

Francois Strydom, Chair
FIDE Ethics Commission
Re: Complaint against GM Evgeny Solozhenkin

Dear Ethics Commission Members:

This is an account of my investigations into the complaints involving the accusation against Ms. Bibisara Assaubayeva regarding her play at the World Youth Chess Championship, U14 Girls' section, in Montevideo, Uruguay, last September. I was originally contacted by CA Hal Bond on 20th September while Round 5 was underway. He said there had been a complaint against Ms. Assaubayeva and asked me to report on her specifically. I provided daily "screening test" reports not only on her games but on all sections of the tournament—the analysis and report generation are all automated except for lacking a secure Web interface so it was little extra trouble for me. Her games always showed screening results in the middle of the "completely normal" range.

I learned of the article in Russian by GM Solozhenkin, around which the present complaint revolves, from an e-mail to the Anti-Cheating Commission on 27th September by member Yuliya Levitan, who asked for my input, whereupon I (a) gave ACC the same non-public, non-indexed links to my reports as I'd given Mr. Bond, and (b) circulated my previous screening results of tournaments where she took part, none of them a cause for suspicion. At that time I had not run a "full test" of her games because there seemed no reason to do so. (The full test takes 4–6 hours on one processor for a typical game versus 10–15 minutes for screening.) After being notified of the current process, I ran full tests not only of the Montevideo games but also all available games of every event she played in the previous 12 months, including several other events to which GM Solozhenkin broadened the accusation. My present reporting of the results has these main purposes:

- (1) To address the question of possible cheating in Montevideo.
- (2) To address the broadened questions of possible cheating in the other events and also to ascertain the relation of Ms. Assaubayeva's level of play as measured by my model's "Intrinsic Performance Rating" (IPR) relative to her FIDE Elo rating—and to the levels of computer programs.
- (3) To provide scientific input on the specific junctures of games highlighted by GM Solozhenkin.

Here are my full-test results, in reverse chronological order from Montevideo. I tested all events in 2017, plus I "cherry-picked" two previous world youth championships in which Ms. Assaubayeva had an impressive score. Below, 'GT' means the number of games tested, and the * for the 2017 EICC signifies that only 2 of her 11 games were available.

The full test reports results in the form of z -scores as used in many other areas of science and fraud detection. Scores between -2.00 and +2.00 are always considered “completely normal”; FIDE requires that any other judgment involve a score outside the range -2.75 to +2.75 that is accompanied by evidence apart from engine tests. My tests always exclude turns 1–8 of any game, moves in repeating sequences, and positions where one side is ahead by more than 3.00 (figuratively, ahead by more than three pawns) and remains so until the end of the game. My full test also excludes book moves after turn 8 by players rated 2300 or higher and moves after turn 70. It uses the player’s rating *after* the tournament—as is proper under the “null hypothesis” of fair play—unless the rating before is higher. The IPR includes book moves but can be stated as IPR_{nb} without them; both figures are rounded to the nearest 05 for players and have error bars typically ± 200 –300 for 9-game sets.

Tournament	Date	Score	Elo Pre	Elo Post	GT	z -score	IPR	IPR_{nb}
Wch U14G	9/2017	8.5/11	2420	2395	11	-0.74	2550	2455
Trieste Open	9/2017	5.0/9	2420	2404	9	0.44	2500	2370
Millennials Match	7/2017	4.0/4	2386	2420	4	-0.50	2420	2395
RUS chW HL	7/2017	6.0/9	2386	2403	9	1.25	2735	2750
EUR Indiv. CC	6/2017	5.5/11	2245	2386	2*	-0.92*	2315	2175
RUS chT U14G	5/2017	6.5/7	2224	2245	7	3.23	2685	2635
Aeroflot Open B	2/2017	3.0/9	2229	2224	9	-1.17	2140	2085
Moscow Open A	1/2017	4.0/9	2177	2229	9	-0.50	2285	2225
Wch U12G	10/2016	10.0/11	2287	2292	11	0.51	2530	2470
Wch U10G	9/2014	10.0/11	1927	1943	11	1.23	2410	2445

The only result outside the completely normal range is from the May event in Sochi. This is the only event from mid-2017 that I have not seen mentioned in an accusation, which is important because outside evidence would be required to make it a positive. I tested it with alternate engines per ACC policy and these results do not corroborate a positive: Komodo 11.2 gives a z -score of 1.54 and an IPR of 2505, while testing with Stockfish 8 gives a z -score of 1.68 and an IPR of 2710. One point is that apart from a draw with the other 2200+ rated player, her six wins were all over opponents more than 400 Elo lower, which may affect the quality of the games in a manner that my model usually does not encounter.

The other z -scores are not only all normal but balanced between positive and negative, the latter meaning that Ms. Assaubayeva performed below the expectation for her rating. The IPRs indicate that she merits a rating in the 2400–2500 range. It is not unusual for a player to have an IPR that is 400 points higher than the rating—the rating system itself was designed around the idea that 200 points was one standard deviation of performance by a human player over the length of a typical tournament (originally as studied in the 1950s and 1960s), so 400 makes “two-sigma” which again is the extent of “normal.” In cases where single players have been sanctioned after my tests the IPR as a rule has been 700–800 points higher and usually above 3000 Elo. The two tournaments from 2016 and 2014 were hand-picked; if I had the games from some of her poor performances we might see 2000 averaging with 2450 to make the 2225 hinted by her level in Moscow last winter. The conclusion is:

There is no indication in the test results that can be distinguished apart from the typical track record of a rapidly improving junior player, alternating some great and some inconsistent performances.

Now to move on to point (3), I review what the test results say about the particular games accused by GM Solozhenkin. I include also an item (<http://chess-news.ru/node/23802>) by IM Sergey Zagrebely because it mentions games played in Montevideo *after* the second round game tied to the complaint. The moves addressed in the articles seem to fall into two categories:

- Moves or sequences of moves that seem computer-like or “too good” and so might indicate cheating.
- Mistakes made by Ms. Assaubayeva that seem strange.

Sometimes the point of the latter is to allege an error of transmission of a destination square or forgetting the followup of a line that might have been communicated some turns earlier. Other times it seems to be just wondering at easy missed wins. All in all, the complainers have not met the burden of showing that they are different from crazy mistakes made by any young player.¹ Hence I will concern myself mostly with the former. Taking them in order of the article “Nothing good”:

- (a) The game Yip-Assaubayeva from the Millennials match. Solozhenkin starts by saying, “The series of moves 16–29 would be from a player rated minimum 2500.” My IPR results averaged from the whole summer show a player of that level. (Incidentally, if I combine the games since June and use a rating of 2420 the z -score is only 0.45; if I add the May youth team event it rises to 1.40.) Then he marvels at the moves 22...Rfc8 and 23...Qa5. The former matches but the latter is regarded in my full test as a slight mistake, Rybka preferring 23...Ng6 and both Komodo 11.2 and Stockfish 8 agree at the respective depth levels of 17 and 19 (in 50-PV mode) which I use in my formal tests. The move 29...Rd8“!” prompts a reaction of “I do not believe!” but Komodo and Stockfish both give it equal value to 29...Nxc2, Stockfish listing 29...Rd8 first but Komodo listing 29...Nxc2 first—while the official Rybka 3 test data makes 29...Rd8 a clear choice with 29...Nxc2 in second place 0.27 behind.
- (b) The game Assaubayeva-Belenka from the RUSch HL in July: At turn 46, she blundered 46.Rc4? and the allegation is that this was supposed to follow 46.Bc5+. After 46...Kc6, however, 47.Rc4? is a waste of time, and Komodo prefers 47.Kf3 even if Black replies 46...Kd7 instead. The official Rybka test file also gives the PV 46.Bc5+ Kd7 47.Kf3. The only support comes from Stockfish’s PV 46.Bc5+ Kd7 47.Rc4 but this is a minority view.
- (c) Janzeli-Assaubayeva, Trieste round 4: The imputation is that Black’s losing blunder 26...Nd7?? was supposed to be 26...Bd7, but all of Rybka, Stockfish, and Komodo

¹GM Yasser Seirawan would not have gone on to win the 1979 World Junior if I hadn’t chosen an idiotic way to recapture a pawn in round 3 of the 1979 US Junior when he had lost his first two games. I was just “on tilt” after squandering a \pm endgame edge.)

prefer 26...Ra6 by wide margins after a few seconds of search in my tests—with Black in deep trouble anyway. (Then follow a bunch of blunder examples that seem to have no other point; I mention the last since it is from Montevideo.)

- (d) Mandukhai-Assaubayeva, Montevideo round 1: Indeed, 22...Ng4+? is a terrible error but it seems to be understandable as “winning” the rook on f1.
- (e) Assaubayeva-Mangalam, Montevideo round 2. The key game. Indeed, turns 18–27 are high quality by White, and if I start from turn 11 after 10...exd5 leaves book, the full test with Rybka on turns 11–27 gives a z -score of 2.56. However, the backup tests with the then-current Komodo and Stockfish versions give only 1.77 and 1.44, respectively, both in the completely-normal range. However, White keeps much the same advantage around 1.50 through turn 35, and if the range 11–35 is used then all the z -scores are completely normal—as they are for the whole game after White’s subsequent misplays, of course. That the screening tests also show high agreement in moves 11–27 is the one piece of evidence in the direction of the complaint. (The posting by Solozhenkin finishes with saying move 50 was strange, but I think being prompted to offer a draw upon seeing how 50...Qf6 unexpectedly hits h4 is completely natural.)
- (f) Turning now to IM Zagrebelny’s posting, he too states about the July RUS-ch HL that “The level of the game of whites in these games is estimated at 2500+”—with which my tests neither disagree nor show as unusual. Then he addresses the round-9 game in Montevideo against Jyothsna, which indeed features many silly missed wins by Black, but my analysis does not find any point to what is said.

The main indication from my review of these arguments is that *the complainers have not attempted to give any analysis that could support their allegations*. Instead, my analysis—using a neutral and automated scientific procedure—mostly argues against them. To the extent that it is any evidence at all, what it exposes is that the complainers have failed to give “due diligence” in backing up their assertions.

Shortly after learning of GM Solozhenkin’s post on 9/27 from Yuliya Levitan, I wrote Hal Bond and included the sentence, “It doesn’t have much more information than you gave me about the incident, but rather a lot of commentary on past games of the kind I’ve often ‘debunked’ in the past.” This was before I ran any tests of the games from other tournaments that were mentioned. In my view, whatever germ there is in the incident during the round-2 game versus Ms. Mangalam does not justify such an expansion to other tournaments when you have nothing more to say—and this kind of expansion with careless game commentary has been enough of a problem in the past that measures exposing this clearly should be taken.

Yours sincerely

Kenneth W. Regan