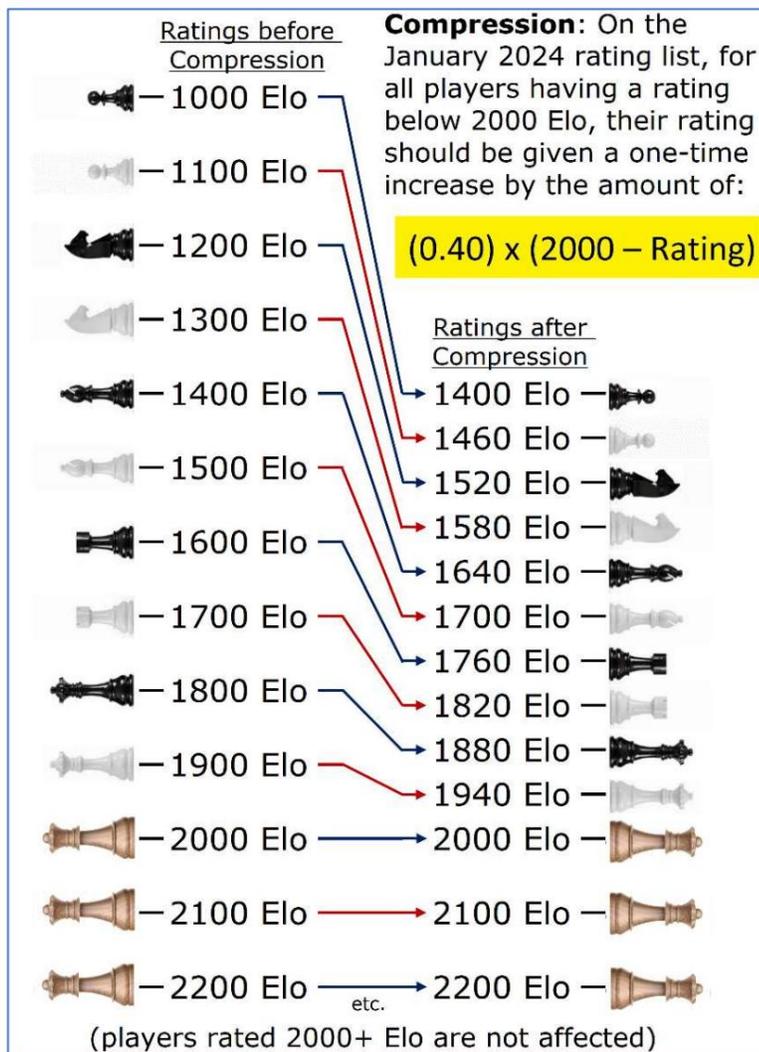
Por ejemplo, mire el peón negro a la izquierda, que representa a cualquiera que tenga una calificación de 1000 Elo antes de la Compresión. sucede. Un jugador así obtendría un aumento de 400 puntos Elo y terminaría con una calificación de 1400 Elo, como se puede ver en el peón negro de la derecha.

Como otro ejemplo, mire el alfil blanco a la izquierda, que representa a cualquiera que tenga una calificación de 1500 Elo antes de que ocurra la compresión. Obtendrían un aumento de 200 puntos Elo y terminarían con una calificación de 1700 Elo, como puede ver en el alfil blanco de la derecha.

Tenga en cuenta también que no hay discontinuidades en la fórmula; por ejemplo, no hay ventaja ni desventaja en tener poco más de 1500 o poco menos de 1500. Si tuvieras 3 jugadores con calificaciones de 1495, 1500 y 1505, respectivamente, recibirían aumentos de calificación únicos de 202, 200 y 198, respectivamente.

Las piezas de ajedrez en este diagrama se proporcionan sólo para ayudarle a seguir las líneas y para ayudar a ilustrar lo que sucedería para los jugadores en varios lugares de la lista de clasificación.

Los jugadores con una calificación justo por debajo de 2000 Elo casi no obtendrían ningún aumento. Y para todos los jugadores que tengan una calificación de Elo de 2000 o superior, su calificación no se verá afectada directamente por esto. Tenga en cuenta que este cambio preservaría el orden actual de todos los que figuran en la lista de clasificación; solo ajustaría sus espacios entre sí. Y finalmente, nadie tendría una calificación inferior a 1400 Elo.



La segunda medida correctiva, las Mejoras de Cálculo, intenta garantizar que la deflación de las calificaciones no vuelva a ocurrir, introduciendo algunos efectos inflacionarios leves en las regulaciones de calificación.

Esto incluye aumentar la calificación mínima de 1000 Elo a 1400 Elo, restaurar la regla de los 400 puntos a su estado anterior para que pueda aplicarse varias veces en un evento y ajustar la fórmula utilizada para calcular las calificaciones iniciales para que incorpore dos empates adicionales contra hipotéticos oponentes con calificación de 1800.

Entiendo que estos son pasos muy importantes a tomar y no esperaré que nadie los apoye sin entender por qué son necesarios. Por eso no sólo propongo estas medidas, sino que también las justifico como parte de esta propuesta. Las medidas correctivas se describen con más detalle en la Parte III, a continuación, pero primero es necesario brindar algunos antecedentes y describir el problema en detalle.

Parte I – Antecedentes

La FIDE mantiene una lista mundial oficial de clasificación de ajedrez, basada en el sistema de clasificación Elo, para todas las partidas reportadas jugadas en controles de tiempo estándar. Cada mes, las calificaciones de los jugadores se actualizan en función de todos los resultados de los torneos de ajedrez informados a la FIDE ese mes, y la FIDE publica en línea la lista de calificaciones actualizada. El sistema Elo garantiza que los jugadores que excedieron su puntuación esperada verán aumentar sus calificaciones, y los jugadores que no alcanzaron su puntuación esperada verán disminuir sus calificaciones. Los nuevos jugadores califican para la lista una vez que han jugado suficientes juegos contra jugadores clasificados existentes, por lo que la lista de clasificación crece cada mes.

Las primeras listas oficiales de clasificación de la FIDE, a principios de los años 1970, incluían sólo jugadores con un índice Elo superior a 2200. Esta restricción se mantuvo durante veinte años, mientras que la lista de clasificación creció veinte veces más, de 600 jugadores a principios de los años 1970 hasta 12.000 jugadores en principios de los años 1990.

Con la reducción del rating mínimo a 2000 Elo en enero de 1993, la FIDE comenzó una nueva fase de veinte años en la que se permitía a jugadores con ratings cada vez más bajos entrar en las listas de rating. El mínimo se redujo nuevamente a 1800 Elo en octubre de 2001 y luego incrementalmente a 1600, 1400 y 1200, hasta que se estableció el mínimo actual de 1000 Elo en agosto de 2012.

Esta reducción gradual del rating mínimo aumentó enormemente el número de ajedrecistas clasificados por la FIDE en todo el mundo. Por ejemplo, había menos de 40.000 jugadores en la lista de clasificación de enero de 2003, en comparación con más de 400.000 jugadores en la lista de enero de 2023. Aquí hay un gráfico que muestra cómo la distribución de jugadores clasificados (agrupados en rangos de 100 puntos Elo) ha cambiado a lo largo de esos veinte años:



Sin embargo, las reducciones de la calificación mínima tuvieron un efecto secundario muy problemático. Muchos jugadores más jóvenes/débiles reciben calificaciones iniciales FIDE muy bajas a una edad muy temprana, durante la etapa de su carrera en la que sus habilidades ajedrecísticas mejoran más rápidamente, y esta combinación de calificaciones iniciales bajas y jugadores que mejoran rápidamente ha producido en última instancia una fuerte presión deflacionaria sobre el sistema general de calificación durante la última década.

Hace veinte o treinta años, todos estábamos preocupados por la "inflación de rating", mientras veíamos que cada vez más jugadores con rating 2700+ llenaban el top 100, pero ya pasó la era de la inflación de rating.

De hecho, esa era inflacionaria probablemente ya había quedado atrás en el momento de esa lista (morada) de enero de 2013 que acabamos de ver. Desde hace muchos años, la "deflación de las calificaciones" ha sido en realidad el mayor problema, aunque no se notará con sólo mirar las listas mensuales de los 100 principales.

En 2014, la FIDE promulgó medidas diseñadas para contrarrestar la deflación del rating en la lista de rating estándar, aumentando la mayoría de los factores K y también manteniendo altos los factores K para los jugadores juveniles más débiles, incluso después de haber jugado muchas partidas puntuadas. Los factores K más altos harían que las calificaciones de esos jugadores aumentaran más rápidamente cuando superaran su expectativa de calificación. La esperanza era que las calificaciones de los jóvenes en mejora se mantuvieran más fácilmente al día con sus habilidades mejoradas, y así sus próximos oponentes no perderían tantos puntos de calificación al enfrentarlos. Sin embargo, esas medidas de 2014 no han demostrado ser suficientes para contrarrestar las presiones deflacionarias.

Parte II – El problema de la deflación

Aunque el sistema Elo todavía funciona razonablemente bien para los jugadores de nivel de gran maestro, el hecho es que el 99% de los jugadores clasificados tienen puntuaciones Elo inferiores a 2400, y existen algunos problemas muy serios con sus puntuaciones.

El grupo de rating activo duplicó su tamaño durante los siete años comprendidos entre 2013 y 2020, gracias principalmente a una gran afluencia de jugadores juveniles más débiles, muchos de los cuales fueron inmediata y significativamente subestimados desde su primera calificación. A medida que estos jugadores comenzaron a jugar con mayor habilidad de lo que sus propias calificaciones predecirían, el funcionamiento natural del sistema Elo comenzó a transferir una gran cantidad de puntos de calificación a estos nuevos jugadores fuera del grupo de calificación existente.

Es relativamente fácil entender cómo se ve la "inflación de calificación": los jugadores obtienen nuevas calificaciones que son demasiado altas (en relación con el grupo de calificación existente) y comienzan a perder sus puntos de calificación sobrantes en el grupo de jugadores establecidos, quienes ahora se vuelven ligeramente más bajos. sobrevalorado, y luego esos jugadores (a su vez) tienen un rendimiento ligeramente inferior a sus calificaciones y distribuyen los puntos sobrantes aún más... y con el tiempo esos puntos sobrantes se extienden por todo el sistema de calificación.

Es un poco más difícil visualizar cómo se ve la "deflación de calificaciones". Más que un "exceso" de puntos de rating, se trata más bien de una "deficiencia" de puntos de rating, pero todo lo demás funciona de manera similar. Por lo tanto, estos nuevos jugadores ingresan al sistema de calificación fuertemente subestimados (en relación con el grupo de calificación existente), con una gran "deficiencia" de puntos de calificación. A medida que superan sus bajas calificaciones y les quitan puntos a sus oponentes, esos oponentes a su vez adquieren una "deficiencia" de puntos de calificación (se vuelven un poco subestimados), y así sucesivamente a lo largo de todo el grupo de calificaciones. En última instancia, la calificación de todos se reduce, aunque la gravedad variará según el grado de conexión de cada uno con los nuevos jugadores (por ejemplo, X juega con Y, quien juega con Z, quien juega con un jugador nuevo). Los grandes maestros son los que están más alejados de los nuevos jugadores y, por tanto, los que están más protegidos del efecto.

Todo lo que acabo de describir sobre la "deflación del rating" realmente le ha sucedido, durante la última década, al sistema de rating Elo estándar de la FIDE. Han entrado nuevos jugadores al sistema, cientos de miles de nuevos jugadores, muchos de ellos con una gran deficiencia de puntos de clasificación, y su impacto general ha sido la reducción de las clasificaciones del grupo de clasificación establecido.

Este efecto deflacionario no habría sido demasiado problemático si el conjunto de calificaciones hubiera sido un grupo cerrado, porque el sistema Elo eventualmente encontraría un equilibrio adecuado, aunque tal vez uno en el que todos tuvieran calificaciones ligeramente más bajas que antes. Sin embargo, el sistema de calificación nunca tuvo tiempo de recuperar el aliento y establecer tal equilibrio. Siempre había muchos jugadores nuevos ingresando al sistema cada mes, obteniendo calificaciones iniciales que eran demasiado bajas e inmediatamente quitando más puntos de calificación a los jugadores establecidos, ¡incluso a oponentes jóvenes que recientemente habían sido jugadores nuevos y subestimados! Es una presión deflacionaria constante que se está aplicando incluso ahora, y que empeora cada año, y no mejorará hasta que se implementen contramedidas significativas.

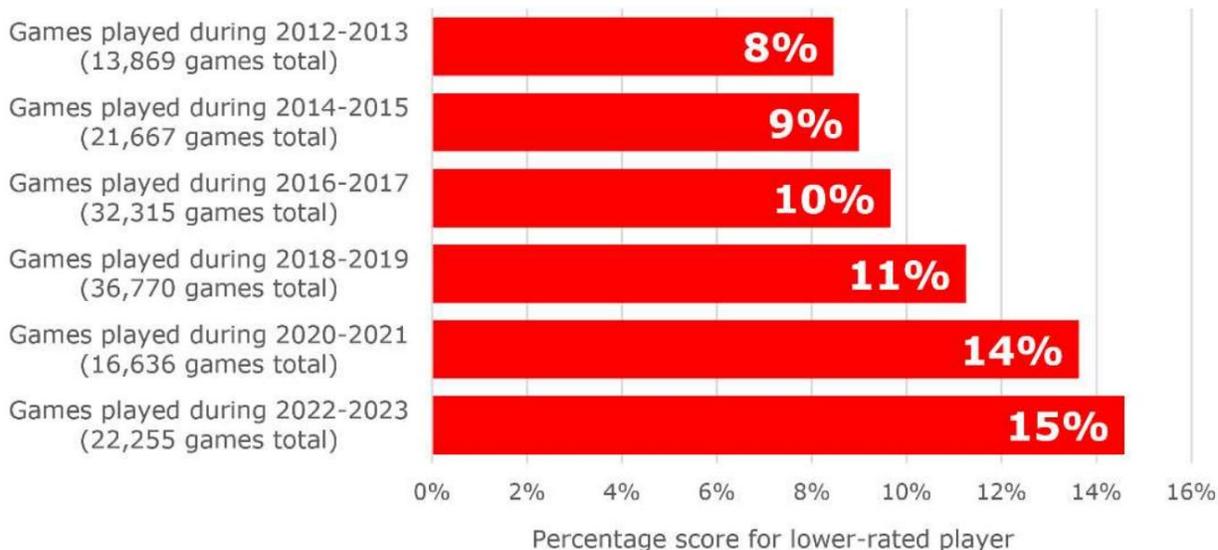
Podríamos preguntarnos si los cierres por el COVID-19 realmente habían ayudado a solucionar este problema, porque la afluencia de nuevos actores se redujo prácticamente a cero. Sin embargo, el sistema Elo todavía no tuvo tiempo de recuperar el aliento y encontrar un equilibrio adecuado, porque las calificaciones no pueden cambiar a menos que se jueguen juegos, y esencialmente no se jugaron juegos entre los jugadores calificados durante los cierres. Entonces el sistema Elo se detuvo durante esos momentos.

De hecho, las cosas empeoraron aún más para el sistema Elo de la FIDE con la aparición del COVID-19 y la drástica reducción de los eventos sobre el tablero. La falta de eventos puntuados no impidió que todos los jugadores jóvenes más débiles mejoraran sus habilidades ajedrecísticas, acercándose cada vez más a la fuerza de juego de los jugadores de nivel maestro. Pero el cierre restringió severamente la capacidad del sistema de clasificación Elo para seguir el ritmo del avance de todos esos jugadores junior más débiles. Debido a esto, parece bastante plausible que una calificación Elo de 1500 hoy, por ejemplo, refleje una mayor proximidad a una habilidad de nivel maestro de lo que habría representado una calificación Elo de 1500 hace cinco o diez años.

Veamos si podemos investigar este último punto objetivamente, en lugar de involucrarnos en especulaciones teóricas. Por ejemplo, ¿qué sucede cuando jugadores con una puntuación de 1.500 se enfrentan a jugadores con una puntuación de 2.100 en el mundo del ajedrez actual? ¿Y es diferente de lo que ocurrió hace cinco o diez años? Veamos la fuerte evidencia empírica de más de 140.000 juegos clasificados jugados desde 2012, entre jugadores con un Elo cercano a 1500 y oponentes con un Elo cercano a 2100. De hecho, la tendencia es bastante clara:

Percentage scores for players rated ~1500 Elo in games vs. opponents rated ~2100 Elo, since 2012

NOTE: includes all games where one player was rated 1400-1599 and other player was rated 2000-2199



Podemos decir con confianza que la brecha en la fuerza de juego se ha reducido enormemente durante la última década entre los jugadores de estos dos grupos (~1500 Elo y ~2100 Elo) en cualquier lista de clasificación determinada.

Se supone que esto no debe suceder en absoluto en el sistema Elo: se supone que una diferencia de X puntos Elo siempre corresponde (en promedio) al mismo puntaje porcentual Y. Sin embargo, es fácil ver que estos jugadores están mucho más cerca en fuerza de juego hoy que hace cinco o diez años.

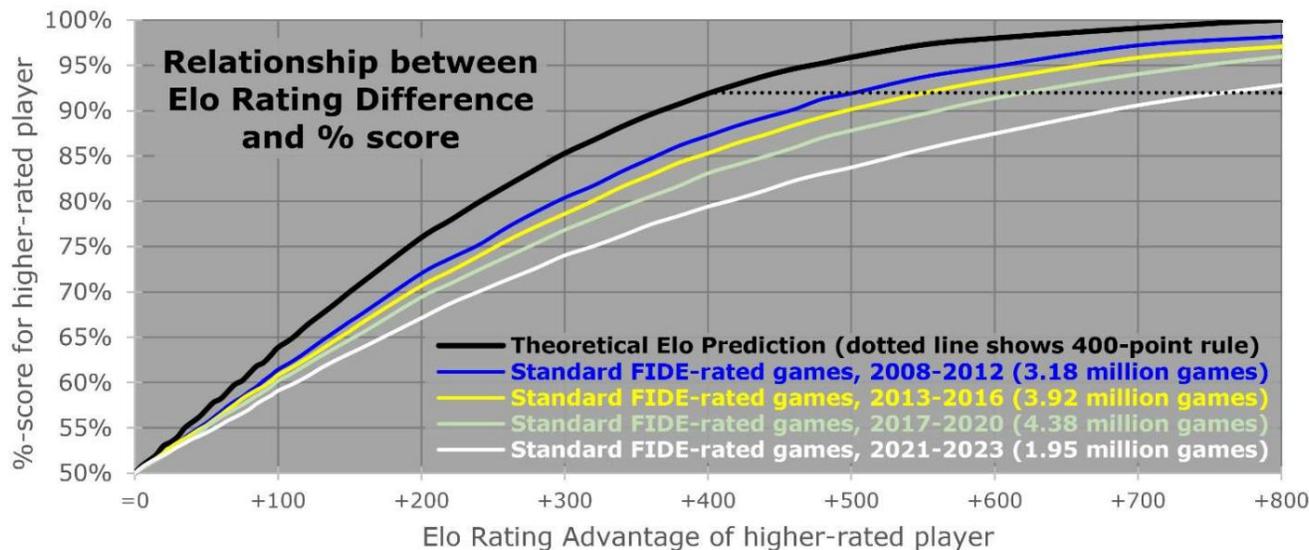
Hace diez años, los jugadores con menor rating anotaban un 8% en esos juegos. Más recientemente, en cambio, han obtenido una puntuación del 15%. Las actuaciones de hoy son lo que se esperaría de jugadores cuya fuerza está separada por 300 puntos Elo, en lugar de los 600 puntos Elo sugeridos por la lista de clasificación. Y, por cierto, la tabla Elo nos diría que los jugadores con menor puntuación deberían obtener alrededor del 2% en estos juegos.

Los espacios entre calificaciones simplemente no tienen el mismo significado que antes. Estos dos grupos de jugadores, que muestran una fuerte evidencia de estar separados por 300 puntos Elo en fuerza de juego real, en cambio están separados por 600 puntos Elo en la lista de clasificación. Esta es la razón por la que el término "expansión de la calificación" podría ser en realidad una descripción más adecuada que "deflación de la calificación". De hecho, todo el conjunto de calificaciones (por debajo del nivel maestro) se extiende cada año, en dirección descendente. Y por lo tanto, una brecha de calificación de 600 puntos Elo en la lista de calificación actual no corresponde a la misma diferencia en la fuerza de juego que existía hace apenas unos años, porque las calificaciones ahora están más separadas de lo que deberían.

Y, por supuesto, una diferencia de calificación de 600 puntos Elo no tiene nada de especial. Este problema está presente en todas las brechas de calificación numérica, pequeñas o grandes. Veamos un rango mucho más amplio, desde una brecha de 0 puntos de calificación Elo donde cada jugador tiene una puntuación esperada del 50%, hasta una brecha de 800 puntos Elo donde el jugador con mayor calificación esencialmente debería obtener una puntuación del 100% según a las mesas Elo. Ese rango cubre más del 99,5% de todos los juegos clasificados.

El siguiente gráfico muestra la relación entre la diferencia de calificación Elo y el puntaje porcentual, en todos los juegos clasificados por la FIDE (en controles de tiempo estándar) desde 2008, donde los jugadores estaban dentro de los 800 puntos Elo entre sí. El eje horizontal indica la ventaja de calificación Elo para

el favorito de clasificación, y el eje vertical indica el puntaje porcentual de ese jugador. La curva negra muestra el puntaje porcentual esperado del favorito, basado en la búsqueda de su diferencia de clasificación en la tabla Elo. Entonces puedes ver en la curva negra que el jugador mejor calificado tiene una puntuación esperada justo por encima del 50% si su ventaja de clasificación es pequeña, hasta una puntuación esperada del 85% con una ventaja de +300 puntos de clasificación, y hasta arriba, esencialmente al 100% con ventajas de clasificación aún mayores.



Otros colores muestran los resultados empíricos, es decir, las puntuaciones medias de los favoritos en juegos puntuados reales, cuando realmente tenían esas ventajas en el rating. Esto muestra todos los juegos clasificados por la FIDE estándar con estas diferencias de clasificación con respecto a los juegos jugados en 2008-2012 (azul), los juegos jugados en 2013-2016 (amarillo), los juegos jugados en 2017-2020 (verde) y, finalmente, los juegos jugados en 2021-2023 (blanco). Esta imagen resume más de 13 millones de juegos estándar clasificados jugados en los últimos 15 años. Y desafortunadamente, pinta un panorama bastante sombrío.

Si el sistema de clasificación hubiera funcionado bien durante todos estos años, entonces las curvas de colores se superpondrían o intersectarían perfectamente con la curva negra. Sin embargo, todas las curvas de colores son menos profundas que la curva negra, lo que indica que los favoritos en clasificación tienen consistentemente un desempeño inferior a sus calificaciones, cualquiera que sea el tamaño de su ventaja en clasificación.

Incluso hace diez o quince años, la situación no era muy buena. La curva azul nos dice que los favoritos en rating tuvieron consistentemente un desempeño inferior en los juegos de 2008-2012. Pero luego las cosas empeoraron aún más, como vemos con la curva amarilla (2013-2016) más baja que la azul, la curva verde (2017-2020) más baja que la amarilla y, finalmente, la curva blanca (2021-2023) siendo la más baja de todo. El ritmo de declive incluso parece estar acelerándose.

Nunca hemos estado más lejos de esa curva negra, esa relación ideal entre la diferencia de clasificación Elo y el puntaje porcentual, que con el sistema de clasificación actual y las calificaciones actuales. La curva blanca es la más superficial de todas, lo que significa que los favoritos de clasificación tienen un desempeño inferior al de antes. Y, por supuesto, esto no es sólo una rareza teórica. De hecho, los jugadores pierden puntos de rating en proporción directa a hasta qué punto su puntuación porcentual promedio cae por debajo de la curva negra. Entonces, en la práctica, esta tendencia significa que los favoritos en rating están perdiendo más y más puntos de rating por juego jugado, cada año.

En mi opinión, la situación está empeorando porque constantemente vemos que nuevos jugadores obtienen calificaciones iniciales demasiado bajas, y luego, cuando juegan partidas calificadas contra jugadores establecidos,

están reduciendo esas calificaciones establecidas al quitarles puntos de calificación a los jugadores establecidos. Y esto, en última instancia, está ampliando todo el conjunto de calificaciones, lo que lleva a calificaciones que exageran las verdaderas diferencias en la fuerza de juego entre los jugadores.

También debemos reconocer que este efecto no se produce sólo entre los jugadores más débiles o sólo entre los jugadores más fuertes. Más bien, está presente en todo el conjunto de calificaciones. Podemos volver a verificar esto construyendo una "tabla resumida cruzada" que abarque varios años completos de datos.

Una tabla cruzada tradicional representa un evento con jugadores individuales mostrados en cada fila y cada columna. Pero esta "tabla resumida cruzada" mostraría el desempeño de grupos de clasificación completos (agrupados en intervalos de 100 puntos Elo) frente a otros grupos de clasificación completos, a lo largo de muchos torneos o incluso muchos años.

Por ejemplo, hagamos una tabla resumida de los juegos jugados entre 2008 y 2012, con los jugadores asignados en grupos dependiendo de los dos primeros dígitos de sus calificaciones Elo. Entonces tendríamos jugadores con una calificación de 1200-1299 en un grupo, jugadores con una calificación de 1300-1399 en otro grupo, y así sucesivamente. Entre 2008 y 2012, los jugadores con una calificación de Elo de 1200-1299 jugaron un total de 726 juegos contra oponentes con una calificación de Elo de 1300-1399, y se esperaba que los jugadores con una calificación más baja anotaran un 37% en esos juegos y en realidad anotaron un 38%, por lo que excedieron su puntuación esperada. puntuación porcentual en +1%. Y a la inversa, los 1300-1399 jugadores anotaron un 1% por debajo de lo esperado en los mismos ju

Podemos mostrar esto en una tabla cruzada resumida, donde codificamos por colores los fondos de las celdas con un degradado azul-rojo, de modo que el rendimiento excesivo se muestra en azul como números positivos, el rendimiento inferior se muestra en rojo como números negativos y las desviaciones mayores del porcentaje esperado. La puntuación se muestra en azules y rojos más oscuros. Una diferencia del 1% con respecto a lo esperado es bastante pequeña, por lo que los fondos rojo y azul son muy pálidos para el primer ejemplo. Las diferencias del 2% o 3% con respecto a las expectativas son progresivamente más oscuras. Idealmente, cada celda sería "0%" y un fondo blanco, si todos tuvieran un desempeño de acuerdo con sus expectativas.

opponent player	vs. 1200-99	vs. 1300-99	vs. 1400-99
1200-99	X	+1%	+3%
1300-99	-1%	X	+2%
1400-99	-3%	-2%	X

Podemos encontrar el grupo 1200-99 en la columna más a la izquierda y leer para ver su ligero +1% de sobrerrendimiento (azul pálido) contra oponentes en el grupo 1300-99 y un mayor +3% de sobrerrendimiento (azul claro medio) contra oponentes en el grupo 1400-99. En ambos casos, los jugadores con una puntuación de 1200 a 99 tuvieron un rendimiento superior al de sus oponentes con una puntuación más alta.

De manera similar, podemos encontrar el grupo 1400-99 en la columna más a la izquierda y leer para ver que tuvieron un rendimiento inferior al -3 % (rojo claro medio) frente a oponentes en el grupo 1200-99 y un rendimiento inferior al -2 % (rojo claro).) contra oponentes del grupo 1300-99.

La diagonal está llena de X, ya que no mostramos lo que sucede cuando un grupo juega contra sí mismo (los números siempre serían 50% esperados y 50% reales). Y tenga en cuenta que la región "inferior izquierda" (las celdas debajo y a la izquierda de la diagonal) representa el desempeño de los favoritos durante los juegos de 2008 a 2012, y la región "superior derecha" (las celdas arriba y

a la derecha de la diagonal) representan el desempeño de los perdedores durante esos mismos juegos. Basándonos en nuestro entendimiento anterior de que incluso en 2008-2012, los favoritos de calificación estaban por debajo de sus puntajes previstos, podemos esperar ver principalmente celdas rojas en la parte inferior izquierda y en su mayoría celdas azules en la parte superior derecha, y al menos esta pequeña El ejemplo sigue esa tendencia.

Por supuesto, la imagen se vuelve mucho más interesante e informativa si mostramos, no solo los grupos con calificaciones inferiores a 1500 y cómo les fue a todos entre sí, sino también a todos los diferentes grupos versus todos los demás grupos. Agruparemos a todos los jugadores con 2500+ Elo en su propio grupo para poder mostrarles cantidades significativas de datos, en lugar de intentar mostrar datos separados para 2500-99, 2600-99, etc. Aquí está la imagen completa para juegos jugados durante el lapso de cinco años desde enero de 2008 hasta diciembre de 2012:

2008-2012	opponent	vs.	vs.	vs.												
	player	1200-99	1300-99	1400-99	1500-99	1600-99	1700-99	1800-99	1900-99	2000-99	2100-99	2200-99	2300-99	2400-99	2500+	
+4.5% in 6,865 games	1200-99		+1%	+3%	+6%	+6%	+7%	+4%	+4%	+2%	+2%	=0%	=0%	=0%	=0%	
+4.8% in 25,544 games	1300-99	-1%		+2%	+4%	+7%	+7%	+4%	+5%	+4%	+4%	+1%	=0%	+2%	=0%	
+5.0% in 60,420 games	1400-99	-3%	-2%		+3%	+5%	+7%	+8%	+7%	+4%	+3%	+1%	+1%	+1%	=0%	
+4.9% in 135,409 games	1500-99	-6%	-4%	-3%		+3%	+6%	+7%	+8%	+6%	+5%	+3%	+2%	+1%	=0%	
+4.0% in 241,595 games	1600-99	-6%	-7%	-5%	-3%		+4%	+6%	+8%	+7%	+6%	+4%	+2%	+2%	+0%	
+2.7% in 385,143 games	1700-99	-7%	-7%	-7%	-6%	-4%		+3%	+6%	+7%	+6%	+5%	+3%	+1%	+1%	
+1.4% in 540,108 games	1800-99	-4%	-4%	-8%	-7%	-6%	-3%		+3%	+5%	+6%	+4%	+4%	+2%	+1%	
+0.0% in 699,217 games	1900-99	-4%	-5%	-7%	-8%	-8%	-6%	-3%		+3%	+4%	+4%	+3%	+2%	+0%	
-1.0% in 801,405 games	2000-99	-2%	-4%	-4%	-6%	-7%	-7%	-5%	-3%		+2%	+3%	+3%	+2%	+1%	
-1.5% in 779,484 games	2100-99	-2%	-4%	-3%	-5%	-6%	-6%	-6%	-4%	-2%		+2%	+2%	+2%	+1%	
-1.3% in 644,501 games	2200-99	=0%	-1%	-1%	-3%	-4%	-5%	-4%	-4%	-3%	-2%		+2%	+2%	+1%	
-1.3% in 453,537 games	2300-99	=0%	=0%	-1%	-2%	-2%	-3%	-4%	-3%	-3%	-2%	-2%		+2%	+1%	
-1.1% in 315,497 games	2400-99	=0%	-2%	-1%	-1%	-2%	-1%	-2%	-2%	-2%	-2%	-2%	-2%		+1%	
-0.8% in 213,368 games	2500+	=0%	=0%	=0%	=0%	-0%	-1%	-1%	-0%	-1%	-1%	-1%	-1%	-1%		

Sorprendentemente, incluso hace 12 o 15 años no había ningún color azul en toda la región inferior izquierda. Las calificaciones se estiraron demasiado incluso en 2008-2012, ya que los favoritos de calificación tuvieron un desempeño inferior en todos estos enfrentamientos que tenían una cantidad razonable de datos, e incluso en los que no tienen muchos datos, solo se mantienen firmes en "=0 %" en las celdas blancas.

Bien, eso fue 2008-2012. Hace mucho tiempo. Avancemos más de una década y generemos nuevamente este mismo tipo de tabla cruzada resumida, pero con datos más recientes, es decir, los juegos de 2021-2023. Entonces, antes que nada, debemos incluir dos grupos de calificación adicionales (jugadores con calificación 1000-99 y jugadores con calificación 1100-99), ya que las calificaciones Elo son más bajas en estos días que antes de 2012. Entonces, hay dos filas más en la parte superior, y dos columnas más a la izquierda. Nos centraremos en comprobar si todavía está todo rojo en la parte inferior izquierda y todo azul en la parte superior derecha. ¿Qué vemos?

2021-2023	opponent	vs.															
	player	1000-99	1100-99	1200-99	1300-99	1400-99	1500-99	1600-99	1700-99	1800-99	1900-99	2000-99	2100-99	2200-99	2300-99	2400-99	2500+
+7.4% in 47,760 games	1000-99		+2%	+7%	+9%	+10%	+10%	+10%	+9%	+8%	+5%	+5%	+4%	+2%	=0%	=0%	
+7.8% in 150,873 games	1100-99	-2%		+4%	+7%	+9%	+12%	+12%	+10%	+9%	+7%	+6%	+5%	+4%	+1%	+1%	=0%
+6.8% in 205,982 games	1200-99	-7%	-4%		+5%	+9%	+12%	+14%	+13%	+12%	+9%	+8%	+6%	+4%	+2%	+1%	=0%
+5.3% in 243,185 games	1300-99	-9%	-7%	-5%		+5%	+11%	+14%	+15%	+14%	+12%	+10%	+9%	+7%	+4%	+3%	+3%
+3.7% in 270,932 games	1400-99	-10%	-9%	-9%	-5%		+6%	+11%	+14%	+15%	+14%	+12%	+9%	+7%	+5%	+2%	+3%
+1.6% in 291,021 games	1500-99	-10%	-12%	-12%	-11%	-6%		+6%	+11%	+14%	+14%	+13%	+10%	+9%	+6%	+4%	+0%
+0.1% in 303,892 games	1600-99	-10%	-12%	-14%	-14%	-11%	-6%		+6%	+10%	+13%	+14%	+13%	+10%	+7%	+5%	+3%
-1.3% in 318,533 games	1700-99	-9%	-10%	-13%	-15%	-14%	-11%	-6%		+5%	+10%	+12%	+11%	+11%	+9%	+6%	+3%
-2.3% in 322,168 games	1800-99	-8%	-9%	-12%	-14%	-15%	-14%	-10%	-5%		+5%	+9%	+10%	+10%	+10%	+9%	+3%
-3.2% in 309,028 games	1900-99	-5%	-7%	-9%	-12%	-14%	-14%	-13%	-10%	-5%		+5%	+8%	+9%	+9%	+7%	+4%
-3.8% in 275,731 games	2000-99	-5%	-6%	-8%	-10%	-12%	-13%	-14%	-12%	-9%	-5%		+4%	+7%	+7%	+7%	+4%
-3.8% in 229,272 games	2100-99	-4%	-5%	-6%	-9%	-9%	-10%	-13%	-11%	-10%	-8%	-4%		+4%	+6%	+6%	+5%
-3.9% in 174,867 games	2200-99	-2%	-4%	-4%	-7%	-7%	-9%	-10%	-11%	-10%	-9%	-7%	-4%		+4%	+4%	+3%
-3.7% in 128,497 games	2300-99	=0%	-1%	-2%	-4%	-5%	-6%	-7%	-9%	-10%	-9%	-7%	-6%	-4%		+3%	+2%
-3.2% in 97,087 games	2400-99	=0%	-1%	-1%	-3%	-2%	-4%	-5%	-6%	-9%	-7%	-7%	-6%	-4%	-3%		+2%
-2.5% in 62,886 games	2500+	=0%	=0%	=0%	-3%	-3%	-0%	-3%	-3%	-3%	-4%	-4%	-5%	-3%	-2%	-2%	

Mmmm, bueno, no sólo el efecto sigue presente en todas partes, sino que, de hecho, el sombreado rojo/azul es mucho más oscuro, ya que muchos de los números llegan a dos dígitos, lo que indica que en muchos casos los favoritos de calificación están por debajo de sus expectativas en 10. % o incluso 15% por juego! Las cosas han empeorado notablemente. Y, de hecho, podríamos haber predicho esto si hubiéramos recordado cuán menos profunda era esa curva blanca en el primer gráfico (para los juegos de 2021-2023), incluso en comparación con cómo había sido la situación solo unos años antes.

La deflación se ha extendido hasta el nivel de maestro e incluso ha comenzado a alcanzar el nivel de gran maestro. La mayoría de los jugadores de élite se enfrentan entre sí, por lo que este efecto deflacionario, que durante años se ha propagado hacia arriba a través del grupo de calificaciones debido a que las nuevas calificaciones eran demasiado bajas, aún no ha afectado completamente a los jugadores de élite. Pero ya viene. Incluso ahora se puede ver un rendimiento inferior al rojo de los grupos 2400-99 y 2500+ contra oponentes con una clasificación Elo cercana a 2000, donde no se veía esto hace 10-15 años. Incluso cerca de la cima, las diferencias en las calificaciones no significan lo que alguna vez significaron.

Sin embargo, dejemos de lado a los grandes maestros y centrémonos únicamente en el 99% restante de los jugadores con calificación FIDE, los que tienen calificaciones Elo en el rango desde 2400 Elo hasta 1000 Elo. Las calificaciones de estos jugadores abarcan 1400 puntos Elo de arriba a abajo. Pero si analizamos cuidadosamente la evidencia de millones de partidas de ajedrez recientes, comparando el desempeño de varios grupos de clasificación con otros grupos de clasificación mediante visualizaciones de datos como estas tablas cruzadas resumidas y otros tipos de análisis, llegamos a una conclusión ineludible.

¡Las diferencias reales en la fuerza de juego en ese grupo, ese 99% de todos los jugadores calificados por la FIDE, solo abarcan 1000 puntos Elo! No 1.400 puntos Elo. Lo que plantea algunas cuestiones muy serias. ¿Los jugadores con la calificación más baja, los que tienen poco más de 1000 Elo? ¿Quizás realmente deberían tener calificaciones más parecidas a 1400 Elo! ¿Y los jugadores obtuvieron alrededor de 1500 Elo? ¿Quizás realmente deberían tener calificaciones más parecidas a 1700 Elo! Eso es lo que sugerirían las cifras de rendimiento.

Los jugadores con una calificación de 2400 no se desempeñan 100 puntos mejor que los jugadores con una calificación de 2300, quienes a su vez no se desempeñan 100 puntos mejor que los jugadores con una calificación de 2200, y así sucesivamente hasta llegar a los jugadores con una calificación de 1000. jugadores. Ninguno de los favoritos en la calificación es capaz de alcanzar el puntaje porcentual que sugerirían sus ventajas Elo publicadas. Es por eso que la región inferior izquierda está tan roja en la última imagen.

El grupo de clasificación Elo estándar de la FIDE está demasiado extendido -demasiado- y hay muy pocos incentivos para que cualquier jugador se enfrente voluntariamente a oponentes de menor puntuación, ya que el resultado más probable para el jugador con mayor puntuación siempre sería una pérdida significativa de puntos de calificación. Y está empeorando cada año que pasa. Honestamente, no veo otra opción factible que comprimir manualmente la parte inferior de la lista de calificaciones para que las calificaciones no estén tan separadas, junto con otras medidas para evitar que todo esto vuelva a suceder.

Parte III – La solución: mejoras en la compresión y el cálculo

En los últimos meses he estado investigando este problema para la FIDE. Desarrollé una herramienta de base de datos que podría replicar los últimos 15 años de cálculos mensuales de clasificación de la FIDE, desde 2008 hasta la actualidad, basándose en los muchos millones de resultados de juegos que en realidad se utilizaron para las listas de clasificación oficiales de la FIDE. Esta herramienta soporta los distintos cambios en la regulación de calificación que se produjeron a lo largo de esos 15 años, como las calificaciones mínimas progresivamente más bajas en 2009 y 2012, los cambios en los factores K en 2011 y 2014, los cambios en la fórmula de calificación inicial en 2014, etc. .

Durante el desarrollo de esta herramienta, verifiqué que podía replicar el cálculo de las listas de calificación históricas reales con bastante precisión. Pero lo más importante es que la herramienta me permite experimentar cómo se habría comportado el sistema de calificación a lo largo de todos esos años si hubieran estado vigentes regulaciones diferentes en ese momento. Obviamente, cualquier cambio real que estemos contemplando se implementaría en el futuro, no en el pasado, pero al menos podemos simular lo que habría sucedido con las calificaciones, basándonos en datos reales del juego, si hubiéramos abordado esta situación hace varios años. Esto nos permite observar algunos de los posibles efectos a largo plazo de cualquier cambio que estemos contemplando, a medida que calculamos calificaciones simuladas cada mes en función de los juegos que realmente sucedieron y, en última instancia, vemos cómo las calificaciones se verían diferentes hoy.

He simulado muchos escenarios diferentes, en paralelo con varias discusiones con miembros de la Comisión de Clasificación de la FIDE, y he elegido la solución recomendada para abordar los problemas que acabo de describir. He simulado escenarios en los que esta solución recomendada se implementó muy recientemente, cuando el grupo de calificaciones era bastante similar al que tenemos hoy a mediados de 2023. Esto me permitió tener una idea clara de los efectos esperados a corto plazo después de los cambios. Además, probé un escenario en el que esta solución se implementó en enero de 2017, lo que me permitió observar tanto los efectos a corto plazo como los efectos a largo plazo durante los próximos 6 años o más de cálculos de calificación simulados. Esto nos da la mejor oportunidad de estimar lo que realmente pasaría con el ranking de la FIDE en los próximos años, si estos cambios entraran en vigor pronto. Es cierto que la aparición de COVID-19 en 2020 plantea desafíos de simulación.

En cualquier caso, propongo que la FIDE emprenda reparaciones significativas en su sistema estándar de calificación Elo mediante la implementación de dos cambios importantes, a los que llamo mejoras en la compresión y el cálculo. Se describe que ocurrieron en enero/febrero de 2024, porque la Compresión debe ocurrir 1 mes antes de las Mejoras de Cálculo, pero también podrían ocurrir en octubre/noviembre de 2023 o marzo/abril de 2024. En cualquier caso, estas reparaciones deben realizarse lo antes posible. Aquí están mis recomendaciones, que creo que son el curso de acción más inteligente de la FIDE en este momento:

III (a) Compresión: En la lista de clasificación de enero de 2024, para todos los jugadores que tengan una clasificación por debajo de 2000 Elo, su clasificación debe recibir un aumento único por la cantidad de:

$$(0,40) \times (2000 - \text{Calificación})$$

El aumento deberá aplicarse después de que los demás cálculos ya se hayan realizado en la lista de calificación de ese mes. Debería suceder para todos los jugadores activos e inactivos con una clasificación de Elo inferior a 2000, y debería suceder independientemente de que hayan jugado algún juego durante ese período de calificación.

La compresión tomaría el grupo de jugadores con calificaciones entre 1000 Elo y 2000 Elo, que es el 85% inferior de los jugadores calificados, y comprimiría sus calificaciones más juntas para que solo abarquen 1400 Elo a 2000 Elo. Aumentaría las calificaciones de todos los jugadores sub-2000 entre 0 y 400 puntos Elo. Los jugadores con una calificación justo por encima de 1000 Elo obtendrían un aumento de casi 400 puntos Elo, los jugadores con una calificación cercana a 1500 Elo obtendrían un aumento de aproximadamente 200 puntos Elo y los jugadores con una calificación justo por debajo de 2000 Elo casi no obtendrían ningún aumento.

Y para todos los jugadores que tengan una clasificación de Elo de 2000 o superior, su clasificación no se verá afectada directamente por esto. Tenga en cuenta que este cambio preservaría el orden actual de todos en la lista de clasificación; solo ajustaría sus espacios entre sí. Y finalmente, nadie tendría una clasificación actual inferior a 1400 Elo.

A continuación se muestran algunos ejemplos representativos para ilustrar mejor cómo funcionaría la compresión:

Ejemplo 1A: Después de calcular las actualizaciones de calificación normales para enero de 2024, el jugador A tiene una calificación Elo de 1200. Debido a que esta calificación está por debajo de 2000 Elo, el jugador A recibe un aumento de calificación único de $(0,40) \times (2000 - 1200) = 320$ puntos Elo. Por lo tanto, el jugador A tendrá una calificación de 1520 Elo en la lista de calificación de enero de 2024.

Ejemplo 1B: El jugador B ya tiene una calificación de 1003 Elo y no tiene ningún juego calificado para enero de 2024.

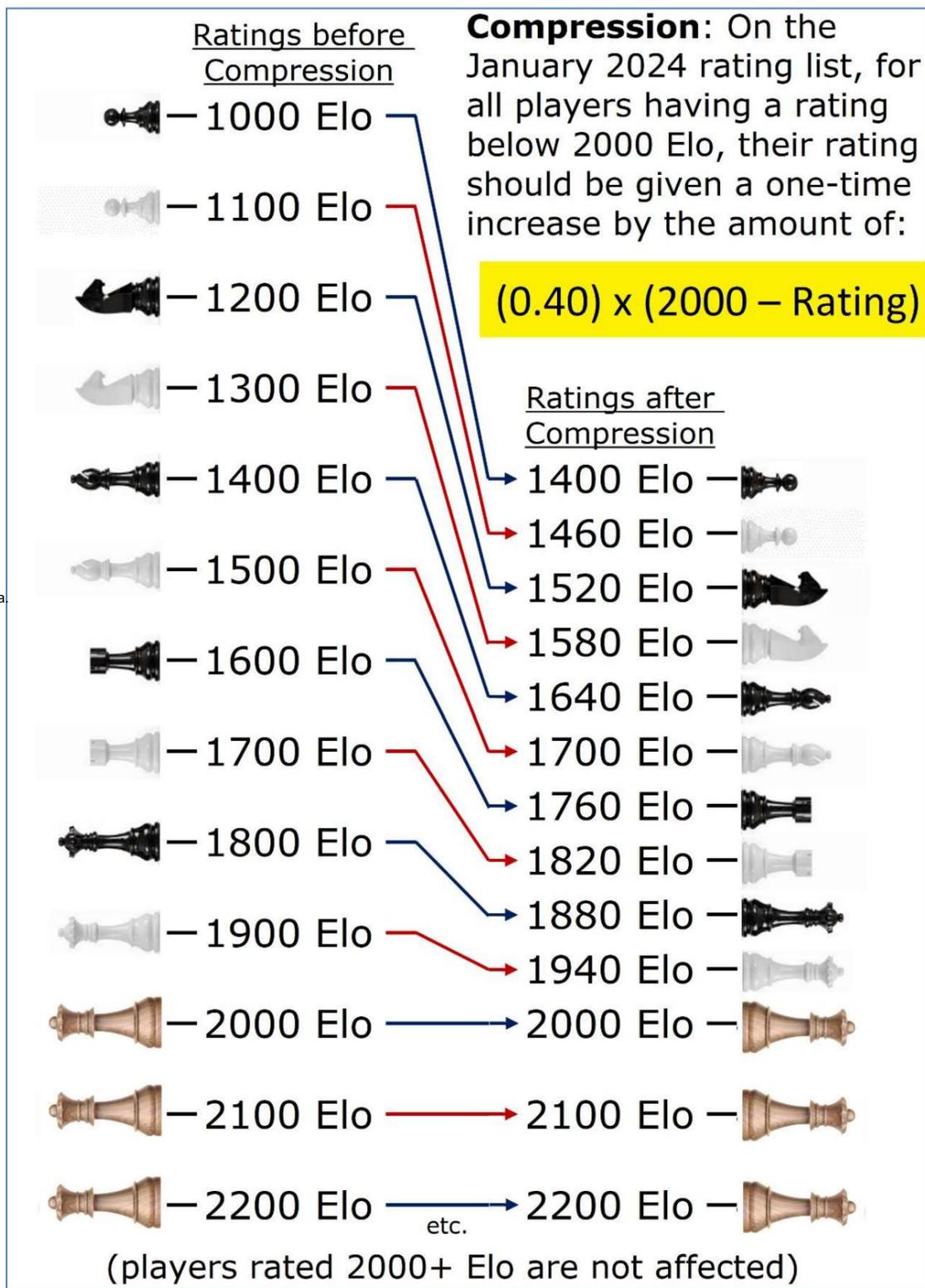
Debido a que esta calificación está por debajo de 2000 Elo, el jugador B recibe una bonificación única aumento de calificación de $(0,40) \times (2000 - 1003) = 398,8$ puntos Elo, redondeado al entero más cercano de 399 puntos Elo. Por lo tanto, el jugador B tendrá una calificación de 1402 Elo en la lista de calificación de enero de 2024.

Ejemplo 1C: El jugador C es un jugador sin clasificación que califica para una clasificación inicial de 1900 Elo para enero de 2024.

Debido a que esta calificación está por debajo de 2000 Elo, y debido a que el aumento se aplica después de los cálculos regulares para la lista de calificación, el jugador C recibe un aumento de calificación único de $(0,40) \times (2000 - 1900) = 40$ puntos Elo.

Por lo tanto, el jugador C tendrá una calificación inicial de 1940 Elo en la lista de calificación de enero de 2024.

Ejemplo 1D: El jugador D es un jugador sin clasificación que aún no califica para una clasificación inicial para la lista de clasificación de enero de 2024. Por lo tanto, el jugador D continúa sin calificar y nunca recibirá un aumento de calificación único, aunque eventualmente puede recibir una calificación inicial por medios normales, en algún momento en el futuro.



Ejemplo 1E: Después de calcular las actualizaciones de calificación normales para enero de 2024, el jugador E tiene una calificación Elo de 2350. Debido a que esta calificación no está por debajo de 2000 Elo, la calificación del jugador E no se ve afectada y el jugador E nunca recibirá un aumento de calificación único. Por lo tanto, el jugador E tendrá una calificación de 2350 Elo en la lista de calificación de enero de 2024.

III (b) Mejoras en el cálculo: a partir del cálculo de la lista de calificación de febrero de 2024, y para todas las listas de calificación futuras, el algoritmo de cálculo de calificación debe cambiarse de algunas maneras importantes:

(1) La calificación Elo mínima será 1400 Elo en lugar de 1000 Elo.

(2) La regla de los 400 puntos volverá a su estado anterior, sin restricciones sobre cuántas veces se puede aplicar a un jugador durante un solo torneo.

(3) Al sumar los resultados de las partidas de un jugador no clasificado contra oponentes clasificados, para efectos de calcular una clasificación inicial, al jugador no clasificado se le acreditarán dos empates adicionales contra oponentes hipotéticos con una clasificación de 1800 Elo, además de sus resultados reales del juego. Todos los requisitos existentes seguirían aplicándose, por ejemplo exigir que los jugadores tengan un total de cinco o más juegos reales contra oponentes clasificados (sin contar los hipotéticos).

(4) Cuando los jugadores no clasificados tengan una puntuación positiva en sus juegos contra oponentes clasificados, su calificación inicial se calculará de la misma manera que para las puntuaciones negativas, siendo una calificación de rendimiento basada en su puntuación porcentual, en lugar de usar la puntuación positiva multiplicada por $(K/2)$. Sin embargo, la calificación inicial máxima que se puede lograr con esta fórmula será 2200 Elo. Si un jugador obtiene una calificación inicial superior a 2200 Elo, recibirá una calificación inicial de 2200 Elo.

(5) Para los jugadores no clasificados que aún no tienen una calificación, todas las actuaciones parciales no calificadas en el pasado, logradas contra oponentes clasificados antes de la Compresión, serán ignoradas y descartadas.

Entonces, todos los jugadores sin calificación tendrán un "nuevo comienzo" en este punto para obtener una nueva calificación.

Aquí hay algunos ejemplos representativos para ilustrar mejor cómo el Cálculo

Las mejoras funcionarían:

Ejemplo 2A: El jugador A es un jugador sin clasificación que obtiene 1 de 4 en su primer evento contra oponentes clasificados, lo cual se informa en la lista de clasificación de febrero de 2024. El jugador A aún no califica para una calificación inicial porque aún no ha acumulado cinco resultados contra oponentes clasificados. Los dos empates hipotéticos descritos en (3) arriba no los empujan por encima del mínimo de cinco juegos. Por lo tanto, el jugador A permanecerá sin clasificar en la lista de clasificación de febrero de 2024.

Ejemplo 2B: El jugador B es un jugador sin clasificación que obtiene 1 de 6 en su primer evento, contra oponentes clasificados que tienen una calificación promedio de 1600 Elo, y este evento se informa para la lista de clasificación de febrero de 2024. Según las regulaciones anteriores, una puntuación del 17% contra una oposición de nivel 1600 habría dado una calificación inicial de 1327 Elo. En lugar de ello, se añaden dos empates hipotéticos contra oponentes de nivel 1800 a los resultados totales del jugador B, a los efectos de este cálculo. Entonces, en cambio, el jugador B es tratado como si hubiera anotado 2 de 8 (25%) contra una oposición de nivel 1650.

Por lo tanto, el jugador B recibe una calificación inicial de 1457 Elo en la lista de clasificación de febrero de 2024.

Ejemplo 2C: El jugador C es un jugador sin clasificación que obtuvo 1 de 3 en su primer evento contra oponentes clasificados que tenían una calificación promedio de 1500 Elo, y este evento se informó para la lista de clasificación de julio de 2023 (antes de la Compresión).

Luego, en enero de 2024 (después de la Compresión), obtienen 0 de 3 en su segundo evento, contra oponentes clasificados que tienen una calificación promedio de 1700, y este evento se informa para la lista de clasificación de febrero de 2024. Los resultados de julio de 2023 se descartan porque ocurrieron antes de la Compresión, por lo que los resultados de enero de 2024 se consideran su primer resultado como jugadores sin clasificación. Dado que obtuvieron una puntuación de cero en ese evento, los resultados se descartan. El jugador C permanecerá sin clasificación en la lista de clasificación de febrero de 2024 y aún no tendrá ningún resultado que cuente para una clasificación inicial.

Ejemplo 2D: El jugador D es un jugador no clasificado sin resultados previos y que tiene dos eventos reportados para la lista de clasificación de febrero de 2024 (después de la Compresión), con una puntuación de 1 de 3 en su primer evento (contra oponentes clasificados que tienen una puntuación promedio de 1500). Elo, y con una puntuación de 0 de 3 en su segundo evento (contra oponentes clasificados con una puntuación media de 1700). Al igual que con el jugador B, a los resultados totales del jugador D se añaden dos empates hipotéticos contra oponentes de nivel 1800, a los efectos de este cálculo. Por lo tanto, se trata al jugador D como si hubiera anotado 2 de 8 (25%) contra oponentes de nivel 1650. El jugador D recibe una calificación inicial de 1457 Elo en la lista de clasificación de febrero de 2024.

Ejemplo 2E: El jugador E tiene una calificación de 1430 en la lista de calificación de enero de 2024 y luego obtiene 0,5 puntos en 7 juegos calificados informados para la lista de calificación de febrero de 2024, donde se esperaba que obtuviera 2,5 puntos. De un rendimiento inferior por 2,0 puntos con $K=40$, el jugador E debería perder 80 puntos Elo. Debido a que esto los colocaría por debajo de la calificación mínima de 1400, el jugador E queda sin calificación para la lista de febrero de 2024 (o se informa que tiene una calificación de cero y luego queda sin calificación).

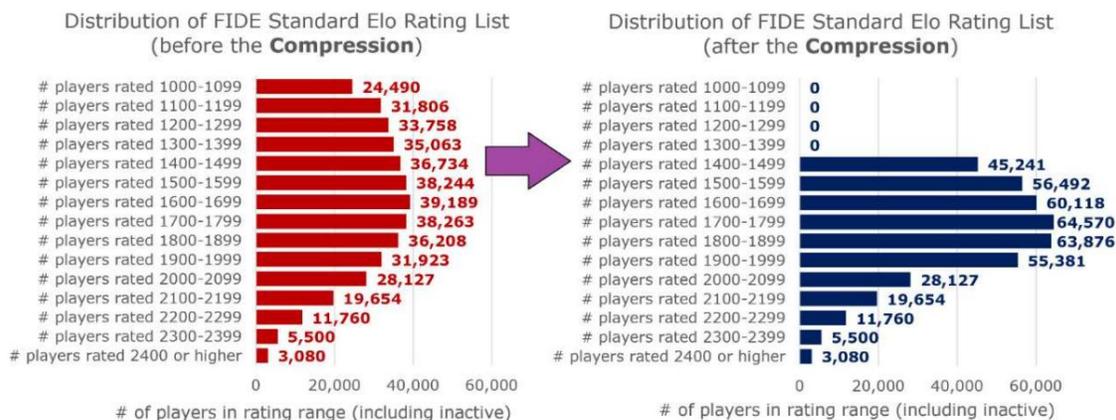
Ejemplo 2F: El jugador F es un jugador sin clasificación que obtiene 5 de 6 en su primer evento, contra oponentes clasificados que tienen una calificación promedio de 1600 Elo, y este evento se informa para la lista de calificación de febrero de 2024. Según las regulaciones anteriores, una puntuación de +4 contra una oposición de nivel 1600 habría dado una calificación inicial de 1680 Elo. En lugar de ello, se añaden dos empates hipotéticos contra rivales de nivel 1800 a los resultados totales del jugador F, a los efectos de este cálculo, y se calcula una calificación de rendimiento. Entonces, en cambio, el jugador F es tratado como si hubiera anotado 6 de 8 (75%) contra una oposición de nivel 1650. Por lo tanto, el jugador F recibe una calificación inicial de 1843 Elo en la lista de calificación de febrero de 2024.

Parte IV – Justificaciones y resultados esperados

Intentemos ahora proyectar y predecir el resultado anticipado si estas correcciones se implementan pronto, además de brindar más fundamentos para ellas.

Obviamente, habría un cambio muy grande de inmediato en las calificaciones de los jugadores, ya que aproximadamente el 85 % de todos los jugadores calificados obtendrían un aumento de calificación de algún tipo gracias a la Compresión, y más de la mitad de esos aumentos serían más de 150 puntos Elo.

Como ejemplo, si la compresión se hubiera aplicado a la lista de clasificación de la FIDE en abril de 2023, aquí hay una comparación antes/después, que muestra cuántos jugadores (incluidos jugadores activos e inactivos) habrían estado en cada rango de 100 puntos Elo, con el rojo mostrando los números antes de la compresión y el azul oscuro mostrando los números después de la compresión:



Algunas características clave de la compresión se pueden ver fácilmente en este gráfico. En primer lugar, de repente no habría jugadores con una calificación inferior a 1400 Elo. Todos los jugadores que anteriormente tenían una calificación de Elo de 1000-1400 recibirían ganancias de calificación suficientes para superar los 1400 Elo. Además, una de las mejoras en los cálculos sería restablecer inmediatamente la calificación mínima a 1400 Elo, por lo que no seguiría habiendo jugadores con una calificación inferior a 1400 Elo en el futuro.

Aunque muchos jugadores recibirían ganancias de calificación, estas ganancias serían menores a medida que miras más alto en el grupo de calificación y, de hecho, la compresión no elevaría la calificación de nadie por encima de 2000 Elo. Habría los mismos 346.000 jugadores clasificados por debajo de 2000 Elo tanto antes como después; es sólo que ahora abarcarían un rango de 600 puntos Elo en lugar de 1000 puntos Elo.

Y finalmente, ninguno de los jugadores con calificación Elo superior a 2000 se vería directamente afectado por la compresión. Habría los mismos 68.000 jugadores que antes, con las mismas valoraciones que antes.

A primera vista, se podría pensar que aumentar la calificación mínima a 1400 Elo excluiría a una gran cantidad de jugadores en el futuro de obtener calificaciones. Después de todo, actualmente hay 125.000 jugadores con calificaciones inferiores a 1400 Elo. ¿Qué pasaría con los jugadores que aún no han alcanzado una clasificación FIDE, pero que son aproximadamente tan buenos como los jugadores con una puntuación de 1250 Elo en la lista de clasificación actual? ¿Quizás deberíamos ser más indulgentes y tener una calificación mínima de 1000 Elo o quizás 1200 Elo?

Es importante recordar que todos los jugadores existentes con calificaciones bajas obtendrían aumentos significativos en sus calificaciones. Los jugadores que actualmente tienen una calificación de 1250 Elo terminarían alrededor de 1550 Elo como resultado de la compresión. Y entonces, si hay un jugador sin calificación de fuerza comparable, enfrentándose a oponentes con una calificación de alrededor de 1550 Elo, entonces el índice de desempeño del jugador sin calificación probablemente también sería de aproximadamente 1550 Elo. No deberían tener problemas para obtener una calificación inicial. Para ser honesto, incluso podría resultar un poco más fácil para dichos jugadores obtener una calificación inicial, ya que otra de las mejoras de cálculo aumentaría los resultados para los jugadores sin calificación en dos empates contra oponentes hipotéticos con una calificación de 1800.

De hecho, creo que sería imprudente hacer algo más que aumentar el rating mínimo hasta 1400 Elo. Como ejemplo adicional, consideremos a los jugadores que actualmente tienen una calificación de Elo de 1000-1050.

Estos son los jugadores peor valorados de la actualidad. Después de la Compresión, estos jugadores habitarían el rango de 1400-1430 Elo. Si luego permitiéramos que nuevos jugadores ingresen a la lista de clasificación con calificaciones justo por encima de 1200 Elo, entonces esos nuevos jugadores serían aproximadamente 200 puntos Elo más débiles que los jugadores con la calificación más baja de la actualidad. Por lo tanto, optar por una calificación mínima de 1200 Elo permitiría la entrada a jugadores considerablemente más débiles que aquellos que se encuentran al final de la lista de clasificación actual. Por el contrario, usar una calificación mínima de 1400 Elo (después de comprimir las calificaciones) estaría más en línea con los estándares actuales. Los únicos jugadores a los que se les impediría obtener una calificación mínima de 1400 serían los jugadores que probablemente tampoco alcanzarían una calificación de 1000 en el sistema actual.

El peligro de optar por una calificación mínima inferior a 1400 Elo, como 1200 Elo, es que los jugadores débiles o en mejora obtendrían calificaciones iniciales que probablemente sean demasiado bajas incluso para su habilidad actual, y comenzarían a obtener grandes cantidades de puntos de calificación. Lejos del grupo establecido mientras sus calificaciones se recuperan, lo que sería otro fuerte efecto deflacionario. Eso es básicamente lo que ha sucedido durante la última década con la calificación mínima baja de 1000 Elo. Sería una gran lástima tomar todas estas medidas para revertir la deflación de la última década, sólo para alentar su regreso tomando una decisión que fomente más deflación.

También se podría pensar, a primera vista, que aumentar las calificaciones de los jugadores más débiles en 400 puntos Elo es un salto demasiado grande, y que tal vez deberíamos conformarnos con algo así como una "compresión del 20%" en lugar de una "compresión del 40%". Esto daría como resultado que solo se agregaran 200 puntos Elo a los jugadores con la calificación más baja, y aumentos proporcionalmente menores para los jugadores con una calificación más alta, hasta 2000 Elo. Esta es una reacción razonable, pero nuevamente advertiría que no se debe hacer nada que no sea una compresión del 40%.

La compresión propuesta se basa en un análisis detallado y extenso que sugiere el verdadero nivel de desempeño de los distintos grupos de calificación. Para tomar como ejemplo a los jugadores con una calificación de Elo de 1000-1100, actualmente hay alrededor de 25,000 de ellos, lo que representa el 6% del grupo de calificación, y la evidencia de su desempeño reciente indica que en realidad son solo 900-1000 puntos más débiles que el típico. Jugador con calificación de 2400, a pesar de tener una calificación de 1300-1400 puntos Elo más baja. Esta conclusión no se extrae sólo de los resultados directos cuando jugadores con una Elo de 1000-1100 se enfrentan a oponentes con una puntuación de 2400, ya que estos enfrentamientos son bastante raros y no son muy informativos cuando ocurren. Más bien, la conclusión se toma del análisis simultáneo de todos los grupos.

actuaciones contra todos los demás grupos, especialmente aquellos clasificados cerca a quienes se enfrentan a menudo. La compresión colocaría a los jugadores con una calificación Elo actual de 1000-1100 en un rango Elo mucho más apropiado de 1400-1460, en línea con su fuerza evidente.

El peligro de optar por una compresión inferior al 40%, como por ejemplo el 20%, es que grandes bloques de jugadores en mejora (particularmente los más débiles) no aumentarían lo suficiente, por lo que continuarían siendo fuertemente subestimados y tirando de ellos. grandes cantidades de puntos de calificación lejos del grupo de calificación establecido al enfrentarlos. Esto ejercería una fuerza deflacionaria significativa sobre el conjunto de calificaciones en general, que es precisamente el efecto que estamos tratando de contrarrestar con todos estos cambios propuestos.

Quizás también se pregunte por qué es necesario agregar dos empates adicionales contra oponentes de nivel 1800 en la fórmula para nuevas calificaciones. Análisis anteriores han indicado que en el actual sistema Elo de la FIDE, los jugadores que reciben calificaciones iniciales relativamente bajas tenderán a comenzar inmediatamente a superar su calificación (y ganar puntos de calificación), mientras que los jugadores que reciben calificaciones iniciales relativamente altas comenzarán inmediatamente a tener un rendimiento inferior a su calificación (y a perder calificación). puntos). Este es un fuerte indicador de que podría haber demasiada variabilidad, demasiada dispersión, en las calificaciones iniciales que los jugadores han estado recibiendo.

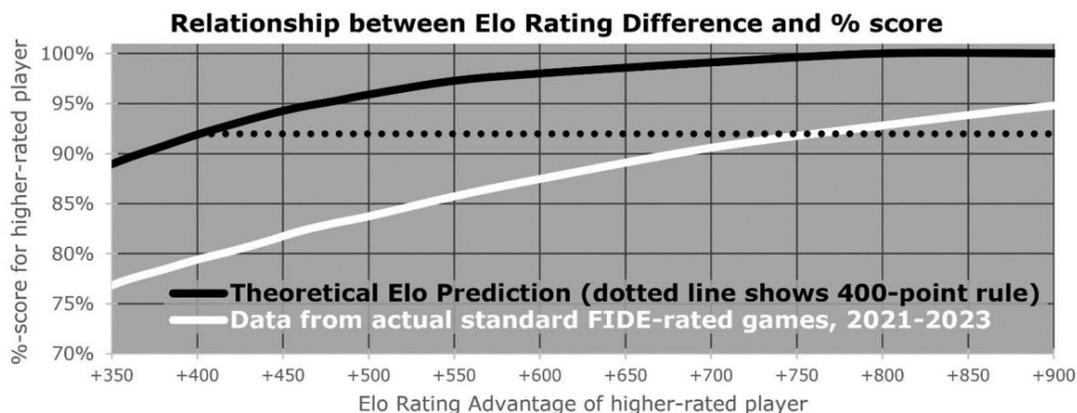
Además, todos los indicios apuntan a que la fórmula de calificación inicial es uno de los principales contribuyentes a la extrema deflación de calificaciones que estamos experimentando actualmente. Dado que gran parte del sistema Elo es un juego de suma cero en el que los puntos de calificación simplemente se intercambian entre jugadores, simplemente no existe otro medio significativo por el cual se puedan introducir grandes cantidades de puntos de calificación en exceso o, por el contrario, grandes deficiencias de puntos de calificación en el sistema. sistema de valoración. La fórmula de calificación inicial es la clave, ya que contribuye en gran medida tanto al problema actual como a la solución propuesta.

Las mejoras en los cálculos, incluidos dos resultados hipotéticos empatados contra una oposición bastante fuerte (1800 Elo), incorporarían un factor inflacionario leve en la fórmula de calificación inicial para ayudar a contrarrestar el factor deflacionario natural que es inevitable cuando la FIDE permite que muchos jugadores junior débiles o en mejora ingrese al grupo de calificación. Si no se añade así algún tipo de contrapeso al efecto deflacionario, es probable que volvamos a estar en el mismo barco dentro de unos años, incluso si implementamos una gran compresión única del conjunto de calificaciones Elo.

Al agregar dos empates hipotéticos al desempeño de cada jugador mientras no está calificado, acercaremos su puntaje porcentual en esos juegos al 50% y, por lo tanto, su puntaje inicial se desviará menos de la fuerza promedio de su oponente y, en última instancia, esto traerá menos variabilidad al juego. calificaciones iniciales. Y empatar contra un oponente con una calificación de 1800 es probablemente un ligero rendimiento excesivo para la mayoría de los jugadores sin calificación, por lo que estos empates adicionales "integrarán" una pequeña cantidad de mejora planificada en las calificaciones iniciales de los nuevos jugadores, de modo que su mejora esperada y natural se mantendrá. no tendría tanto impacto sobre el conjunto de calificaciones establecido como ocurre hoy.

Nota especial sobre la regla de los 400 puntos: La regla de los 400 puntos pone un límite del 92% en la puntuación esperada para un jugador que tiene una ventaja de calificación extrema (por encima de 400 puntos Elo) en un juego calificado. La FIDE implementó recientemente un cambio regulatorio para que un jugador no pueda beneficiarse de la regla de los 400 puntos más de una vez por evento.

La relación entre la diferencia de calificación Elo y la puntuación porcentual ya se ha mostrado en esta propuesta, pero analicemos una parte particular de ella, es decir, la parte donde la ventaja Elo oscila entre +350 y +900. Compararemos la curva negra (que muestra la predicción teórica de Elo) con la curva blanca (la evidencia real de juegos recientes de 2021-2023 donde la diferencia de Elo estaba entre 350 y 900).



Al observar los datos reales del juego, podemos ver que actualmente, los jugadores con ventajas de calificación de 400 puntos ni siquiera obtienen el 80% en dichos juegos, y mucho menos el 92%, y se necesita algo así como una ventaja de calificación de 650-700 puntos Elo hasta que podamos esperar razonablemente que el favorito del rating obtenga una puntuación tan alta como el 90%.

Además, en el sistema de clasificación actual, en aquellos eventos en los que son favoritos con más de 400 puntos Elo en más de un juego durante el evento, dichos jugadores con mayor puntuación esperarían perder puntos de clasificación en más del 70% de esos eventos, incluso si la regla de los 400 puntos se aplicara plenamente a todos los juegos aplicables (no sólo a uno).

Aunque parece bastante probable que las mejoras de compresión y cálculo propuestas generen mejoras drásticas, obviamente no hay garantía de que esto suceda. Por el momento, en el sistema de clasificación actual, la baraja está muy marcada por los favoritos en la clasificación. El gráfico anterior nos dice que incluso con el 400-

Si se aplicara plenamente la regla de los puntos, los favoritos de calificación aún esperarían perder puntos de calificación en juegos donde la diferencia de calificación sea de 700 puntos Elo o menos.

Por lo tanto, parece contraproducente en el entorno actual penalizar a los favoritos de rating extremo por enfrentarse a jugadores de rating más bajo. Mi evaluación es que la regla de los 400 puntos es actualmente una herramienta útil para animar a los jugadores más fuertes a enfrentarse a jugadores de menor rating, y deberíamos restablecer el comportamiento en el que la regla de los 400 puntos se aplica a todos los juegos, incluso si ya hay un favorito en el rating. se beneficia de ello en otras partes del torneo. Esto también introduce un efecto inflacionario menor que ayuda ligeramente a contrarrestar la deflación en el sistema.

Sin embargo, si estos cambios propuestos realmente tienen el efecto deseado, entonces recomendaría que la regla de los 400 puntos se reduzca a una regla de 500 o 600 puntos, o que se elimine por completo. Quizás se pueda reevaluar después de un año.

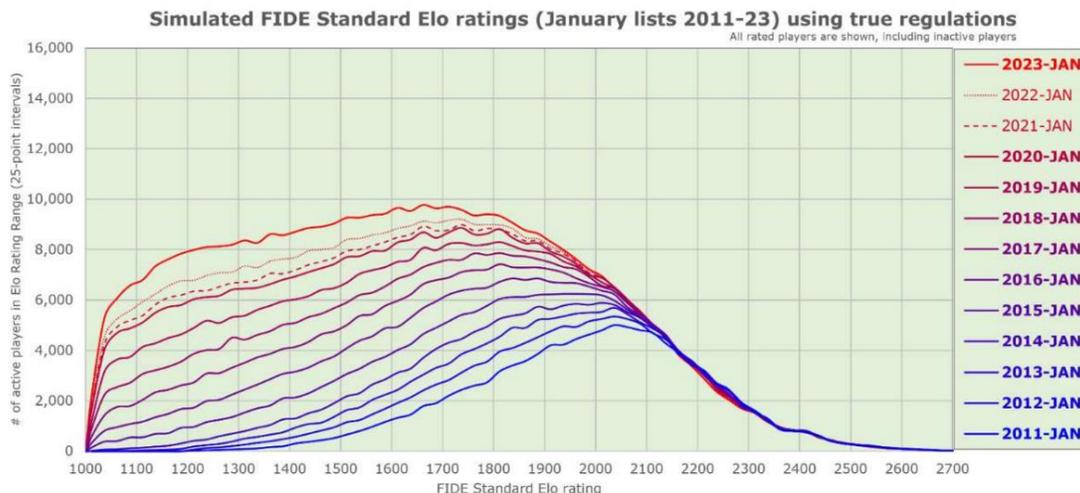
Parte V – Simulaciones

He probado exhaustivamente estas diversas correcciones mediante simulación. La simulación más útil que ejecuté es probablemente el escenario en el que pretendemos que ambas correcciones se implementaron a principios de 2017, lo que significa que la compresión se implementó para la lista de calificación de enero de 2017 y las mejoras de cálculo se implementaron para la calificación de febrero de 2017. lista. Esto no sucedió en la vida real, por supuesto, pero podemos ver qué habría sucedido con las calificaciones si estas correcciones realmente hubieran ocurrido en 2017 y luego todos los millones de juegos de la vida real fueran jugados posteriormente por jugadores dentro de ese simulado. sistema de valoración.

En primer lugar, debemos establecer el comportamiento básico, es decir, lo que sucedió en la vida real sin estos cambios. Al ejecutar el simulador con las regulaciones históricas reales que estuvieron vigentes desde 2008 hasta 2023, y luego contar cuántos jugadores calificados había de varias calificaciones en cada lista de enero, podemos trazarlos gráficamente e intentar comprender cómo se distribuye la calificación. Los jugadores han cambiado con el tiempo en los doce años transcurridos desde 2011.

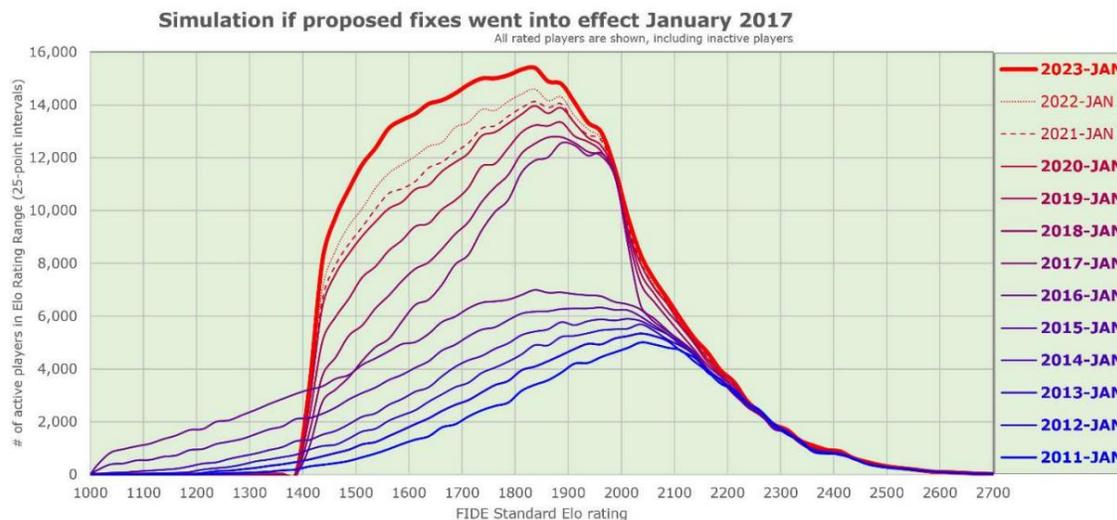
En la siguiente representación de histograma, se cuenta el número de jugadores en cada rango de 25 puntos Elo en cada lista de enero y luego se representa en una curva separada para ese año. Por lo tanto, la curva azul inferior representa la distribución de jugadores calificados de enero de 2011, la curva superior

La curva roja representa la distribución de jugadores calificados de enero de 2023, y también se muestran todos los años intermedios, en varios colores degradados rojo, violeta y azul.

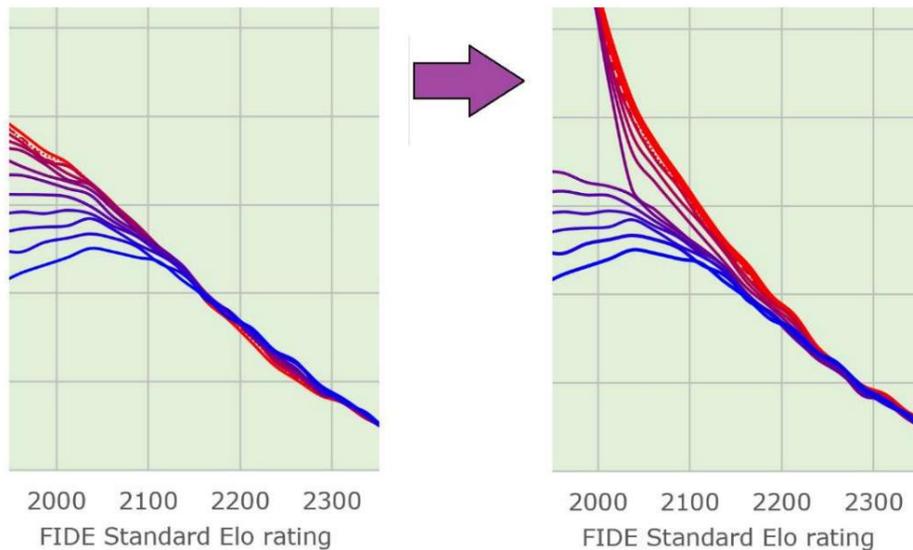


Las líneas aumentan cada año en la mitad izquierda de la imagen, lo que indica que cada año había muchos más jugadores con una clasificación inferior a 2100. El grupo de clasificación en general creció de unos 100.000 jugadores en 2011 a casi 400.000 en 2023. Pero a pesar de todo este crecimiento, podemos ver en la mitad derecha de la imagen que esencialmente no ha habido crecimiento en la cantidad de jugadores con calificación Elo 2100+. Al principio, esto realmente no tiene sentido. Seguramente con tantos jugadores nuevos obteniendo calificaciones, ¿algunos de ellos son maestros? La explicación, por supuesto, es que cualquier crecimiento modesto que haya habido entre los jugadores fuertes ha sido anulado por la deflación, al perder injustamente puntos de rating los jugadores más fuertes. Esencialmente, el conjunto de calificaciones ha ido creciendo hacia la izquierda en lugar de hacia arriba.

En la simulación en la que pretendemos que las mejoras de compresión y cálculo se implementaron en enero/febrero de 2017, nada es diferente en este gráfico hasta 2016, por lo que las líneas azules se verán muy parecidas. Sin embargo, en 2017 las cosas empiezan a cambiar mucho. La compresión reúne la mayor parte del conjunto de calificaciones y luego la fórmula de calificación inicial comienza a empujar más calificaciones nuevas hacia el medio. La calificación mínima se mantiene en 1400 Elo en lugar de 1000 Elo. Entonces, seis años después, hay una distribución de jugadores calificados de aspecto muy diferente, concretamente la línea roja gruesa que indica la lista de clasificación simulada de enero de 2023:



Una característica obvia de la compresión es que conduce a muchos más jugadores con una calificación cercana a 2000 Elo, incluso si no produce inmediatamente ninguna calificación adicional por encima de 2000 Elo. Es inevitable que algunos de esos jugadores avancen a 2100 Elo o incluso a 2200 Elo. Una vez más podemos esperar un crecimiento modesto en el número de jugadores de nivel maestro. Acerquémonos a la región de 2000-2300 Elo en las dos últimas imágenes.



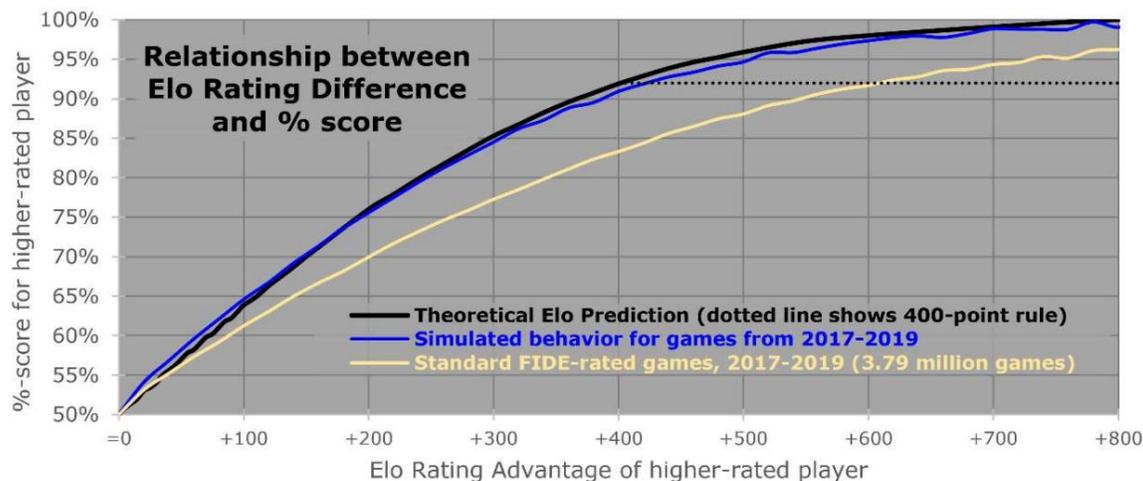
Podemos ver en la imagen del lado izquierdo (lo que realmente sucedió) que las líneas rojas no subieron más que las líneas azules de años anteriores, lo que significa que no hubo crecimiento en esta parte del grupo de calificaciones. Sin embargo, en la imagen del lado derecho (la simulación si la compresión hubiera ocurrido en 2017), vemos en las líneas rojas que aumentan un poco cada año, que una vez más habría cierto crecimiento en el número de jugadores con una calificación de 2100-2200, y una pequeña cantidad de crecimiento en aquellos clasificados entre 2200 y 2300. Esto no es un crecimiento artificial, sino una representación más precisa de lo que realmente ha estado sucediendo con la distribución de las fortalezas del juego. Las mejoras en la compresión y el cálculo son un intento de alinear las calificaciones con las fortalezas evidentes de los jugadores.

Estos cambios probablemente conducirían a un modesto aumento en el número de jugadores titulados en los próximos años, pero estos títulos serían bien merecidos, nuevamente porque las calificaciones (con suerte) se están alineando más con las fortalezas de juego demostradas de los jugadores. La intención es revertir una década de deflación, en lugar de aumentar artificialmente la calificación de alguien o conceder títulos inmerecidos. La compresión solo afecta directamente las calificaciones por debajo de 2000 Elo, e incluso los jugadores un poco por debajo de eso solo ganarían una pequeña cantidad de puntos de calificación. En realidad, mi análisis apuntó a una compresión que llegaba aún más alto en la lista de clasificación, pero sentí que detenerla en 2000 Elo minimizaría cualquier efecto no deseado en el sistema de títulos o en los jugadores profesionales. La deflación restante en el sistema aún puede contrarrestarse en última instancia mediante las mejoras en los cálculos; Simplemente hará falta más tiempo para que esto suceda.

También podemos usar esta simulación para determinar si el desempeño de los jugadores calificados estaría más en línea con su expectativa de Elo (es decir, usando sus calificaciones simuladas). Podemos tomar las calificaciones simuladas de todos esos jugadores y una vez más trazarlas contra las curvas negras que indican su puntaje porcentual esperado. Esto nos dirá si los favoritos de calificación realmente están alcanzando sus puntajes Elo esperados en la simulación.

En el siguiente gráfico, la curva azul indica cómo se habrían desempeñado los favoritos en 2017-2019 con estas calificaciones simuladas (antes de la pandemia), mientras que la curva dorada indica cómo se desempeñaron realmente los favoritos en 2017-2019 (antes de la pandemia) con su real

Clasificaciones publicadas por la FIDE:



La curva azul se superpone/interseca casi perfectamente con la curva negra, que es esencialmente el objetivo. Esto refleja datos de calificaciones simuladas durante 2017-2019. Incluso durante la simulación, los cierres del juego sobre el tablero durante la pandemia de COVID habrían tenido un efecto adverso en el sistema de calificación en 2020-2022, causando que las calificaciones de los juniors en mejora se quedaran por detrás de su verdadera mejora. Es por eso que nos centramos en el rendimiento solo en los juegos de 2017.

2019 en el gráfico anterior, que representa los primeros tres años después de la Compresión y la Mejoras en el cálculo.

Y, finalmente, podemos consultar la tabla resumida de esos años simulados de 2017 a 2019 y ver si los favoritos de calificación realmente se están desempeñando más cerca de sus expectativas en todo el conjunto de calificaciones. Como puede ver a continuación, el sombreado rojo/azul es mucho más pálido y los números están mucho más cerca de cero que lo que sucedió en la vida real durante 2017-2019. De hecho, las cifras parecen mucho mejores que incluso en el período 2008-2012.

2017-2019 (simulated)	opponent player	vs.	vs.										
		1400-99	1500-99	1600-99	1700-99	1800-99	1900-99	2000-99	2100-99	2200-99	2300-99	2400-99	2500+
+1.6% in 131,598 games	1400-99	0%	+0%	+2%	+2%	+2%	+2%	+1%	+1%	+1%	=0%	=0%	=0%
+0.8% in 416,053 games	1500-99	-0%	0%	-0%	+1%	+2%	+2%	+1%	+1%	+1%	+1%	+0%	=0%
+0.5% in 612,034 games	1600-99	-2%	+0%	0%	+0%	+1%	+1%	+1%	+2%	+2%	+1%	+1%	+1%
+0.0% in 797,100 games	1700-99	-2%	-1%	-0%	0%	+0%	+1%	+0%	+1%	+2%	+1%	+1%	+0%
-0.5% in 932,399 games	1800-99	-2%	-2%	-1%	-0%	0%	-0%	-0%	+1%	+1%	+2%	+1%	+1%
-0.2% in 957,268 games	1900-99	-2%	-2%	-1%	-1%	+0%	0%	-1%	+0%	+1%	+2%	+2%	+0%
+0.7% in 809,355 games	2000-99	-1%	-1%	-1%	-0%	+0%	+1%	0%	+1%	+2%	+3%	+2%	+1%
+0.2% in 581,912 games	2100-99	-1%	-1%	-2%	-1%	-1%	-0%	-1%	0%	+2%	+3%	+3%	+1%
-0.5% in 418,565 games	2200-99	-1%	-1%	-2%	-2%	-1%	-1%	-2%	-2%	0%	+2%	+3%	+2%
-1.3% in 272,061 games	2300-99	=0%	-1%	-1%	-1%	-2%	-2%	-3%	-3%	0%	+2%	+2%	+1%
-1.5% in 193,010 games	2400-99	=0%	-0%	-1%	-1%	-1%	-2%	-2%	-3%	-3%	0%	+2%	+1%
-1.2% in 123,845 games	2500+	=0%	=0%	-1%	-0%	-1%	-0%	-1%	-1%	-2%	-1%	0%	+1%

Hay muchas razones para esperar que estas medidas sean muy efectivas para revertir los efectos de una década o más de deflación de calificaciones, además de proporcionar fuertes contramedidas para evitar que esto vuelva a suceder. Esto conducirá a un sistema de clasificación Elo estándar de la FIDE que funcionará mucho mejor que en la actualidad.