

## Do PDGA Ratings Differ between F and M?

By Steve West Disc Golf, LLC

October 18, 2023

I tried to find out if an F player would be as good as an M player with the same rating.

The reason this might be possible is that there are limited opportunities for F and M to play on the same course. This is mostly a good thing, in that it means F players may be getting a layout better suited to their skill set. However, it does tend to separate the calculation of ratings for F players from M. This could cause ratings drift; an F rating may not represent the same level of skill as the numerically equal M rating.

By Level of Skill, I mean the expected score of a player at that rating. This sidesteps any complications that could come from using round ratings as the measure of skill.

I excluded non-North American events because European ratings may be different because of being isolated. Also, from what I've seen F and M often play the same course over there.

From 1,337 North American courses where both F and M played the same layout during a sanctioned event, I found 120 course layouts where there were at least 32 F players and at least 32 M players. If these, 8 had some bad data, so I used data from the other 112. (Foreshadowing: This is more data than I expected to find, indicating potentially more cross-pollination in the ratings calculations.)

Because ratings are calculated as a linear function of scores, the expected score should be a linear function of rating. I solved for the two-parameter best fit of round scores to player rating, then solved again with a four-parameter model which accounts for the sex of the player. (All players who had ever played in any F division anywhere in the data were included as Fs, even if they were entered into an M division for an event.)

The Fisher statistical tests for whether sex mattered were significant for all but two of the events. In other words, the fit for the four-parameter model was almost always a lot better than could be expected from random chance if sex did not actually matter.

HOWEVER, that is a different than the original question of whether F or M ratings are inflated or deflated. To look at this, I compared the expected scores from the four-parameter model (which includes sex) to the expected scores from the two-parameter model (which ignores sex). Doing the comparison this way - rather than comparing directly to actual scores - allowed for an easier-to-calculate test for significance.

As a check, we should see that when the four-parameter F expected scores are higher than the two-parameter expected scores, M expected scores are lower, and vice versa. This turns out to be the case. See the chart at the end. The F and M do not exactly mirror each other because they have different distributions of player ratings.

If the F ratings are actually inflated, we should see that the expected scores from the four-parameter model are higher than the expected scores from the two-parameter model. This was the case for two-thirds of the courses. Which is not a lot better than a 50/50 coin flip.

The average amount by which an F expected score was higher was 0.59 throws per round with a standard deviation of 1.17. If there was no inflation at all, there would have been a 31% chance the extra 0.59 throws per round could have been observed as a result of random chance. Which indicates a lack of statistical significance. From the M point of view, the values were -0.61, 1.23, and 32%.

**Conclusion:** While there are some hints that F ratings are inflated by about 6 ratings points (0.59 throws at 10 points per throw), it varies so much from course to course that it is not statistically significant.

