Women's Chess Titles: The Unanticipated Effects of Institutional Intervention on Female Participation in Competitive Chess

Netflix: The Queen's Gambit

Abstract

The goal of this project was to identify how the introduction of women's titles contributed to the growth of women's chess using a simple random sample of the dataset Top Women Chess Players from Kaggle. The age and rating of female chess players were used to assess the growth of women's chess over the last century. It was found that more women played chess professionally after the release of women's titles in 1980. Tukey's HSD revealed that players with certain women's titles hold a comparable rating to women with the corresponding universal titles making the low required rating of women's titles versus universal titles less relevant. This means women's titles do not create separate rating categories for women as intended, but instead constitute a women's league. While women's titles once fostered the growth of women's chess, the segregation of women in chess an unnecessary and potentially harmful institution.

Background and Significance

Chess is played recreationally by millions of people from every corner of the world, but men have long dominated the competitive chess scene. Active male players outnumber females 16:1, and only a handful of women have broken the global top 100 (Bilalić et al. 2009). This disparity is linked to a long history of the societal exclusion of women in tournaments and events as chess was seen as a "men's game." However, women's chess has evolved in more recent years possibly due to the emergence of incredible female players such as Judit Polgar and Hou Yifan. With the success of the Netflix series *The Queen's Gambit* in 2020, this topic has never been more relevant.

Chess titles such as Grandmaster (GM) are available to players who have achieved a minimum skill level and are essential for anyone seeking a career in chess. Women-specific titles were made popular around 1980 by FIDE, the governing body of chess. The titles featured a 200-point lower rating requirement for female players, making titles more available to them. However, women's titles are often interpreted as another product of chess's history of sexism by lowering the bar for women and isolating them into a separate category (iChess.net n.d.). FIDE intended to promote chess to women and encourage women to improve at chess. This research project will uncover whether the introduction of women's titles actually accomplished this. It is hypothesized that the introduction of women's chess titles brought more women to the game but did not increase the average skill level. By observing trends in standard rating differences between titles, this project will explore the necessity of women's titles to the future growth of chess and question if they should be abolished in the name of equality.

Methods

The dataset Top Women Chess Players was taken from Kaggle online database ("Top Women Chess Players" n.d.). The dataset contained information on 8553 active and inactive top women chess players. It was sorted by standard FIDE rating (highest to lowest above 1800 Elo) as updated in August 2020. Simple random sampling was conducted to generate a sample of 500 subjects, and subjects with unknown data were removed leaving 481 total subjects.

The age of players was the quantitative predictor variable used to represent time. Age was determined from the year of birth using the Excel function age = (2022 - year of birth). Rating was the quantitative response variable. An Elo rating system calculates the relative skill of all players. Standard rating represents the player's skill level in classical chess, in which games may last multiple hours. Rapid chess is played using a faster time format and has its own rating scale. Titles are awarded to strong chess players by FIDE based on rating. Abbreviations and required ratings of all titles are found in Table 5 of the Appendix. Descriptive statistics mean and median were used to analyze the standard rating of different titles.

Section 1 of the results section focuses on the analysis of age and its influence on women's chess growth. A one-population mean t-test was conducted to determine if the average age of women chess players is less than 40 years old. A frequency histogram of age was used to determine the claim for this test. Scatterplots were used to determine if there is a linear relationship between age and rating (standard rating and rapid rating). Best-fit regression modeling was conducted to further analyze the linear relationship between age and the ratings. The correlation coefficient, estimated regression parameters and coefficient of determinations were included in the results. Normal QQ-plots and residual plots were presented to check the regression assumptions for models with observable trends. In Section 2, summary statistics of quantitative variable standard rating across the different categories of the qualitative variable title are compared. Boxplots were made to compare distribution of players within each women's title with the corresponding universal title. One-way ANOVA test followed by Tukey's HSD test was utilized to determine which titles have significantly different mean standard ratings.

Results

Section 1: Statistical Analysis of Influence of Time (Age) on Women's Chess Growth

The histogram in Figure 1a demonstrates the distribution of age of female players. Figure 1a shows that there are more younger female chess players (aged 20-40) compared to the older generations (aged 40-80). More specifically, we see that the modal group of players is in the age group 35-45, corresponding to players born in the late 70s to early 80s. Based on this analysis of the histogram, a one-population mean t-test was conducted to determine if the mean age of female chess players is less than 40 (Table 1b). The t-test confirmed the hypothesis that the mean age of women chess players is less than 40 years old (p-value = 0.0002).

In Figure 2 and Figure 3, age is used as a predictor for standard rating and rapid rating, respectively. No linear trend was detected in the scatterplot of standard rating and age. On the other hand, the correlation between age and rapid rating is 0.2170, which shows that age has a small positive linear relationship with rapid rating. Figure 3 shows that the estimated slope of the best-fit regression line equation between rapid rating and age is 3.0412 indicating that as age increases by a year, the average rapid rating increases by 3.0412. Since the correlation between age and rapid rating seemed somewhat significant, a best-fit regression model was conducted. Figure 4 shows a summary of the model check of the regression model shown in Figure 3. The resulting regression model with age as a predictor only explains 4.7% of the variability in rapid rating since the r^2 value for this linear model is only 0.047 (Figure 4a). Figure 4b shows the normal probability plot of the residuals for the regression model between rapid rating and age. The normality plot showed a skewed to the right distribution, so normality was not satisfied for rapid rating vs age. Based on Figure 4c, the residuals fail to satisfy the homoscedasticity and independence of the residual because the data points are not equally distributed. Therefore, there is no evidence to suggest that women improved or worsened at chess over time.

Section 2: Statistical Analysis of Mean Standard Rating between Each Chess Title

Table 5 presents some summary statistics of standard rating among titles including the mean and median standard ratings for each of the titles. Table 5 also includes the full title representing the acronyms used throughout the paper, and the required rating for each title. The mean standard ratings are often lower than the rating requirement for each title because most players' standard rating decreases after they have achieved a title. Grandmaster is the highest title a player can achieve because it requires the highest rating. This is followed by International Master, FIDE Master, and Candidate Master. The required rating to obtain each woman's title is 200 rating points lower than the universal title counterpart. For example, the rating requirement of a FIDE Master is 2300, but a Woman FIDE Master is only 2100. Therefore, we should expect to see that the mean rating of universal titles is around 200 points greater than the women's titles. However, the data demonstrates that this is not always the case. GMs and IMs retain their rating separation, but the mean standard rating of FMs is only 139 points greater than WFMs. Further, the median standard rating held by women CMs and WCMs is separated by only 66.5 points. Since a great majority of titled women hold WFM or WCM titles (figure 6), these numbers are made even more significant.

Four boxplots were created to compare the distribution of players' standard ratings between each pairing of universal title and woman's title. Figure 7 shows the distribution of GMs vs WGMs and Figure 8 shows the distribution of IMs vs WIMs. In both figures, we see that the upper quartile of standard rating in WGMs and WIMs is below the lower quartile of standard rating in GMs and IMs. However, Figures 9 and 10 show significant interquartile overlap in standard rating distribution, particularly between CMs and WCMs. Further analysis in a one-way ANOVA and Tukey's HSD test was necessary to determine if the overlap shows statistical significance. Figure 11a in the appendix shows the output of the one-way ANOVA statistical

analysis of mean standard rating across the different titles. The ANOVA test supported the claim that there is a significant difference in the mean standard rating across most titles (p-value=0). However, Tukey's test (Figure 11b) revealed that not all pairs of titles are significantly different from each other. Notably, the pairs FM-WFM (p-value=0.9732) and CM-WCM (p-value=0.8358) have very high p-values. This means the standard rating of players with universal titles FM and CM is not significantly different from their counterparts WFM and WCM. The pairs GM-WGM and IM-WIM were found to have significantly different mean standard ratings ($p < 0.001$).

Discussion and Conclusions

Figure 1 supports the hypothesis that women's titles increased the number of women in chess. The mean age of female chess players corresponds to girls born after the late 70s and early 80s. This generation would be the first to consider aiming for a woman-specific title, as the first Woman FIDE Master title was awarded in 1980. However, there is no evidence to suggest that female chess titles are the only reason for the spike in women's chess popularity. Another contributing factor could have been GM Susan Polgar who played in the 80s and inspired many young women with her ability to compete against men at the highest level. Only the timing of the spike suggests women's titles were a major contributing factor to the growth of woman's chess. We see from Figures 2 and 3 that the year a player was born is an incredibly weak predictor of rating, with no significant correlation. The lower expectations of women's titles did not appear to affect average female skill level. In conclusion, the original hypothesis is supported because women's titles were likely a major factor in the attraction of woman to chess but did not influence overall chess talent.

Women's titles did the short-term job of encouraging females to play chess because women's chess grew rapidly after the new titles were introduced. The Woman FIDE Master title quickly became the most common title for women (Figure 6). However, according to the statistical analysis conducted in Figure 11, the rating of women who hold the WFM title is *not* significantly different than the rating of women who hold the FM title. In fact, a portion of WFMs and FMs overlap in rating (Figure 9). The actual skill gap between a WFM and an FM is only 139 rating points on average, which is much smaller than the 200 required rating points separating the titles (Table 5). CMs are even closer to WCMs, with only a 66.5 median standard rating disparity. Amazingly, women's titles fail to fully filter women into separate rating categories because WFMs and WCMs are competing at a similar level as their universal title counterparts. This calls into question the necessity of women's titles entirely.

Deliberately lowering the bar for women was originally established as a means for encouraging female participation but may now be recognized as the source of institutionalized sexism in chess. WFM and WCM titles do not offer the inclusion they promised but instead isolate women into a separate league. The increasingly similar average rating of women's titles to universal titles only offers further proof of their own decreasing practicality. Separate leagues based on gender should have no place in an intellectual game when gender is not a determinant of intelligence. Due to the nature of the sport, women's titles perpetuate the stereotype that women have less natural chess ability, which may cause young female players to have negative interpretations of their own intellect.

Now that the short-term benefits of women's titles have come to fruition, it is time to reconsider the purpose of women's titles in today's world and their effect on the future of chess. The long-term harm of assumed stereotypes could start to recoil the recent growth that we have seen in women's chess and become a detriment to the game. Chess is a great equalizer that connects people from every corner of the world. Its beauty lies in that two people of any gender, race, or age become immediately humbled to equality at the starting position on the chess board. The removal of women's titles would bring equality back to a game that depends on it.

References

Bilalić, Merim, Kieran Smallbone, Peter McLeod, and Fernand Gobet. 2009. “Why Are (the Best) Women so Good at Chess? Participation Rates and Gender Differences in Intellectual Domains.” *Proceedings of the Royal Society B: Biological Sciences* 276 (1659): 1161–65. https://doi.org/10.1098/rspb.2008.1576.

iChess.net, iChess net. n.d. “Chess Titles: History of Women Chess Titles Explained.” Accessed June 24, 2022. https://www.ichess.net/blog/history-women-chess-titles/.

Soflytics Corp., Rguroo User’s Guide, Anaheim, CA, Version March 2021.

“Top Women Chess Players.” n.d. Accessed March 16, 2022. https://kaggle.com/vikasojha98/top-women-chess-players.

vwest. 2017. “Should Women’s Chess Titles Be Eliminated?” US Chess.Org. May 25, 2017. https://new.uschess.org/news/should-womens-chess-titles-be-eliminated.

Section 1: Statistical Analysis of Influence of Time (Age) on Women's Chess Growth

Figure 1: Number of Female Chess Players Across Generations and t-Test

Figure 1a: Histogram of Frequency of Age

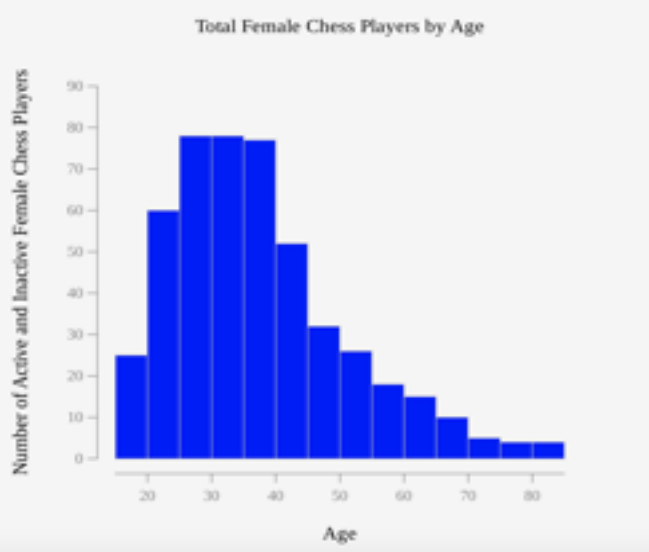


Table 1b: Test of Hypothesis: t-Test

Test of Hypothesis: t-Test
Age

Alternative Hypothesis H_a: Mean of 'Age' is less than 40
5% lower critical value in units of data = 38.99511

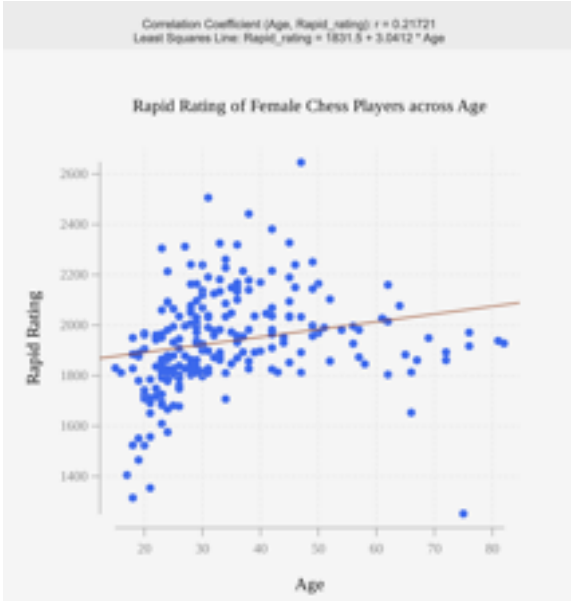
Sample Mean	Std Error	Obs t Stat	DF	5% t-Lower Critical	P-value
37.8223	0.609755	-3.57141	483	-1.64801	0.000195317

Test is significant at 5% level.

Figure 1: The mean age of women's chess players was determined to be 37.82, with a heavy trend toward younger generations. One population mean inference was conducted. The sample data support the claim that the mean age of women chess players is <40 (p-value=0.0002).

Figure 2: Standard Rating Across Age

Figure 3: Rapid Rating Across Age



Figures 2 and 3: Little to no correlation is observed in standard rating over age (r=0.16674). A small positive linear correlation may exist in rapid rating over age (r=0.21721), so a best-fit regression modeling was conducted.

Figure 4: Best-fit Linear Regression Model

Table 4a: Model Summary: Coefficient of Determination

Model formula: Rapid_rating ~ Age

Residual Standard Error	DF	R-Squared	Adjusted R-Squared	Correlation (r)
190.433	226	0.0471807	0.0429647	0.217211

Figure 4b: Normal Probability Plot

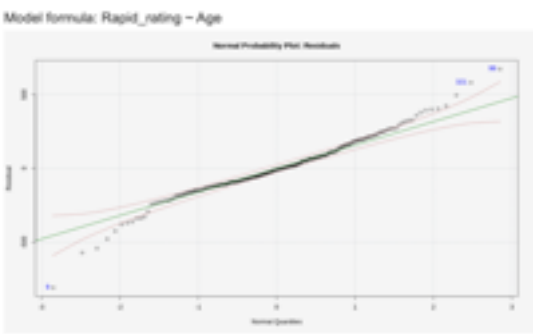


Figure 4c: Residual Versus Fit

Figure 4: Coefficient of determination (r^2) was determined to be 0.047, which shows that age only explains 4.7% of the variability in rapid rating. The normal probability plot is skewed to the right, and the distribution in the residual versus fit plot shows non-random patterns, so the data does not fulfill independence and homoscedasticity assumptions of the residuals.

Section 2: Statistical Analysis of Mean Standard Rating between each Chess Title

Table 5: Titles Statistics: Required Rating Compared to Actual Rating of Female Players

Title	*FIDE Required Standard Rating*	*Median Standard Rating*	*Mean Standard Rating*
Candidate Master (CM)	*2200*	*1996.5*	*2017.9*
Woman Candidate Master (WCM)	*2000*	*1930*	*1929.9*
FIDE Master (FM)	*2300*	*2216.5*	*2188.1*
Woman FIDE Master (WFM)	*2100*	*2062*	*2049.1*
International Master (IM)	*2400*	*2370*	*2354.1*
Woman International Master (WIM)	*2200*	*2149*	*2129.2*
Grandmaster (GM)	*2500*	*2475*	*2482.8*
Woman Grandmaster (WGM)	*2300*	*2248.5*	*2245.4*

Figure 5: A summary statistic was taken for the entire dataset to accurately represent all titles. The median standard rating between WCM and CM is only separated by 66.5 rating points despite a 200-point gap in the required standard rating between the two. Mean can be heavily influenced by outliers in this case, but still shows a <100 rating point difference. Players barely reach the rating requirement for a title and then drop down in rating afterward, which explains why every mean standard rating is below the FIDE rating requirement to achieve the title (vwest 2017).

Figure 6: Distribution of all Titled Female Chess Players

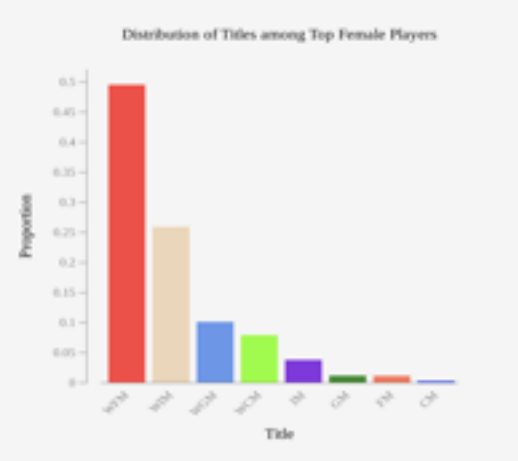


Figure 6: The graph shows that WFM is the most prevalent title in the dataset, followed by WIM, WGM, WCM, and the universal titles. Women obtain women's titles more frequently than universal titles.

Figure 7: Boxplot of Standard Rating distribution between GM and WGM

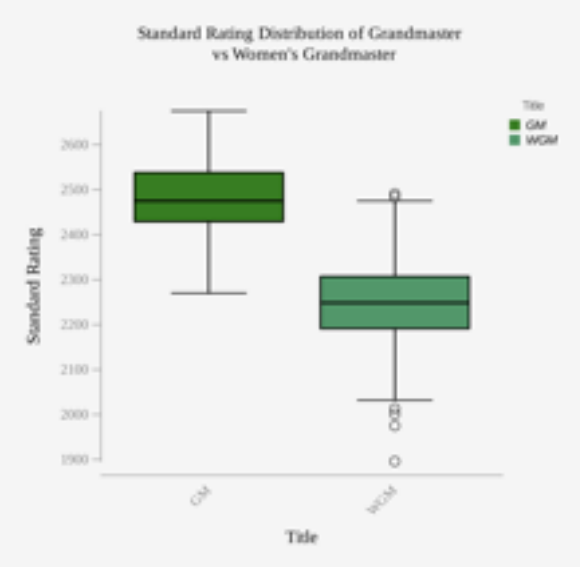


Figure 8: Boxplot of Standard Rating distribution between IM and WIM

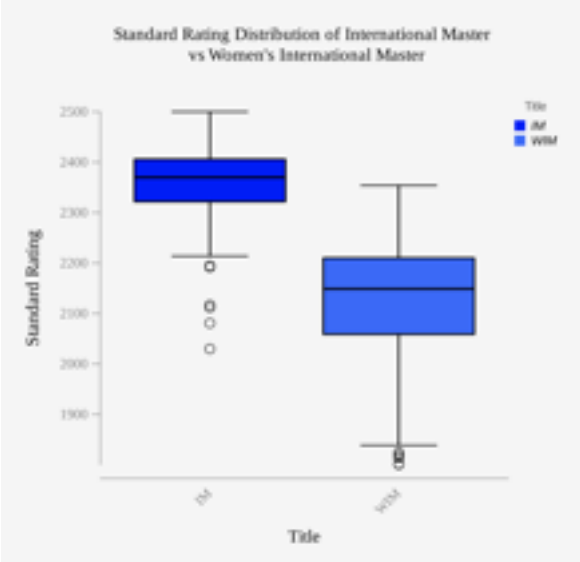
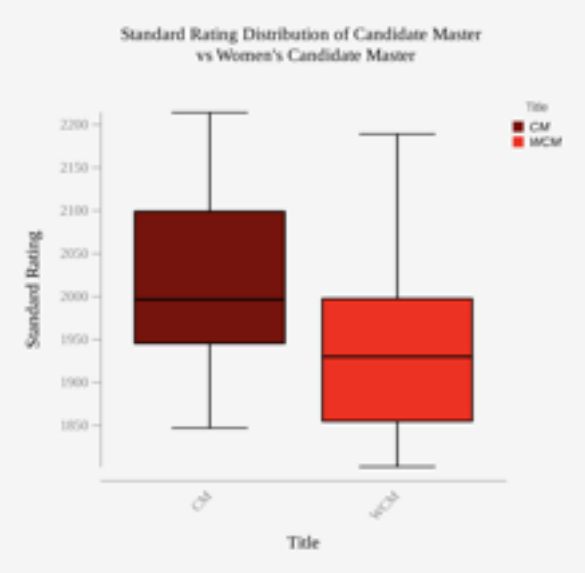


Figure 9: Boxplot of Standard Rating distribution between FM and WFM

Figure 10: Boxplot of Standard Rating distribution between CM and WCM



Figures 7-10: The titles remain separated into their rating categories. However, there is a lot of overlap in the distribution between women's titles and universal titles. All boxplots have some intersection in rating distribution, but FM-WFM and CM-WCM's box distributions are more similar to each other. A one-way ANOVA test and Tukey's HSD test was conducted to determine if these observed similarities and differences are significant.

Table 11: One-Way ANOVA test and Tukey's HSD to Determine Significance Between Mean Standard Rating and each Title.

Table 11a: ANOVA Table

Model: Standard_Rating ~ Title
H0: The means for all levels are equal

Source	DF	Sum of Squares	Mean Square	F Value	Pr>F	BFB
Title	7	2.44261e+06	348944	30.5222	0	Inf
Residual	195	2.22933e+06	11432.5			

Title is significant at 5% significance level.

Table 11b: Tukey's HSD: Multiple Comparison of Means

Table of 95% family-wise confidence level

Difference Levels	Mean	Lower Limit	Upper Limit	Adjusted p-value
FM-CM	15.3333	-362.977	393.644	1
GM-CM	483.5	117.202	849.798	0.00190249
IM-CM	270.833	-83.0438	624.71	0.27491
WCM-CM	-158.556	-495.16	178.049	0.835815
WFM-CM	-47.6286	-376.812	281.554	0.999844
WGM-CM	133.684	-202.454	469.822	0.92559
WIM-CM	28.1277	-302.966	359.221	0.999996
GM-FM	468.167	217.938	718.396	1.02727e-06
IM-FM	255.5	23.833	487.167	0.0194563
WCM-FM	-173.889	-378.2	30.4222	0.158094
WFM-FM	-62.9619	-254.8	128.877	0.97316
WGM-FM	118.351	-85.1907	321.892	0.633227
WIM-FM	12.7943	-182.304	207.893	0.999999
IM-GM	-212.667	-424.149	-1.18461	0.0476316
WCM-GM	-642.056	-823.158	-460.953	0
WFM-GM	-531.129	-698.033	-364.224	0
WGM-GM	-349.816	-530.05	-169.582	3.44769e-07
WIM-GM	-455.372	-626.014	-284.731	9.85545e-13
WCM-IM	-429.389	-583.834	-274.944	8.35998e-14
WFM-IM	-318.462	-455.983	-180.94	6.50042e-10
WGM-IM	-137.149	-290.574	16.2761	0.117293
WIM-IM	-242.706	-384.74	-100.672	1.15861e-05
WFM-WCM	110.927	27.3473	194.507	0.00174819
WGM-WCM	292.24	184.477	400.002	4.12004e-13
WIM-WCM	186.683	96.8697	277.497	5.43054e-08
WGM-WFM	181.313	99.6323	262.993	3.44938e-09
WIM-WFM	75.7562	18.2576	133.255	0.00195719
WIM-WGM	-105.557	-194.625	-16.4878	0.00848085

Figure 11: The sample data support the claim that there is a significant difference in rating between the titles (p-value=0). However, some specific combinations of titles are not significantly different according to Tukey's Comparison of Means. Notable pairs include WCM-CM (p-value=0.8358) and WFM-FM (p-value=0.9732). These relationships are not significantly different in mean standard rating according to the statistical analysis.