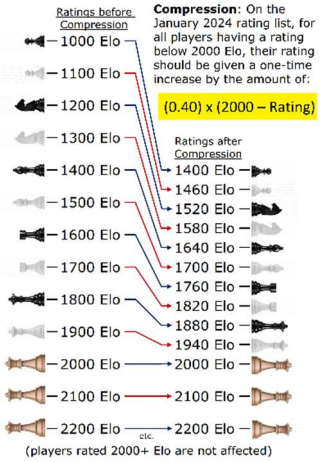
Compression and Calculation Improvements: Supplemental Report

-- by Jeff Sonas (29 October 2023)

I know this is a long document! I am hoping that most readers will at least read the first four pages, because page 3 summarizes my current opinion on several pertinent questions, and page 4 provides many useful and interesting excerpts of responses from the chess community. The remaining 43 pages (**Appendix #1** through **Appendix #7**) provide additional detail intended mainly for readers who are interested in diving more deeply into some of these topics.

Background



In July of 2023 I submitted a proposal to FIDE, in which I recommended a Compression to 1400 Elo. In this compression, the FIDE standard ratings of all players rated 1000-1999 Elo would be given a one-time increase according to a simple formula, with players near 1000 Elo gaining almost 400 Elo points and players near 2000 Elo gaining almost no Elo points.

This formula would cause the Elo range that currently spans 1000 Elo points (from 1000-2000 Elo) to be compressed into a range of 600 Elo points instead (spanning 1400-2000 Elo). I recommended that FIDE at the same time should increase the minimum rating from 1000 Elo to 1400 Elo, revert the 400-point-rule back to how it was a few years ago, and implement several changes to how initial ratings are calculated.

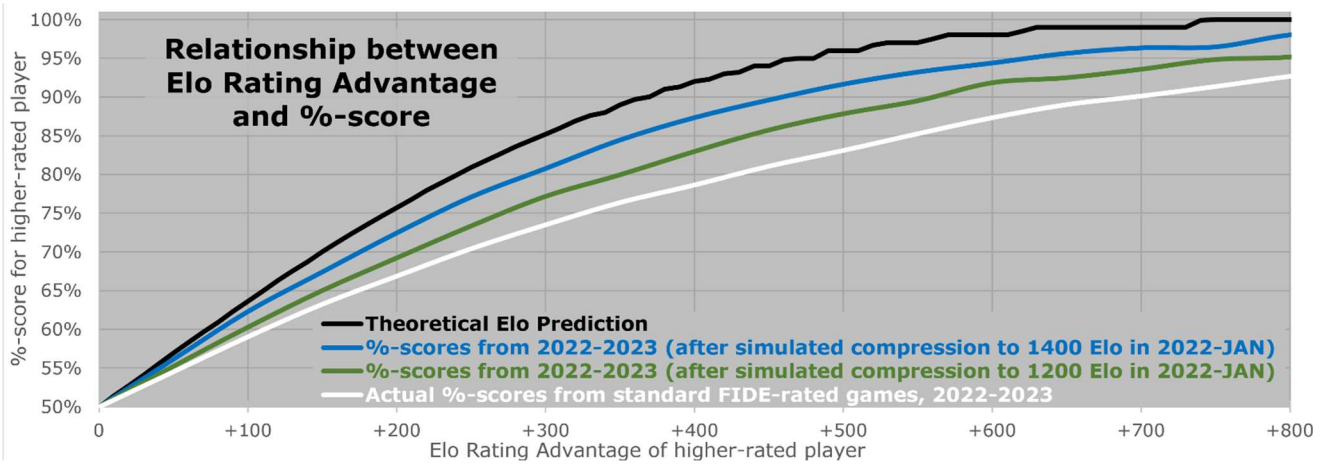
FIDE published the proposal later that month (see https://www.fide.com/news/2538) and encouraged the chess community to express their comments, suggestions, ideas, and calculations regarding this proposal, during the months of July, August, and September. This public discussion period has now closed. I have carefully read all submitted feedback and considered several of the important points that were raised, and I have also performed additional analysis myself in the meantime. I can't come close to responding to all of the individual points that were raised, but I would like to reference several notable comments, on page 4 and also while presenting additional analysis in the various appendices of this document.

Executive Summary

A big problem in the current FIDE Standard Elo system is that over the past decade, ratings have become way too spread out, and the Elo tables aren't really functioning very well anymore. Rating favorites are constantly struggling to reach their Elo expectations, and an overly pessimistic formula for players' initial ratings has ultimately pulled tremendous quantities of rating points away from established rated players. I have described this effect as "rating deflation". And I think the best way to tackle this, without revolutionizing the Elo system, is to compress the ratings back together and also make the initial formula less pessimistic.

My primary recommendation is the "Compression to 1400 Elo", which I believe would solve about two-thirds of the problem all at once. A different approach, with some support within FIDE, would be a lesser intervention: a "Compression to 1200 Elo". I believe this would solve about one-third of the problem all at once.

This can be illustrated easily by the graphic at the top of the next page:



The black line shows the traditional Elo expected score for various amounts of Elo rating advantage, and the white line shows what rating favorites have actually been scoring in standard chess since January of 2022. There's a huge gap between the black line (theory) and the white line (actual practice); rating favorites are falling way short of their Elo expectation.

If we had done a "compression" of player ratings at the start of the year 2022, then player ratings would have been closer together, and so it would have been easier for rating favorites to achieve their Elo expectation. The green line above shows us how rating favorites would have done during the years of 2022-2023 if a "Compression to 1200 Elo" had happened at the start of 2022. Evidently this would have closed the gap by about one-third.

Or even better, if we had done a "Compression to 1400 Elo" at the start of 2022, then the blue line shows us how those rating favorites would have done during the years 2022-2023. In this case, it would seemingly close the gap by about two-thirds. Of course, we are projecting into the future, based on simulations from the recent past, but I think the simulations are reliable.

Practically speaking, we can't solve the problem completely by one of these compressions, because ratings are too spread out among the 2000+ Elo crowd as well (though less so), and it seems unwise to do a compression that reaches to the master level. Nevertheless, even closing the gap by two-thirds would be a fantastic improvement, if we can accomplish this in real life.

My Latest Recommendations

In the rest of this document, I discuss various topics surrounding the proposed compression and calculation improvements. These are the seven main topics covered in this document:

#1. Recommended compression amount
#2. Has the deflation reached elite players yet?
#3. Should we expect a compression to cause inflation?
#4. Should we revert the 400-point-rule to its earlier form?
#5. Treatment of lower-rated player for 400-point-rule
#6. Should we compress rapid/blitz now as well?
#7. Should we require 5 games, or 9 games, for your initial rating?

For the most part, my recommendations have not changed since July. On the next page you will find a summary of my current position on these seven main topics. For each numbered Summary Point, there is also a corresponding numbered appendix later in the document, where I go into far more detail. In this way, you can either quickly review my current thoughts and then stop reading after 3 or 4 pages, or you can dive deeper into the corresponding appendices.

Summary Point #1. Recommended compression amount

I still believe that **we should compress the rating pool up to a minimum of 1400 Elo**. This is what I recommended in my original proposal, which also included evidence (from simulations) that the compression and calculation improvements would bring expected scores more into line with actual performance. However, I did not address the question of why I specifically suggest a Compression to 1400 Elo and not to some other level, such as 1200 Elo. A detailed analysis of this topic can be found in **Appendix #1**, on pages 5-15 of this document.

Summary Point #2. Has the deflation reached elite players yet?

I believe that **the rating deflation has already reached the ranks of elite players**. Recent trends indicate that even at the IM/GM level, player ratings are dropping. The deflation has almost certainly reached the top 1,000 players already (that's an Elo rating slightly below 2500), and probably even higher. I believe there is some urgency to take action on this (via the proposed compression), because if we only apply direct corrective measures to players rated below 2000 Elo, it will take time for the corrective effect to propagate up to the IM/GM level. More evidence of this deflation among master-level players can be found in **Appendix #2**, on pages 16-21 of this document.

Summary Point #3. Should we expect a compression to cause inflation?

I believe that **a compression to 1400 Elo will not result in any sort of inflation.** There are many complex factors at play in a rating system, some of which tend to increase ratings (we could call these "inflationary factors") and others which tend to decrease ratings (we could call these "deflationary factors"). A compression to 1400 Elo is certainly an inflationary factor, aimed at counteracting a large amount of the deflation that has happened in the past, but the compression is very unlikely to result in overall rating inflation. In fact, I do not expect it even to counteract the full amount of deflation that has already happened; we will still be left in a state where there is some deflation, albeit much less deflation. The "calculation improvements" in my proposal, especially the introduction of two additional hypothetical draws against 1800-strength opposition when calculating ratings for new players, introduce a mild inflationary effect that ought to further work against the remaining deflation.

To make such judgments, we can look at simulations based on real game results, giving us a sense of what we can expect to happen in the first years after a compression. More discussion about these simulations, and what they predict about inflation, can be found in **Appendix #3**, on pages 22-29 of this document.

Summary Point #4. Should we revert the 400-point-rule to its earlier form?

I still believe that **we should revert the 400-point-rule to its pre-2022 state**, where any rating difference larger than 400 Elo points would instead be treated as a difference of 400 Elo points, no matter how many times this happens during an event. Admittedly, any compression would lessen the need for a 400-point-rule, because player ratings would move closer together and thus there would be fewer matchups having such a large Elo difference. Nevertheless, from analyzing the performances of several rating groups against each other, across several different scenarios and simulations, it appears to me that mathematically, the (reverted) 400-point-rule is still needed. A more detailed analysis of this topic can be found in **Appendix #4**, on pages 30-39 of this document.

Summary Point #5. Treatment of lower-rated player for 400-point-rule

There seems to have been an unanticipated change that was made to the rating calculation when the newer 400-point-rule was implemented in January 2022. I believe that this unanticipated change should be reversed, and thus **whenever a player faces a rating disadvantage of more than 400 Elo points, it should be treated as a rating disadvantage of 400 Elo points**, with the consequent expected score of 0.08 for that player. A more detailed discussion of this topic can be found in **Appendix #5**, on pages 40-42 of this document.

Summary Point #6. Should we compress rapid/blitz now as well?

I believe that if we are eventually going to compress rapid/blitz ratings similarly to how we compress standard ratings, then **we ought to compress rapid/blitz ratings at the same time that we compress standard ratings**. We recently identified some potential problems that could happen if standard ratings were compressed right away, and then rapid/blitz ratings were compressed at a later date. Also, I have done some basic analysis of the general behavior of the rapid and blitz rating systems, enough to indicate that they are behaving quite similarly to the standard rating system, and would likely benefit from the same changes being contemplated for the standard rating system. This topic is discussed in **Appendix #6**, on pages 43-45 of this document.

Summary Point #7. Should we require 5 games, or 9 games, for your initial rating?

A few people have asked me to reconfirm some earlier analysis from a decade ago, as to whether increasing the games requirement from 5 games back up to 9 games (for an initial rating) would actually help the rating system. I believe that **we should continue to require only 5 games played** in order to receive an initial rating. This is discussed in **Appendix #7**, on pages 46-47 of this document.

Selected Chess Community Feedback

A lot of people from the chess community responded to FIDE's July request for feedback regarding the proposed changes. There were some excellent points made, and some very useful suggestions, and I have extracted several of those for inclusion here (below) in order to amplify their ideas. I hope that FIDE will be able to consider these ideas and others, in the near future:

(a) IM Douglas Bryson: *"FIDE uses variable Ks such that a junior will win more points than an adult loses. However the system has a problem with regard to junior v junior games since they have the same K and will merely swap points rather than raise the level of the players."*

(b) IM Douglas Bryson: *"Junior Additions: In the Chess Scotland (CS) system the longest established procedure (introduced in 1976) for compensating opponents is to add a range of extra points to a junior's grade before calculating expected scores for the opponent. The range of points varies with age. In contrast to FIDE and other rating systems which do not utilise junior additions it means points are added into the rating pool even when games are played between juniors. e.g. a 1000 rated junior plays another 1000 rated junior and they draw – both players will gain points in CS processing because the expected score for each player will be lower than 0.5 even though their grades are identical. In FIDE such a game would produce no increase. However we know juniors will on average have improved over any time period but that will not be reflected with an improved FIDE rating."*

(c) Vlad Ghita: *"The FIDE rating system has underperformed URS in predicting individual game results for quite some time already. In order to validate the health of the proposed changes, I suggest that the QC considers displaying the URS as an 'auxiliary rating' on the FIDE rating page of each player for the period 1 January 2023 - 31 December 2024, then make a final report."*

(d) Marc Gamble: *"Another approach would be to move away from the traditional Elo rating system to something more like Glicko which accounts for inactivity and rapid improvement by applying a multiplier to the score of less active or less established players."*

(e) Vlad Ghita: *"The compression (and the calculation improvements) means that K=40 is not needed for juniors anymore, as their ratings are far more accurate to begin with, so I suggest reducing it to the optimal value of 32, or even 30."*

(f) Alistair Compton: *"I further propose that even after a one off adjustment has been made, that future individual games between adults and players under 18 years of age should be rated using a differing K factor. The junior player should still have the game rated using their K factor of 40, but the adult should have the game rated using a K factor of 5. I believe this would reduce the effect of the significantly underrated juniors deflating the adults grade, while still allowing the junior to reach their correct rating quickly."*

(g) GM Keith Arkell: *"At the moment we have a crazy situation where, for example, inactive 2500 players can be weaker than active 2400 players, due to the 12 years of deflation. To remedy this I propose increasing the point where k-factor reduces to 10 from 20 from 2400 to 2600. This way players will come more quickly back to their correct rating."*

(h) GM David Smerdon: *"FIDE should return to players the Elo rating points that were 'lost' as a result of the January 2022 'single-upgrade' rule change to the 400-point rule."*

(i) Alexander Kalenov: *"Modern computing power is ready to maintain rating systems in which all games played in a tournament are counted, regardless of whether the player has a rating or not."*

Appendix #1: Recommended compression amount

The compression to 1400 Elo would be an immense change. It would add approximately 70 million rating points into the standard Elo rating system. We would be doing this in response to the clear and compelling evidence, from millions of game results across recent years, that player ratings in the Elo system have become far too spread out, to the point that differences in ratings are no longer very reflective of the true differences in strength among players.

Some people such as **GM Keith Arkell** felt that a compression to 1400 Elo didn't go far enough:

> I agree that a floor of 1400 would go some way towards reducing the stretchedness of the ratings range. I would prefer a minimum of 1600, but certainly not less than 1400 if you want this to have any significant impact.

Conversely, there seems to be some thought within FIDE that a compression to 1400 Elo is too severe, and instead a milder compression to 1200 Elo might be better, with perhaps additional compression another time. For example, **GM and FIDE CEO Emil Sutovsky** in an interview:

> The question is if we should go, let's say 400 points or 300 or 200 points: that's a mathematical situation, as there are other things to be considered, for example, titles. Also, an immediate change may cause inflation. We saw deflation for such a long time, now we have to be careful not to compensate for it in one go: maybe we should spread the points over time, months maybe, even years. These technical decisions are important and very difficult to measure.
>
> I think we should take a balanced decision to see how it reflects on the community, and how players feel about it, and then adjust: it's not a huge drama if in one year we see that the K factor should be 30, instead of 20 or 40. We might raise the rating floor to 1200 starting next year and then 1300. The simplest would be just to take one decision and stick to it and protect it. We often are criticized but I think that's actually the smart thing to do: adjust and then think what can be done better. It's easy to think "we are FIDE, we can decide and go" but that's not the way to go – we are actually listening, we hear many proposals and ideas, and of course, we can't be perfect.

If we were to only do a compression to 1200 Elo, taking the Elo range that is currently 1000-2000 and compressing it so that it only spans 800 Elo points, and then establishing a minimum Elo rating of 1200, it would still be a very large change, adding 35 million rating points into the system. Because it seems to be an alternative that FIDE is seriously considering, I have made sure to include the compression to 1200 Elo (not just the compression to 1400 Elo) as an option in my various simulations, investigations, and graphs presented later in this document.

Nevertheless, although there would be nothing wrong with doing a compression to 1200 Elo, and it certainly would help, I am still in favor of the full compression to 1400 Elo instead. Believe it or not, my calculations suggest that the true level of overall deflation is more like 110 million Elo points, and reaches all the way up to elite players. So even my full proposal (compression to 1400 Elo, which adds about 70 million Elo points) would only address about 64% of the problem, and a halfway "compression to 1200 Elo" (which adds about 35 million Elo points) would only address about 32% of the problem. So I don't think that a compression to 1200 Elo will do nearly enough.

Before going any further into the question of why I think a compression to 1400 Elo would be best, I first want to reference some things said by **GM Bartlomiej Macieja** in his feedback, because I decided to rework some of my calculations after reading these comments:

> 4) In your proposal, you suggest adding extra points to players solely based on their rating. This proposal can probably be dramatically improved by adding the age factor. I don't expect 60-year old players in that very rating group underperforming significantly. Why would adding them up to 400 points make the system more reliable?
> Have you tried making analyses for the same rating ranges but for different age groups?
>
> 5) It brings us to the next point. If FIDE is to adjust ratings of young rapidly progressing players so the new ratings better reflect the expected performance, then also an adjustment (in the opposite direction) is needed for inactive old players. With the same explanation.

I feel that I should emphasize something here: the purpose of the compression is certainly not to *"adjust ratings of young rapidly progressing players so the new ratings better reflect the expected performance"*. If we were trying to do that, then we would need to intervene repeatedly, by raising players' ratings on a regular basis to try and keep up with their age-related improvements. Instead, this compression is intended as a one-time intervention, to try and repair the damage to the rating system that has been inflicted by more than a decade of compounding deflation. It is not the current rapidly progressing junior players that I am concerned with. Rather, I worry about the ratings of all the past opponents of improving juniors from years ago, and the subsequent opponents of those opponents, and so on. There is an extraordinary amount of deflation that has propagated around to virtually the entire rating pool.

Nevertheless, this feedback from GM Macieja raises an important point that I should have addressed previously.

We all know that many junior players have low ratings and perform above their Elo expectation, simply because they are early in their career and still improving. Thus if junior players face higher-rated opposition, we are unsurprised when the junior players perform better (on average) than predicted, and take rating points away from their higher-rated opponents. And similarly, we all know that many older players have high ratings and perform below their Elo expectation, because they are late in their career and declining. Thus if older players face lower-rated opposition, we are unsurprised when the older players perform worse (on average) than predicted, and lose rating points to their lower-rated opponents.

In both cases, we have an alternative explanation for why the true difference in strength between opponents is actually less than what their ratings show. Is it possible that all I have "detected" is that junior players often improve and older players often decline? If so, it would be quite unfortunate for me to raise such a fuss, since we already knew that much about junior players and older players!

In fact, I am claiming that the ratings of **all** players are too spread out, not just the junior and older players. Which raises a very important question. Namely, is there a way we could measure the true extent of the "rating deflation" by only focusing on players in a stable age range, to avoid any problematic bias introduced by improving junior players and declining older players?

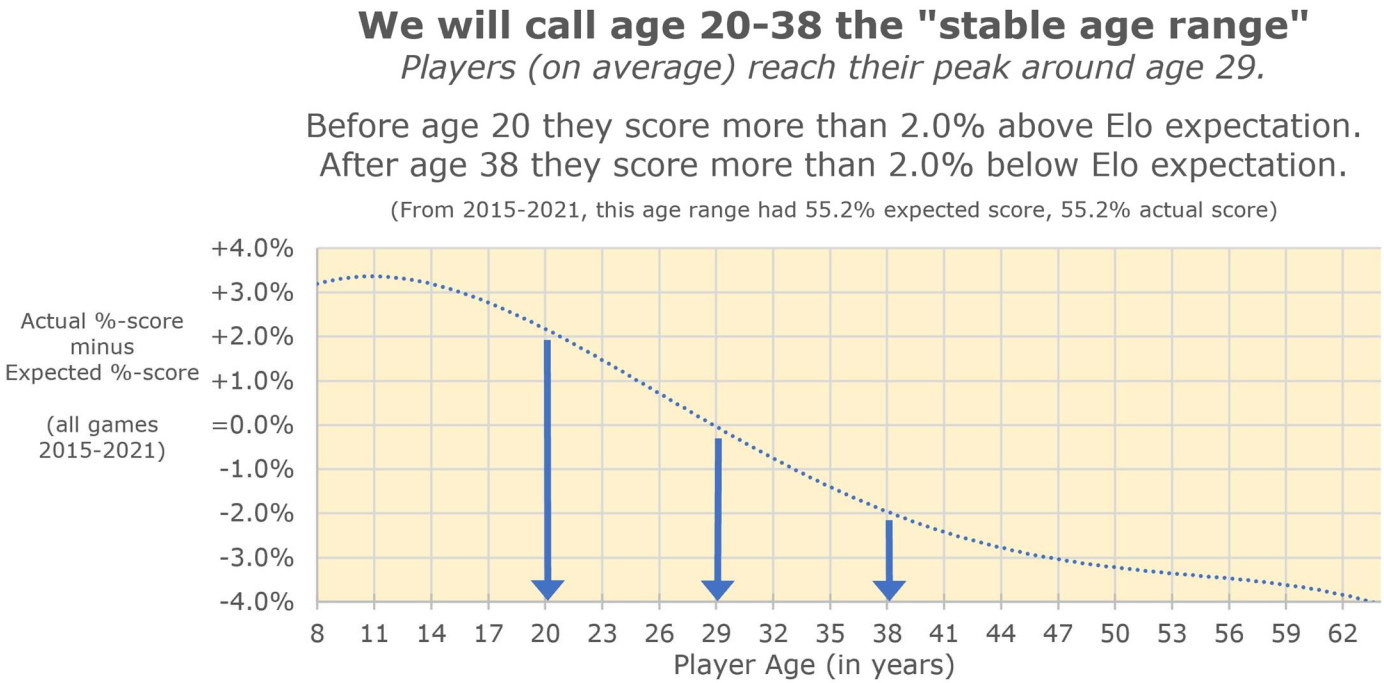
That leads to another important question. Namely, what is the "stable age range"? Maybe we can just look at those players and see if their ratings are indeed too stretched apart right now?

To investigate this question, I decided to partition players into three different age ranges: the younger "improving" group, the older "declining" group, and the middle "stable" group. Then I would look only at recent games played among players in the stable group, when checking whether we do see universal evidence of ratings having become stretched too far apart.

Okay, well, where exactly are the boundaries between "improving" ages and "stable" ages and "declining" ages? Since I was going to ultimately analyze games played during 2022-2023 and look for signs of deflation and ratings being spread out too much, I used data from a slightly earlier time period (2015-2021) to help me assess where these boundaries should be defined.

I looked at all rated games played from 2015-JAN through 2021-DEC, for players of each integer age, and I calculated the average expected score and average actual score across those games for each age. Then, I subtracted the expected score from the actual score to see whether players of each age were overperforming or underperforming relative to their rating, and I drew a line to represent the average level of overperformance or underperformance as players age.

The results are plotted in the following graphic, where the X-axis tells you the player age, and the Y-axis tells you how many percentage points the players of that age scored (per game) above/below their expectation. So this is "per game" rather than "per player", of course.

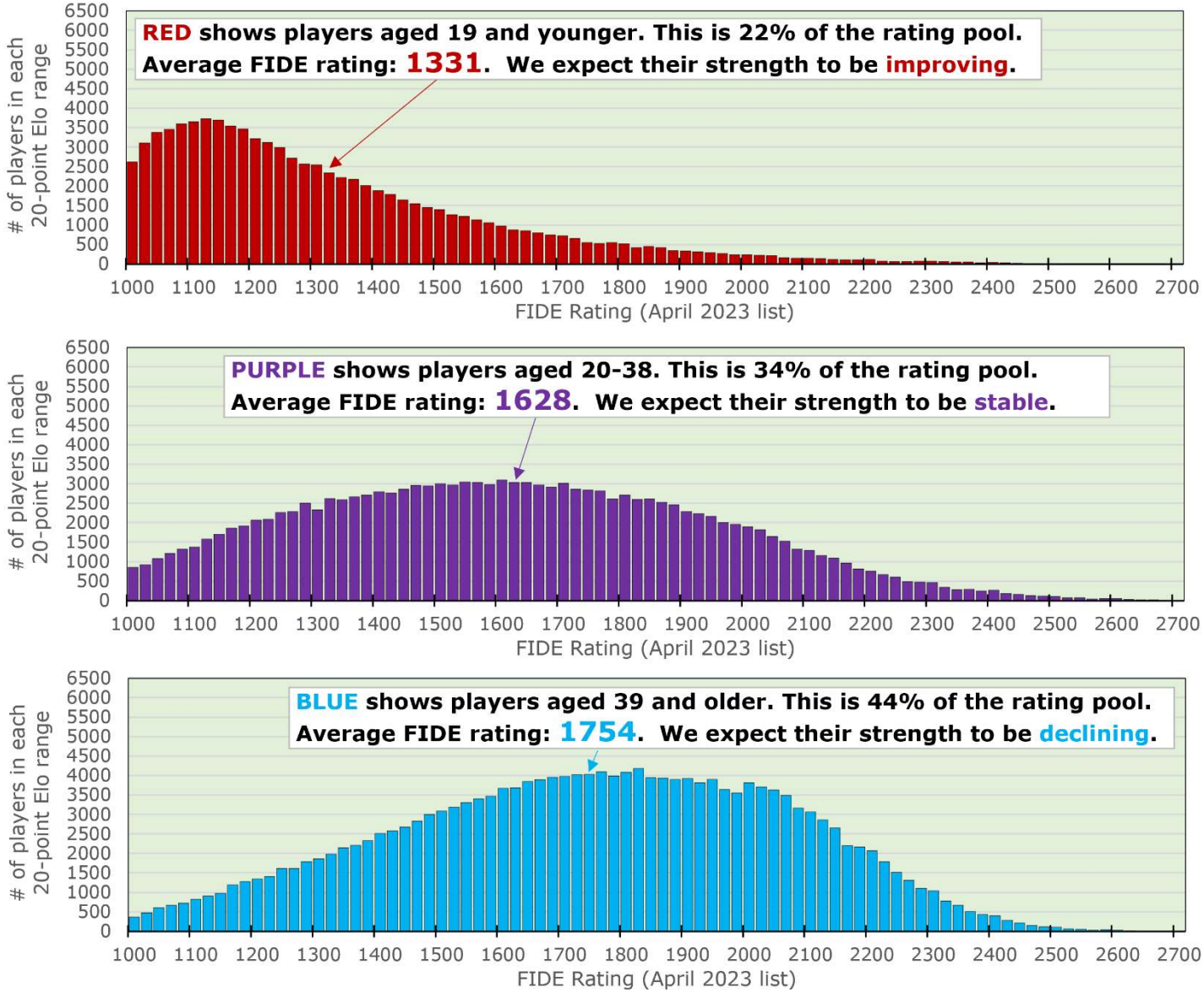


For example, players aged 35-37 years had an average expected score of about 55.5% during those seven years, while actually scoring only 53.7%, meaning they underperformed their rating expectation by 1.8%. Interestingly, Professor Arpad Elo in his 1978 book stated that *"Peak performance is attained around age 36"*, though by now it is clear that in today's game – half a century later - players reach their rating peak at a much younger age.

We can see from this graphic that younger players outperform their rating expectation (and thus gain rating points) up to about age 29, after which they start losing rating points. So the peak is

now reached (on average) at age 29. We can also define a "stable age range" from age 20-38, which is the age range where players tend to perform closest to their expectation (within +/- 2 percentage points). As a double-check, we can see that this entire group, spanning age 20-38, did indeed achieve stable results across those seven years of data (2015-2021), with an overall expected score of 55.2% matching their actual score of 55.2%.

Armed with this new definition of a "stable age range", we can then break apart the FIDE standard rating pool into three groups, based on age. An "improving" group, including all players younger than 20, is shown in red at the top of the next graphic. A "declining" group, including all players older than 38, is shown in blue at the bottom. And a "stable" group, which is all players aged 20-38, is shown in purple in the middle.



In each case, a histogram view shows you the distribution of player ratings (from that age group) on the April 2023 FIDE standard rating list. Each vertical bar indicates the number of players in a different 20-point Elo range. For example, in the two leftmost red bars, we can see that there were about 2,600 players younger than 20 who were rated 1000-1019, about 3,100 players younger than 20 who were rated 1020-1039, and so on.

It is clear that these are three very different populations of players, with the younger players (in red) tending to have much lower ratings than the older players (in blue). My contention is that large amounts of deflation are present throughout the entire pool of rated players. If this were true, we would expect to see, even among the players in the purple "stable" group, that there is strong evidence of their ratings being too spread apart. Therefore, in the next section, where I will go through some examples of how I arrived at the recommended compression to 1400 Elo, my analysis will exclude games involving players from the younger age group or the older age group. Instead, I will only consider games played during 2022-2023 between two players from the (purple) stable age group, when both players were age 20-38. Unfortunately, this does exclude a lot of game data, but it likely will give us a clearer sense of the level of deflation that is currently present throughout the rating pool.

In my original proposal, I produced a lot of evidence that ratings had become far too spread apart over the years. But I didn't really demonstrate how I had arrived at my final recommended numbers for the compression that would counteract this effect; I knew it would take up a lot of space to explain! Finally the time has come to devote a few pages to this effort.

My assertion is that the range of players currently rated 1000-2000 Elo, a range spanning 1,000 Elo points of ratings, actually spans only 600 Elo points' worth of actual playing strength. And that is why I suggested compressing that range of player ratings by 40%, so that it only spans a range from 1400-2000 Elo, or 600 Elo points.

Let us now dive into the underlying justification for my claim of *"600 Elo points' worth of actual playing strength"*. We will try a few different approaches, and remember that for all of this analysis, we are only considering players in the "stable age range" of age 20-38, and looking at very recent game data (games played in 2022-2023).

First of all, let's take a very broad view. We will establish a few rating groups based on 200-point Elo intervals: 1000-1199, 1200-1399, and so on. And then we will see what happened on average, in all the games from 2022-2023, when a player from one group faced an opponent from the next-higher-rated group, so when a 1000-1199 player faced a 1200-1399 opponent, or a 1200-1399 player faced a 1400-1599 opponent, etc.:

All games from 2022-2023 when both players aged 20-38 and are in adjacent FIDE Elo Rating Groups

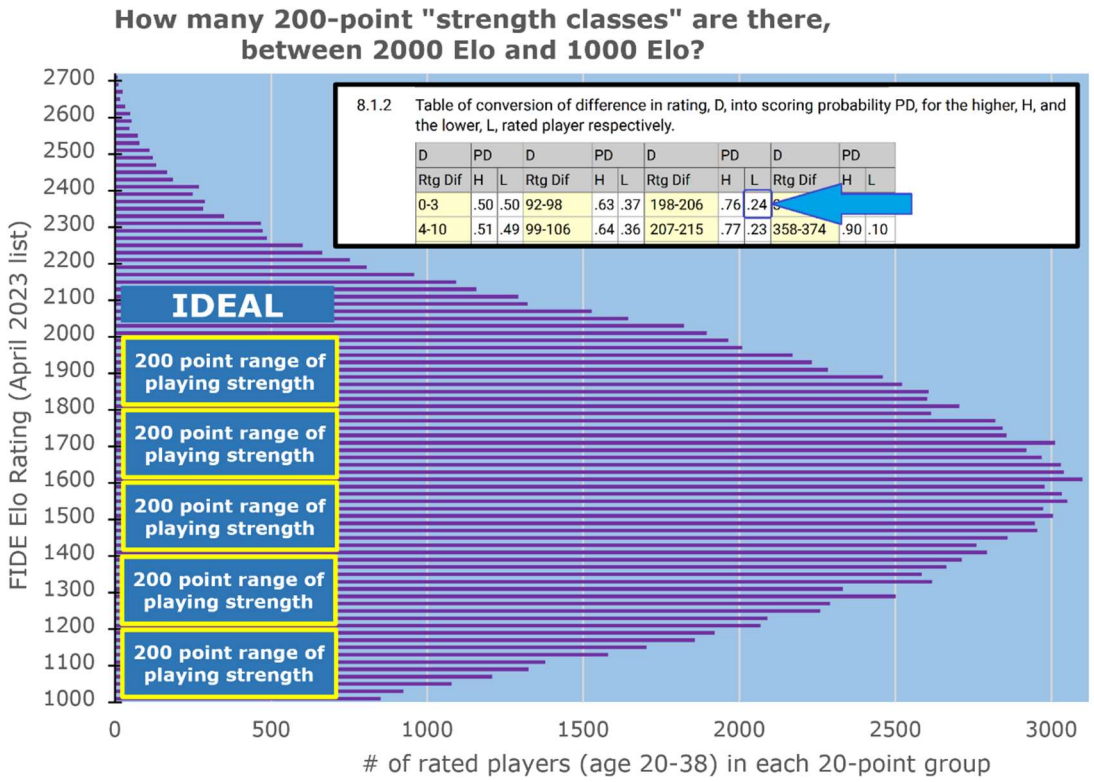
FIDE Elo Rating Group (lower-rated player)	FIDE Elo Rating Group (higher-rated player)	# games played (between groups)	Average FIDE Rating Difference (Elo points)	Average Expected %-score (for lower-rated player)	Average Actual %-score (for lower-rated player)	Empirical Difference in Playing Strength (Elo points)	Suggested Compression Percentage
1000 to 1199	**1200 to 1399**	1,436	-196 (1121 vs. 1317)	25%	34%	-117	**40%**
1200 to 1399	**1400 to 1599**	3,761	-195 (1306 vs. 1502)	25%	34%	-117	**40%**
1400 to 1599	**1600 to 1799**	5,946	-191 (1507 vs. 1699)	26%	34%	-117	**39%**
1600 to 1799	**1800 to 1999**	8,370	-188 (1709 vs. 1898)	26%	34%	-117	**38%**
1800 to 1999	**2000 to 2199**	10,093	-186 (1908 vs. 2094)	26%	32%	-133	**28%**
2000 to 2199	**2200 to 2399**	10,251	-182 (2109 vs. 2291)	27%	31%	-141	**23%**
2200 to 2399	**2400 to 2599**	9,038	-166 (2311 vs. 2478)	29%	30%	-149	**10%**
2400 to 2599	**2600 to 2799**	2,830	-134 (2509 vs. 2644)	32%	36%	-102	**24%**

As you can see from the data highlighted in yellow, when we look at all these games from 2022-2023 between players from adjacent rating groups (where both players were from the stable age

group and rated below 2000 Elo), the lower-rated players were outrated by almost 200 Elo points on average – more specifically 196, 195, 191, and 188 - translating to an expected score of 25% or 26%. However, the data highlighted in blue tells us that the lower-rated players have actually been scoring 34% consistently, which is what we would expect from a rating difference slightly under 120 Elo points (more specifically, 117). In order to reduce a rating difference of almost 200 Elo points, down to a rating difference of almost 120 Elo points, you have to compress the ratings more closely together by about 40%. And in each of these four slices, we see the 39% or 40% compression recommended in the rightmost blue column.

Thus, we could reasonably conclude that the 1,000-point-range of Elo ratings only represents a range of 600 Elo points in actual playing strength, among players rated below 2000 Elo.

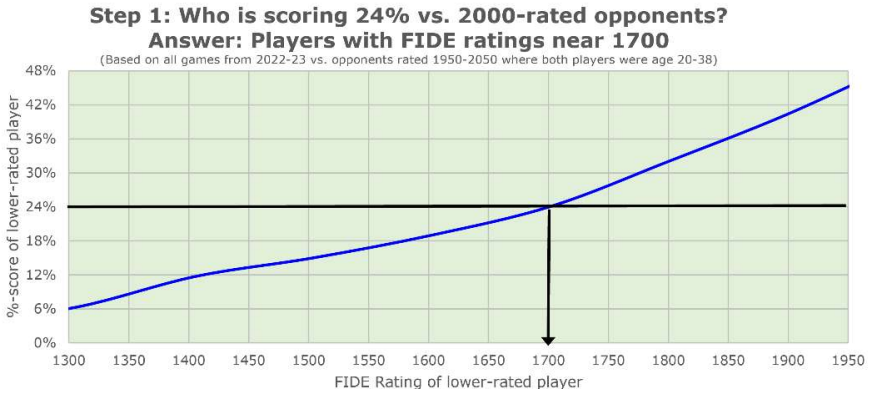
Or we could turn this around and look at it from the other direction. We can ask, how many different "class intervals" of playing strength are there, between 2000 Elo and 1000 Elo? In theory, because that range spans 1,000 Elo points, there ought to be five different class intervals of 200 Elo points each (in terms of playing strength, not just published Elo ratings):



So, if you are 200 Elo points weaker than someone else, you ought to score about 24% against them (that's what the FIDE rating handbook tells us, in the excerpt from 8.1.2 shown above, where the arrow is pointing). In theory, the 1800-rated player should score 24% against the 2000-rated opponent, and the 1600-rated player should score 24% against the 1800-rated opponent, and so on.

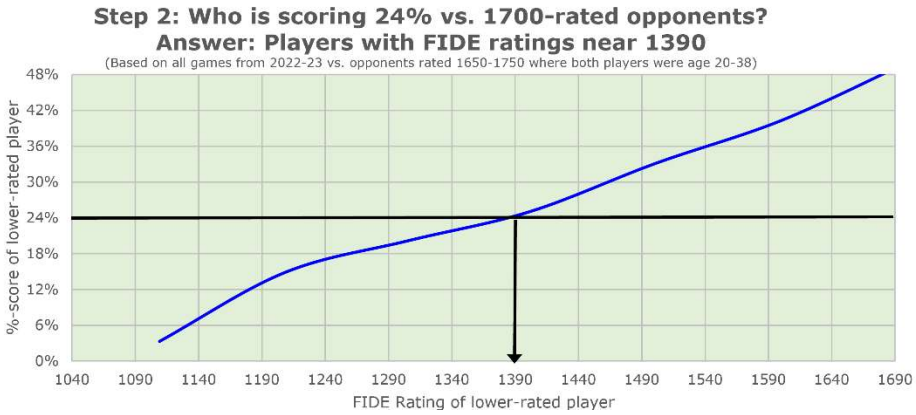
However, what does the data show? This leads to our first question – which players are one strength class below the 2000-rated players? In other words, who is currently scoring 24% in their games against 2000-level opponents?

To get a convincing amount of data, we can look at all games played in 2022-2023 among the stable age group, where the higher-rated player had an Elo rating from 1950 to 2050. Then we can plot the relationship between the rating of the lower-rated player and their percentage score, and thereby determine who exactly is scoring 24% against 2000-level opposition:



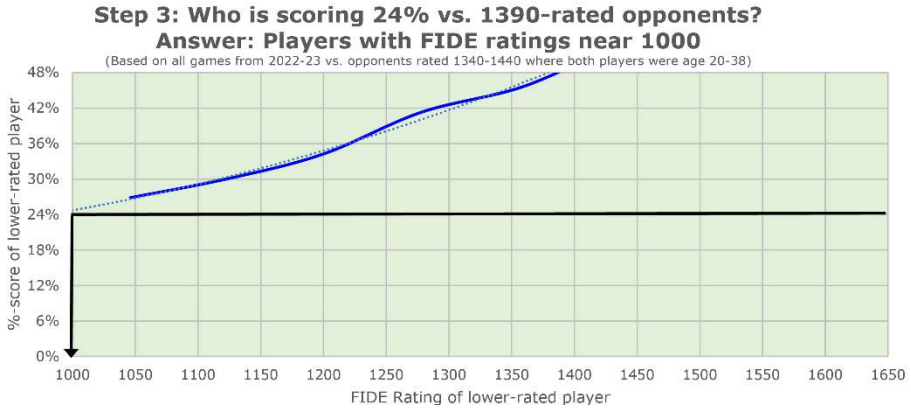
The blue curve tells us what percentage score is being achieved by players of each rating, against opponents rated 1950-2050 Elo. The horizontal black line tells us where the 24% score is. By finding the intersection of the two, and then following the black arrow down to the X-axis number labels, we find that, in fact, it is players rated around 1700 Elo who are actually scoring 24% against 2000-rated opponents.

Okay, so that is one class down from 2000 Elo (in terms of playing strength). Now we are able to ask the next question – who is scoring 24% against 1700-rated opponents?



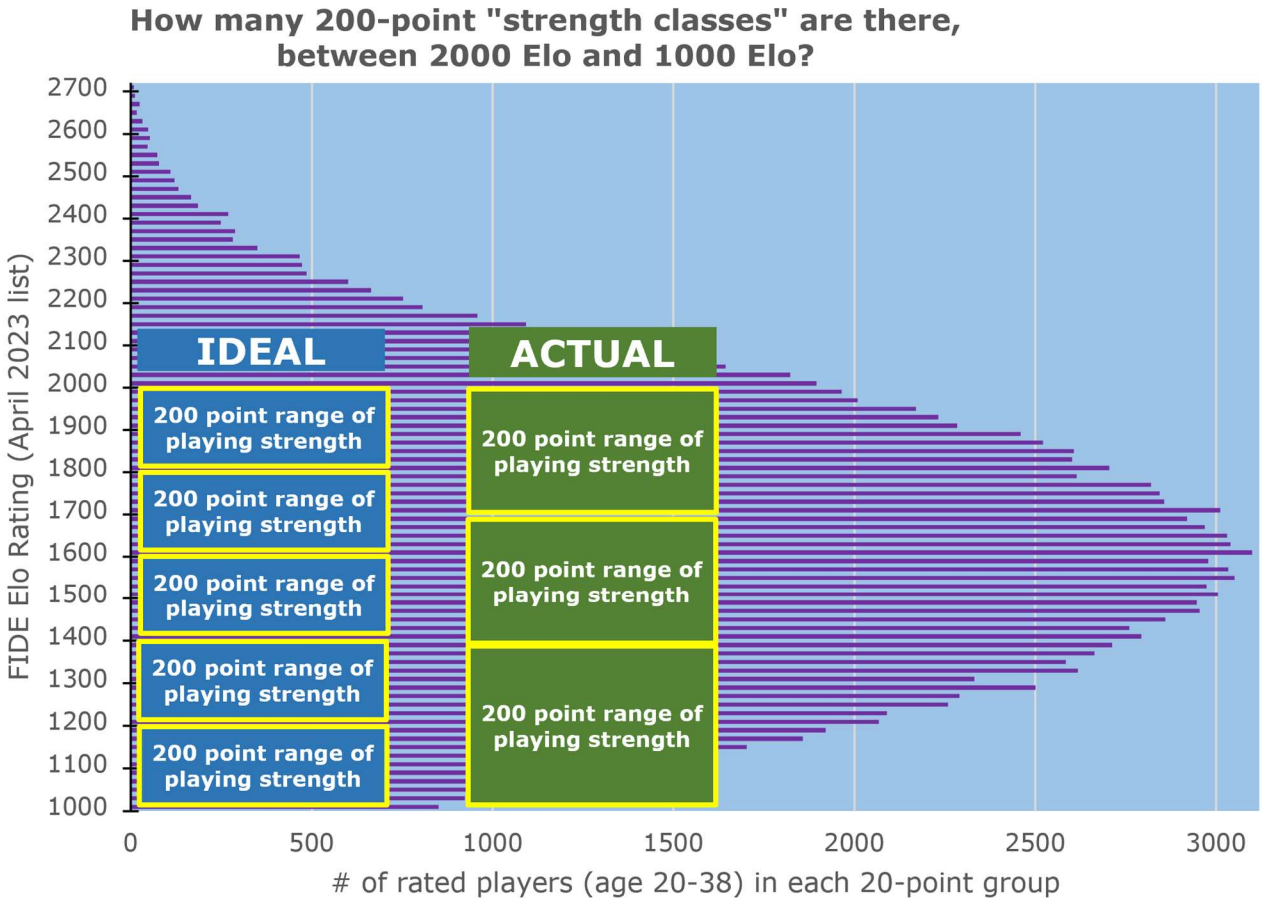
Evidently it is the players rated around 1390 Elo who are scoring 24% against 1700-rated opponents, and thus are empirically about 400 points weaker (in playing strength) than a 2000-rated player, or two classes of playing strength below.

And then finally, who is it that currently scores 24% against 1390-rated opponents? Let's see!



Actually, in this case, there's really nobody who is scoring as poorly as 24% against 1390-rated opponents, but from projecting the results of players rated near the bottom of the rating list, we can estimate that it would be players currently rated almost exactly 1000 Elo, that would be scoring 24% against 1390-rated opponents. So we have our answer.

In fact, there currently seem to be three (not five) class intervals, covering 200 points of playing strength within each interval, spanning the range from players rated 2000 Elo down to players rated 1000 Elo. So again, this points us toward a compression that will bring that range of players to span only 600 Elo points of rating, to match their evident range of playing strength:



Perhaps you are already convinced by now. However, there is one other method I have used for estimating what an ideal compression should look like, which also pushed me toward the simple linear form of the actual formula I have proposed.

We will assign each rated player into a 100-point rating group, based on their rating at the time each game was played. So we will have a group of players rated 1000-99, and another group rated 1100-99, and so on up to 1900-99. And we will lump all the players rated 2000 Elo or higher into a single group that we will call 2000+. Then we will look at all the games played from 01-JAN-2022 through 31-MAR-2023 (that's the latest game data I've imported from FIDE), and consider only the games played between players from different rating groups where both players were in the stable age group from 20-38. If we do all that, these are the results we get:

2022-2023 (using actual FIDE ratings, with 2000+ grouped together)

FIDE Elo Rating Group	# games played (vs. other groups) *	Average FIDE Rating		Expected %-score	Actual %-score	Actual % minus Expected %
		Player	Opponent			
1000-99	1,075	1060	1439	13.51%	23.02%	**+9.5%**
1100-99	2,524	1154	1483	17.76%	25.97%	**+8.2%**
1200-99	4,422	1254	1532	22.79%	30.39%	**+7.6%**
1300-99	6,610	1353	1559	30.27%	35.14%	**+4.9%**
1400-99	8,527	1453	1615	34.69%	38.57%	**+3.9%**
1500-99	10,335	1551	1662	39.25%	41.77%	**+2.5%**
1600-99	12,744	1652	1728	42.46%	43.62%	**+1.2%**
1700-99	15,035	1751	1795	45.41%	45.87%	**+0.5%**
1800-99	16,540	1851	1869	47.76%	47.12%	**-0.6%**
1900-99	17,844	1950	1939	50.44%	49.45%	**-1.0%**
2000+	23,592	2159	1817	84.21%	77.90%	**-6.3%**

* Includes all FIDE-rated games played from 01-JAN-2022 through 31-MAR-2023 where both players were age 20-38 and in different FIDE Elo Rating Groups

The rightmost column is color-coded with a blue/red gradient, where strong overperformance (actual score greater than expected score) is shown in a blue highlight, strong underperformance (actual score less than expected score) is shown in a red highlight, and numbers further away from zero are shown in darker highlight shades. So the strongest overperformance is shown in the darkest blue cell, which is the first group (those players rated 1000-99), who had an expected score of 13.5% but an actual score of 23% in such games, an overperformance by 9.5 percentage points.

We also see the characteristic behavior, since the top cells are blue and the bottom cells are red, where lower-rated players are scoring above their expectation (overperformance gets color-coded blue), and higher-rated players are scoring below their expectation (underperformance gets color-coded red), indicating again that the evident playing strengths of these players are clustered more closely together than their published ratings would claim.

My goal was then to determine an integer Elo amount for each group, that could be added to all the players' Elo ratings in that group, to more accurately reflect their performance and evident playing strength. It turns out to be possible, by an iterative procedure, to find an optimal solution, simultaneously calculating rating adjustments for each of these groups. In that optimal solution, each group's actual results are consistent with their adjusted ratings (and the adjusted ratings of their opponents). I won't bore you with the calculation details, but I will state the one caveat that the calculation is "anchored" by the assumption that players rated 2000+ will not be adjusted, and so everyone is calculated relative to that. Let's skip ahead to the stable solution:

2022-2023 (after many iterations have converged)										
FIDE Elo Rating Group	# games played (vs. other groups) *	Average FIDE Rating (original)		Expected %-score (original)	Adjustment	Average FIDE Rating (adjusted)		Expected %-score (adjusted)	Actual %-score	Actual % minus Expected %
		Player	Opponent			Player	Opponent			
1000-99	1,075	1060	1439	13.5%	**+378**	1438	1670	23.03%	23.02%	**=0.0%**
1100-99	2,524	1154	1483	17.8%	**+336**	1490	1697	25.93%	25.97%	**=0.0%**
1200-99	4,422	1254	1532	22.8%	**+305**	1559	1727	30.36%	30.39%	**=0.0%**
1300-99	6,610	1353	1559	30.3%	**+262**	1615	1744	35.16%	35.14%	**=0.0%**
1400-99	8,527	1453	1615	34.7%	**+226**	1679	1781	38.51%	38.57%	**+0.1%**
1500-99	10,335	1551	1662	39.3%	**+187**	1738	1811	41.78%	41.77%	**=0.0%**
1600-99	12,744	1652	1728	42.5%	**+146**	1798	1856	43.55%	43.62%	**+0.1%**
1700-99	15,035	1751	1795	45.4%	**+111**	1862	1901	45.85%	45.87%	**=0.0%**
1800-99	16,540	1851	1869	47.8%	**+74**	1925	1953	47.14%	47.12%	**=0.0%**
1900-99	17,844	1950	1939	50.4%	**+47**	1997	2005	49.47%	49.45%	**=0.0%**
2000+	23,592	2159	1817	84.2%	**=0**	2159	1908	77.93%	77.90%	**=0.0%**

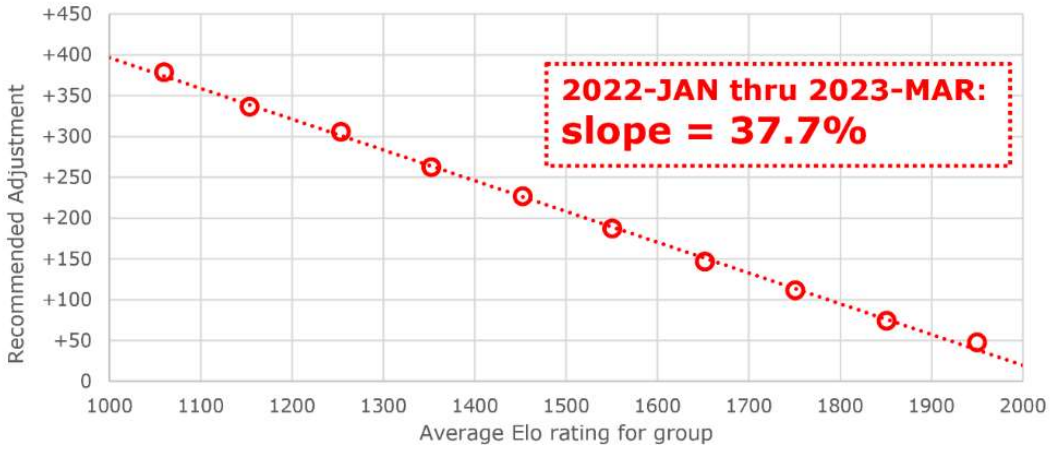
* Includes all FIDE-rated games played from 01-JAN-2022 through 31-MAR-2023 where both players were age 20-38 and in different FIDE Elo Rating Groups

Because all ratings and rating adjustments have to be integers, there is no exact perfect solution, but you can see that the rightmost column is very close to all zeroes now, indicating each group is performing very much in accordance with their own adjusted ratings and the adjusted ratings of their opponents.

Also look at the yellow columns. In the column entitled "Adjustment", we see the set of rating adjustments for each 100-point Elo group that yields expected scores that are as close as possible to the group's actual score. And again we can see that ratings of players near 1000 Elo should be increased by almost 400 Elo points, and the ratings of players near 2000 Elo should barely be increased at all, which again leads to the conclusion that this whole range of players only spans 600 Elo points of evident playing strength.

This suggests we could take all the players rated 1000-99 (the table tells us their average rating was 1060 Elo, on a per-game basis) and add +378 Elo to their ratings, and we could take all the players rated 1100-99 (their average rating was 1154 Elo), and add +336 Elo to their ratings, and so on. If we did this, we would reach a self-consistent set of ratings where each group's performance is consistent with their rating. Of course, it wouldn't really be fair to have a big discontinuity at integer multiples of 100 Elo, so a smoother formula would be needed. Nevertheless, these calculations point us to the general shape/scope of compression needed.

We can plot out these adjustments on an X-Y grid, showing the points at (1060, +378) and (1154, +336) and so on, and in fact all the points between 1000 Elo and 2000 Elo line up quite nicely into a straight line:



It seems that the relationship between rating and recommended rating adjustment is quite linear, at least for players rated 2000 Elo or below, and this is why I was able to propose such a simple formula for the compression of ratings. If we tried to extend this to higher-rated players, to account for variable levels of deflation even above 2000 Elo, the adjustment would no longer be linear, and it would start to get complex. And also, as I said in the original proposal, I wanted to stay away from the title system and ratings of professional players, as much as possible.

It's also worth noting that the slope of the above trendline is only 37.7%, not quite 40%. However, this data only covers a 15-month timespan from 2022-JAN through 2023-MAR, so on average it gives a snapshot of where things were about a year ago. And the effect is clearly worsening with the passage of time. For example, if we were to perform the same analysis except using games data from 2019-2021 instead, we would see a suggested slope of only 32%.

Thus I feel quite comfortable in asserting once again that **we should apply a linear compression formula to the ratings of players rated 1000-1999 Elo, so that their resultant span is only 600 Elo rating points rather than 1,000 Elo rating points.** Players rated 2000 Elo or higher would not have their ratings directly affected by the compression, and so the 600 points would range from 2000 Elo down to 1400 Elo. We could compactly describe this as a "compression to 1400 Elo ".

Also note that we could have done that last analysis in rating groups of 50 Elo points wide, or 200 Elo points wide, and the overall recommendation of a linear correction would not significantly change, in either its linearity or its overall magnitude. In fact, the numbers don't change a huge amount even if we included games played by those improving juniors and declining seniors in our dataset after all, though it would give a slope more like 42% (instead of 38%) in that trendline from the last graphic above. So, excluding the juniors and seniors does yield a lesser estimate for the overall stretchedness of ratings, but not by a huge amount.

Appendix #2: Has the deflation reached elite players yet?

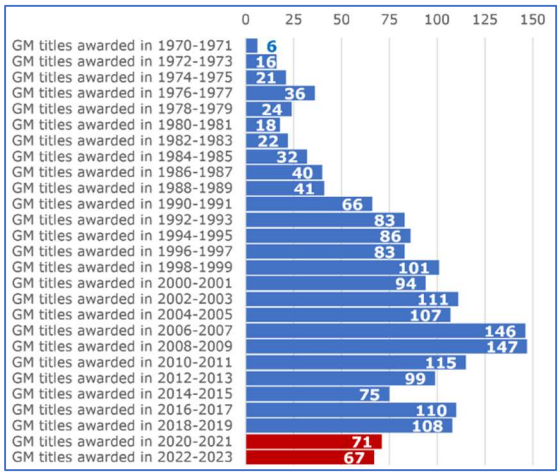
One of the strongest responses from the chess community - against the proposal - was from **GM Michal Krasenkow**, including this statement:

> I am strongly against the proposed changes of the rating system. The main thing is not to cause harm. At the moment I don't see any considerable deflation: the number of players scoring GM norms and obtaining GM titles every year has not changed much in recent years.

This is certainly a reasonable position to take, if you are not yet convinced that anything is wrong. Of course, grandmasters would be the last ones affected by any deflationary effects that originate from the initial ratings given to new players, because those effects would propagate slowly through the rating pool and take a long time to reach the very top. In fact it would take an extra-long time, because GM's are quite insular with regard to which opponents they face.

For a top group that should be growing (at least modestly) in size, the first symptoms of deflation would be that the group's size stagnates and fails to increase. And then it would start shrinking. That's where I think we are now, for even top Elo groups.

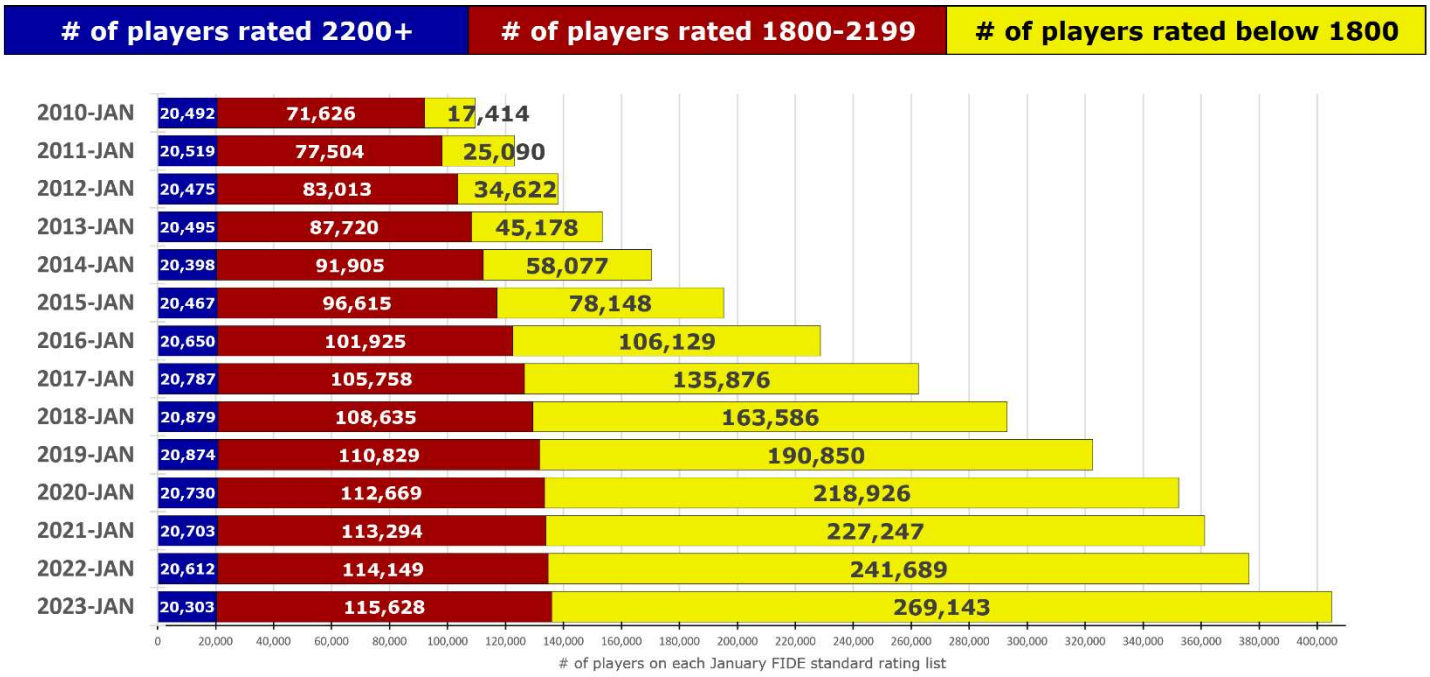
I had never tried counting up the number of GM titles assigned each year until seeing GM Krasenkow's comments, so I did some research and assembled this graphic:



The last two bars, spanning 2020-2023, are marked in red because they are not very meaningful, the 2020-2021 bar being affected by COVID shutdowns, and the 2022-2023 bar being incomplete since we aren't done with 2023 yet. In any event, GM Krasenkow is correct that we haven't seen much change in the number of players being awarded GM titles. Indeed, we had 218 GM titles awarded in 2002-2005 (54.5 per year), and 436 GM titles awarded in 2008-2015 (54.5 per year), and 218 GM titles awarded in 2016-2019 (54.5 per year). Nevertheless, it is a bit concerning when an increasingly-large base of strong players is not leading to more grandmasters. As an example, we had about 44,000 players rated 2000-2399 in January 2004 and about 66,000 players rated 2000-2399 in January 2019, but no increase in the rate of grandmaster titles being awarded.

Let's look at some more data, starting with the big picture. In the past dozen years, FIDE has seen tremendous growth in the number of players rated below 1800, moderate growth in the number of players rated 1800-2199, and a long period of equilibrium, followed by actual shrinkage recently, in the number of master-level players (those rated 2200+). My primary explanation for this shrinkage is that ratings are dropping due to strong deflationary effects finally impacting the master level.

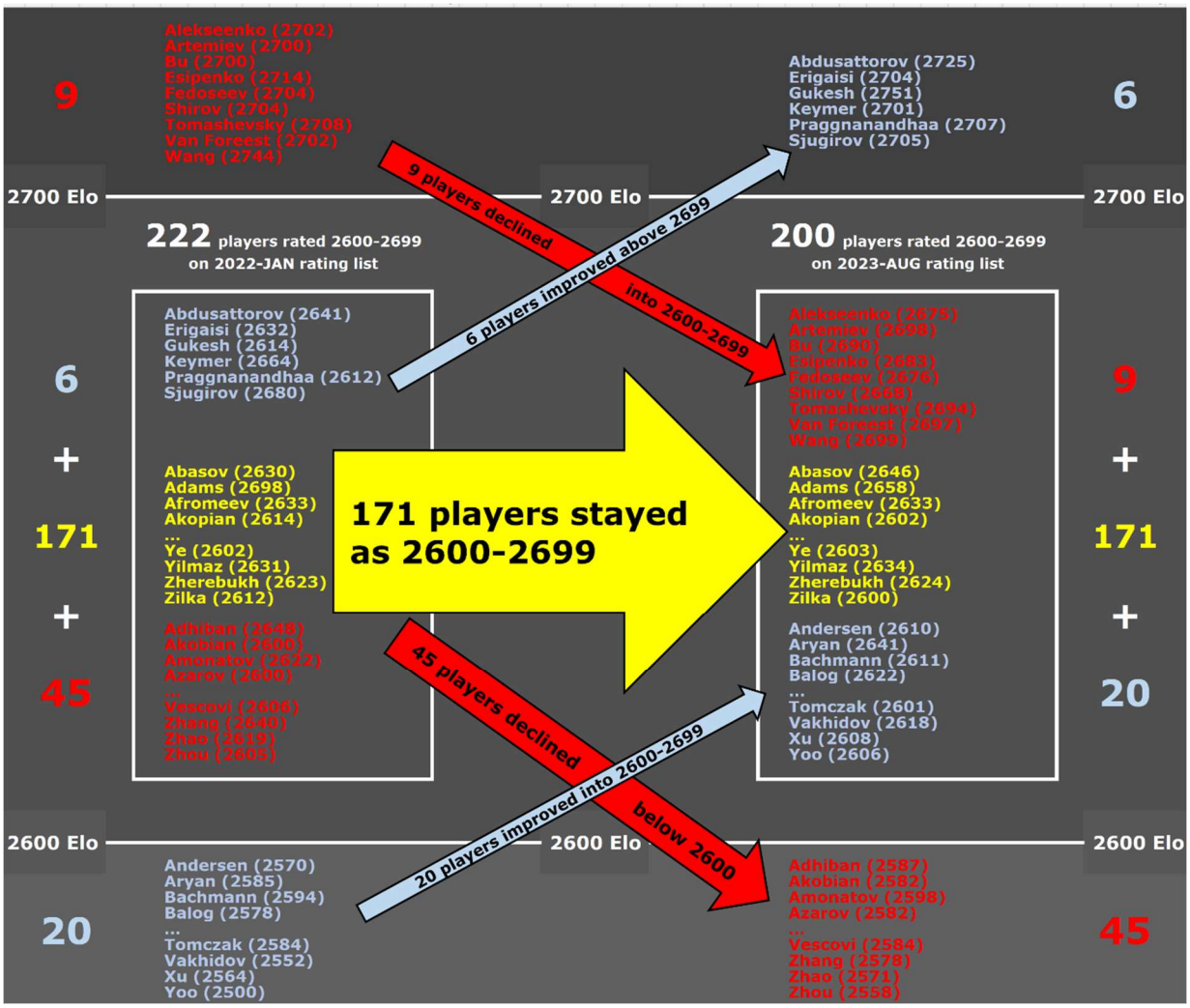
We can see the evidence of these different levels of growth/shrinkage, by counting up the number of players on each January rating list going back to 2010. We will count a player in the blue group if they are rated 2200+, in the red group if they are rated 1800-2199, and the yellow group if they are rated below 1800. This includes all players, whether active or inactive.



Of particular concern is the bottom left corner of that picture, where we see that despite all this growth, the number of players rated 2200+ has actually declined every year since 2018, and the decline seems to be accelerating. For example, between January 2022 and January 2023, their ranks dropped by slightly more than 300 players, to an overall size (20,303 players) that has not been this low in more than a dozen years. We will return to that "300 players" number in a few pages.

Now let's take a break from the big numbers, and instead look at individual players. We shall first focus on the Elo range from 2600-2699, and see which top players moved into this group, or out of this group, during a recent 20-month stretch from the January 2022 list until the August 2023 list. For this analysis, I'm not looking at the rating lists in between; I am just looking (for each player) at whether they were in the 2600-2699 group in January 2022, and also whether they were in the 2600-2699 group in August 2023.

In the next couple of graphics, a player's name will be color-coded blue if they have improved into a different 100-point group (crossing above a boundary like 2600 Elo or 2700 Elo), or they will be color-coded red if they have declined into a different 100-point group (dropping below a boundary like 2600 Elo or 2500 Elo).



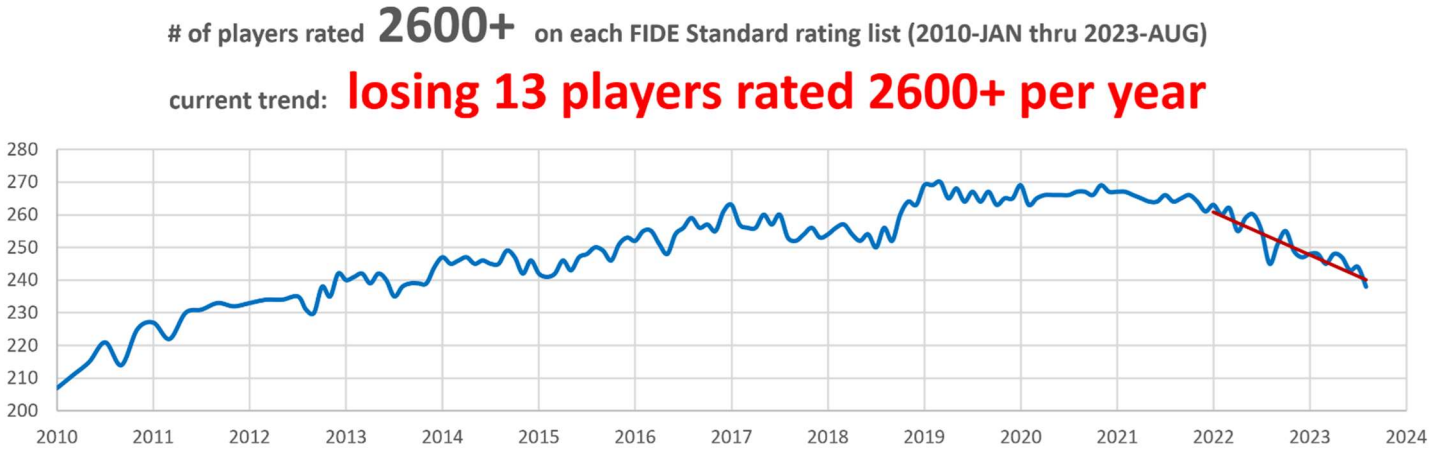
On the left we see player ratings from January 2022, and on the right we see player ratings from August 2023. Near the top, you can see a horizontal white line representing the 2700 Elo level. The upward-pointing blue arrow that crosses the white line shows the movement of players who improved above 2700 between these two lists. The downward-pointing red arrow shows the movement of players who dropped below 2700 during the same time period.

From comparing the ratings of 2022-JAN against 2023-AUG, we can see that nine players (color-coded in red) dropped below that 2700 boundary: Alekseenko, Artemiev, Bu, Esipenko, Fedoseev, Shirov, Tomashevsky, Van Foreest, and Wang. And six players (color-coded in blue) climbed above that 2700 boundary: Abdusattorov, Erigaisi, Gukesh, Keymer, Praggnanandhaa, and Sjugirov. So at first glance, you might think that the group rated 2600-2699 had grown in size over that time, since nine players dropped into it and only six players climbed out of it.

However, we must also consider the lower boundary. Notice the other horizontal white line, near the bottom of the picture, representing the 2600 Elo level. The arrows and listings tell us that 45 red players dropped below it, out of the 2600-2699 group (Adhiban, Akobian, …, Zhou) and only 20 blue players climbed above it into the 2600-2699 group (Andersen, Aryan, …, Yoo). And of course, let's not overlook the 171 players (including inactive players) who remained in the 2600-2699 group on both rating lists, although of course they didn't cause the group to get any bigger or smaller. Those players are shown in yellow.

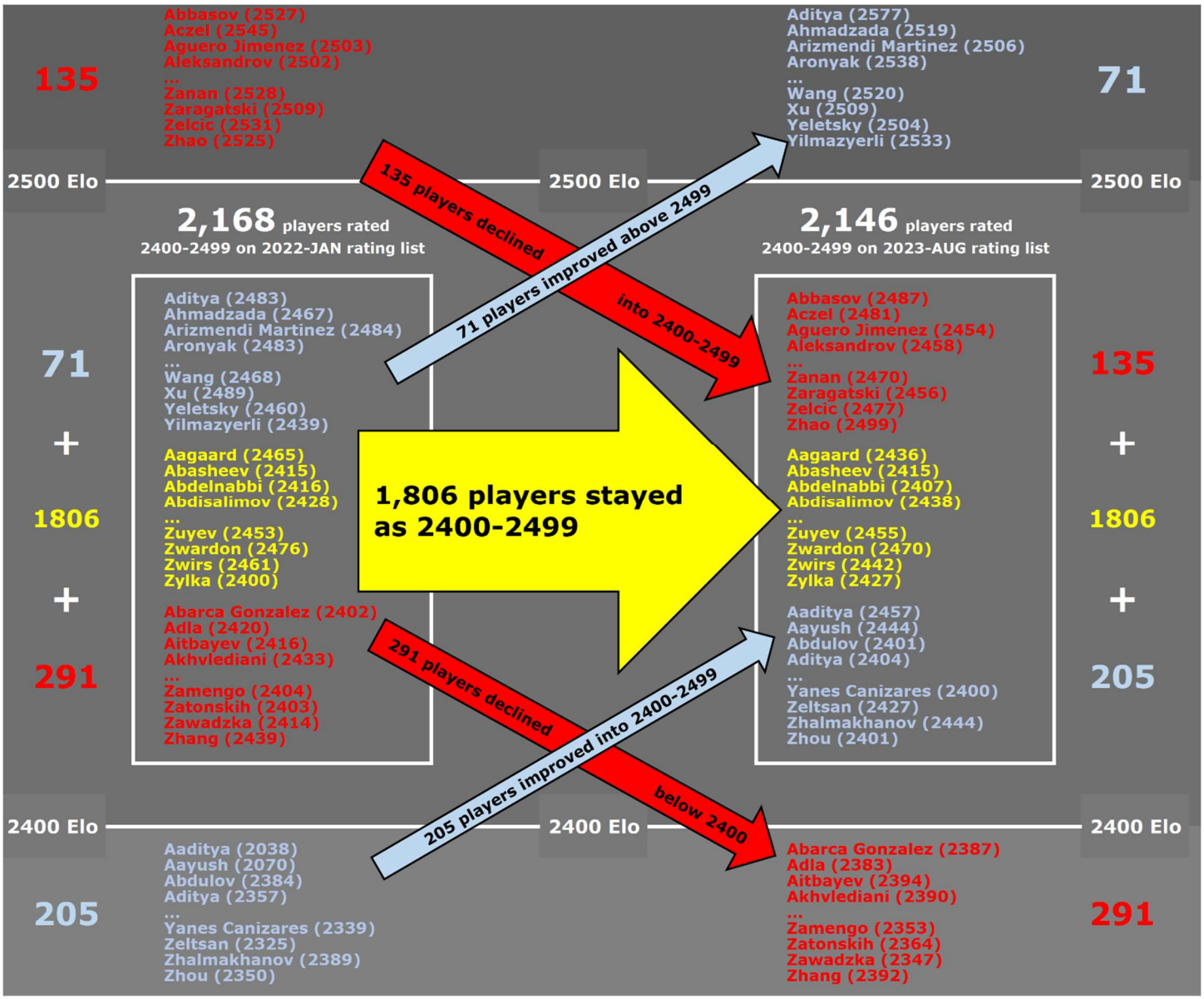
So even though the net gain to the group was positive based on crossings of the 2700 line, this "growth" was more than outweighed by what happened at the 2600 line. And overall, the number of players rated 2600-2699 dropped from 222 to 200 players during these 20 months.

We are losing approximately one player from the 2600+ group each month, and this trend has been steadily continuing for more than two years. This is easily seen by simply counting how many players were rated 2600+ on each recent rating list. We can draw a clear trendline for the past 20 months (shown below as a slanting red line over on the right, for the years 2022-23)



On the November 2020 rating list, there were 269 players rated 2600+, and I am quite confident that we will never again see that many on one list, unless significant steps are taken to counteract this ongoing rating deflation. By the time of the October 2023 rating list, 35 months later, there were 35 fewer players rated 2600+ (a total of 234 players). We have to go all the way back to 2012 before we find a rating list with so few players rated 2600+.

We see this same effect happening for several adjacent Elo groups at the master level, where in each group, there is a net increase at the upper boundary, which is then more than outweighed by players dropping below the lower boundary. As an example, let's step down 200 points and consider the 2400-2499 group instead:



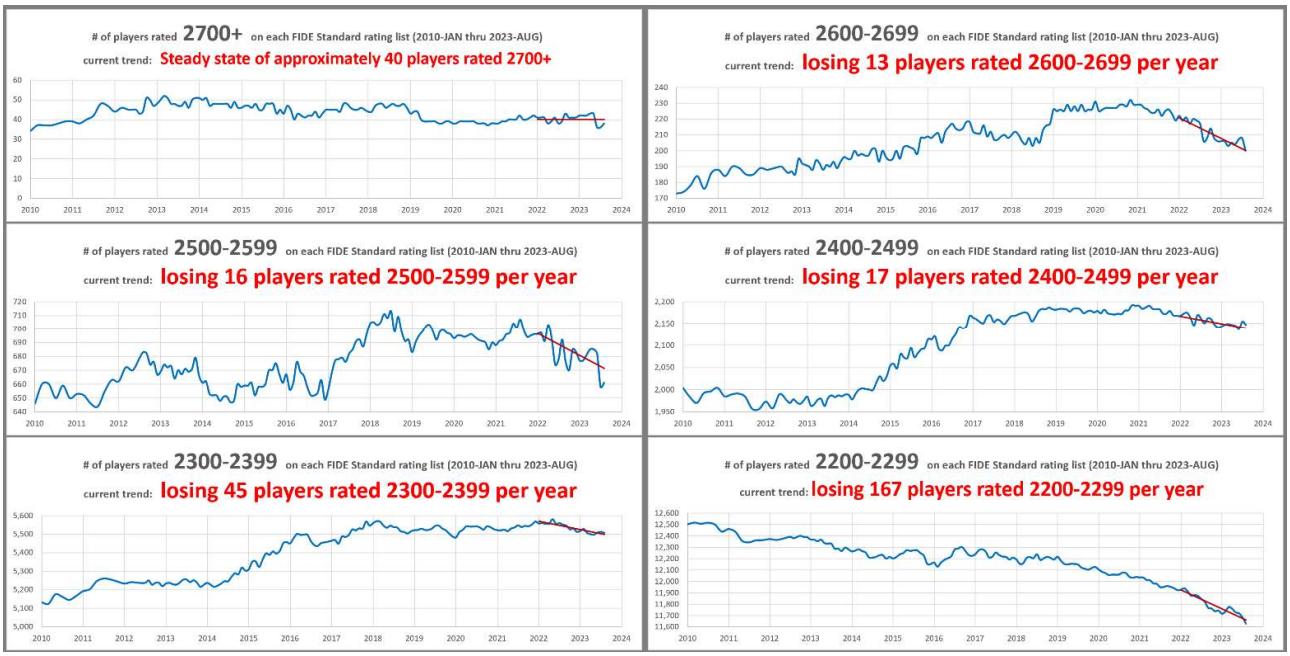
Again there is some amount of "growth" in this group of players rated 2400-2499, in the sense that there were more red players falling into the 2400-2499 group from above the 2500 Elo line, than there were blue players improving above the 2500 Elo line (135 red players entering the 2400-2499 group from above, and 71 blue players leaving the group by improving above 2500).

But this is more than outweighed by what's happening at the bottom of the group, where there were 291 red players declining out of the 2400-2499 group by falling below 2400 Elo, with only 205 blue players improving into the 2400-2499 group by increasing to 2400+ Elo. So we again see the numbers driven by players dropping below boundary lines, and overall a group that has shrunk in size (from 2,168 players down to 2,146 players).

Indeed, this effect is happening for several Elo groups in a row. We've already seen that there's more players dropping below 2600 than players improving above 2600 (a net loss of 13 players per year). Similar things are happening to all the Elo groups above 1900. They are all shrinking, and there's nowhere those exiting players can be going to, other than a mass exodus downward to lower-rated Elo groups.

Rather than looking at direct transitions and specific player listings anymore, let's take a more general look and just show, for each 100-point Elo group, how its size has grown or shrunk over time. Also, in each case I have drawn a (red) trendline for the data since January 2022, showing what the current trend is for each group, during those 20 months since then.

Let's look at the first several rating groups, starting from the most elite group, where it actually seems to be somewhat stable - for now! - with about 40 players each month in the 2700+ "Super GM" group. But several adjacent groups below it are all shrinking.



Each month, there are some players who improve sufficiently enough that their rating crosses over the 2200 Elo boundary, plus a very rare player getting an initial rating in the 2200's. And each month, there are other players who decline sufficiently enough that their rating falls below the 2200 Elo boundary. But this is not a state of steady equilibrium, because the second group is generally bigger; there's about 25 more players each month dropping below 2200 Elo than there are players climbing above 2200 Elo (or getting initial ratings above 2200 Elo). So overall we are losing almost 300 players a year out of the group of players rated 2200+ Elo, which is consistent with that blue-red-yellow bar graph we saw back on page 17, indicating about 300 fewer players rated 2200+ Elo in January 2023, than there were just a year before in January 2022.

In my opinion, this is very strong evidence that deflation has indeed reached the master levels. And it will continue to get worse unless we address it with strong countermeasures such as my proposed Compression to 1400 Elo. By adding a large amount of Elo rating points into the rating pool, we will be easing the deflationary pressure that is being exerted from below upon the master-level players. Instead of shrinkage in the size of the top groups, after some time we would actually see some modest growth, commensurate with the growing population of strong players in the general rating pool. This "modest growth" is addressed in the next appendix.

Appendix #3: Should we expect a compression to cause inflation?

Much of my analysis is concerned with what has happened in the past. However, a more pressing question is: what will happen in the future, if we make these proposed changes?

Once again, some of the feedback from **GM Michal Krasenkow** is of particular interest to this question of what would happen after a compression. He is very concerned about inflation:

> A one-time increase of ratings of numerous players will lead to a real explosion of inflation, which is much worse than any kind of deflation. High ratings and titles will become meaningless.

Of course, we do not know what conditions will apply in the future chess world and the future FIDE pool of rated players, and so it is challenging to make predictions and projections. I certainly do understand and appreciate GM Krasenkow's concern: perhaps introducing a strong inflationary factor would indeed lead to lots of players receiving undeserved high ratings and achieving undeserved titles? A key question, of course, is what is meant by "undeserved". And we must make a measured assessment of how much we actually do expect ratings to change.

My contention is that players rated below 2000 Elo do not currently "deserve" to have ratings so far below 2000 Elo. I think the evidence of their performance, as a group, is sufficient to prove that they are actually stronger than their rating says, and this is the purpose of the compression.

Let's take one example. Most FIDE title eligibility will not be directly affected by the proposed "Compression to 1400 Elo", because nobody will be directly pushed above 2000 Elo by the compression. However, there will likely be some cases where the compression does lead directly to the award of a Woman FIDE Master (WFM) or Woman Candidate Master (WCM) title.

Although a Woman FIDE Master (WFM) title is normally dependent upon the player crossing over the 2100 Elo boundary, you can still qualify for the direct WFM title if you have medaled in certain events, or scored 65% in an Olympiad, as long as you have reached a 1900 Elo rating at some point in your career. And certainly some players would be pushed above the 1900 Elo boundary by a compression that affects all players rated below 2000 Elo.

From analyzing rating lists and titles, it appears to me that there are approximately 750 female players for whom all three of the following criteria are true:

(1) currently rated between 1833 Elo and 1899 Elo; and
(2) not already qualified for the WFM title; and
(3) have never reached a 1900 Elo rating in their career.

Because of their current rating in the 1833-1899 Elo range, all of these women would be increased into the 1900-1940 Elo range by the proposed full compression. My estimate from that same analysis is that 3-4% of these players have indeed achieved the necessary medal or Olympiad score. So once their rating crosses 1900 Elo, they would be directly awarded the WFM title. Thus I believe that approximately 22-30 female players (which is 3-4% of those ~750 players) would directly become WFM's if a "Compression to 1400 Elo" were to occur.

Would that be "undeserved"? My contention is that such a player has already demonstrated that their actual playing strength is indeed 1900+ Elo, thanks to their demonstrated ability to reach a FIDE rating in the range of 1833-1899 Elo in the current chess world with the current level of Elo deflation. As evidence, if we take those 750 players and look at how they have performed during 2022-2023 against opponents rated 2000+ Elo, we find that as a group, they have a performance rating of 1915. That, to me, means that they would indeed deserve their new ratings in the range of 1900-1940 Elo. And so they would also deserve their direct title. We would not be giving undeserved gifts of Elo rating bonuses or titles to all these players; rather, we would be attempting to fix past injustices that were caused by years of rating deflation.

Similarly, there might be a concern that players could achieve undeserved titles, even IM or GM titles, by getting high scores against opponents who don't deserve their high ratings. But again, the purpose of the compression is to relocate player ratings into the Elo range that they seem to actually deserve, based on the evidence of results by all players of comparable strength. So I feel quite confident that the new ratings players would end up with, or any titles directly (or indirectly) attributable to the Compression to 1400 Elo, would indeed be "deserved".

Nevertheless, there is still the concern raised by GM Krasenkow and others, that this large infusion of rating points into the rating system might lead to an *"explosion of inflation"*. How can we respond to this concern?

My answer is this: we can try to project what a realistic outcome would be, by applying the proposed changes within a realistic simulation. We could either use some sort of fake data, or we could use real-world data (taken from the past). Because of the complexity of the chess world and chess events, I think that any sort of hypothetical events or game results using fake game outcomes will not provide a very realistic picture. That is why I prefer simulations using actual historical events and game results. In the simulation, we can replay the rating calculation through to the present, under the assumption that a compression and the associated calculation improvements had actually been rolled out in the past. For example, if we pretend that the calculation happened in 2017, we can see what would subsequently happen to player ratings if they subsequently had all the same game results that actually occurred in real life since 2017.

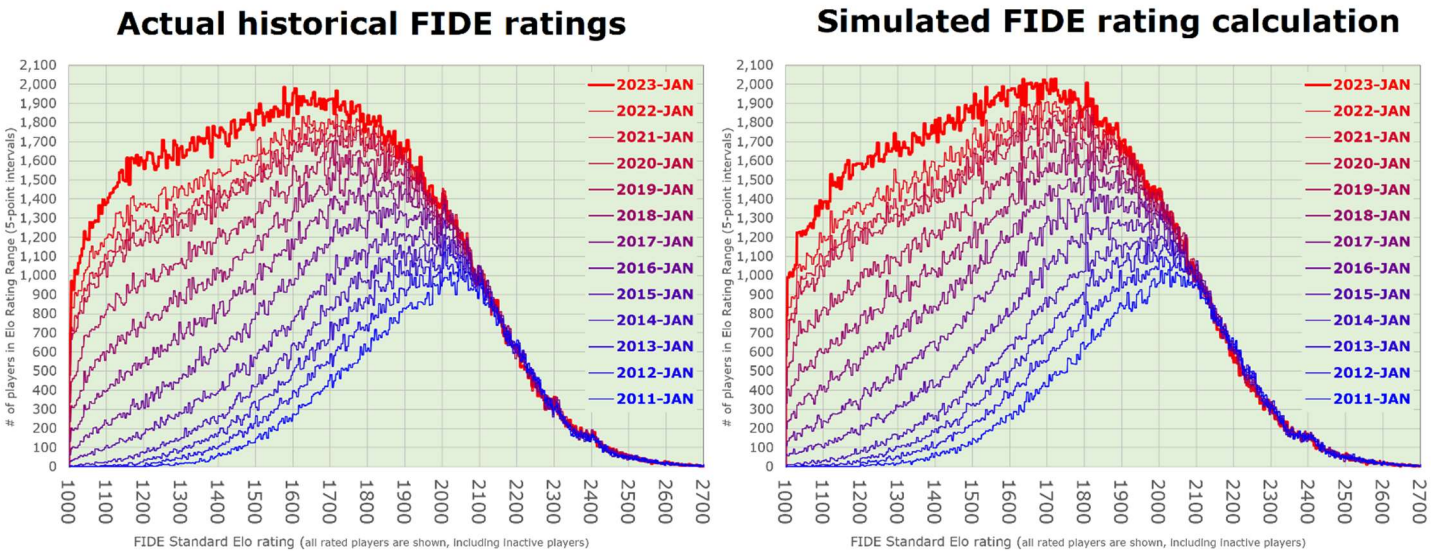
Of course, the simulation is complicated by the fact that recent years included a pandemic, drastically affecting the chess world and the behavior of chess players. In our simulation as well, many players would have an extended pause in over-the-board play in 2020-2021 (due to COVID), and ratings would be affected by that. That's why I am leaning toward the simulations based on a compression happening in early 2022, as the most useful indication of what would happen if we were to do a compression soon (such as my proposed date of January 2024).

A major focus of my project with FIDE this year was to develop a rating simulator, one that could recreate the entirety of the FIDE rating calculations from 2008 up to the present, incorporating all of the various regulation changes over the years (K-factors, minimum Elo rating, initial rating formula, frequency of rating updates, etc.) Eventually this could be used in powerful ways to simulate the behavior of the rating system under alternative regulations such as I am proposing with the compression and calculation improvements. But before getting that far, a key test of this simulator would be to see if it could replicate the actual historical rating calculations that happened in real life. That way we could be confident that we have captured the current system behavior in our simulation.

I did not skip ahead to experimenting with alternative regulations until I was first satisfied with the correspondence between the actual historical rating calculations, and my simulation of them.

I developed a technique to show the distribution of the FIDE rating list for several years all in the same image, so that we can see how many players were rated 1000-1004 Elo, how many were rated 1005-1009 Elo, etc., all the way up to the top, as a line graph. The distribution of the rating list from January 2011 gets shown in a blue line, the distribution of the rating list from January 2023 gets shown in a thick red line, and January rating lists for interim years are shown in appropriate shades of purple lines, to span the whole range from 2011's blue to 2023's red.

Since we have a nice way to view 12 years' worth of rating list distributions, we can do a side-by-side comparison of the real historical ratings, against the simulated historical ratings from the rating simulator, and see if they are roughly equivalent. It took a long time and a lot of programming stops and starts to reach this point, but eventually I reached the point where the line graphs were extremely similar, and I could feel confident that I had captured almost every aspect of the rating calculation with great accuracy. In the following pictures, the left side shows the actual distributions of historical ratings, and the right side shows my simulations of ratings:



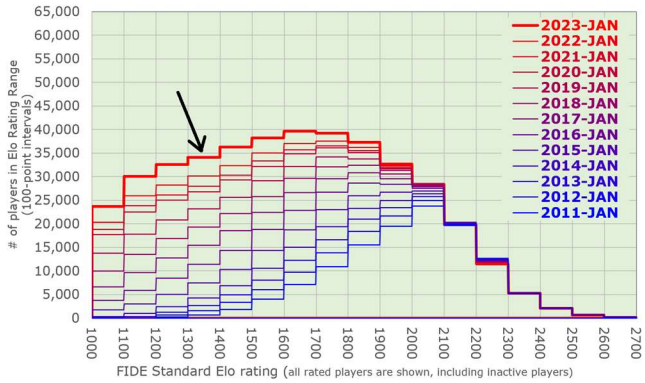
Please note that the simulation used no FIDE rating lists since 2008. It started from the January 2008 FIDE rating list, considered all of the historical game results that were reported each rating period, and tried to simulate the entirety of the monthly rating list calculations all the way up to the present, in 2023.

We can now use the one on the right as a sort of baseline simulation. We can try different K-factors, different formulas for the initial ratings, etc., and see how the distributions of ratings would be different. In each case, we can compare the revised simulation against the baseline simulation, and thereby improve our understanding of the impact of the changes under consideration. We could even try more significant changes, such as a compression.

In those above graphs, the number of players at each 5-point Elo interval is shown (1000-1004 Elo, 1005-1009 Elo, etc., all the way up to 2695-2699 Elo). So the "curves" are in fact very jagged. Although this does reveal some interesting features, I eventually decided that it was too detailed, and in some cases became hard to interpret, especially where there is no growth.

So instead, I settled upon a view that used 100-point-intervals, accompanied by a data chart listing each year how many players there were in each 100-point-interval. So this can be thought of as the final baseline, which experimental approaches can then be compared against:

(Baseline) Simulation using actual historical regulations



	1000-1099	1100-1199	1200-1299	1300-1399	1400-1499	1500-1599	1600-1699	1700-1799	1800-1899	1900-1999	2000-2099	2100-2199	2200-2299	2300-2399	2400-2499	2500-2599	2600-2699	2700-2799	2800-2899	Total
2023-JAN	23,716	30,025	32,556	34,099	36,276	38,215	39,652	39,197	37,291	32,686	28,380	19,816	11,485	5,241	2,067	634	179	33	3	411,551
2022-JAN	20,330	25,924	28,192	30,158	32,326	35,047	37,047	37,549	36,228	32,218	28,515	19,846	11,717	5,287	2,086	649	202	30	2	383,353
2021-JAN	18,810	23,809	26,095	28,000	30,315	33,309	35,706	36,562	35,556	31,914	28,510	19,996	11,811	5,232	2,123	652	198	29	3	368,630
2020-JAN	17,721	22,498	24,988	26,780	29,340	32,188	34,837	36,104	35,081	31,677	28,557	20,018	11,865	5,190	2,129	655	196	31	4	359,859
2019-JAN	13,764	17,792	20,834	23,178	25,634	29,170	32,173	34,174	33,747	31,330	28,336	20,116	12,012	5,213	2,135	661	196	34	5	330,504
2018-JAN	9,951	13,478	16,844	19,297	22,142	25,815	29,649	32,088	32,441	30,546	28,149	20,199	12,013	5,304	2,119	669	199	39	4	300,946
2017-JAN	6,587	9,675	12,745	15,487	18,516	22,441	26,581	29,546	30,828	29,575	27,902	20,250	12,101	5,259	2,152	646	216	36	5	270,548
2016-JAN	3,757	5,879	8,476	11,451	14,371	18,785	22,868	26,718	28,661	28,422	27,510	20,299	12,083	5,324	2,125	679	206	41	3	237,658
2015-JAN	1,771	3,048	5,041	7,498	10,342	14,328	18,754	23,003	25,965	26,742	26,928	20,286	12,251	5,248	2,134	672	200	43	4	204,258
2014-JAN	355	983	2,398	4,264	6,837	10,561	15,036	19,397	23,274	24,952	26,358	20,106	12,384	5,271	2,086	683	199	50	3	175,197
2013-JAN	81	306	1,275	2,700	4,931	8,082	12,293	16,642	20,964	23,399	25,791	19,958	12,515	5,285	2,089	686	197	45	4	157,243
2012-JAN	0	0	645	1,605	3,317	6,006	9,707	13,850	18,355	21,657	24,915	19,889	12,536	5,299	2,078	684	191	42	4	140,780
2011-JAN	0	0	217	672	1,864	4,037	7,132	10,912	15,573	19,499	23,778	19,657	12,655	5,209	2,116	673	193	36	4	124,227

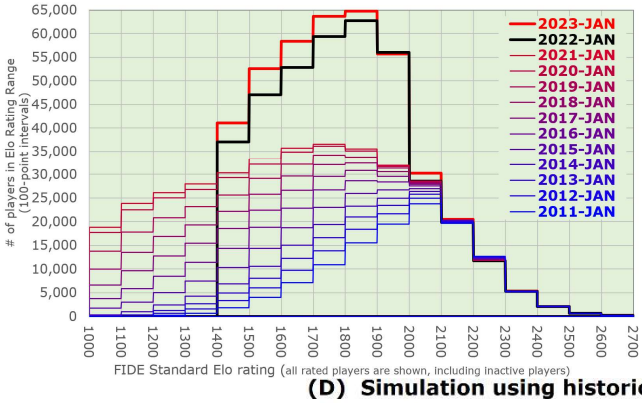
As a very simple example of how to read this, look at the highest red line that's drawn between 1300 Elo and 1400 Elo (it has a black arrow pointing at it). It is a horizontal line, not quite as high vertically as the 35,000 gridline. That tells us that on that red rating list (2023-JAN), there were not quite 35,000 players rated from 1300 to 1399. If you want more precise numbers, you can look over in the data tables where the purple arrow is. Under 1300-1399, for 2023-JAN, we see that indeed, there were 34,099 such players, matching what was shown graphically.

Okay, we can finally now start talking about the simulations. First of all, what can we expect to happen in the first year or so after a Compression to 1200 Elo or a Compression to 1400 Elo?

In reality, of course, such a compression would happen no earlier than January 2024, but in my simulations, I tried to see what would have happened to the rating system if the compression had happened in the past. Scenarios (A) and (B) were for the compression happening in January of 2017, which ought to give us a good sense of what ratings might look like five or six years after a compression, but let's start with scenarios (C) and (D). These were for the compression happening in January of 2022. This is after the main COVID shutdowns had happened, and still gives us 15 months of actual game results to see what would have happened to the simulated ratings after the compression. And it would be a very similar rating pool to what we have in real life, since it was only a couple of years ago, so the simulation ought to be quite realistic.

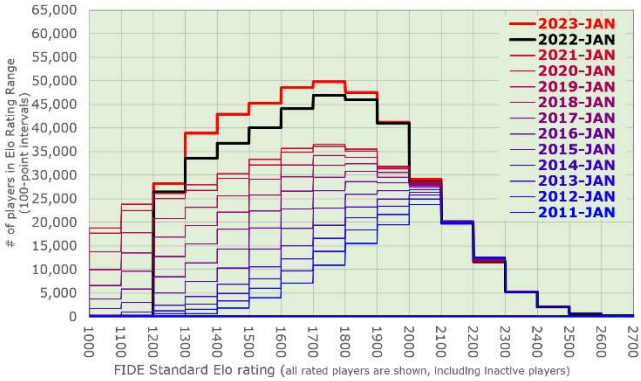
Here is the comparison of how the rating distributions would differ among these three scenarios:

(C) Simulation using historical regulations through the end of 2021, and then 40% compression with 1400 Elo minimum (and calculation improvements) on 2022-JAN



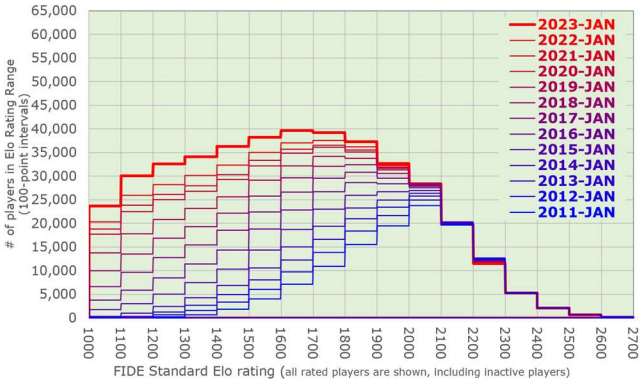
	1000-1099	1100-1199	1200-1299	1300-1399	1400-1499	1500-1599	1600-1699	1700-1799	1800-1899	1900-1999	2000-2099	2100-2199	2200-2299	2300-2399	2400-2499	2500-2599	2600-2699	2700-2799	2800-2899	Total
2023-JAN	0	0	0	0	41,086	52,602	58,354	63,654	64,796	55,703	30,254	20,487	11,769	5,331	2,093	638	183	33	3	406,986
2022-JAN	0	0	0	0	37,039	47,028	52,863	59,347	62,711	56,031	28,515	19,846	11,717	5,287	2,086	649	202	30	2	383,353
2021-JAN	18,810	23,809	26,095	28,000	30,315	33,309	35,706	36,562	35,556	31,914	28,510	19,996	11,811	5,232	2,123	652	198	29	3	368,630
2020-JAN	17,721	22,498	24,988	26,780	29,340	32,188	34,837	36,104	35,081	31,677	28,557	20,018	11,865	5,190	2,129	655	196	31	4	359,859
2019-JAN	13,764	17,792	20,834	23,178	25,634	29,170	32,173	34,174	33,747	31,330	28,336	20,116	12,012	5,213	2,135	661	196	34	5	330,504
2018-JAN	9,951	13,478	16,844	19,297	22,142	25,815	29,649	32,088	32,441	30,546	28,149	20,199	12,013	5,304	2,119	669	199	39	4	300,946
2017-JAN	6,587	9,675	12,745	15,487	18,516	22,441	26,581	29,546	30,828	29,575	27,902	20,250	12,101	5,259	2,152	646	216	36	5	270,548
2016-JAN	3,757	5,879	8,476	11,451	14,371	18,785	22,868	26,718	28,661	28,422	27,510	20,299	12,083	5,324	2,125	679	206	41	3	237,658
2015-JAN	1,771	3,048	5,041	7,498	10,342	14,328	18,754	23,003	25,965	26,742	26,928	20,286	12,251	5,248	2,134	672	200	43	4	204,258
2014-JAN	355	983	2,398	4,264	6,837	10,561	15,036	19,397	23,274	24,952	26,358	20,106	12,384	5,271	2,086	683	199	50	3	175,197
2013-JAN	81	306	1,275	2,700	4,931	8,082	12,293	16,642	20,964	23,399	25,791	19,958	12,515	5,285	2,089	686	197	45	4	157,243
2012-JAN	0	0	645	1,605	3,317	6,006	9,707	13,850	18,355	21,657	24,915	19,889	12,536	5,299	2,078	684	191	42	4	140,780
2011-JAN	0	0	217	672	1,864	4,037	7,132	10,912	15,573	19,499	23,778	19,657	12,655	5,209	2,116	673	193	36	4	124,227

(D) Simulation using historical regulations through the end of 2021, and then 20% compression with 1200 Elo minimum (and calculation improvements) on 2022-JAN



	1000-1099	1100-1199	1200-1299	1300-1399	1400-1499	1500-1599	1600-1699	1700-1799	1800-1899	1900-1999	2000-2099	2100-2199	2200-2299	2300-2399	2400-2499	2500-2599	2600-2699	2700-2799	2800-2899	Total
2023-JAN	0	0	28,223	38,906	42,892	45,298	48,568	49,773	47,505	41,242	29,112	20,139	11,613	5,281	2,086	637	183	33	3	411,494
2022-JAN	0	0	26,451	33,596	36,784	40,099	44,123	46,926	46,015	41,025	28,515	19,846	11,717	5,287	2,086	649	202	30	2	383,353
2021-JAN	18,810	23,809	26,095	28,000	30,315	33,309	35,706	36,562	35,556	31,914	28,510	19,996	11,811	5,232	2,123	652	198	29	3	368,630
2020-JAN	17,721	22,498	24,988	26,780	29,340	32,188	34,837	36,104	35,081	31,677	28,557	20,018	11,865	5,190	2,129	655	196	31	4	359,859
2019-JAN	13,764	17,792	20,834	23,178	25,634	29,170	32,173	34,174	33,747	31,330	28,336	20,116	12,012	5,213	2,135	661	196	34	5	330,504
2018-JAN	9,951	13,478	16,844	19,297	22,142	25,815	29,649	32,088	32,441	30,546	28,149	20,199	12,013	5,304	2,119	669	199	39	4	300,946
2017-JAN	6,587	9,675	12,745	15,487	18,516	22,441	26,581	29,546	30,828	29,575	27,902	20,250	12,101	5,259	2,152	646	216	36	5	270,548
2016-JAN	3,757	5,879	8,476	11,451	14,371	18,785	22,868	26,718	28,661	28,422	27,510	20,299	12,083	5,324	2,125	679	206	41	3	237,658
2015-JAN	1,771	3,048	5,041	7,498	10,342	14,328	18,754	23,003	25,965	26,742	26,928	20,286	12,251	5,248	2,134	672	200	43	4	204,258
2014-JAN	355	983	2,398	4,264	6,837	10,561	15,036	19,397	23,274	24,952	26,358	20,106	12,384	5,271	2,086	683	199	50	3	175,197
2013-JAN	81	306	1,275	2,700	4,931	8,082	12,293	16,642	20,964	23,399	25,791	19,958	12,515	5,285	2,089	686	197	45	4	157,243
2012-JAN	0	0	645	1,605	3,317	6,006	9,707	13,850	18,355	21,657	24,915	19,889	12,536	5,299	2,078	684	191	42	4	140,780
2011-JAN	0	0	217	672	1,864	4,037	7,132	10,912	15,573	19,499	23,778	19,657	12,655	5,209	2,116	673	193	36	4	124,227

(Baseline) Simulation using actual historical regulations



	1000-1099	1100-1199	1200-1299	1300-1399	1400-1499	1500-1599	1600-1699	1700-1799	1800-1899	1900-1999	2000-2099	2100-2199	2200-2299	2300-2399	2400-2499	2500-2599	2600-2699	2700-2799	2800-2899	Total
2023-JAN	23,716	30,025	32,556	34,099	36,276	38,215	39,652	39,197	37,291	32,686	28,380	19,816	11,485	5,241	2,067	634	179	33	3	411,551
2022-JAN	20,330	25,924	28,192	30,158	32,326	35,047	37,047	37,549	36,228	32,218	28,515	19,846	11,717	5,287	2,086	649	202	30	2	383,353
2021-JAN	18,810	23,809	26,095	28,000	30,315	33,309	35,706	36,562	35,556	31,914	28,510	19,996	11,811	5,232	2,123	652	198	29	3	368,630
2020-JAN	17,721	22,498	24,988	26,780	29,340	32,188	34,837	36,104	35,081	31,677	28,557	20,018	11,865	5,190	2,129	655	196	31	4	359,859
2019-JAN	13,764	17,792	20,834	23,178	25,634	29,170	32,173	34,174	33,747	31,330	28,336	20,116	12,012	5,213	2,135	661	196	34	5	330,504
2018-JAN	9,951	13,478	16,844	19,297	22,142	25,815	29,649	32,088	32,441	30,546	28,149	20,199	12,013	5,304	2,119	669	199	39	4	300,946
2017-JAN	6,587	9,675	12,745	15,487	18,516	22,441	26,581	29,546	30,828	29,575	27,902	20,250	12,101	5,259	2,152	646	216	36	5	270,548
2016-JAN	3,757	5,879	8,476	11,451	14,371	18,785	22,868	26,718	28,661	28,422	27,510	20,299	12,083	5,324	2,125	679	206	41	3	237,658
2015-JAN	1,771	3,048	5,041	7,498	10,342	14,328	18,754	23,003	25,965	26,742	26,928	20,286	12,251	5,248	2,134	672	200	43	4	204,258
2014-JAN	355	983	2,398	4,264	6,837	10,561	15,036	19,397	23,274	24,952	26,358	20,106	12,384	5,271	2,086	683	199	50	3	175,197
2013-JAN	81	306	1,275	2,700	4,931	8,082	12,293	16,642	20,964	23,399	25,791	19,958	12,515	5,285	2,089	686	197	45	4	157,243
2012-JAN	0	0	645	1,605	3,317	6,006	9,707	13,850	18,355	21,657	24,915	19,889	12,536	5,299	2,078	684	191	42	4	140,780
2011-JAN	0	0	217	672	1,864	4,037	7,132	10,912	15,573	19,499	23,778	19,657	12,655	5,209	2,116	673	193	36	4	124,227

The downside to having the simulated compression in January of 2022 is that we don't get very many years to look at, post-compression. But we can at least get a really good sense of what would change a year later, since the rating pool in 2022 was so similar to today's rating pool. The compression (2022-JAN) is shown in black, and one year later (2023-JAN) is shown in thick red.

Certainly we would see large increases right away in the number of players rated 1900-1999 Elo, or 1800-1899 Elo, or really any of the ranges below 2000 Elo. That is inevitable from the compression, since we had ten groups of players being squished together into six (or eight) groups of players instead. What is perhaps less immediately obvious, is that the compression would also cause eventual increases in the number of players rated 2000-2099 Elo, or 2100-2199 Elo, even just a year or two later. This is because we would have more players with ratings

increased up near 2000 Elo and more able to cross over, and also because some of the players rated over 2000 Elo would now find it easier to gain rating points than before, from their games against players who had their ratings increased. This is all by design, trying to gradually counteract the effects previously caused by the deflation. It is an inflationary effect, yes, but it does not necessarily lead to overall inflation. It is aimed simply at counteracting most – but not all - of the existing deflation.

Instead of straining your eyes from trying to read those little data tables, I will just summarize for you what is happening in these three scenarios, within the first year, in all the groups rated 2000 Elo or higher. By seeing the player counts on the January 2022 list, and then on the simulated January 2023 list, one year after the compression, we can hopefully get a very good idea of what might happen in the first year after a future compression (presumably in January of 2024):

Projected changes - 1 year after compression

	# of players rated **2000-2099**	# of players rated **2100-2199**	# of players rated **2200-2299**	# of players rated **2300-2399**	# of players rated **2400-2499**	# of players rated **2500-2599**	# of players rated **2600+**
	2022-JAN → 2023-JAN	2022-JAN → 2023-JAN	2022-JAN → 2023-JAN	2022-JAN → 2023-JAN	2022-JAN → 2023-JAN	2022-JAN → 2023-JAN	2022-JAN → 2023-JAN
Scenario C: Compression to 1400 in 2022-JAN	28,515 → 30,254	19,846 → 20,487	11,717 → 11,769	5,287 → 5,331	2,086 → 2,093	649 → 638	234 → 219
	+1,739	**+641**	**+52**	**+44**	**+7**	**-11**	**-15**
Scenario D: Compression to 1200 in 2022-JAN	28,515 → 29,112	19,846 → 20,139	11,717 → 11,613	5,287 → 5,281	2,086 → 2,086	649 → 637	234 → 219
	+597	**+293**	**-104**	**-6**	**=0**	**-12**	**-15**
Baseline Simulation	28,515 → 28,380	19,846 → 19,816	11,717 → 11,485	5,287 → 5,241	2,086 → 2,067	649 → 634	234 → 215
	-135	**-30**	**-232**	**-46**	**-19**	**-15**	**-19**

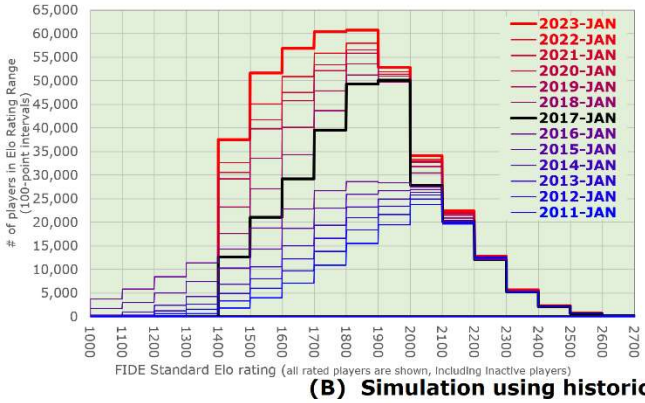
As you can see from the uniform redness of the bottom row (the "Baseline Simulation"), what we have been seeing in real life in the past couple of years is shrinkage of all these high-rated groups spanning 100 Elo points each. We also encountered this in Appendix 2 when we were looking for evidence of deflation at the master level.

A compression to 1200 Elo (the middle row) would start to tackle some of that shrinkage, though it wouldn't reach very far in the first year. A more aggressive compression to 1400 Elo would reverse a lot of that shrinkage, into modest growth at the master level even after a year. But only modest growth, not massive inflation, and there is a lot of growth in those groups that probably should have happened in the past decade, which would be gradually being corrected. Also, there would still mostly be the same shrinkage in the GM-level groups of 2500-2599 Elo and 2600+ Elo, within that first year. The extra points will take a while to propagate upward.

The last thing to look at, in our projections, would be a longer-term estimate. In order to simulate what might happen after five years, we would need to schedule our simulated compression back in January of 2017. A good estimate for the COVID years of 2020-2021 is that approximately one year's worth of games was played, across two calendar years. So we can kind of treat that two-year span like one year's worth of events. Then, with a simulated compression in January of 2017, we can see how the simulated rating list would look in January of 2023, and perhaps that's a good estimate of how things will change in five years after any compression that we perform in 2024. With the caveat that COVID caused chaos, of course.

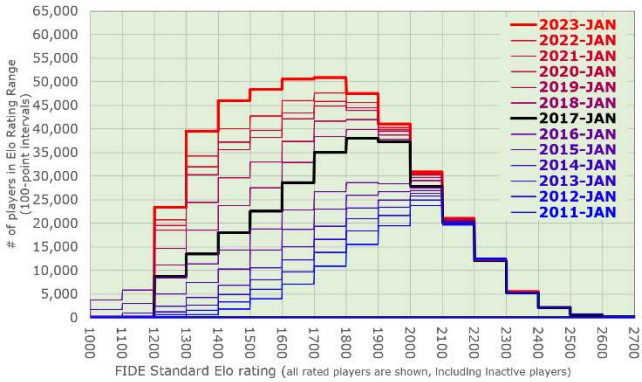
So this takes us to scenarios (A) and (B), which are simulated compressions to 1400 Elo and 1200 Elo, respectively, in January of 2017:

(A) Simulation using historical regulations through the end of 2016, and then 40% compression with 1400 Elo minimum (and calculation improvements) on 2017-JAN



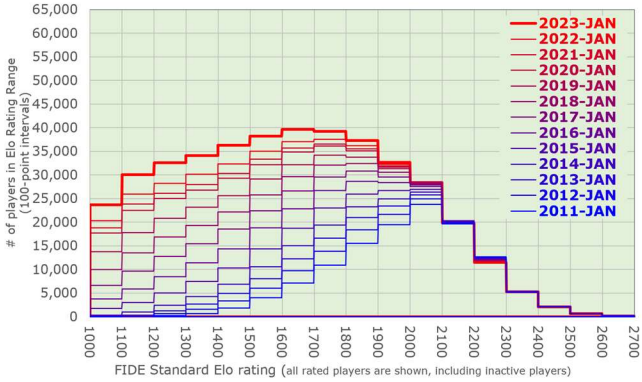
	1000-1099	1100-1199	1200-1299	1300-1399	1400-1499	1500-1599	1600-1699	1700-1799	1800-1899	1900-1999	2000-2099	2100-2199	2200-2299	2300-2399	2400-2499	2500-2599	2600-2699	2700-2799	2800-2899	Total
2023-JAN	0	0	0	0	37,580	51,671	56,853	60,409	60,703	52,812	34,093	22,390	12,843	5,736	2,304	721	217	38	3	398,373
2022-JAN	0	0	0	0	32,687	45,129	50,879	55,895	57,929	51,942	33,377	22,057	12,812	5,710	2,249	702	222	33	2	371,625
2021-JAN	0	0	0	0	30,588	41,777	47,560	53,391	56,494	51,394	33,041	22,011	12,773	5,617	2,253	686	216	36	3	357,840
2020-JAN	0	0	0	0	29,194	39,829	45,815	52,199	55,896	51,005	32,755	21,873	12,767	5,519	2,235	692	209	32	4	350,024
2019-JAN	0	0	0	0	23,281	33,581	40,141	47,893	53,549	50,347	31,809	21,471	12,562	5,413	2,191	680	201	35	5	323,159
2018-JAN	0	0	0	0	17,609	27,164	34,320	43,639	51,261	49,763	30,486	20,899	12,336	5,357	2,130	673	199	39	4	295,879
2017-JAN	0	0	0	0	12,702	21,050	29,258	39,536	49,294	50,141	27,902	20,250	12,101	5,259	2,152	646	216	36	5	270,548
2016-JAN	3,757	5,879	8,476	11,451	14,371	18,785	22,868	26,718	28,661	28,422	27,510	20,299	12,083	5,324	2,125	679	206	41	3	237,658
2015-JAN	1,771	3,048	5,041	7,498	10,342	14,328	18,754	23,003	25,965	26,742	26,928	20,286	12,251	5,248	2,134	672	200	43	4	204,258
2014-JAN	355	983	2,398	4,264	6,837	10,561	15,036	19,397	23,274	24,952	26,358	20,106	12,384	5,271	2,086	683	199	50	3	175,197
2013-JAN	81	306	1,275	2,700	4,931	8,082	12,293	16,642	20,964	23,399	25,791	19,958	12,515	5,285	2,089	686	197	45	4	157,243
2012-JAN	0	0	645	1,605	3,317	6,006	9,707	13,850	18,355	21,657	24,915	19,889	12,536	5,299	2,078	684	191	42	4	140,780
2011-JAN	0	0	217	672	1,864	4,037	7,132	10,912	15,573	19,499	23,778	19,657	12,655	5,209	2,116	673	193	36	4	124,227

(B) Simulation using historical regulations through the end of 2016, and then 20% compression with 1200 Elo minimum (and calculation improvements) on 2017-JAN



	1000-1099	1100-1199	1200-1299	1300-1399	1400-1499	1500-1599	1600-1699	1700-1799	1800-1899	1900-1999	2000-2099	2100-2199	2200-2299	2300-2399	2400-2499	2500-2599	2600-2699	2700-2799	2800-2899	Total
2023-JAN	0	0	23,433	39,570	46,008	48,363	50,572	50,915	47,576	41,090	30,923	21,061	12,121	5,495	2,180	672	204	35	3	420,221
2022-JAN	0	0	20,730	34,232	40,053	42,789	45,996	47,654	45,591	40,285	30,534	20,971	12,196	5,443	2,168	681	211	31	2	389,567
2021-JAN	0	0	19,595	32,012	37,204	39,760	43,407	45,854	44,511	39,844	30,391	20,968	12,237	5,374	2,184	669	205	31	3	374,249
2020-JAN	0	0	18,559	30,331	35,630	38,198	42,165	44,855	43,904	39,510	30,255	20,875	12,275	5,320	2,185	670	200	31	4	364,967
2019-JAN	0	0	14,678	24,473	29,627	33,011	37,359	41,666	41,958	38,702	29,733	20,678	12,252	5,304	2,155	668	199	34	5	332,502
2018-JAN	0	0	11,144	18,550	23,764	27,579	32,901	38,400	39,914	37,816	29,046	20,452	12,163	5,316	2,120	673	199	39	4	300,080
2017-JAN	0	0	8,753	13,552	18,086	22,619	28,645	34,993	38,074	37,259	27,902	20,250	12,101	5,259	2,152	646	216	36	5	270,548
2016-JAN	3,757	5,879	8,476	11,451	14,371	18,785	22,868	26,718	28,661	28,422	27,510	20,299	12,083	5,324	2,125	679	206	41	3	237,658
2015-JAN	1,771	3,048	5,041	7,498	10,342	14,328	18,754	23,003	25,965	26,742	26,928	20,286	12,251	5,248	2,134	672	200	43	4	204,258
2014-JAN	355	983	2,398	4,264	6,837	10,561	15,036	19,397	23,274	24,952	26,358	20,106	12,384	5,271	2,086	683	199	50	3	175,197
2013-JAN	81	306	1,275	2,700	4,931	8,082	12,293	16,642	20,964	23,399	25,791	19,958	12,515	5,285	2,089	686	197	45	4	157,243
2012-JAN	0	0	645	1,605	3,317	6,006	9,707	13,850	18,355	21,657	24,915	19,889	12,536	5,299	2,078	684	191	42	4	140,780
2011-JAN	0	0	217	672	1,864	4,037	7,132	10,912	15,573	19,499	23,778	19,657	12,655	5,209	2,116	673	193	36	4	124,227

(Baseline) Simulation using actual historical regulations



	1000-1099	1100-1199	1200-1299	1300-1399	1400-1499	1500-1599	1600-1699	1700-1799	1800-1899	1900-1999	2000-2099	2100-2199	2200-2299	2300-2399	2400-2499	2500-2599	2600-2699	2700-2799	2800-2899	Total
2023-JAN	23,716	30,025	32,556	34,099	36,276	38,215	39,652	39,197	37,291	32,686	28,380	19,816	11,485	5,241	2,067	634	179	33	3	411,551
2022-JAN	20,330	25,924	28,192	30,158	32,326	35,047	37,047	37,549	36,228	32,218	28,515	19,846	11,717	5,287	2,086	649	202	30	2	383,353
2021-JAN	18,810	23,809	26,095	28,000	30,315	33,309	35,706	36,562	35,556	31,914	28,510	19,996	11,811	5,232	2,123	652	198	29	3	368,630
2020-JAN	17,721	22,498	24,988	26,780	29,340	32,188	34,837	36,104	35,081	31,677	28,557	20,018	11,865	5,190	2,129	655	196	31	4	359,859
2019-JAN	13,764	17,792	20,834	23,178	25,634	29,170	32,173	34,174	33,747	31,330	28,336	20,116	12,012	5,213	2,135	661	196	34	5	330,504
2018-JAN	9,951	13,478	16,844	19,297	22,142	25,815	29,649	32,088	32,441	30,546	28,149	20,199	12,013	5,304	2,119	669	199	39	4	300,946
2017-JAN	6,587	9,675	12,745	15,487	18,516	22,441	26,581	29,546	30,828	29,575	27,902	20,250	12,101	5,259	2,152	646	216	36	5	270,548
2016-JAN	3,757	5,879	8,476	11,451	14,371	18,785	22,868	26,718	28,661	28,422	27,510	20,299	12,083	5,324	2,125	679	206	41	3	237,658
2015-JAN	1,771	3,048	5,041	7,498	10,342	14,328	18,754	23,003	25,965	26,742	26,928	20,286	12,251	5,248	2,134	672	200	43	4	204,258
2014-JAN	355	983	2,398	4,264	6,837	10,561	15,036	19,397	23,274	24,952	26,358	20,106	12,384	5,271	2,086	683	199	50	3	175,197
2013-JAN	81	306	1,275	2,700	4,931	8,082	12,293	16,642	20,964	23,399	25,791	19,958	12,515	5,285	2,089	686	197	45	4	157,243
2012-JAN	0	0	645	1,605	3,317	6,006	9,707	13,850	18,355	21,657	24,915	19,889	12,536	5,299	2,078	684	191	42	4	140,780
2011-JAN	0	0	217	672	1,864	4,037	7,132	10,912	15,573	19,499	23,778	19,657	12,655	5,209	2,116	673	193	36	4	124,227

This time the black lines (representing the compression) will be earlier, and we can see what happens for several years after the compression, by inspecting the red lines from subsequent years (typically higher than the black lines). It may be easiest to grasp this by just looking again at the summary table covering the baseline scenario and the other two:

Projected changes - 5 years after compression

	# of players rated **2000-2099**	# of players rated **2100-2199**	# of players rated **2200-2299**	# of players rated **2300-2399**	# of players rated **2400-2499**	# of players rated **2500-2599**	# of players rated **2600+**
	2017-JAN → 2023-JAN	2017-JAN → 2023-JAN	2017-JAN → 2023-JAN	2017-JAN → 2023-JAN	2017-JAN → 2023-JAN	2017-JAN → 2023-JAN	2017-JAN → 2023-JAN
Scenario A: Compression to 1400 in 2017-JAN	27,902 → 34,093	20,250 → 22,390	12,101 → 12,843	5,259 → 5,736	2,152 → 2,304	646 → 721	257 → 258
	+6,191	**+2,140**	**+742**	**+477**	**+152**	**+75**	**+1**
Scenario B: Compression to 1200 in 2017-JAN	27,902 → 30,923	20,250 → 21,061	12,101 → 12,121	5,259 → 5,495	2,152 → 2,180	646 → 672	257 → 242
	+3,021	**+811**	**+20**	**+236**	**+28**	**+26**	**-15**
Baseline Simulation	27,902 → 28,380	20,250 → 19,816	12,101 → 11,485	5,259 → 5,241	2,152 → 2,067	646 → 634	257 → 215
	+478	**-434**	**-616**	**-18**	**-85**	**-12**	**-42**

As you can see, even after five full years, there is not some big inflation and high ratings becoming meaningless, nor is there a huge increase in players eligible for the top titles. Essentially the compression will have reversed the shrinkage (that had been caused by the ongoing deflation) in all of these Elo groups, and restored growth to the lower groups such as 2000-2099 Elo and 2100-2199 Elo.

For example, instead of annually losing 17 players rated 2400-2499 Elo, which is what really happened since 2017 (we are viewing that as effectively five years, so a total of "-85" means losing 17 such players per year), it is either a very modest increase of 5-6 players per year rated 2400-2499 Elo (if the compression was to 1200 Elo) or a somewhat larger increase of 30 players per year (if the compression was to 1400 Elo).

Or you can even look up at the 2600+ Elo level. As we have seen elsewhere, this group is shrinking. The compression to 1400 Elo would essentially stabilize the size of that top group, as you can see by the blue +1 in the upper right corner, compared to the -15 (losing 3 of those players a year) after a compression to 1200 Elo, or the -42 (losing 8 of those players a year) if no compression takes place. Please also note that this "losing 8 players rated 2600+ per year" is an average stretching back all the way to 2017; we have already seen in Appendix #2 (on Page 19) that more recently, we are losing 13 such players per year.

If we accept where we currently are, as the ideal place, then the introduction of any sort of inflationary factor might feel like a bad move. But I don't think the rating system is currently in an ideal place – far from it! - and it appears to me that a Compression to 1400 Elo would move the rating system into a much better place. **I don't see any reason to expect inflation to result from a Compression to 1400 Elo.**

Appendix #4: Should we revert the 400-point-rule to its earlier form?

The general direction of my analysis for the 400-point-rule is to look at various matchups among different rating groups, across different time periods, or even under some simulations (like what would happen after a compression), to see which version of the rule seems to work best. It should not surprise you to hear that there's no single answer that works perfectly for all possible scenarios throughout history. Mathematically speaking, in some cases it seems to work best to use the older 400-point-rule, but in other cases it works better to use the current "1-upgrade" rule, or even no rule at all.

My criteria for determining what works "better" in a particular scenario, is to see which version of the rule yields expected scores that are closest to the actual score achieved by the higher-rated players. So I will check whether the average actual score achieved by the higher-rated player is closer to the 92% from the older 400-point-rule expectation, or to the expected score from the current "1-upgrade" rule, or even using an expected score with no 400-point-rule. I have developed a color-coding scheme to help readers try to make sense of it all.

I have created a new term "Extreme-Elo-Difference" to help me in speaking clearly about the 400-point-rule. It's important to remember that nothing special has to be done for the 400-point-rule when the rating gap is smaller than 412 Elo points. It's only once we hit a rating difference of 412 Elo or higher, that the 400-point-rule does anything, because that's when the expected scores for the higher rated player would normally become 93% instead of 92% (see the relevant part of the Elo expectancy table from the FIDE Handbook, below).

8.1.2 Table of conversion of difference in rating, D, into scoring probability PD, for the higher, H, and the lower, L, rated player respectively.

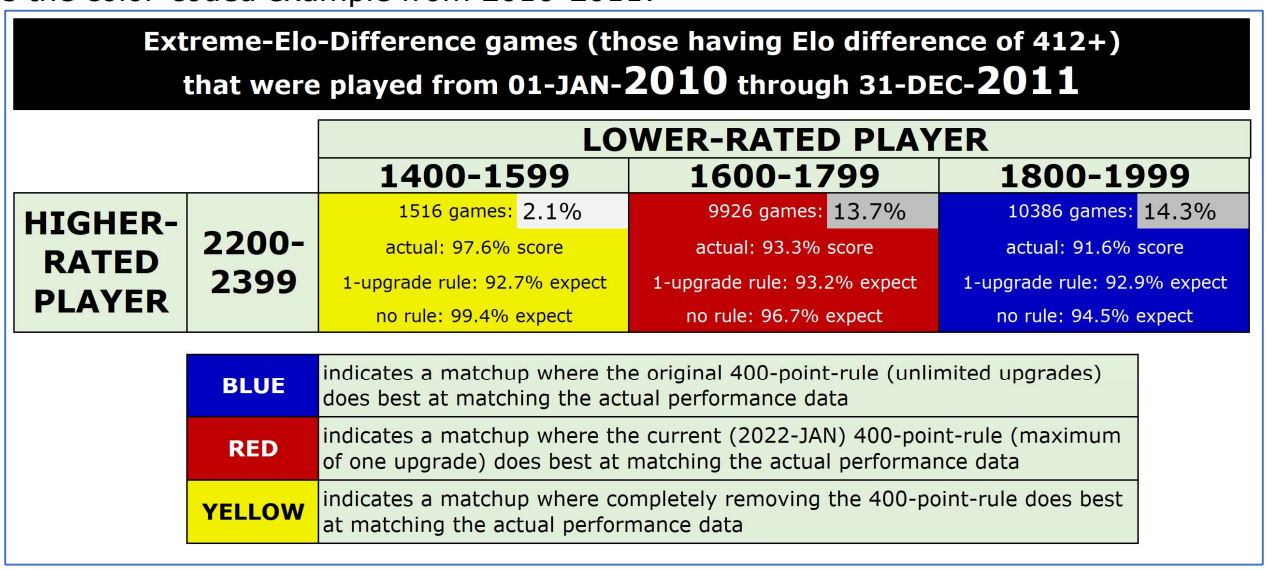
D	PD	
Rtg Dif	H	L
345-357	.89	.11
358-374	.90	.10
375-391	.91	.09
392-411	.92	.08
412-432	.93	.07
433-456	.94	.06

I will be using the term "Extreme-Elo-Difference games" to refer to precisely those games having 412+ Elo rating difference. And the data used for analysis in this appendix is restricted to just those Extreme-Elo-Difference games. I developed a color-coding scheme to help represent the 400-point-rule options visually, and it will be important for you to understand the meaning of the colors. To understand this color-coding scheme, let's start with an example.

The 400-point-rule came into existence in mid-2009 (previously it was a 350-point-rule), and we can look at its historical behavior in the first two full years (2010-2011) that it was in effect.

We will start by assigning all players into Elo groups spanning 200 Elo points (1000-1199, 1200-1399, etc.) and then looking at all the Extreme-Elo-Difference games played between members of these various groups. For each matchup of two different Elo groups, we will check the average score achieved by the higher-rated-players, as well as what the expected score would have been under various flavors of the 400-point-rule (including the scenario where there was no 400-point-rule).

Here is the color-coded example from 2010-2011:



Extreme-Elo-Difference games (those having Elo difference of 412+) that were played from 01-JAN-2010 through 31-DEC-2011

		LOWER-RATED PLAYER		
		1400-1599	1600-1799	1800-1999
HIGHER-RATED PLAYER	2200-2399	1516 games: 2.1% actual: 97.6% score 1-upgrade rule: 92.7% expect no rule: 99.4% expect	9926 games: 13.7% actual: 93.3% score 1-upgrade rule: 93.2% expect no rule: 96.7% expect	10386 games: 14.3% actual: 91.6% score 1-upgrade rule: 92.9% expect no rule: 94.5% expect

BLUE	indicates a matchup where the original 400-point-rule (unlimited upgrades) does best at matching the actual performance data
RED	indicates a matchup where the current (2022-JAN) 400-point-rule (maximum of one upgrade) does best at matching the actual performance data
YELLOW	indicates a matchup where completely removing the 400-point-rule does best at matching the actual performance data

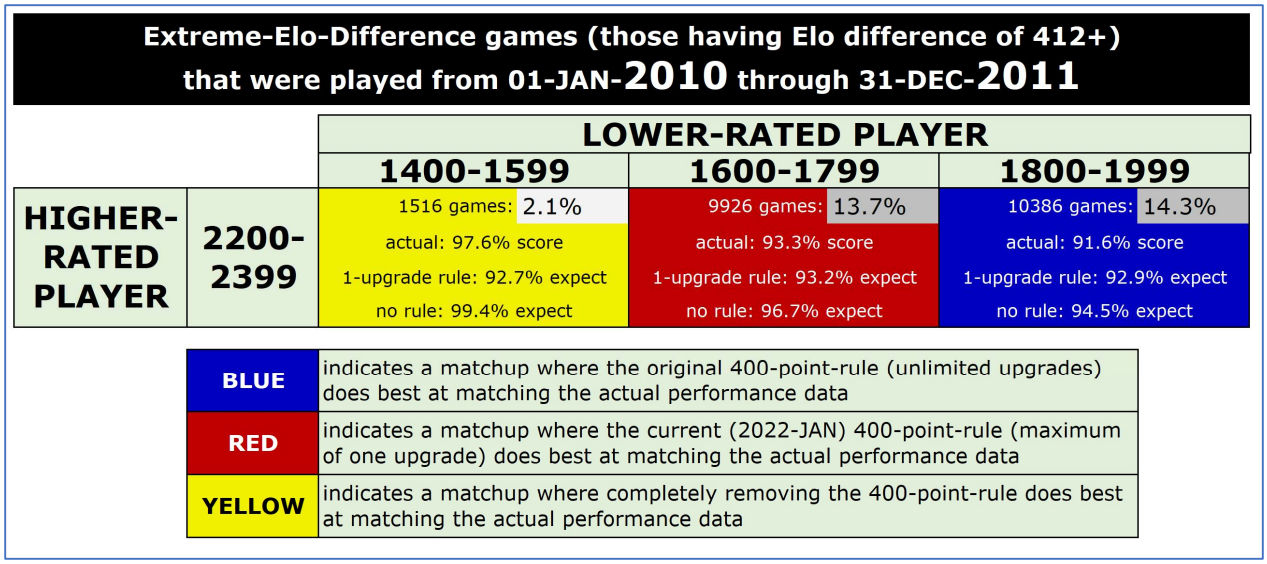
Although it covers more than 20,000 games, this example shows only a few cells of data, specifically three different categories of lower-rated opponents, where the higher-rated player was rated 2200-2399 during these Extreme-Elo-Difference games from 2010 and 2011.

Let's first look at the blue cell on the right. The row label of "2200-2399", lined up with the column label of "1800-1999", tells us we are looking at the averages (during 2010-2011) when a 2200-2399 player faced an 1800-1999 opponent in Extreme-Elo-Difference games. Remember, those are precisely the games where the rating differences was 412+ Elo points, and so any games where the Elo rating difference was 0-411 are excluded from the calculation.

We see in that blue cell that there were 10,386 Extreme-Elo-Difference games played during 2010-2011 between one player rated 2200-2399 and another player rated 1800-1999, representing 14.3% of all Extreme-Elo-Difference games during that time period. The text *"actual: 91.6% score"* tells us that the higher-rated player scored 91.6% overall in these games.

Then we look lower down in the blue cell to expected scores. Of course, for the original 400-point-rule, all of these Extreme-Elo-Difference games would have an expected score of 92.0%, no matter what the ratings were, and so we don't need to use up real estate showing that in each cell. But we see that if the "1-upgrade rule" had been in place (that's the current rule as adopted in 2022), then the overall expected score for all these games would be 92.9%. And finally, if there had been no 400-point-rule at all for these games, then the higher-rated-player would have had an average expected score of 94.5%.

Overall, the color-coding assigns a background color of blue to the cell, rather than red or yellow. Why is that? Well, for this matchup, the original 400-point-rule would have yielded an expected score that is closest to the actual score that was achieved (since 91.6% is very close to 92.0%) in these games. The other two approaches would have given expected scores that were significantly too high (92.9% and 94.5%) in comparison to the 91.6% score that was actually achieved by higher-rated players. So that's why this matchup is categorized as blue, because for this matchup the original 400-point-rule would have worked best. Let's show the same picture again for this example, on the next page, and look at another data cell:



Extreme-Elo-Difference games (those having Elo difference of 412+) that were played from 01-JAN-2010 through 31-DEC-2011

		LOWER-RATED PLAYER		
		1400-1599	1600-1799	1800-1999
HIGHER-RATED PLAYER	2200-2399	1516 games: 2.1% actual: 97.6% score 1-upgrade rule: 92.7% expect no rule: 99.4% expect	9926 games: 13.7% actual: 93.3% score 1-upgrade rule: 93.2% expect no rule: 96.7% expect	10386 games: 14.3% actual: 91.6% score 1-upgrade rule: 92.9% expect no rule: 94.5% expect

BLUE	indicates a matchup where the original 400-point-rule (unlimited upgrades) does best at matching the actual performance data
RED	indicates a matchup where the current (2022-JAN) 400-point-rule (maximum of one upgrade) does best at matching the actual performance data
YELLOW	indicates a matchup where completely removing the 400-point-rule does best at matching the actual performance data

In fact, not all matchups will be color-coded blue like that one over on the right. For example, let's look at the red cell's numbers. In the column labeled "1600-1799", we see that when the 2200-2399 Elo players faced opponents rated 1600-1799 in Extreme-Elo-Difference games, the favorites' actual score was 93.3%, which is very close to the 93.2% that would have been the average expected score if the "1-upgrade" rule had been in effect. The 93.3% actual score is further away from the 92% of the traditional 400-point-rule, or the 96.7% if there were no rule at all. So that cell is color-coded red, to signify that the current "1-upgrade rule" would have worked best for that matchup.

And for the third example, look leftward again to the column labeled "1400-1599". This is the most one-sided of these three categories of matchups, and indeed the higher-rated player scored 97.6% in the 1,516 Extreme-Elo-Difference games that were played between these groups. Although the Elo tables would predict an even higher expected score (99.4%), that is easily the closest match to that 97.6% actual score, and so the cell is color-coded yellow, indicating that the elimination of the 400-point-rule would work best for this matchup. The other options for expected score would be far away at 92% (for the original 400-point-rule) or 92.7% (for the "1-upgrade rule").

Also notice that at the top of each data cell, it tells you not just how many games were played, but what percentage of all Extreme-Elo-Difference games are being shown in that cell. So we can see that the blue and red cells cover 14.3% and 13.7%, respectively, of all Extreme-Elo-Difference games played during 2010-2011, whereas the yellow one is only 2.1% of such games. To convey these magnitudes visually, the larger percentages are given a darker-gray background, whereas the smaller percentages are lighter-shaded or even white. So anytime you see a cell with a white background in its upper-right-corner, that means that it doesn't cover very many games.

And finally, there will sometimes be cases where the blue (original 400-point-rule) and the red (current 1-upgrade rule) are both "best", since they work about equally well. In those cases, the cell will be colored a bluish-purple "indigo" if blue is a little better, or a reddish-purple "burgundy" if red is a little better. And similarly, the yellowish-orange "tangerine" and reddish-orange "marigold" would convey that both the red (current 1-upgrade rule) and the yellow (no 400-point-rule) work about equally well. Essentially, if the best main option (red, blue, or yellow) is no more than 0.5 percentage points better than the second-best main option (based on how far from actual score they both are), that would trigger the purple or orange coloring.

Okay, let's finally look at a full grid with more than just three example data cells! First we will stay in 2010-2011 and look at all the matchups. Overall there were more than 70,000 games:

(A) -- Extreme-Elo-Difference games (those having Elo difference of 412+) that were played from 01-JAN-2010 through 31-DEC-2011

HIGHER-RATED PLAYER \ LOWER-RATED PLAYER	1000-1199	1200-1399	1400-1599	1600-1799	1800-1999	2000-2199	2200-2399	Total
1400-1599	22 games: 0.0% actual: 81.8% score 1-upgrade rule: 92.2% expect no rule: 95.3% expect							22 games: 0.0% actual: 81.8% score 1-upgrade rule: 92.2% expect no rule: 95.3% expect
1600-1799	46 games: 0.1% actual: 91.3% score 1-upgrade rule: 92.5% expect no rule: 97% expect	1699 games: 2.3% actual: 87.7% score 1-upgrade rule: 92.1% expect no rule: 94.4% expect						1745 games: 2.4% actual: 87.8% score 1-upgrade rule: 92.1% expect no rule: 94.4% expect
1800-1999	13 games: 0.0% actual: 80.8% score 1-upgrade rule: 92.5% expect no rule: 99.3% expect	3131 games: 4.3% actual: 90.9% score 1-upgrade rule: 92.3% expect no rule: 96.7% expect	6938 games: 9.6% actual: 88.3% score 1-upgrade rule: 92.4% expect no rule: 94.4% expect					10082 games: 13.9% actual: 89.1% score 1-upgrade rule: 92.4% expect no rule: 95.1% expect
2000-2199	5 games: 0.0% actual: 100% score 1-upgrade rule: 92% expect no rule: 100% expect	876 games: 1.2% actual: 96.5% score 1-upgrade rule: 92.4% expect no rule: 99.5% expect	8470 games: 11.7% actual: 92.2% score 1-upgrade rule: 92.7% expect no rule: 96.7% expect	12390 games: 17.1% actual: 90% score 1-upgrade rule: 92.7% expect no rule: 94.4% expect				21741 games: 29.9% actual: 91.1% score 1-upgrade rule: 92.7% expect no rule: 95.5% expect
2200-2399		122 games: 0.2% actual: 99.6% score 1-upgrade rule: 92.3% expect no rule: 100% expect	1516 games: 2.1% actual: 97.6% score 1-upgrade rule: 92.7% expect no rule: 99.4% expect	9926 games: 13.7% actual: 93.3% score 1-upgrade rule: 93.2% expect no rule: 96.7% expect	10386 games: 14.3% actual: 91.6% score 1-upgrade rule: 92.9% expect no rule: 94.5% expect			21950 games: 30.2% actual: 92.8% score 1-upgrade rule: 93% expect no rule: 95.9% expect
2400-2599		14 games: 0.0% actual: 100% score 1-upgrade rule: 92.6% expect no rule: 100% expect	176 games: 0.2% actual: 97.7% score 1-upgrade rule: 92.6% expect no rule: 100% expect	1438 games: 2.0% actual: 98.9% score 1-upgrade rule: 93.2% expect no rule: 99.4% expect	7993 games: 11.0% actual: 95.6% score 1-upgrade rule: 93.5% expect no rule: 96.7% expect	5836 games: 8.0% actual: 93.8% score 1-upgrade rule: 93.1% expect no rule: 94.4% expect		15457 games: 21.3% actual: 95.3% score 1-upgrade rule: 93.3% expect no rule: 96.1% expect
2600-2799				40 games: 0.1% actual: 100% score 1-upgrade rule: 92.6% expect no rule: 100% expect	284 games: 0.4% actual: 99.6% score 1-upgrade rule: 93% expect no rule: 99.2% expect	1054 games: 1.5% actual: 96.9% score 1-upgrade rule: 93.6% expect no rule: 96% expect	253 games: 0.3% actual: 93.5% score 1-upgrade rule: 92.7% expect no rule: 93.7% expect	1633 games: 2.2% actual: 96.9% score 1-upgrade rule: 93.4% expect no rule: 96.3% expect
Total	86 games: 0.1% actual: 87.8% score 1-upgrade rule: 92.4% expect no rule: 97.1% expect	5842 games: 8.0% actual: 91% score 1-upgrade rule: 92.3% expect no rule: 96.5% expect	17102 games: 23.5% actual: 91.2% score 1-upgrade rule: 92.6% expect no rule: 96% expect	23794 games: 32.8% actual: 91.9% score 1-upgrade rule: 92.9% expect no rule: 95.7% expect	18663 games: 25.7% actual: 93.4% score 1-upgrade rule: 93.2% expect no rule: 95.5% expect	6890 games: 9.5% actual: 94.3% score 1-upgrade rule: 93.2% expect no rule: 94.6% expect	253 games: 0.3% actual: 93.5% score 1-upgrade rule: 92.7% expect no rule: 93.7% expect	72630 games: 100.0% actual: 92.3% score 1-upgrade rule: 92.9% expect no rule: 95.7% expect

LEGEND:	
BLUE	indicates a matchup where the original 400-point-rule (unlimited upgrades to 92%) does best at matching the actual performance results
RED	indicates a matchup where the current 400-point rule (maximum of 1 upgrade to 92%) from 2022 does best at matching the actual performance results
YELLOW	indicates a matchup where completely removing the 400-point-rule does best at matching the actual performance results
INDIGO	indicates a matchup where the original 400-point-rule works best, but the current rule (maximum of 1 upgrade to 92%) is almost as good
BURGUNDY	indicates a matchup where the current rule (maximum of 1 upgrade to 92%) works best, but the original 400-point-rule is almost as good
TANGERINE	indicates a matchup where completely removing the 400-point-rule works best, but the current rule (maximum of 1 upgrade to 92%) is almost as good
MARIGOLD	indicates a matchup where the current rule (maximum of 1 upgrade to 92%) works best, but completely removing the 400-point-rule is almost as good
	indicates a matchup where there were fewer than 5 games played with a rating difference of 412+

Some definite trends emerge. There's a lot of yellow in the bottom half, indicating that for the titled players, often having no 400-point-rule at all would have yielded an expected score most closely matching their actual score (a dozen years ago). This is consistent with the general understanding that back in 2010-2011, top players were able to exploit the 400-point-rule and its relatively low 92% expected score. For instance, we can see in the rightmost "Total" column that players rated 2400-2599 Elo scored 95.3% across 15,000+ Extreme-Elo-Difference games, and this was 21.3% of all Extreme-Elo-Difference games. And it's color-coded yellow, indicating no rule would have been best. Thus the exploitation by top players was indeed problematic.

Another color you see a lot of, in the top half, is "Indigo", the bluish-purple color. This particular color-coding tells us that when relatively weak players had a big rating advantage, the original 400-point rule was indeed best, but the current "1-upgrade" rule would have been almost as good. This makes sense if you think about it. If you were rated about 1700 Elo back then, it would have been pretty rare to face any opponents rated well below 1300 Elo, let alone more than one such opponent in a single event. If you are never facing more than one opponent rated 412+ Elo points below you in the same tournament, then there's no difference between the current "1-upgrade" rule and the original 400-point-rule. So throughout these graphics, you will often see Indigo in the uppermost region of the image, especially when the lowest range of the rating pool was sparsely populated.

Several cells are blanked out with a light green background and a diagonal slash through them, indicating there were 5 or fewer Extreme-Elo-Difference games in that matchup. This would be true either if the rating groups are too close together (such as 1800-1999 versus 1600-1799, which could never have more than a 399-point rating difference), or if there just weren't many games played at such an extreme matchup (such as 2600-2799 versus 1000-1199).

Overall, I think it would have been quite reasonable for someone to have been critical of the behavior of the 400-point-rule during 2010-2011. It was indeed being exploited heavily by titled players, and in fact the games with players rated 2200+ Elo made up more than half of all

Extreme-Elo-Difference games, back then. Admittedly, that was a dozen years ago. So let's take a big step forward in time - six years forward, in fact - and look at data from the years 2016-2017 instead, using the exact same types of calculations and color-coding logic:

(B) -- Extreme-Elo-Difference games (those having Elo difference of 412+) that were played from 01-JAN-2016 through 31-DEC-2017

HIGHER-RATED PLAYER \ LOWER-RATED PLAYER	1000-1199	1200-1399	1400-1599	1600-1799	1800-1999	2000-2199	2200-2399	Total
1400-1599	17217 games: 6.6% actual: 85.9% score 1-upgrade rule: 92.3% expect no rule: 94.5% expect							17217 games: 6.6% actual: 85.9% score 1-upgrade rule: 92.3% expect no rule: 94.5% expect
1600-1799	24776 games: 9.5% actual: 89.6% score 1-upgrade rule: 92.8% expect no rule: 97% expect	27293 games: 10.5% actual: 85.8% score 1-upgrade rule: 92.9% expect no rule: 94.6% expect						52069 games: 20.0% actual: 87.6% score 1-upgrade rule: 92.8% expect no rule: 95.8% expect
1800-1999	7668 games: 3.0% actual: 95.1% score 1-upgrade rule: 92.7% expect no rule: 99.5% expect	31953 games: 12.3% actual: 89.7% score 1-upgrade rule: 93.4% expect no rule: 97.1% expect	29942 games: 11.5% actual: 86.3% score 1-upgrade rule: 93.1% expect no rule: 94.6% expect					69563 games: 26.8% actual: 88.8% score 1-upgrade rule: 93.2% expect no rule: 96.3% expect
2000-2199	2029 games: 0.8% actual: 97.9% score 1-upgrade rule: 92.6% expect no rule: 100% expect	7398 games: 2.8% actual: 95% score 1-upgrade rule: 93.8% expect no rule: 99.5% expect	28894 games: 11.1% actual: 90.6% score 1-upgrade rule: 93.8% expect no rule: 96.9% expect	24206 games: 9.3% actual: 88.3% score 1-upgrade rule: 93.2% expect no rule: 94.5% expect				62527 games: 24.1% actual: 90.5% score 1-upgrade rule: 93.5% expect no rule: 96.4% expect
2200-2399	450 games: 0.2% actual: 99.6% score 1-upgrade rule: 92.6% expect no rule: 100% expect	1222 games: 0.5% actual: 98% score 1-upgrade rule: 94.1% expect no rule: 100% expect	4396 games: 1.7% actual: 96.2% score 1-upgrade rule: 94.4% expect no rule: 99.5% expect	17254 games: 6.6% actual: 92% score 1-upgrade rule: 94% expect no rule: 96.9% expect	13661 games: 5.3% actual: 89.4% score 1-upgrade rule: 93.3% expect no rule: 94.6% expect			36983 games: 14.2% actual: 91.8% score 1-upgrade rule: 93.8% expect no rule: 96.5% expect
2400-2599	57 games: 0.0% actual: 100% score 1-upgrade rule: 92.3% expect no rule: 100% expect	168 games: 0.1% actual: 99.4% score 1-upgrade rule: 92.8% expect no rule: 100% expect	671 games: 0.3% actual: 98.6% score 1-upgrade rule: 93.3% expect no rule: 100% expect	2682 games: 1.0% actual: 97.2% score 1-upgrade rule: 94.2% expect no rule: 99.5% expect	9682 games: 3.7% actual: 93.9% score 1-upgrade rule: 94.2% expect no rule: 96.7% expect	6005 games: 2.3% actual: 92.3% score 1-upgrade rule: 93.4% expect no rule: 94.4% expect		19265 games: 7.4% actual: 94.1% score 1-upgrade rule: 93.9% expect no rule: 96.5% expect
2600-2799			17 games: 0.0% actual: 100% score 1-upgrade rule: 92% expect no rule: 100% expect	94 games: 0.0% actual: 99.5% score 1-upgrade rule: 92.7% expect no rule: 100% expect	377 games: 0.1% actual: 97.9% score 1-upgrade rule: 94% expect no rule: 99.3% expect	1405 games: 0.5% actual: 95.5% score 1-upgrade rule: 93.7% expect no rule: 96.1% expect	390 games: 0.2% actual: 92.4% score 1-upgrade rule: 92.8% expect no rule: 93.8% expect	2284 games: 0.9% actual: 95.6% score 1-upgrade rule: 93.6% expect no rule: 96.4% expect
Total	52197 games: 20.1% actual: 89.6% score 1-upgrade rule: 92.6% expect no rule: 96.7% expect	68035 games: 26.2% actual: 88.9% score 1-upgrade rule: 93.3% expect no rule: 96.4% expect	63920 games: 24.6% actual: 89.1% score 1-upgrade rule: 93.5% expect no rule: 96% expect	44236 games: 17.0% actual: 90.3% score 1-upgrade rule: 93.6% expect no rule: 95.8% expect	23720 games: 9.1% actual: 91.4% score 1-upgrade rule: 93.7% expect no rule: 95.5% expect	7410 games: 2.9% actual: 92.9% score 1-upgrade rule: 93.4% expect no rule: 94.7% expect	390 games: 0.2% actual: 92.4% score 1-upgrade rule: 92.8% expect no rule: 93.8% expect	259908 games: 100.0% actual: 89.7% score 1-upgrade rule: 93.3% expect no rule: 96.1% expect

LEGEND:	
BLUE	indicates a matchup where the original 400-point-rule (unlimited upgrades to 92%) does best at matching the actual performance results
RED	indicates a matchup where the current 400-point rule (maximum of 1 upgrade to 92%) from 2022 does best at matching the actual performance results
YELLOW	indicates a matchup where completely removing the 400-point-rule does best at matching the actual performance results
INDIGO	indicates a matchup where the original 400-point-rule works best, but the current rule (maximum of 1 upgrade to 92%) is almost as good
BURGUNDY	indicates a matchup where the current rule (maximum of 1 upgrade to 92%) works best, but the original 400-point-rule is almost as good
TANGERINE	indicates a matchup where completely removing the 400-point-rule works best, but the current rule (maximum of 1 upgrade to 92%) is almost as good
MARIGOLD	indicates a matchup where the current rule (maximum of 1 upgrade to 92%) works best, but completely removing the 400-point-rule is almost as good
	indicates a matchup where there were fewer than 5 games played with a rating difference of 412+

Here we still see a lot of yellow near the bottom, but it's mainly just for games where the higher-rated player was IM/GM level (2400+ Elo). So this indicates that the original 400-point-rule was still being successfully exploited by the strongest players. But whereas in 2010-2011 we saw more than 20% of Extreme-Elo-Difference games involved players rated 2400+ Elo, by 2016-2017 that percentage had been cut in half (actually it was below 9%). In contrast, 77% of the Extreme-Elo-Difference games, back in 2016-2017, had both players rated below 2200 Elo.

We see a very strong blue presence, indicating that the original 400-point-rule was quite appropriate in 2016-2017. Take note of the percentages of games in the rightmost column. The lower ranks of players were filling in heavily, and so there were far more Extreme-Elo-Difference games played at lower Elo levels than before. For example, more than half of all such games involved a matchup of two players rated below 2000 Elo. And for an overall conclusion about the appropriateness of the 400-point-rule options, you can look in the bottom right corner of each graphic. In this case, it tells us that in almost 260,000 games having a rating difference of 412+ Elo, the higher-rated player was scoring slightly below 90%. And it is color-coded blue, indicating that the unlimited 400-point-rule (with its universal 92% expected score) would have worked best in that era. By contrast, the "1-upgrade-rule" would have yielded an expected score above 93%, far too optimistic an expectation for extreme rating favorites in that era.

So essentially in the six years from 2010-2011 until 2016-2017, the composition of the rating pool changed somewhat, and already it had become harder to exploit the 400-point-rule. In addition, I would say that any unfairness associated with the highest-rated players exploiting the rule, was more than offset by the real need for something like a 400-point-rule among the many untitled players who were large rating favorites in games with extreme Elo differences, and who struggled to meet their overly optimistic Elo expectations in those games.

And now we can jump forward another six years and look at the current situation, based on the games played in 2022 and 2023 (I've only counted up through the first quarter of 2023, which is the last time I imported full data from FIDE):

(C) -- Extreme-Elo-Difference games (those having Elo difference of 412+) that were played from 01-JAN-2022 through 31-MAR-2023

HIGHER-RATED PLAYER \ LOWER-RATED PLAYER	1000-1199	1200-1399	1400-1599	1600-1799	1800-1999	2000-2199	2200-2399	Total
1400-1599	15458 games: 8.2% actual: 82% score 1-upgrade rule: 92.4% expect no rule: 94.6% expect							15458 games: 8.2% actual: 82% score 1-upgrade rule: 92.4% expect no rule: 94.6% expect
1600-1799	23439 games: 12.5% actual: 85.6% score 1-upgrade rule: 93.1% expect no rule: 97.1% expect	21836 games: 11.6% actual: 79.9% score 1-upgrade rule: 93.2% expect no rule: 94.7% expect						45275 games: 24.1% actual: 82.9% score 1-upgrade rule: 93.1% expect no rule: 95.9% expect
1800-1999	7647 games: 4.1% actual: 90.8% score 1-upgrade rule: 93.3% expect no rule: 99.5% expect	26479 games: 14.1% actual: 84.4% score 1-upgrade rule: 94% expect no rule: 97.2% expect	19466 games: 10.4% actual: 80.5% score 1-upgrade rule: 93.4% expect no rule: 94.7% expect					53592 games: 28.5% actual: 83.9% score 1-upgrade rule: 93.7% expect no rule: 96.6% expect
2000-2199	1923 games: 1.0% actual: 95.5% score 1-upgrade rule: 93.2% expect no rule: 100% expect	6900 games: 3.7% actual: 90.6% score 1-upgrade rule: 94.6% expect no rule: 99.6% expect	18925 games: 10.1% actual: 85.8% score 1-upgrade rule: 94.4% expect no rule: 97.2% expect	12980 games: 6.9% actual: 83.2% score 1-upgrade rule: 93.5% expect no rule: 94.6% expect				40728 games: 21.7% actual: 86.2% score 1-upgrade rule: 94.1% expect no rule: 96.9% expect
2200-2399	421 games: 0.2% actual: 98.2% score 1-upgrade rule: 93% expect no rule: 100% expect	1334 games: 0.7% actual: 95.6% score 1-upgrade rule: 94.5% expect no rule: 100% expect	3722 games: 2.0% actual: 92.3% score 1-upgrade rule: 95.3% expect no rule: 99.6% expect	9896 games: 5.3% actual: 87.9% score 1-upgrade rule: 94.6% expect no rule: 97.1% expect	7071 games: 3.8% actual: 86% score 1-upgrade rule: 93.5% expect no rule: 94.6% expect			22444 games: 12.0% actual: 88.7% score 1-upgrade rule: 94.4% expect no rule: 96.9% expect
2400-2599	37 games: 0.0% actual: 97.3% score 1-upgrade rule: 92.4% expect no rule: 100% expect	196 games: 0.1% actual: 96.7% score 1-upgrade rule: 93.3% expect no rule: 100% expect	529 games: 0.3% actual: 96.7% score 1-upgrade rule: 94.5% expect no rule: 100% expect	1552 games: 0.8% actual: 95% score 1-upgrade rule: 94.8% expect no rule: 99.4% expect	4701 games: 2.5% actual: 91% score 1-upgrade rule: 94.4% expect no rule: 96.7% expect	2513 games: 1.3% actual: 90.1% score 1-upgrade rule: 93.3% expect no rule: 94.3% expect		9528 games: 5.1% actual: 91.9% score 1-upgrade rule: 94.2% expect no rule: 96.8% expect
2600-2799			17 games: 0.0% actual: 100% score 1-upgrade rule: 92% expect no rule: 100% expect	36 games: 0.0% actual: 97.2% score 1-upgrade rule: 93.6% expect no rule: 100% expect	146 games: 0.1% actual: 95.9% score 1-upgrade rule: 94.6% expect no rule: 99.3% expect	442 games: 0.2% actual: 93.8% score 1-upgrade rule: 93.8% expect no rule: 96.1% expect	89 games: 0.0% actual: 92.1% score 1-upgrade rule: 92.7% expect no rule: 93.5% expect	731 games: 0.4% actual: 94.3% score 1-upgrade rule: 93.8% expect no rule: 96.7% expect
Total	48925 games: 26.1% actual: 85.8% score 1-upgrade rule: 92.9% expect no rule: 96.8% expect	56746 games: 30.2% actual: 83.7% score 1-upgrade rule: 93.8% expect no rule: 96.6% expect	42659 games: 22.7% actual: 84.1% score 1-upgrade rule: 94% expect no rule: 96.3% expect	24464 games: 13.0% actual: 85.9% score 1-upgrade rule: 94% expect no rule: 95.9% expect	11918 games: 6.3% actual: 88.1% score 1-upgrade rule: 93.9% expect no rule: 95.5% expect	2955 games: 1.6% actual: 90.7% score 1-upgrade rule: 93.4% expect no rule: 94.6% expect	89 games: 0.0% actual: 92.1% score 1-upgrade rule: 92.7% expect no rule: 93.5% expect	187756 games: 100.0% actual: 85% score 1-upgrade rule: 93.6% expect no rule: 96.4% expect

LEGEND:	
BLUE	indicates a matchup where the original 400-point-rule (unlimited upgrades to 92%) does best at matching the actual performance results
RED	indicates a matchup where the current 400-point rule (maximum of 1 upgrade to 92%) from 2022 does best at matching the actual performance results
YELLOW	indicates a matchup where completely removing the 400-point-rule does best at matching the actual performance results
INDIGO	indicates a matchup where the original 400-point-rule works best, but the current rule (maximum of 1 upgrade to 92%) is almost as good
BURGUNDY	indicates a matchup where the current rule (maximum of 1 upgrade to 92%) works best, but the original 400-point-rule is almost as good
TANGERINE	indicates a matchup where completely removing the 400-point-rule works best, but the current rule (maximum of 1 upgrade to 92%) is almost as good
MARIGOLD	indicates a matchup where the current rule (maximum of 1 upgrade to 92%) works best, but completely removing the 400-point-rule is almost as good
	indicates a matchup where there were fewer than 5 games played with a rating difference of 412+

Now we see that in the current era, even the IM/GM level of player is not able to exploit the 400-point-rule very well. Certainly, for the ridiculously one-sided games, the 2400+ Elo group is still scoring near 100% percent, but that's only a tiny handful of games now. In fact, barely 500 games' worth are color-coded yellow here, out of an overall total of almost 188,000 Extreme-Elo-Difference games. Just about everyone is scoring below 92% in these matchups lately, and so now the blue color-coding dominates this graphic, indicating that at present, the 400-point rule is clearly the best-performing approach to determining the expected score of these games.

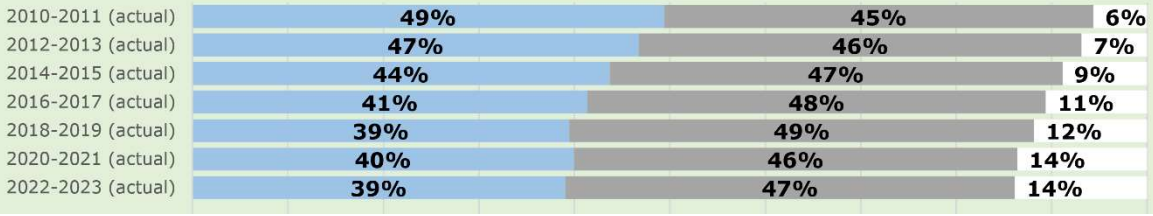
So if we were only considering the 400-point-rule in isolation, this would be a very easy choice, mathematically speaking; the overwhelming evidence shows that we should revert the 400-point-rule back to its "blue" form, as it was prior to 2022.

However, we are also seriously contemplating a compression of the ratings of players who are rated below 2000 Elo. This will move player ratings closer together, in a direct attempt to bring expectations more in line with actual results. And of course, this will directly affect the effectiveness of the 400-point-rule. It might even make the 1-upgrade rule (as enacted in 2022-JAN) more appealing, or perhaps we should even eliminate the 400-point-rule in any form. So it's time to check those color-coding schemes for the simulations, and see what they tell us!

However, one thing to mention first: over the years, the rating pool has stretched out and we have also seen a large influx of mostly lower-rated players. That would suggest that we've been seeing more and more games with extreme rating differences, and indeed this is true. We can easily produce a bar chart showing, for each two-year-period, what percentage of rated games had an expectation of 50%-70% for the higher-rated player, versus a 71%-92% expectation, versus the category we are focused on right now, the Extreme-Elo-Difference games where the Elo table would give an expectation of 93%-100%. Let's look at how the game percentages have changed over time, with blue, gray, and white coloring for those three categories:

Games with large rating differences have become more and more frequent.
This trend would be reversed by a Compression to 1200 or to 1400.

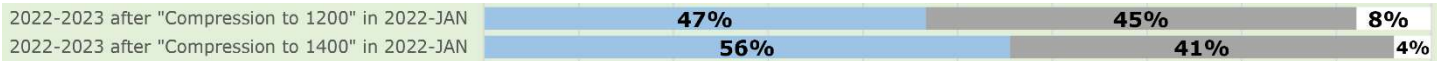
BLUE	shows the % of games with rating difference 0-153 (e.g. 50% to 70% expectation)
GRAY	shows the % of games with rating difference 154-411 (e.g. 71% to 92% expectation)
WHITE	shows the % of games with rating difference 412+ (e.g. 93% or higher expectation)



So we can see, from looking at the white bars over on the right, that originally (in 2010-2011) only 6% of rated games had a rating difference as extreme as 412+ Elo points, but more recently (such as in 2020-2023) that number has more than doubled, to the point where currently, 14% of all rated games have a rating difference of 412+.

However, the 400-point-rule would become less impactful if we were also doing a compression of lower-rated players (those currently rated 1000-2000 Elo). A compression pushes ratings closer together for many players, leading to a chess world with far fewer games having extreme rating differences. So those white bars ought to get a lot smaller in the case of a compression.

At this stage, I would say the two finalists for how much to compress, are either to a minimum of 1400 Elo (as stated in my July proposal) or a half-compression, to a minimum of 1200 Elo instead. I have simulated both of these compressions as though they happened in January of 2022, and then looked to see what would have happened to rated matchups during 2022-2023 after that (using simulated ratings). And indeed, those white numbers would drop significantly, with far fewer Extreme-Elo-Difference games.



Regardless, it's still important to get the 400-point-rule behaving as well as possible. So the final test would be to see what happens in these two simulations. Firstly, if we compressed to a minimum of 1200 Elo, so that 8% of all games were Extreme-Elo-Difference, what happens to our color-coded chart about the 400-point-rule options, for the games that were played subsequently? We can pretend that all the ratings really did change from the compression at the start of 2022, and then see what would have happened in all the games where the (simulated) ratings were different by 412+ Elo points. As we've seen, there would be fewer Extreme-Elo-Difference games than what really happened, but there would still have been a bunch of them.

It will certainly shift a bit away from blue and toward a bit more yellow and red, but would it shift sufficiently to change our conclusions about which approach works best? Let's see:

(D) -- Extreme-Elo-Difference games (those having Elo difference of 412+) that were played from 01-JAN-2022 through 31-MAR-2023 (after a simulated "Compression to 1200" on 01-JAN-2022)

HIGHER-RATED PLAYER \ LOWER-RATED PLAYER	1000-1199	1200-1399	1400-1599	1600-1799	1800-1999	2000-2199	2200-2399	Total
1400-1599								
1600-1799		13976 games: 13.1% actual: 86.3% score 1-upgrade rule: 92.4% expect no rule: 94.5% expect						13976 games: 13.1% actual: 86.3% score 1-upgrade rule: 92.4% expect no rule: 94.5% expect
1800-1999		17080 games: 16.0% actual: 89% score 1-upgrade rule: 92.9% expect no rule: 96.8% expect	17783 games: 16.6% actual: 85.8% score 1-upgrade rule: 93% expect no rule: 94.6% expect					34863 games: 32.6% actual: 87.3% score 1-upgrade rule: 92.9% expect no rule: 95.7% expect
2000-2199		3100 games: 2.9% actual: 94.6% score 1-upgrade rule: 93.3% expect no rule: 99.5% expect	14508 games: 13.6% actual: 89.2% score 1-upgrade rule: 93.9% expect no rule: 96.9% expect	11361 games: 10.6% actual: 86% score 1-upgrade rule: 93.3% expect no rule: 94.6% expect				28969 games: 27.1% actual: 88.5% score 1-upgrade rule: 93.6% expect no rule: 96.3% expect
2200-2399		606 games: 0.6% actual: 97.1% score 1-upgrade rule: 93.1% expect no rule: 100% expect	2664 games: 2.5% actual: 94.2% score 1-upgrade rule: 94.7% expect no rule: 99.5% expect	8814 games: 8.2% actual: 90.2% score 1-upgrade rule: 94.4% expect no rule: 97% expect	7146 games: 6.7% actual: 86.7% score 1-upgrade rule: 93.4% expect no rule: 94.5% expect			19230 games: 18.0% actual: 89.7% score 1-upgrade rule: 94% expect no rule: 96.5% expect
2400-2599		58 games: 0.1% actual: 96.6% score 1-upgrade rule: 92.6% expect no rule: 100% expect	372 games: 0.3% actual: 96.9% score 1-upgrade rule: 93.6% expect no rule: 100% expect	1178 games: 1.1% actual: 96.3% score 1-upgrade rule: 94.6% expect no rule: 99.4% expect	4841 games: 4.5% actual: 92% score 1-upgrade rule: 94.3% expect no rule: 96.7% expect	2611 games: 2.4% actual: 90.3% score 1-upgrade rule: 93.3% expect no rule: 94.3% expect		9060 games: 8.5% actual: 92.3% score 1-upgrade rule: 94% expect no rule: 96.5% expect
2600-2799			5 games: 0.0% actual: 100% score 1-upgrade rule: 92% expect no rule: 100% expect	34 games: 0.0% actual: 100% score 1-upgrade rule: 92.5% expect no rule: 100% expect	163 games: 0.2% actual: 95.7% score 1-upgrade rule: 94.5% expect no rule: 99.2% expect	471 games: 0.4% actual: 93.4% score 1-upgrade rule: 93.8% expect no rule: 96.1% expect	90 games: 0.1% actual: 92.2% score 1-upgrade rule: 92.7% expect no rule: 93.9% expect	763 games: 0.7% actual: 94.1% score 1-upgrade rule: 93.8% expect no rule: 96.6% expect
Total		34820 games: 32.6% actual: 88.6% score 1-upgrade rule: 92.7% expect no rule: 96.1% expect	35332 games: 33.1% actual: 87.9% score 1-upgrade rule: 93.5% expect no rule: 96% expect	21387 games: 20.0% actual: 88.3% score 1-upgrade rule: 93.8% expect no rule: 95.8% expect	12150 games: 11.4% actual: 88.9% score 1-upgrade rule: 93.8% expect no rule: 95.4% expect	3082 games: 2.9% actual: 90.8% score 1-upgrade rule: 93.4% expect no rule: 94.6% expect	90 games: 0.1% actual: 92.2% score 1-upgrade rule: 92.7% expect no rule: 93.5% expect	106861 games: 100.0% actual: 88.4% score 1-upgrade rule: 93.3% expect no rule: 95.9% expect

LEGEND:	
BLUE	indicates a matchup where the original 400-point-rule (unlimited upgrades to 92%) does best at matching the actual performance results
RED	indicates a matchup where the current 400-point rule (maximum of 1 upgrade to 92%) from 2022 does best at matching the actual performance results
YELLOW	indicates a matchup where completely removing the 400-point-rule does best at matching the actual performance results
INDIGO	indicates a matchup where the original 400-point-rule works best, but the current rule (maximum of 1 upgrade to 92%) is almost as good
BURGUNDY	indicates a matchup where the current rule (maximum of 1 upgrade to 92%) works best, but the original 400-point-rule is almost as good
TANGERINE	indicates a matchup where completely removing the 400-point-rule works best, but the current rule (maximum of 1 upgrade to 92%) is almost as good
MARIGOLD	indicates a matchup where the current rule (maximum of 1 upgrade to 92%) works best, but completely removing the 400-point-rule is almost as good
	indicates a matchup where there were fewer than 5 games played with a rating difference of 412+

For starters, after a compression to 1200 Elo minimum, we would no longer have any players rated 1000-1199 Elo, and so the leftmost column is completely cleared out. And it becomes impossible for a player rated 1400-1599 Elo to ever have a rating advantage of 412+ Elo points, and so the topmost row is also cleared out. But overall, the compression does not change the picture too significantly.

It's still true that the top-colored row is Indigo (because after a compression to 1200 Elo, players rated 1600-1799 Elo would struggle to find more than one extremely-low-rated opponent in the same event) and that the strongest players can still dominate the one-sided matchups. But again, only a tiny percentage of games are color-coded yellow, and the rightmost column indicates strongly that blue is the optimal solution for almost the whole spectrum of higher-rated players. So if we do a compression to a minimum of 1200 Elo, then the best choice would still seem to be reverting to the original 400-point-rule. Overall, the extreme rating favorites would still be scoring below 89% on average (you can confirm this by looking at the stats in the bottom-right blue cell).

Finally, what about my proposed approach, namely a compression to a minimum of 1400 Elo? This is the largest compression under consideration, and would lead to the most severe shrinkage of typical rating differences. But again the question is, would it be enough to change our assessment as to what the fairest version of 400-point-rule would be? Let's see what would have happened in 2022-2023 if a compression to a minimum 1400 Elo had happened at the start of 2022:

(E) -- Extreme-Elo-Difference games (those having Elo difference of 412+) that were played from 01-JAN-2022 through 31-MAR-2023 (after a simulated "Compression to 1400" on 01-JAN-2022)

HIGHER-RATED PLAYER \ LOWER-RATED PLAYER	1000-1199	1200-1399	1400-1599	1600-1799	1800-1999	2000-2199	2200-2399	Total
1400-1599								
1600-1799								
1800-1999			11013 games: 21.5% actual: 90.9% score 1-upgrade rule: 92.3% expect no rule: 94.3% expect					11013 games: 21.5% actual: 90.9% score 1-upgrade rule: 92.3% expect no rule: 94.3% expect
2000-2199			7668 games: 15.0% actual: 92.6% score 1-upgrade rule: 93% expect no rule: 96.6% expect	8593 games: 16.8% actual: 90.2% score 1-upgrade rule: 93% expect no rule: 94.5% expect				16261 games: 31.7% actual: 91.3% score 1-upgrade rule: 93% expect no rule: 95.5% expect
2200-2399			1246 games: 2.4% actual: 95.9% score 1-upgrade rule: 93.5% expect no rule: 99.5% expect	6799 games: 13.3% actual: 92.2% score 1-upgrade rule: 94% expect no rule: 96.8% expect	6765 games: 13.2% actual: 89.5% score 1-upgrade rule: 93.3% expect no rule: 94.5% expect			14810 games: 28.9% actual: 91.3% score 1-upgrade rule: 93.6% expect no rule: 96% expect
2400-2599			136 games: 0.3% actual: 97.1% score 1-upgrade rule: 93.1% expect no rule: 100% expect	859 games: 1.7% actual: 97.3% score 1-upgrade rule: 94.1% expect no rule: 99.4% expect	4847 games: 9.5% actual: 92.5% score 1-upgrade rule: 94% expect no rule: 96.5% expect	2581 games: 5.0% actual: 91% score 1-upgrade rule: 93.3% expect no rule: 94.3% expect		8423 games: 16.4% actual: 92.6% score 1-upgrade rule: 93.8% expect no rule: 96.2% expect
2600-2799				23 games: 0.0% actual: 100% score 1-upgrade rule: 92.3% expect no rule: 100% expect	159 games: 0.3% actual: 96.2% score 1-upgrade rule: 94.1% expect no rule: 99.1% expect	473 games: 0.9% actual: 93.7% score 1-upgrade rule: 93.8% expect no rule: 96.1% expect	92 games: 0.2% actual: 94% score 1-upgrade rule: 92.7% expect no rule: 93.5% expect	747 games: 1.5% actual: 94.4% score 1-upgrade rule: 93.7% expect no rule: 96.5% expect
Total			20063 games: 39.1% actual: 91.9% score 1-upgrade rule: 92.7% expect no rule: 95.6% expect	16274 games: 31.8% actual: 91.4% score 1-upgrade rule: 93.5% expect no rule: 95.7% expect	11771 games: 23.0% actual: 90.8% score 1-upgrade rule: 93.6% expect no rule: 95.4% expect	3054 games: 6.0% actual: 91.4% score 1-upgrade rule: 93.3% expect no rule: 94.6% expect	92 games: 0.2% actual: 94% score 1-upgrade rule: 92.7% expect no rule: 93.5% expect	51254 games: 100.0% actual: 91.5% score 1-upgrade rule: 93.2% expect no rule: 95.5% expect

LEGEND:	
BLUE	indicates a matchup where the original 400-point-rule (unlimited upgrades to 92%) does best at matching the actual performance results
RED	indicates a matchup where the current 400-point rule (maximum of 1 upgrade to 92%) from 2022 does best at matching the actual performance results
YELLOW	indicates a matchup where completely removing the 400-point-rule does best at matching the actual performance results
INDIGO	indicates a matchup where the original 400-point-rule works best, but the current rule (maximum of 1 upgrade to 92%) is almost as good
BURGUNDY	indicates a matchup where the current rule (maximum of 1 upgrade to 92%) works best, but the original 400-point-rule is almost as good
TANGERINE	indicates a matchup where completely removing the 400-point-rule works best, but the current rule (maximum of 1 upgrade to 92%) is almost as good
MARIGOLD	indicates a matchup where the current rule (maximum of 1 upgrade to 92%) works best, but completely removing the 400-point-rule is almost as good
	indicates a matchup where there were fewer than 5 games played with a rating difference of 412+

First of all, we have to eliminate another entire left column, since there would no longer be any players rated 1200-1399 Elo. And we eliminate another top row, since there would no longer be any players rated 1600-1799 Elo having any games with a rating advantage of 412+ points. We are getting fewer and fewer color-coded boxes!

Nevertheless, it's still true that the strong conclusion (see the colors of the right-hand totals column) is that the original 400-point-rule would still be the most appropriate rule to use, for the performance data that we are seeing. Yes, players rated 2600+ could exploit it somewhat, but only to the tune of a 94% score in all games when they have an extreme Elo rating advantage, and that score is far from a sure thing. And anyway, only 1.5% of Extreme-Elo-Difference games would involve any players rated 2600+, whereas more than half of all Extreme-Elo-Difference games would be played between non-master-level players (specifically, more than 53% of Extreme-Elo-Difference games would have the rating favorite being rated below 2200 Elo).

Because so many of our left columns and top rows were getting blanked out into useless green boxes (due to the compressions), I tried a slightly different way of looking at the color-coding scheme for this final scenario. Namely, I switched over to rating groups that were only 100 Elo points wide, rather than being 200 Elo points wide. I also eliminated the all-green rows from the top, and the all-green columns from the left. So we get a lot more boxes shown, and more detail conveyed, albeit fewer games represented in each box. Nevertheless, the story does not change too much:

(E2) -- Extreme-Elo-Difference games (those having Elo difference of 412+) that were played from 01-JAN-2022 through 31-MAR-2023 (after a simulated "Compression to 1400" on 01-JAN-2022) and looking at 100-Elo-point groups rather than 200-Elo-point

HIGHER-RATED PLAYER \ LOWER-RATED PLAYER	1400-1499	1500-1599	1600-1699	1700-1799	1800-1899	1900-1999	2000-2099	2100-2199	2200-2299	Total
1800-1899	2006 games: 5.9% actual: 90.5% score 1-upgrade rule: 92.1% expect no rule: 93.6% expect									2006 games: 5.9% actual: 90.5% score 1-upgrade rule: 92.1% expect no rule: 93.6% expect
1900-1999	4744 games: 9.3% actual: 91.3% score 1-upgrade rule: 92.3% expect no rule: 95.2% expect	3463 games: 6.8% actual: 90% score 1-upgrade rule: 92.4% expect no rule: 93.7% expect								8207 games: 16.0% actual: 91% score 1-upgrade rule: 92.4% expect no rule: 94.6% expect
2000-2099	1352 games: 2.6% actual: 95% score 1-upgrade rule: 92.7% expect no rule: 97.6% expect	4153 games: 8.1% actual: 90.9% score 1-upgrade rule: 93% expect no rule: 95.4% expect	2516 games: 4.9% actual: 88.9% score 1-upgrade rule: 92.7% expect no rule: 93.7% expect							8021 games: 15.6% actual: 91% score 1-upgrade rule: 92.8% expect no rule: 95.3% expect
2100-2199	563 games: 1.1% actual: 97% score 1-upgrade rule: 92.8% expect no rule: 99.1% expect	1600 games: 3.1% actual: 93.2% score 1-upgrade rule: 93.5% expect no rule: 97.9% expect	3913 games: 7.6% actual: 91.4% score 1-upgrade rule: 93.3% expect no rule: 95.4% expect	2164 games: 4.2% actual: 89.4% score 1-upgrade rule: 92.8% expect no rule: 93.7% expect						8240 games: 16.0% actual: 91.6% score 1-upgrade rule: 93.2% expect no rule: 95.7% expect
2200-2299	263 games: 0.5% actual: 97.5% score 1-upgrade rule: 92.8% expect no rule: 99.9% expect	627 games: 1.2% actual: 94.2% score 1-upgrade rule: 93.9% expect no rule: 99% expect	1505 games: 2.9% actual: 92.7% score 1-upgrade rule: 94.2% expect no rule: 97.8% expect	3303 games: 6.4% actual: 90.6% score 1-upgrade rule: 93.6% expect no rule: 95.4% expect	1890 games: 3.7% actual: 89.7% score 1-upgrade rule: 92.9% expect no rule: 93.7% expect					7588 games: 14.6% actual: 91.4% score 1-upgrade rule: 93.5% expect no rule: 95.8% expect
2300-2399	98 games: 0.2% actual: 98.5% score 1-upgrade rule: 92.5% expect no rule: 100% expect	258 games: 0.5% actual: 97.3% score 1-upgrade rule: 93.7% expect no rule: 99.9% expect	645 games: 1.3% actual: 95.9% score 1-upgrade rule: 94.4% expect no rule: 99.1% expect	1346 games: 2.6% actual: 93.7% score 1-upgrade rule: 94.5% expect no rule: 97.9% expect	2941 games: 5.7% actual: 90.7% score 1-upgrade rule: 93.7% expect no rule: 95.6% expect	1934 games: 3.8% actual: 87.4% score 1-upgrade rule: 93% expect no rule: 93.7% expect				7222 games: 14.1% actual: 91.2% score 1-upgrade rule: 93.7% expect no rule: 96% expect
2400-2499	24 games: 0.0% actual: 95.8% score 1-upgrade rule: 92.7% expect no rule: 100% expect	98 games: 0.2% actual: 97.4% score 1-upgrade rule: 93.1% expect no rule: 100% expect	254 games: 0.5% actual: 96.1% score 1-upgrade rule: 93.7% expect no rule: 99.9% expect	477 games: 0.9% actual: 97.4% score 1-upgrade rule: 94.4% expect no rule: 99.1% expect	1204 games: 2.3% actual: 94.6% score 1-upgrade rule: 94.4% expect no rule: 97.9% expect	2785 games: 5.4% actual: 90.5% score 1-upgrade rule: 93.9% expect no rule: 95.4% expect	1148 games: 2.2% actual: 90.2% score 1-upgrade rule: 92.9% expect no rule: 93.7% expect			5990 games: 11.7% actual: 92.2% score 1-upgrade rule: 93.8% expect no rule: 96.1% expect
2500-2599		12 games: 0.0% actual: 95.8% score 1-upgrade rule: 94% expect no rule: 100% expect	28 games: 0.1% actual: 100% score 1-upgrade rule: 92.6% expect no rule: 100% expect	100 games: 0.2% actual: 99% score 1-upgrade rule: 93.9% expect no rule: 99.8% expect	243 games: 0.5% actual: 97.1% score 1-upgrade rule: 93.8% expect no rule: 99% expect	615 games: 1.2% actual: 95.8% score 1-upgrade rule: 94.4% expect no rule: 97.8% expect	1045 games: 2.0% actual: 91.7% score 1-upgrade rule: 93.7% expect no rule: 95.2% expect	388 games: 0.8% actual: 91.2% score 1-upgrade rule: 93% expect no rule: 93.8% expect		2433 games: 4.7% actual: 93.6% score 1-upgrade rule: 93.8% expect no rule: 96.3% expect
2600-2699			6 games: 0.0% actual: 100% score 1-upgrade rule: 92% expect no rule: 100% expect	17 games: 0.0% actual: 100% score 1-upgrade rule: 92.5% expect no rule: 100% expect	44 games: 0.1% actual: 98.9% score 1-upgrade rule: 93.1% expect no rule: 99.7% expect	115 games: 0.2% actual: 95.2% score 1-upgrade rule: 94.5% expect no rule: 98.9% expect	175 games: 0.3% actual: 96% score 1-upgrade rule: 94.5% expect no rule: 97.6% expect	288 games: 0.6% actual: 92.2% score 1-upgrade rule: 93.4% expect no rule: 95.1% expect	88 games: 0.2% actual: 94.3% score 1-upgrade rule: 92.7% expect no rule: 93.4% expect	733 games: 1.4% actual: 94.5% score 1-upgrade rule: 93.7% expect no rule: 96.5% expect
Total	9852 games: 19.2% actual: 92.4% score 1-upgrade rule: 92.4% expect no rule: 95.5% expect	10211 games: 19.9% actual: 91.4% score 1-upgrade rule: 92.9% expect no rule: 95.6% expect	8867 games: 17.3% actual: 91.4% score 1-upgrade rule: 93.4% expect no rule: 95.8% expect	7407 games: 14.5% actual: 91.4% score 1-upgrade rule: 93.6% expect no rule: 95.7% expect	6322 games: 12.3% actual: 91.5% score 1-upgrade rule: 93.6% expect no rule: 95.6% expect	5449 games: 10.6% actual: 90.1% score 1-upgrade rule: 93.6% expect no rule: 95.2% expect	2369 games: 4.6% actual: 91.3% score 1-upgrade rule: 93.4% expect no rule: 94.6% expect	685 games: 1.3% actual: 91.7% score 1-upgrade rule: 93.2% expect no rule: 94.4% expect	91 games: 0.2% actual: 94.5% score 1-upgrade rule: 92.7% expect no rule: 93.5% expect	51254 games: 100.0% actual: 91.5% score 1-upgrade rule: 93.2% expect no rule: 95.5% expect

LEGEND:	
BLUE	indicates a matchup where the original 400-point-rule (unlimited upgrades to 92%) does best at matching the actual performance results
RED	indicates a matchup where the current 400-point rule (maximum of 1 upgrade to 92%) from 2022 does best at matching the actual performance
YELLOW	indicates a matchup where completely removing the 400-point-rule does best at matching the actual performance results
INDIGO	indicates a matchup where the original 400-point-rule works best, but the current rule (maximum of 1 upgrade to 92%) is almost as good
BURGUNDY	indicates a matchup where the current rule (maximum of 1 upgrade to 92%) works best, but the original 400-point-rule is almost as good
TANGERINE	indicates a matchup where completely removing the 400-point-rule works best, but the current rule (maximum of 1 upgrade to 92%) is almost as
MARIGOLD	indicates a matchup where the current rule (maximum of 1 upgrade to 92%) works best, but completely removing the 400-point-rule is almost as
	indicates a matchup where there were fewer than 5 games played with a rating difference of 412+

So even if we do the most severe compression, up to a minimum of 1400 Elo, the performance data still indicates pretty convincingly that reverting to the original 400-point-rule would be the best choice. There's lots of blue down the right Total edge and the bottom Total edge, and overall the higher-rated players would still be scoring less than 92% on average, in all these Extreme-Elo-Difference games.

For players rated 2500 Elo or above, the most accurate option would be to keep things as they have been since 2022 (with the "1-upgrade rule"), but even those players would not manage to reach a 95% score in their games with an extreme Elo rating advantage, and anyway, that's barely 6% of all the Extreme-Elo-Difference games even after the compression. Far more weighty is the evidence of actual percentage scores below 92% for the Extreme-Elo-Difference games having both players rated below 2400 Elo, and that would be well above 80% of all Extreme-Elo-Difference games.

It is true that the whole point of the compression is to reduce rating differences, so it may seem odd that the original 400-point-rule would still be needed. I think the answer lies in the fact that we are not addressing all of the deflation, even with the compression. It would be appropriate mathematically to lower the ratings of the highest-rated players (everyone above 2000) by some amount, bringing them closer to the weaker players, but this would be a big practical problem, and thus we are instead relying on the Elo system itself to sort them out over a longer period of time, aided by the additional rating points being added into the rating pool below 2000 Elo, and the additional draws being added to initial ratings to resist the deflation's return.

I should point out that I also tried this analysis under the scenario where the compressions to 1200 Elo or 1400 Elo (and the correction improvements) happened at the start of 2017, rather than the start of 2022, and it's still true that even years after the compression, the reversion of the 400-point-rule to its previous state is still the recommended action, based on this same color-coding methodology.

Thus I am quite comfortable in concluding that, independent of any other decisions, **we should revert the 400-point-rule to its pre-2022 state**, where any rating difference larger than 400 Elo points would instead be treated as a difference of 400 Elo points.

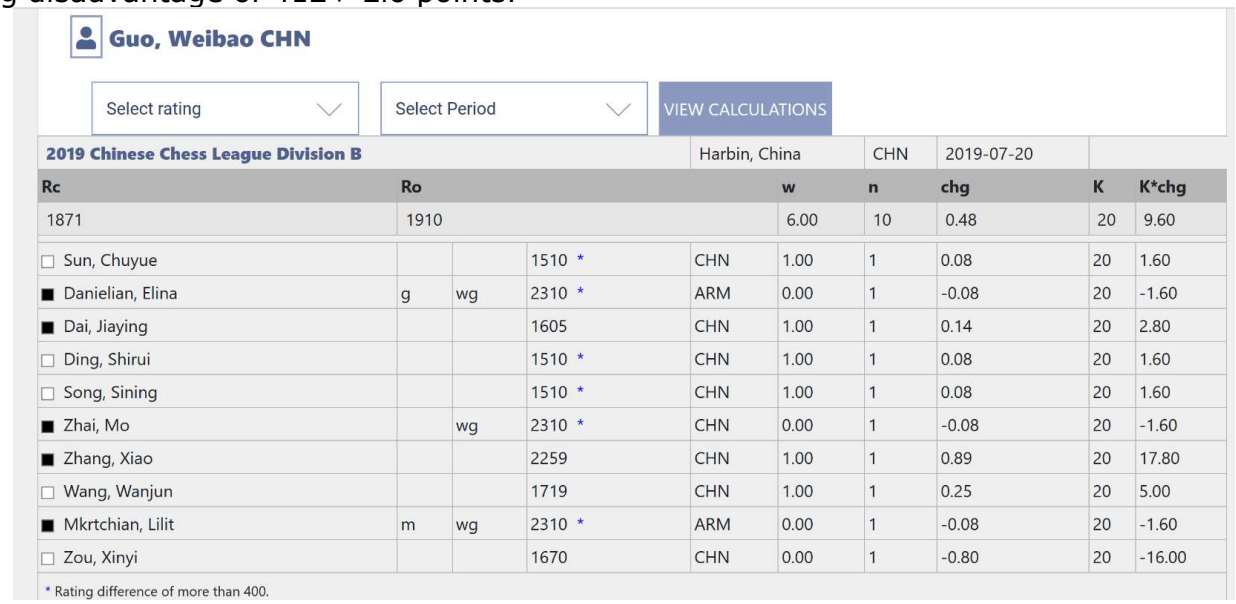
Appendix #5: Treatment of lower-rated player for 400-point-rule

For many years (prior to 2022) the 400-point-rule was worded quite simply:

> 8.54 A difference in rating of more than 400 points shall be counted for rating purposes as though it were a difference of 400 points.

This rule was applied equally to the players who had rating advantages of 400+ points (giving them an expected score of 0.92 for those games) and also to the players who had rating disadvantages of 400+ points (giving them an expected score of 0.08 for those games). In fact, some people ended up serving as both types of player during the same event! These players serve as good illustrations of the 400-point-rule, since their tournament performance page for that event (from the FIDE website) can show, in one screenshot, how the 400-point-rule functions from the perspectives of both the higher-rated player and also the lower-rated player.

Below is a particularly extreme case from 2019, where a single player had three games in one event with a rating advantage of 412+ Elo points, as well as three games in the same event with a rating disadvantage of 412+ Elo points!



Guo, Weibao CHN

Select rating | Select Period | VIEW CALCULATIONS

2019 Chinese Chess League Division B				Harbin, China		CHN	2019-07-20		
Rc	**Ro**				**w**	**n**	**chg**	**K**	**K*chg**
1871	1910				6.00	10	0.48	20	9.60
☐ Sun, Chuyue			1510 *	CHN	1.00	1	0.08	20	1.60
■ Danielian, Elina	g	wg	2310 *	ARM	0.00	1	-0.08	20	-1.60
■ Dai, Jiaying			1605	CHN	1.00	1	0.14	20	2.80
☐ Ding, Shirui			1510 *	CHN	1.00	1	0.08	20	1.60
☐ Song, Sining			1510 *	CHN	1.00	1	0.08	20	1.60
■ Zhai, Mo		wg	2310 *	CHN	0.00	1	-0.08	20	-1.60
■ Zhang, Xiao			2259	CHN	1.00	1	0.89	20	17.80
☐ Wang, Wanjun			1719	CHN	1.00	1	0.25	20	5.00
■ Mkrtchian, Lilit	m	wg	2310 *	ARM	0.00	1	-0.08	20	-1.60
☐ Zou, Xinyi			1670	CHN	0.00	1	-0.80	20	-16.00

* Rating difference of more than 400.

In this event from 2019, the player (Guo Weibao) had a rating of 1910, and faced three opponents who were rated more than 400 points below her: Sun Chuyue (rated 1020), Ding Shirui (rated 1463), and Song Sining (rated 1219). Due to the 400-point-rule, Guo Weibao was given an expected score of 0.92 for each of those games (she won all three games) and thus with her K=20, she ultimately gained 4.80 rating points across those three games combined. This was a correct application of the 400-point-rule as of 2019. You can see that all three of these opponents have an asterisk next to their rating, indicating the use of the 400-point-rule.

Amusingly, Guo Weibao also faced three opponents who were rated more than 400 points higher than her in that same event: Elina Danielian (rated 2417), Zhai Mo (rated 2351), and Lilit Mkrtchian (rated 2395). Due to the 400-point-rule, Guo Weibao was given an expected score of 0.08 for each of those games (she lost all three games) and thus with her K=20, she ultimately lost 4.80 rating points across those three games combined. This, again, was a correct application of the 400-point-rule as of 2019. And you can see that all three of these opponents have an asterisk next to their ratings as well, indicating the use of the 400-point-rule.

Now let's jump forward a few years to a very analogous situation but in late 2022. By this point, the new "1-upgrade rule" had gone into effect, based on the following regulation:

B. Permanent Commissions / 02. FIDE Rating Regulations (Qualification Commission) / FIDE Rating Regulations effective from 1 January 2022 /

8.3.1 For each game played against a rated player, determine the difference in rating between the player and their opponent, D.

A difference in rating of more than 400 points shall be counted for rating purposes as though it were a difference of 400 points. In any tournament, a player may benefit from only one upgrade under this rule, for the game in which the rating difference is greatest.

We once again see a player who was faced with three opponents having much lower ratings, and three other opponents having much higher ratings, all within the same event in late 2022:

Sheka, Nikolay RUS

Select rating | Select Period | VIEW CALCULATIONS

XVIII Gubernator's Cup Ugra 2022				Khanty-Mansiysk	RUS	2022-11-17		
Rc	**Ro**			**w**	**n**	**chg**	**K**	**K*chg**
2059	1911			5.00	9	1.57	38	59.66
■ Bocharov, Dmitry	g		2526	RUS	1.00	1	0.98	38	37.24
□ Ponkratov, Pavel	g		2593	RUS	0.00	1	-0.01	38	-0.38
■ Yeletsky, Ivan	f		2491	RUS	0.00	1	-0.02	38	-0.76
□ Yurov, Nikita			1431	RUS	1.00	1	0.05	38	1.90
■ Obolentseva, Alexandra		wg	2330	RUS	0.00	1	-0.07	38	-2.66
□ Bochkovskiy, Mikhail Kon			1511 *	RUS	1.00	1	0.08	38	3.04
■ Lupolov, Boris			2088	RUS	0.00	1	-0.27	38	-10.26
□ Sattarov, Dinis			1435	RUS	1.00	1	0.05	38	1.90
■ Ufimtsev, Egor			2128	RUS	1.00	1	0.78	38	29.64

* Rating difference of more than 400.

In this event from 2022, the player (Nikolay Sheka) had a rating of 1911, and faced three opponents who were rated more than 400 points below him: Nikita Yurov (rated 1431), Mikhail Kon Bochkovskiy (rated 1074), and Dinis Sattarov (rated 1435). Due to the changes to the 400-point-rule, Nikolay Sheka was only "upgraded" to an expected score of 0.92 for the lowest-rated opponent (Mikhail Kon Bochkovskiy) whereas for the other two heavily-outrated players, he was given an expected score taken from the Elo table (0.95 for each game). With an effective K-factor of 38 (due to the number of games played), and an expected score of 2.82 rather than 2.76 for these three games, Nikolay Sheka ultimately gained 2.28 fewer rating points than he would have from the earlier 400-point-rule, across these three games. Again, this was a correct implementation of the 400-point-rule as of 2022. And you can see that only one of those three opponents has an asterisk by their rating, indicating an application of the current 400-point-rule.

Finally, this takes us to the problematic scenario. Nikolay Sheka had three opponents rated more than 400 points above him: Dmitry Bocharov (rated 2526), and Ivan Yeletsky (rated 2491), and Alexandra Obolentseva (rated 2330). Like all Extreme-Elo-Difference games since 2022-JAN, none of those games had the 400-point-rule applied to the rating calculation of the lower-rated player. Thus his expected score in the three games combined was treated as 0.11, rather than the combined 0.24 that it would have been from the original 400-point-rule. And so, with an effective K-factor of 38, Nikolay Sheka ultimately gained 4.94 more rating points than he would have from the earlier 400-point-rule, from these three games. This does not look correct to me. Evidently the 400-point-rule is no longer being applied for the lower-rated players.

This calculation behavior seems to have started at the beginning of 2022. I would assume this is either an unintentional programming error that was implemented at the same time as the rollout of the "maximum of one upgrade" rule, or else an intentional choice to skip the 400-point-rule for players having an extreme rating disadvantage. Perhaps the reasoning was that having a rating disadvantage of more than 400 points does not count as "a rating difference of more than 400 points", since it is a negative number? I don't know the answer, but either way it seems like this should be changed back to how it used to work prior to 2022. Although it is an inflationary effect (and so it helps a bit with the deflation), I still don't think this behavior is in accordance with the regulations.

Unfortunately, if it was an unintentional programming error, it has already affected a large number of games. We have already seen, in the previous appendix for the 400-point-rule, that currently 14% of all rated games are mismatched enough that they involve a rating difference of 412+ Elo points. According to the color-coding graphs in the previous appendix, specifically scenario (C) on page 34 of this document, there were 187,756 such games, just in a 15-month period starting in January 2022 and ending in April 2023. Surely it's up to more than 250,000 extreme Elo difference games by now, that have been played since January 2022.

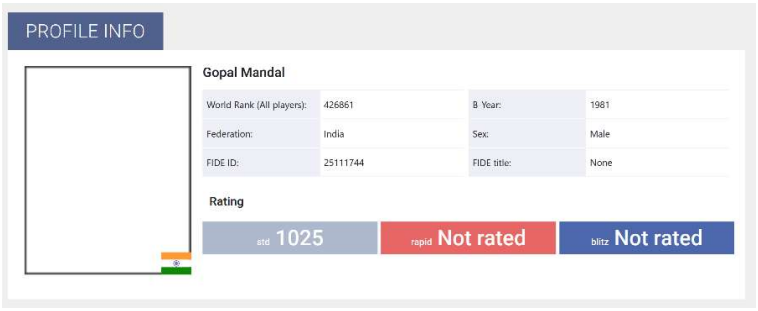
And in all of those games, the regulations imply that the lower-rated player should have had an expected score of 0.08, but instead they were presumably given an expected score between 0.00 and 0.07. It seems like we should fix this, as part of whatever changes are ultimately rolled out at the start of the year. I don't know whether we necessarily need to retroactively fix it, but we should at least fix it moving forward.

Appendix #6: Should we compress rapid/blitz now as well?

Most of my analyses and simulations have been confined to FIDE's standard Elo rating system rather than its rapid or blitz systems. For the most part, the three systems are unrelated and do not affect each other, and decisions about whether to do a compression could be made independently for each of the three systems.

However, there is one scenario where they are certainly connected. If a player participates in a rapid or blitz event, and does not already have an applicable rapid/blitz rating for the event, but they do already have a standard rating, then their standard rating is used as their new rapid/blitz rating for the event. This leads to a possible problem if there is a scenario where compressions happen at different times. Let's pretend that a compression to 1400 Elo happens in January 2024 to standard ratings, like I have suggested, and that there is eventually a decision to do similar rapid/blitz compressions that happen in January 2025. And let's first consider a player (Khan Samayara is a great example) who is currently rated 1025 in both standard and rapid:

After a Compression to 1400 Elo for standard ratings, Khan Samayara would gain 390 points to her standard rating, and so she would have a standard rating of 1415 Elo and a rapid rating of 1025 Elo. If she doesn't play much during 2024 (perhaps one or two standard events, and one or two rapid events), then maybe she still has a standard 1415 Elo and a rapid 1025 Elo at the end of the year. Then we do a Compression to 1400 Elo for rapid ratings, and her rapid rating would also gain 390 points. So at that point she would have a standard rating of 1415 Elo and a rapid rating of 1415. This all sounds perfectly fine, since the current evidence is that she's roughly the same strength at both standard and rapid. But now consider a different player (Gopal Mandal is a great example), who also has a 1025 standard rating but doesn't yet have a rapid rating:

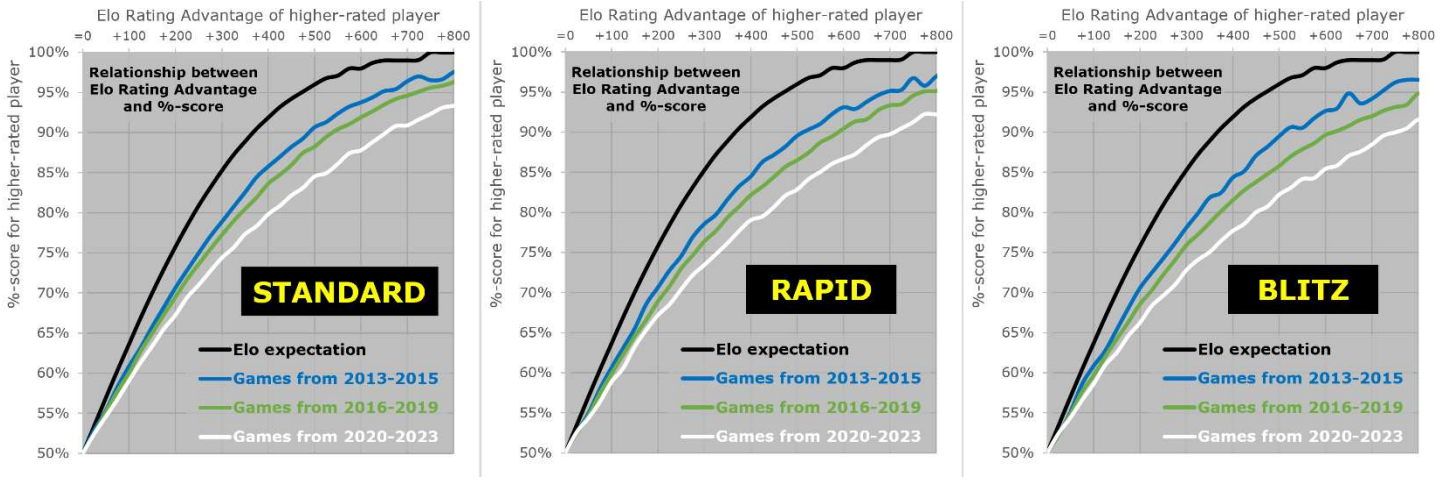


After a Compression to 1400 Elo for standard ratings, Gopal Mandal would gain 390 points to his standard rating, giving him a standard rating of 1415 Elo. If he then participates in a rapid event sometime later in 2024, without having a rapid rating, his rapid rating will get initialized to 1415. He might still be facing others with comparable strength to him, but if so then they would

have (uncompressed) rapid ratings down near 1000 Elo. Further, when that rapid compression happens at the start of 2025, his rapid rating of 1415 would get increased by 234 Elo points, giving him a rapid rating of 1649, even though other players of comparable rapid strength would now get rapid ratings around 1415. His rapid rating would essentially have been "double-compressed". This really seems problematic, and of course the problem came from his rapid rating (within a rating system that had not been compressed) getting initialized using his standard rating (within a rating system that had been compressed). It would be far cleaner if all ratings were compressed at the same time, so that such a thing could not happen.

Of course, as I said previously, the bulk of my analyzing and simulating has been with the FIDE standard rating system, rather than the FIDE rapid or blitz rating systems. But nevertheless, the rapid and blitz regulations were recently adjusted to be very similar to standard, and so any conclusions that apply to standard chess ratings are likely to apply equally well to rapid and blitz chess ratings. So it may indeed be wise to implement the same changes to rapid and blitz, that are being made to standard.

However, it first seems prudent to do some basic comparisons of standard versus rapid and blitz, in order to make sure they really are behaving similarly. For example, at the heart of the justification for these changes, we looked at a line graph illustrating the relationship between Elo rating advantage and percentage score for the higher-rated player, where there was a black line showing the Elo expectation, and a white line showing that the actual performance of rating favorites in recent games has fallen far below the black line's prediction. Would such line graphs look similar if we made one for standard, one for rapid, and one for blitz?



The answer is clearly yes, all three tell quite the same story. In these historical graphs, we used a blue line showing how rating favorites did in 2013-2015, a green line showing how they did in 2016-2019, and a white line showing how they did in 2020-2023. And in all three rating systems, the rating favorites are struggling to reach their Elo expectations, and it's getting worse each year, with the latest data from 2020-2023 being shallowest of all.

Similarly, we drilled deeper into this situation for standard chess ratings (in the original proposal) by looking at red and blue heat maps, where we could see, for each matchup of players from one 100-point Elo group against another 100-point Elo group, whether players were overperforming (blue) or underperforming (red) in comparison to their Elo expectations. When we see primarily red (underperformance) below and to the left of the main diagonal, and blue (overperformance) above and to the right of the main diagonal, that's our special sign telling us that ratings everywhere are too stretched apart.

Let's look briefly at the heat maps for standard rated chess:

Standard 2013-2019	opponent	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	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+
+5.7% in 124,558 games	**1000-99**		**+2%**	**+5%**	**+7%**	**+7%**	**+8%**	**+6%**	**+5%**	**+4%**	**+3%**	**+2%**	**+2%**	**+1%**	**=0%**	**=0%**	**=0%**
+6.0% in 359,124 games	**1100-99**	**-2%**		**+4%**	**+6%**	**+7%**	**+9%**	**+8%**	**+7%**	**+5%**	**+4%**	**+3%**	**+2%**	**+1%**	**+0%**	**=0%**	**=0%**
+5.1% in 505,359 games	**1200-99**	**-5%**	**-4%**		**+4%**	**+7%**	**+9%**	**+10%**	**+9%**	**+7%**	**+5%**	**+4%**	**+3%**	**+1%**	**+0%**	**+1%**	**=0%**
+3.9% in 673,685 games	**1300-99**	**-7%**	**-6%**	**-4%**		**+4%**	**+7%**	**+10%**	**+10%**	**+9%**	**+7%**	**+5%**	**+4%**	**+3%**	**+2%**	**+1%**	**+0%**
+2.7% in 841,968 games	**1400-99**	**-7%**	**-7%**	**-7%**	**-4%**		**+5%**	**+8%**	**+9%**	**+10%**	**+8%**	**+7%**	**+5%**	**+4%**	**+2%**	**+1%**	**=0%**
+1.8% in 1,003,996 games	**1500-99**	**-8%**	**-9%**	**-9%**	**-7%**	**-5%**		**+5%**	**+8%**	**+9%**	**+9%**	**+8%**	**+6%**	**+4%**	**+3%**	**+2%**	**+0%**
+0.9% in 1,184,219 games	**1600-99**	**-6%**	**-8%**	**-10%**	**-10%**	**-8%**	**-5%**		**+4%**	**+7%**	**+8%**	**+8%**	**+7%**	**+5%**	**+4%**	**+2%**	**+1%**
-0.0% in 1,339,260 games	**1700-99**	**-5%**	**-7%**	**-9%**	**-10%**	**-9%**	**-8%**	**-4%**		**+4%**	**+6%**	**+8%**	**+7%**	**+6%**	**+5%**	**+3%**	**+2%**
-0.9% in 1,448,689 games	**1800-99**	**-4%**	**-5%**	**-7%**	**-9%**	**-10%**	**-9%**	**-7%**	**-4%**		**+4%**	**+6%**	**+7%**	**+6%**	**+5%**	**+4%**	**+1%**
-1.5% in 1,474,280 games	**1900-99**	**-3%**	**-4%**	**-5%**	**-7%**	**-8%**	**-9%**	**-8%**	**-6%**	**-4%**		**+3%**	**+5%**	**+6%**	**+5%**	**+4%**	**+2%**
-1.9% in 1,388,122 games	**2000-99**	**-2%**	**-3%**	**-4%**	**-5%**	**-7%**	**-8%**	**-8%**	**-8%**	**-6%**	**-3%**		**+3%**	**+4%**	**+5%**	**+4%**	**+2%**
-2.2% in 1,179,901 games	**2100-99**	**-2%**	**-2%**	**-3%**	**-4%**	**-5%**	**-6%**	**-7%**	**-7%**	**-7%**	**-5%**	**-3%**		**+2%**	**+4%**	**+3%**	**+2%**
-2.1% in 895,473 games	**2200-99**	**-1%**	**-1%**	**-1%**	**-3%**	**-4%**	**-4%**	**-5%**	**-6%**	**-6%**	**-6%**	**-4%**	**-2%**		**+2%**	**+3%**	**+2%**
-2.2% in 624,814 games	**2300-99**	**=0%**	**-0%**	**-0%**	**-2%**	**-2%**	**-3%**	**-4%**	**-5%**	**-5%**	**-5%**	**-5%**	**-4%**	**-2%**		**+2%**	**+1%**
-1.8% in 445,034 games	**2400-99**	**=0%**	**=0%**	**-1%**	**-1%**	**-1%**	**-2%**	**-2%**	**-3%**	**-4%**	**-4%**	**-4%**	**-3%**	**-3%**	**-2%**		**+1%**
-1.4% in 295,830 games	**2500+**	**=0%**	**=0%**	**=0%**	**-0%**	**=0%**	**-0%**	**-1%**	**-2%**	**-1%**	**-2%**	**-2%**	**-2%**	**-2%**	**-1%**	**-1%**	

Standard 2020-2023	opponent	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	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+
+7.1% in 57,570 games	**1000-99**		**+2%**	**+6%**	**+8%**	**+9%**	**+10%**	**+10%**	**+9%**	**+7%**	**+5%**	**+5%**	**+3%**	**+2%**	**+1%**	**=0%**	**=0%**
+7.7% in 177,866 games	**1100-99**	**-2%**		**+4%**	**+7%**	**+9%**	**+12%**	**+12%**	**+10%**	**+9%**	**+7%**	**+5%**	**+5%**	**+3%**	**+1%**	**+1%**	**=0%**
+6.6% in 243,244 games	**1200-99**	**-6%**	**-4%**		**+5%**	**+8%**	**+12%**	**+13%**	**+13%**	**+12%**	**+9%**	**+7%**	**+6%**	**+4%**	**+2%**	**+1%**	**=0%**
+5.2% in 287,149 games	**1300-99**	**-8%**	**-7%**	**-5%**		**+5%**	**+10%**	**+13%**	**+14%**	**+14%**	**+12%**	**+9%**	**+8%**	**+6%**	**+3%**	**+3%**	**+2%**
+3.6% in 324,167 games	**1400-99**	**-9%**	**-9%**	**-8%**	**-5%**		**+6%**	**+11%**	**+13%**	**+15%**	**+14%**	**+11%**	**+9%**	**+6%**	**+5%**	**+2%**	**+3%**
+1.6% in 351,789 games	**1500-99**	**-10%**	**-12%**	**-12%**	**-10%**	**-6%**		**+6%**	**+10%**	**+13%**	**+14%**	**+13%**	**+10%**	**+9%**	**+6%**	**+4%**	**+1%**
+0.1% in 371,733 games	**1600-99**	**-10%**	**-12%**	**-13%**	**-13%**	**-11%**	**-6%**		**+5%**	**+10%**	**+12%**	**+13%**	**+12%**	**+10%**	**+6%**	**+5%**	**+3%**
-1.2% in 389,169 games	**1700-99**	**-9%**	**-10%**	**-13%**	**-14%**	**-13%**	**-10%**	**-5%**		**+5%**	**+9%**	**+12%**	**+11%**	**+10%**	**+8%**	**+5%**	**+3%**
-2.1% in 397,555 games	**1800-99**	**-7%**	**-9%**	**-12%**	**-14%**	**-15%**	**-13%**	**-10%**	**-5%**		**+5%**	**+8%**	**+10%**	**+10%**	**+9%**	**+8%**	**+3%**
-3.0% in 381,482 games	**1900-99**	**-5%**	**-7%**	**-9%**	**-12%**	**-14%**	**-14%**	**-12%**	**-9%**	**-5%**		**+5%**	**+7%**	**+9%**	**+9%**	**+7%**	**+4%**
-3.7% in 339,210 games	**2000-99**	**-5%**	**-5%**	**-7%**	**-9%**	**-11%**	**-13%**	**-13%**	**-12%**	**-8%**	**-5%**		**+4%**	**+6%**	**+7%**	**+6%**	**+4%**
-3.7% in 281,808 games	**2100-99**	**-3%**	**-5%**	**-6%**	**-8%**	**-9%**	**-10%**	**-12%**	**-11%**	**-10%**	**-7%**	**-4%**		**+4%**	**+6%**	**+6%**	**+4%**
-3.7% in 215,503 games	**2200-99**	**-2%**	**-3%**	**-4%**	**-6%**	**-6%**	**-9%**	**-10%**	**-10%**	**-10%**	**-9%**	**-6%**	**-4%**		**+3%**	**+4%**	**+3%**
-3.5% in 156,420 games	**2300-99**	**-1%**	**-1%**	**-2%**	**-3%**	**-5%**	**-6%**	**-6%**	**-8%**	**-9%**	**-9%**	**-7%**	**-6%**	**-3%**		**+3%**	**+2%**
-3.1% in 116,816 games	**2400-99**	**=0%**	**-1%**	**-1%**	**-3%**	**-2%**	**-4%**	**-5%**	**-5%**	**-8%**	**-7%**	**-6%**	**-6%**	**-4%**	**-3%**		**+2%**
-2.4% in 75,793 games	**2500+**	**=0%**	**=0%**	**=0%**	**-2%**	**-3%**	**-1%**	**-3%**	**-3%**	**-3%**	**-4%**	**-4%**	**-4%**	**-3%**	**-2%**	**-2%**	

The top image shows the results from standard rated games played from 2013-2019, and the bottom image shows the results from standard rated games played from 2020-2023. In each case we see the red colors in the bottom left triangular region, the blue colors in the upper right triangular region, and the colors have become deeper in recent years (they are darker for 2020-2023 than for 2013-2019). So across the whole rating pool, rating favorites were already falling well short of their rating expectation, even five or ten years ago, and it's gotten worse in recent years. All of this was covered in more detail in my original proposal, if you want to read more about this way of illustrating the data (via the red and blue heat maps).

What about rapid and blitz? Well, as you might expect, the situation over there is very similar. We can generate comparable pairs of heat maps for rapid (left, below) and blitz (right, below):

Rapid 2013-2019	opponent	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	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+
+5.4% in 56,355 games	**1000-99**		**+1%**	**+4%**	**+6%**	**+8%**	**+8%**	**+8%**	**+6%**	**+5%**	**+4%**	**+3%**	**+2%**	**+0%**	**=0%**	**=0%**	**=0%**
+8.7% in 188,380 games	**1100-99**	**-1%**		**+3%**	**+6%**	**+9%**	**+10%**	**+10%**	**+8%**	**+7%**	**+5%**	**+4%**	**+3%**	**+2%**	**+2%**	**+0%**	**=0%**
+6.0% in 261,726 games	**1200-99**	**-4%**	**-3%**		**+4%**	**+7%**	**+10%**	**+11%**	**+11%**	**+10%**	**+7%**	**+6%**	**+4%**	**+2%**	**+1%**	**+1%**	**+0%**
+5.0% in 323,735 games	**1300-99**	**-6%**	**-6%**	**-4%**		**+4%**	**+8%**	**+11%**	**+12%**	**+11%**	**+9%**	**+7%**	**+6%**	**+4%**	**+2%**	**+1%**	**=0%**
+3.7% in 376,912 games	**1400-99**	**-8%**	**-9%**	**-7%**	**-4%**		**+5%**	**+9%**	**+11%**	**+12%**	**+11%**	**+9%**	**+6%**	**+5%**	**+3%**	**+3%**	**+1%**
+2.5% in 431,796 games	**1500-99**	**-8%**	**-10%**	**-10%**	**-8%**	**-5%**		**+5%**	**+8%**	**+11%**	**+11%**	**+10%**	**+9%**	**+6%**	**+4%**	**+3%**	**+1%**
+0.9% in 484,364 games	**1600-99**	**-8%**	**-10%**	**-11%**	**-11%**	**-9%**	**-5%**		**+4%**	**+8%**	**+9%**	**+10%**	**+9%**	**+7%**	**+5%**	**+3%**	**+2%**
-0.4% in 529,967 games	**1700-99**	**-6%**	**-8%**	**-11%**	**-12%**	**-11%**	**-8%**	**-4%**		**+4%**	**+7%**	**+8%**	**+8%**	**+8%**	**+5%**	**+4%**	**+2%**
-1.5% in 539,063 games	**1800-99**	**-5%**	**-7%**	**-10%**	**-11%**	**-12%**	**-11%**	**-8%**	**-4%**		**+4%**	**+6%**	**+8%**	**+7%**	**+6%**	**+4%**	**+3%**
-2.4% in 526,359 games	**1900-99**	**-4%**	**-5%**	**-7%**	**-9%**	**-11%**	**-11%**	**-9%**	**-7%**	**-4%**		**+4%**	**+6%**	**+6%**	**+6%**	**+5%**	**+3%**
-3.1% in 483,356 games	**2000-99**	**-3%**	**-4%**	**-6%**	**-7%**	**-9%**	**-10%**	**-10%**	**-8%**	**-6%**	**-4%**		**+3%**	**+4%**	**+5%**	**+5%**	**+4%**
-3.4% in 406,034 games	**2100-99**	**-2%**	**-3%**	**-4%**	**-6%**	**-6%**	**-9%**	**-9%**	**-8%**	**-8%**	**-6%**	**-3%**		**+3%**	**+4%**	**+4%**	**+4%**
-3.4% in 312,691 games	**2200-99**	**-0%**	**-2%**	**-2%**	**-4%**	**-5%**	**-6%**	**-7%**	**-8%**	**-7%**	**-6%**	**-4%**	**-3%**		**+2%**	**+3%**	**+4%**
-3.0% in 217,718 games	**2300-99**	**=0%**	**-2%**	**-1%**	**-2%**	**-3%**	**-4%**	**-5%**	**-5%**	**-6%**	**-6%**	**-5%**	**-4%**	**-2%**		**+3%**	**+4%**
-2.9% in 142,465 games	**2400-99**	**=0%**	**-0%**	**-1%**	**-1%**	**-3%**	**-3%**	**-4%**	**-4%**	**-5%**	**-5%**	**-4%**	**-3%**	**-3%**	**-2%**		**+3%**
-3.4% in 103,183 games	**2500+**	**=0%**	**=0%**	**-0%**	**=0%**	**-1%**	**-1%**	**-2%**	**-2%**	**-3%**	**-3%**	**-4%**	**-4%**	**-4%**	**-4%**	**-3%**	

Blitz 2013-2019	opponent	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	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 23,792 games	**1000-99**		**+2%**	**+7%**	**+9%**	**+9%**	**+10%**	**+10%**	**+7%**	**+7%**	**+4%**	**+3%**	**+1%**	**+2%**	**+1%**	**=0%**	**=0%**
+8.5% in 69,948 games	**1100-99**	**-2%**		**+5%**	**+8%**	**+10%**	**+12%**	**+12%**	**+11%**	**+8%**	**+7%**	**+5%**	**+5%**	**+3%**	**+3%**	**+1%**	**=0%**
+7.5% in 139,694 games	**1200-99**	**-7%**	**-5%**		**+5%**	**+9%**	**+11%**	**+13%**	**+12%**	**+11%**	**+8%**	**+7%**	**+5%**	**+3%**	**+2%**	**+2%**	**+0%**
+6.6% in 185,693 games	**1300-99**	**-9%**	**-8%**	**-5%**		**+5%**	**+10%**	**+13%**	**+13%**	**+13%**	**+10%**	**+8%**	**+7%**	**+4%**	**+3%**	**+1%**	**+1%**
+5.2% in 238,415 games	**1400-99**	**-9%**	**-10%**	**-9%**	**-5%**		**+5%**	**+9%**	**+12%**	**+13%**	**+12%**	**+10%**	**+8%**	**+5%**	**+4%**	**+2%**	**+1%**
+3.8% in 289,758 games	**1500-99**	**-10%**	**-12%**	**-11%**	**-10%**	**-5%**		**+5%**	**+9%**	**+12%**	**+12%**	**+11%**	**+9%**	**+7%**	**+5%**	**+2%**	**+2%**
+2.4% in 342,736 games	**1600-99**	**-10%**	**-12%**	**-13%**	**-13%**	**-9%**	**-5%**		**+5%**	**+9%**	**+11%**	**+11%**	**+10%**	**+7%**	**+6%**	**+4%**	**+3%**
+0.8% in 401,167 games	**1700-99**	**-7%**	**-11%**	**-12%**	**-13%**	**-12%**	**-9%**	**-5%**		**+4%**	**+8%**	**+9%**	**+9%**	**+8%**	**+6%**	**+4%**	**+3%**
-0.4% in 457,391 games	**1800-99**	**-7%**	**-8%**	**-11%**	**-13%**	**-13%**	**-12%**	**-9%**	**-4%**		**+5%**	**+7%**	**+8%**	**+8%**	**+7%**	**+5%**	**+3%**
-1.8% in 491,835 games	**1900-99**	**-4%**	**-7%**	**-8%**	**-10%**	**-12%**	**-12%**	**-11%**	**-8%**	**-5%**		**+4%**	**+6%**	**+7%**	**+6%**	**+5%**	**+4%**
-2.6% in 493,640 games	**2000-99**	**-3%**	**-5%**	**-7%**	**-8%**	**-10%**	**-11%**	**-11%**	**-9%**	**-7%**	**-4%**		**+3%**	**+5%**	**+6%**	**+5%**	**+4%**
-3.0% in 450,674 games	**2100-99**	**-1%**	**-5%**	**-5%**	**-7%**	**-8%**	**-9%**	**-10%**	**-9%**	**-8%**	**-6%**	**-3%**		**+3%**	**+4%**	**+5%**	**+4%**
-3.1% in 365,767 games	**2200-99**	**-2%**	**-3%**	**-3%**	**-4%**	**-5%**	**-7%**	**-7%**	**-8%**	**-8%**	**-7%**	**-5%**	**-3%**		**+3%**	**+4%**	**+4%**
-3.0% in 270,204 games	**2300-99**	**-1%**	**-3%**	**-2%**	**-3%**	**-4%**	**-5%**	**-6%**	**-6%**	**-7%**	**-6%**	**-6%**	**-4%**	**-3%**		**+2%**	**+4%**
-2.8% in 178,034 games	**2400-99**	**=0%**	**-1%**	**-2%**	**-1%**	**-2%**	**-2%**	**-4%**	**-5%**	**-5%**	**-5%**	**-5%**	**-4%**	**-2%**	**-2%**		**+3%**
-3.8% in 134,308 games	**2500+**	**=0%**	**=0%**	**-0%**	**-1%**	**-1%**	**-2%**	**-3%**	**-3%**	**-3%**	**-4%**	**-4%**	**-4%**	**-4%**	**-4%**	**-3%**	

Rapid 2020-2023	opponent	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	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+
+6.2% in 58,327 games	**1000-99**		**+2%**	**+5%**	**+7%**	**+9%**	**+10%**	**+10%**	**+8%**	**+7%**	**+6%**	**+6%**	**+4%**	**+1%**	**=0%**	**=0%**	**=0%**
+7.6% in 186,030 games	**1100-99**	**-2%**		**+3%**	**+6%**	**+9%**	**+12%**	**+13%**	**+11%**	**+10%**	**+8%**	**+7%**	**+5%**	**+3%**	**+2%**	**+3%**	**=0%**
+7.4% in 241,032 games	**1200-99**	**-5%**	**-3%**		**+5%**	**+9%**	**+12%**	**+14%**	**+14%**	**+13%**	**+10%**	**+9%**	**+7%**	**+6%**	**+2%**	**+1%**	**+0%**
+5.5% in 265,965 games	**1300-99**	**-7%**	**-6%**	**-5%**		**+5%**	**+9%**	**+13%**	**+15%**	**+15%**	**+13%**	**+10%**	**+8%**	**+5%**	**+4%**	**+2%**	**+1%**
+3.6% in 270,985 games	**1400-99**	**-9%**	**-9%**	**-9%**	**-5%**		**+6%**	**+10%**	**+14%**	**+15%**	**+15%**	**+13%**	**+10%**	**+8%**	**+6%**	**+3%**	**+1%**
+1.5% in 266,299 games	**1500-99**	**-10%**	**-12%**	**-12%**	**-9%**	**-6%**		**+6%**	**+11%**	**+13%**	**+15%**	**+14%**	**+12%**	**+10%**	**+7%**	**+4%**	**+2%**
-1.0% in 261,841 games	**1600-99**	**-10%**	**-13%**	**-14%**	**-13%**	**-10%**	**-6%**		**+6%**	**+10%**	**+12%**	**+14%**	**+12%**	**+11%**	**+8%**	**+6%**	**+3%**
-2.6% in 251,646 games	**1700-99**	**-8%**	**-11%**	**-14%**	**-15%**	**-14%**	**-11%**	**-6%**		**+6%**	**+10%**	**+12%**	**+12%**	**+12%**	**+9%**	**+6%**	**+4%**
-4.3% in 233,512 games	**1800-99**	**-7%**	**-10%**	**-13%**	**-15%**	**-15%**	**-13%**	**-10%**	**-6%**		**+5%**	**+9%**	**+10%**	**+11%**	**+10%**	**+8%**	**+5%**
-5.4% in 212,193 games	**1900-99**	**-6%**	**-8%**	**-10%**	**-13%**	**-15%**	**-15%**	**-12%**	**-10%**	**-5%**		**+5%**	**+8%**	**+10%**	**+11%**	**+9%**	**+5%**
-6.2% in 175,081 games	**2000-99**	**-6%**	**-7%**	**-9%**	**-10%**	**-13%**	**-14%**	**-14%**	**-12%**	**-9%**	**-5%**		**+5%**	**+7%**	**+9%**	**+9%**	**+7%**
-6.2% in 138,370 games	**2100-99**	**-4%**	**-5%**	**-7%**	**-8%**	**-10%**	**-12%**	**-12%**	**-12%**	**-10%**	**-8%**	**-5%**		**+4%**	**+7%**	**+7%**	**+6%**
-6.4% in 99,506 games	**2200-99**	**-1%**	**-3%**	**-6%**	**-5%**	**-8%**	**-10%**	**-11%**	**-12%**	**-11%**	**-10%**	**-7%**	**-4%**		**+5%**	**+6%**	**+7%**
-6.0% in 63,735 games	**2300-99**	**=0%**	**-2%**	**-2%**	**-4%**	**-6%**	**-7%**	**-8%**	**-9%**	**-10%**	**-11%**	**-9%**	**-7%**	**-5%**		**+4%**	**+7%**
-4.9% in 40,902 games	**2400-99**	**=0%**	**-3%**	**-1%**	**-2%**	**-3%**	**-4%**	**-6%**	**-6%**	**-8%**	**-9%**	**-9%**	**-7%**	**-6%**	**-4%**		**+6%**
-5.8% in 30,120 games	**2500+**	**=0%**	**=0%**	**-0%**	**-1%**	**-1%**	**-2%**	**-3%**	**-4%**	**-5%**	**-5%**	**-7%**	**-6%**	**-7%**	**-7%**	**-6%**	

Blitz 2020-2023	opponent	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	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+
+8.2% in 23,904 games	**1000-99**		**+2%**	**+6%**	**+9%**	**+12%**	**+13%**	**+12%**	**+10%**	**+8%**	**+9%**	**+8%**	**+5%**	**+3%**	**=0%**	**=0%**	**=0%**
+9.9% in 86,058 games	**1100-99**	**-2%**		**+5%**	**+8%**	**+12%**	**+15%**	**+15%**	**+14%**	**+12%**	**+10%**	**+9%**	**+6%**	**+4%**	**+2%**	**+2%**	**=0%**
+9.7% in 122,350 games	**1200-99**	**-6%**	**-5%**		**+6%**	**+11%**	**+15%**	**+16%**	**+16%**	**+16%**	**+12%**	**+11%**	**+9%**	**+6%**	**+4%**	**+3%**	**+1%**
+8.1% in 143,840 games	**1300-99**	**-9%**	**-8%**	**-6%**		**+6%**	**+11%**	**+15%**	**+17%**	**+17%**	**+15%**	**+13%**	**+10%**	**+8%**	**+6%**	**+3%**	**+2%**
+5.9% in 166,385 games	**1400-99**	**-12%**	**-12%**	**-11%**	**-6%**		**+7%**	**+11%**	**+16%**	**+17%**	**+17%**	**+14%**	**+12%**	**+9%**	**+8%**	**+4%**	**+4%**
+3.3% in 179,578 games	**1500-99**	**-13%**	**-15%**	**-15%**	**-11%**	**-7%**		**+6%**	**+11%**	**+15%**	**+16%**	**+15%**	**+13%**	**+10%**	**+7%**	**+5%**	**+2%**
+1.2% in 192,100 games	**1600-99**	**-12%**	**-15%**	**-16%**	**-15%**	**-11%**	**-6%**		**+6%**	**+11%**	**+14%**	**+15%**	**+13%**	**+11%**	**+9%**	**+7%**	**+4%**
-1.1% in 199,980 games	**1700-99**	**-10%**	**-14%**	**-16%**	**-17%**	**-16%**	**-11%**	**-6%**		**+6%**	**+10%**	**+12%**	**+13%**	**+12%**	**+10%**	**+7%**	**+4%**
-2.9% in 207,112 games	**1800-99**	**-8%**	**-12%**	**-16%**	**-17%**	**-17%**	**-15%**	**-11%**	**-6%**		**+5%**	**+9%**	**+11%**	**+12%**	**+10%**	**+6%**	**+5%**
-4.2% in 207,593 games	**1900-99**	**-9%**	**-10%**	**-12%**	**-15%**	**-17%**	**-16%**	**-14%**	**-10%**	**-5%**		**+5%**	**+8%**	**+9%**	**+9%**	**+8%**	**+6%**
-5.3% in 189,672 games	**2000-99**	**-8%**	**-9%**	**-11%**	**-13%**	**-14%**	**-15%**	**-15%**	**-12%**	**-9%**	**-5%**		**+4%**	**+7%**	**+9%**	**+8%**	**+6%**
-5.8% in 159,561 games	**2100-99**	**-5%**	**-6%**	**-9%**	**-10%**	**-12%**	**-13%**	**-13%**	**-13%**	**-11%**	**-8%**	**-4%**		**+3%**	**+6%**	**+7%**	**+7%**
-5.7% in 123,462 games	**2200-99**	**-3%**	**-4%**	**-6%**	**-8%**	**-9%**	**-10%**	**-11%**	**-12%**	**-12%**	**-9%**	**-7%**	**-3%**		**+4%**	**+6%**	**+7%**
-5.6% in 80,021 games	**2300-99**	**=0%**	**-2%**	**-4%**	**-6%**	**-8%**	**-7%**	**-9%**	**-10%**	**-10%**	**-9%**	**-9%**	**-6%**	**-4%**		**+3%**	**+5%**
-4.8% in 50,919 games	**2400-99**	**=0%**	**-2%**	**-3%**	**-3%**	**-4%**	**-5%**	**-7%**	**-7%**	**-6%**	**-8%**	**-8%**	**-7%**	**-6%**	**-3%**		**+4%**
-5.4% in 33,341 games	**2500+**	**=0%**	**=0%**	**-1%**	**-2%**	**-4%**	**-2%**	**-4%**	**-4%**	**-5%**	**-6%**	**-6%**	**-7%**	**-7%**	**-5%**	**-4%**	

In fact, it's probably slightly worse for rapid/blitz, as you can tell from the colors being even more intense than they were for standard. But essentially the same story is being told. Ratings are clearly too stretched out in these FIDE rating systems as well, and this evidence is enough to convince me that a compression is in order. I recommend that **we do the same Compression to 1400 Elo, for all three rating systems, at the same time**. Even if rapid and blitz might merit slightly more compression, I think it's best to keep to the same numbers and timing across all three rating systems, so that we don't run into problematic timing like in the earlier examples (on page 43) with the two players rated 1025.

Appendix #7: Should we require 5 games, or 9 games, for your initial rating?

A few people have mentioned the change that went into effect in 2014, where the required number of games in order to get an initial rating was reduced from 9 games to 5 games. On the surface, that might seem like an unwise tradeoff, since in return for getting a lot more players, we were perhaps getting ratings that were a lot more random.

However, I did my usual sort of simulation for FIDE back in 2012-2013 – although it was more rudimentary back then - when the various changes for 2014 were being contemplated. And in fact it was those simulations that led to the decision to reduce the minimum from 9 games to 5 games. We expected to see that we should increase from 9 games to 11 or 12 games, but in fact the opposite was true, once I actually performed the simulation and checked the results.

In essence, what the simulations revealed back then was that even though players would indeed get a less accurate initial rating, there would be a very important side benefit to the 5-game-minimum. Namely, it would get those new players into the rating pool as soon as possible, and so games played against them would get rated. More data overall leads to better ratings overall, and a more stable rating system. So it seemed wisest to get those players into the Elo system right away, instead of holding off in an attempt to get their initial ratings to be a little bit better.

In his recent comments, **GM Bartlomiej Macieja** referenced this earlier decision and asked if I had performed similar analysis recently:

> 6) The other huge problem is the proper evaluation of the initial rating. You mention that many players get their initial ratings too low (some too high). Meanwhile, nowadays it is sufficient to play 5 rated games to have an initial rating. Obviously, based on 5 games, the initial rating will be pretty random. In the past, a higher number of games was required to ensure already the very first rating is at least somewhat reliable. Till 30 June 2014, 9 games were required.
> I recall you explaining that based on your simulations as fewer games as sooner the system would stabilize. It would be good to repeat the simulations.
>
> Also, an additional question is if there is an overall cost. A scenario in which a player gets to his "real" rating faster, but at the same somewhat damages the entire system is possible. If there are so many new players, the effect may be accumulating.

Others have asked similar questions. So I went ahead and ran one last simulation, which asks the question: what would ratings look like today if we had left that 9-game-minimum unchanged in 2014? Otherwise everything about the regulations used during the simulation would match exactly with the baseline simulation.

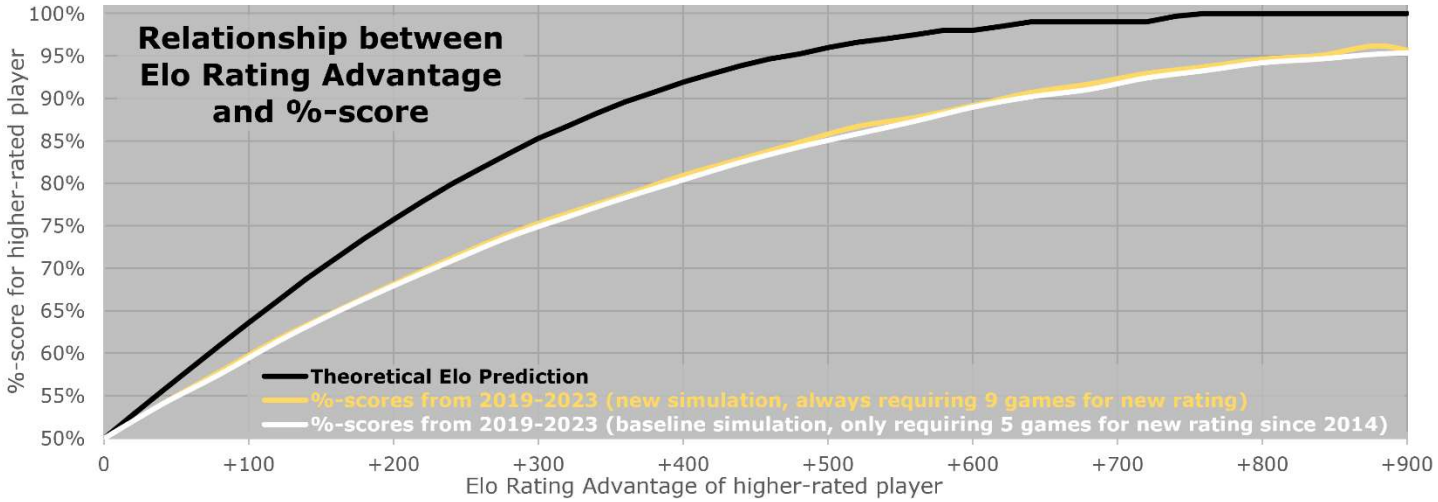
After running that simulation, I can tell you the biggest difference: we would have a lot fewer rated players today, if we had been requiring 9 games for all those years. In fact, there would have been 75,000 fewer rated players once we reached April 2023 (which is as far as my simulation went). So instead of the 404,000 rated players that we had in real life at that point, there would only have been 329,000 rated players.

Now, we can't really compare the accuracy of ratings for those 75,000 players between the two simulations, because in one of the scenarios (the one where we require 9 games for a rating) those 75,000 players wouldn't even have a rating yet. But we can at least check the impact on the overall rating system, by looking at the games played among the other 329,000 players, namely the ones who would have ratings in both simulations. This would let us look for exactly

what GM Macieja references – is the overall system being damaged if we include all those extra players with ratings based only on a few games played?

There were 3.5 million standard rated games played among those 329,000 players, from 2019-2023, and we can draw one of those line graphs again, where the black line shows the theoretical Elo prediction for rating favorites, and the white line shows the rating favorites' percentage scores (from 2019-2023) using ratings calculated from the real regulations. Also, we will also draw a gold line to show how rating favorites would have done with the simulated ratings in those same rated games from 2019-2023.

And remember, this is just for the games played among the 329,000 players who had ratings in both simulations.



The white and gold lines almost perfectly overlap, meaning that there would not be any great difference in how rating favorites perform. The gold is slightly better, but only marginally so. So essentially there would be no real impact on the stability of the other players' ratings, one way or the other, due to a nine game requirement for an initial rating. It would not have resulted in better ratings overall. Mostly what it would have accomplished would be to keep many tens of thousands of players from getting FIDE ratings.

Nevertheless, I think it was a good idea to re-run this analysis, since the FIDE rating pool is far different today from how it was a decade ago, and many things that were true back then might not be true anymore today.