# Unveiling Gender Inequality in American High School Golf: A Time Series Analysis and Future Prediction

Han Wu*

Chengdu Foreign Language School, Sichuan, 610031, China

*Corresponding author: 1910821118@mail.sit.edu.cn

**Abstract.** Gender inequality remains a significant and persistent issue in golf. While some scholars have studied its historical origins, current status, and measures, there is a dearth of research focused on high school golf. Few studies have explored the specific challenges female high school golfers face, accurate predictions for future trends are lacking. This study uses a time series chart to visualize the male-female ratio of high school golf players in the United States from 2010/2011 to 2021/2022. The data is then fitted into an ARIMA model to forecast future trends. The visualization of the data reveals a significant increase in female participation from 2010/2011 to 2021/2022, although it remains lower than that of males. Despite some progress over the past 12 years, true gender equality in high school golf has not yet been achieved. The predictions indicate a decline in the female participation ratio from 2022/2023 to 2033/2034, highlighting the insufficient efforts to promote gender equality in sports. This lack of effort hinders sustained and stable progress in female participation. The results of this study have brought about a greater awareness of the severity of gender inequality in high school golf. This has prompted stakeholders to take stronger action to prevent the exacerbation of gender inequality issues in high school sports and collectively build a more inclusive sports environment free from gender bias and discrimination.

**Keywords:** Gender inequality; golf; high school sports; time series; ARIMA model.

## 1. Introduction

Gender inequality is a pervasive problem in the domain of sports, where disparities between male and female athletes persist across various disciplines [1]. Throughout recorded history and recent studies, numerous instances can be found where women encountered issues of equality in athletic opportunities [2]. Gender disparity is particularly pronounced in golf. The sport has a noticeable imbalance, with approximately 80% male players and only 20% female players. This highlights the challenges women face in gaining acceptance in golf [3]. Persistent negative stereotypes and double standards towards female golfers have been deeply ingrained in the sport, leading to lower participation rates and resource allocation for women [4]. The gender gap in high school golf participation is particularly concerning, with the number of boys participating nearly doubling that of girls in the 2017-2018 academic year, surpassing most other sports [5]. Therefore, analyzing the male-female participation ratio in high school golf and advocating for gender equality in sports holds substantial importance. This paper analyzes historical trends in the male-female ratio among American high school students and utilizes an Autoregressive Integrated Moving Average (ARIMA) model to forecast future trends. It also advocates for gender equality in sports and provides insights for the formulation of future policies in this area.

Previous studies have investigated the history of gender inequality in golf and its impact on the present day. Since the 18th century, the British introduced golf globally, along with their cultural tradition of exclusive gentlemen's clubs, leading to the exclusion of women from both clubhouses and golf practice [3]. Unfortunately, this historical perception has perpetuated stereotypes, misconceptions, social biases, and gender discrimination in the sport [4]. Even today, golf continues to be affected by exclusionary practices, making it challenging for women to achieve equality in terms of access, participation, employment, and decision-making. According to Haig-Muir, discrimination against women golfers is as long as the history of the sport itself [6, 7]. The period from 2010 to 2016 still reflects a noticeable gender disparity in golf participation on a global scale [8]. The 2014 England Golf membership survey reports that females constitute only 15% of members in England, with junior girls accounting for a mere 1% of members [9]. Based on the aforementioned research and data, gender disparity in golf has been a longstanding and severe issue, especially within the youth demographic.

With the evolving mindset, some institutional measures have been introduced to promote gender equality in sports. Title IX, a federal law, was passed to prevent gender discrimination in high school and college sports programs. This law has been instrumental in increasing the participation of girls in school sports by 500% within a decade of its implementation [10]. However, despite significant progress, challenges such as rule discrepancies, nontraditional scheduling, underqualified coaching staff, and a lack of interscholastic teams still exist, indicating a lack of equality [10]. Research has also shown that girls' high school golf programs face obstacles such as lack of recognition, uniforms, equipment, and support, which contribute to the gender disparity in the sport. Therefore, Gender inequality in high school golf remains a persistent issue that may contravene Title IX, which mandates equal sports opportunities for both boys and girls in public school.

Despite the wealth of knowledge generated by previous research, substantial gaps persist. While there is a considerable body of research on gender disparities in sports in general, there is a lack of studies on the unique challenges and experiences that female high school golfers face. Additionally, previous research has mainly focused on retrospective analyses that explain historical trends and current disparities but offer limited capability to predict future patterns. To address these shortcomings, this research utilizes a time series chart to accurately visualize the male-female ratio of high school golf players in the United States from 2010/2011 to 2021/2022. Furthermore, it developed an ARIMA model to forecast the female-to-male ratio among American high school golf players in the coming years. This study contributes to a better understanding of the historical and current gender ratio in high school golf while providing insights into future development trends. It advocates for gender equality in high school golf and provides guidance for future policy formulation.

## 2. Methods

### 2.1. Data Source

The data for this literature is collected from Statista, including the number of participants in high school golf in the United States from 2010/2011 to 2021/2022, by gender.

### 2.2. Data Processing

The missing values filling for statistics in 2019/2020 and 2020/2021 is performed on the collected data through the historical averaging method. By taking the average of two adjacent data, missing values can be estimated based on the historical pattern of the data.

In addition, to provide a more intuitive representation of the relationship between the number of male and female participants in golf from 2010/2011 to 2021/2022, this study transformed the original data into the form of the ratio of female to male participants.

### 2.3. Variable Selection

The statistics are there for the research for time series, figuring out the main rules of the changing trend, which are plotted below in Fig 1. Based on the yearly time series plot of the ratio of female to male participants in American high schools, it can be concluded that the ratio gradually increased from 2010/2011 to 2018/2019. The increase during the period from 2012/2013 to 2018/2019 was steeper than that of the period from 2010/2011 to 2012/2013. However, there was a slight decline in the ratio from 2018/2019 to 2021/2022. Overall, the ratio showed a significant growth from 2010/2011 to 2021/2022, indicating a noticeable increase in the proportion of female high school golf participants compared to the total participants. Although there has been a clear improvement in gender equality in American high school golf over the 12 years, it has not yet reached the desired state of absolute equality, as the ratio of female high school golf participants to male participants remains below 56%.

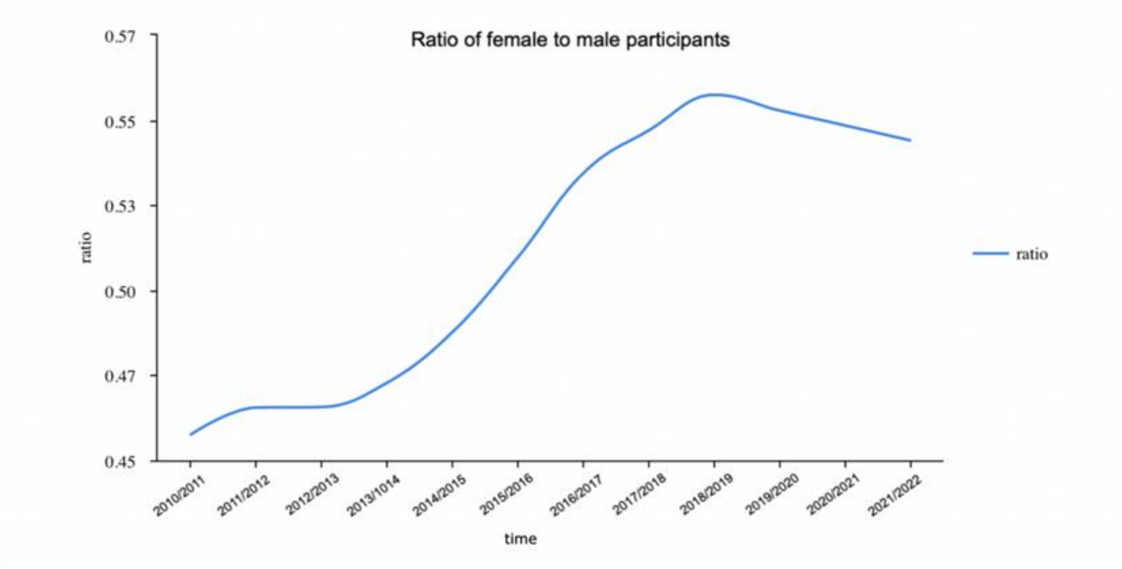


**Fig. 1** Ratio of female to male high school golf participants from 2010/2011 to 2021/2022

### 2.4. Research Method

To begin with, the data is analyzed visually to identify the original trends. Next, an Augmented Dickey-Fuller (ADF) test is conducted to confirm the stationarity of the time series and determine the appropriate differencing order (d) for the ARIMA model. Autocorrelation Function (ACF) plots and Partial Autocorrelation Function (PACF) plots are generated to evaluate the significance of autocorrelation in the time series. By examining these plots, an approximate suitable value for p, which represents the number of autoregressive (AR) terms, and q, the number of moving average (MA) terms, in the ARIMA model can be selected. A Ljung-Box portmanteau test is then performed to accurately assess the significance of autocorrelation. After evaluating multiple combinations of (p, d, q) values and selecting the best ones based on the smallest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values, the data is fitted to an ARIMA model. Finally, the model's suitability and fit are evaluated using the Ljung-Box test for autocorrelation in the residuals.

## 3. Results and Discussion

Determining the stationarity of time series data from its time series plot alone is not accurate. To identify the appropriate value of "d" in the ARIMA model, the ADF test was employed in this study as it is a robust method of assessing the stationarity of time series data. The ADF test assumes the null hypothesis that the series is non-stationary. If the p-value is below 0.1 (or 0.05 as the threshold), the null hypothesis is rejected at the 0.1 significance level, indicating the stationarity of the series.

**Table 1.** Ratio-ADF test table

| Differencing order | t | p | Critical value | | |
|---|---|---|---|---|---|
| | | | 1% | 5% | 10% |
| 0 | -4.139 | 0.001 | -4.939 | -3.478 | -2.844 |

The results of the ADF test are presented in Table 1. For a differencing order of 0, the p-value is calculated as 0.001 (<0.1), leading to the rejection of the null hypothesis and concluding that the original series is stationary. Therefore, the time series satisfies one of the criteria for the ARIMA model, namely, the stability of the data's temporal characteristics, as it does not exhibit significant trends or seasonal variations over time. Furthermore, there is no need to apply differencing to the original time series to mitigate or diminish non-stationary effects within the ARIMA model. Consequently, the value of "d" in the ARIMA model is set to 0.

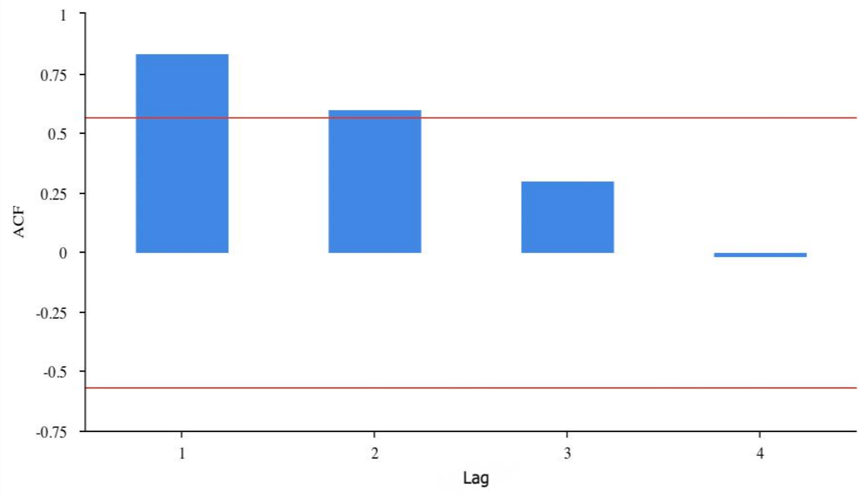


**Fig. 2** The ACF plot of original data

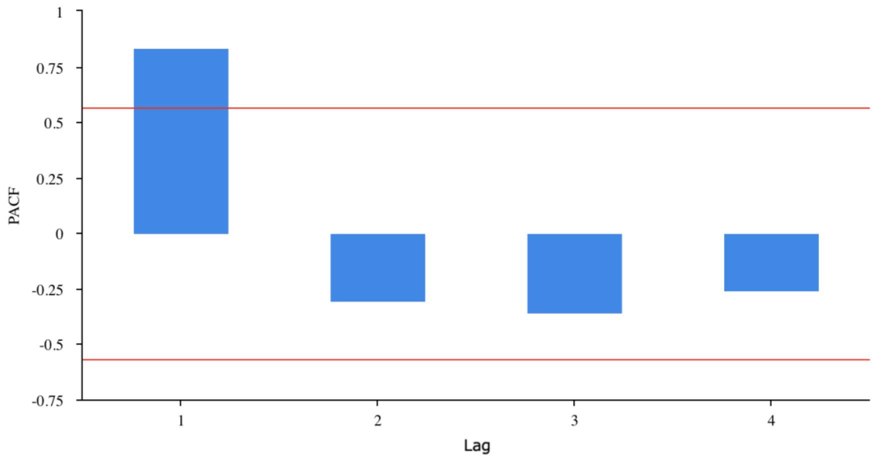


**Fig. 3** The PACF plot of original data

To get a rough idea about the autocorrelation of a time series and determine the approximate values of p and q in ARIMA, this study plotted the ACF and PACF graphs. Based on the results shown in Fig 2 and Fig 3 above, it can be inferred that the time series is somewhat autocorrelated.

If the ACF graph cuts off at lag q (where ACF becomes zero after a certain lag) and the PACF graph doesn't cut off, the ARIMA model can be simplified to MA(q). On the other hand, if the PACF graph cuts off at lag p and the ACF graph doesn't cut off, the ARIMA model can be simplified to AR(p). Therefore, we can conclude that the approximate value of p is around 2 and the approximate value of q is around 0.

**Table 2**. AFC and PACF result table

| Lag order | AC value | PAC value | Q statistic | *p value* |
|---|---|---|---|---|
| lag1 | 0.833 | 0.833 | 10.598 | 0.001 |
| lag2 | 0.601 | -0.304 | 16.665 | 0.000 |
| lag3 | 0.301 | -0.354 | 18.359 | 0.000 |
| lag4 | -0.014 | -0.255 | 18.363 | 0.001 |

After analyzing the ACF and PACF graphs, we conducted the Ljung-Box test to investigate the autocorrelation of the time series more accurately. The Ljung-Box test for autocorrelation assumes the null hypothesis of no serial correlation. If the corresponding p-value is less than 0.05, the null hypothesis could be rejected, indicating the presence of serial correlation. Conversely, if the p-value is greater than 0.05, it suggests that the series has little or no serial correlation, thus supporting the null hypothesis. The results of the test are shown in the Table 3 above. Since p1, p2, p3, and p4 are all less than 0.05, the study rejected the null hypothesis and concluded that the time series exhibits some degree of autocorrelation. This also satisfies another criterion for the use of the ARIMA model, providing a reasonable justification for its application (Table 3).

**Table 3**. AR(2) parameter table

| Term | Sign | Coefficient | Stansard error | z value | p value | 95% CI |
|---|---|---|---|---|---|---|
| Constant term | c | 0.050 | 0.027 | 1.867 | 0.062 | -0.002 ~ 0.102 |
| AR parameter | α1 | 1.754 | 0.174 | 10.086 | 0.000 | 1.413 ~ 2.095 |
| | α2 | -0.853 | 0.171 | -4.990 | 0.000 | -1.188 ~ -0.518 |

Based on all of the judgments above, the study believes that the data can be fitted by the ARIMA model, and the order of the model is represented by 3 integers, (p,d,q). Based on the rough estimation of the values for q, d, and p, as well as multiple attempts and selection in this study, the most suitable combination of q, d, and p with the lowest AIC ( -74.917) and BIC (-72.978) values was found to be (2, 0, 0), corresponding to an AR(2) model. The parameters of the model are in Table 4. The model equation is $y(t) = 0.050 + 1.754y(t-1) - 0.853y(t-2)$.

**Table 4.** Model Q statistic table

| items | statistic value | *p* value |
|---|---|---|
| Q1 | 0.249 | 0.618 |
| Q2 | 0.973 | 0.615 |
| Q3 | 1.101 | 0.777 |
| Q4 | 1.480 | 0.830 |
| Q5 | 1.693 | 0.890 |
| Q6 | 2.356 | 0.884 |
| Q7 | 2.926 | 0.892 |
| Q8 | 3.742 | 0.880 |
| Q9 | 4.350 | 0.887 |
| Q10 | 5.193 | 0.878 |

In order to examine the rationality and applicability of the model, the study tested the residual. The ARIMA model requires the residuals to be white noise, meaning that there is no autocorrelation in the residuals. This can be tested using the Q statistic for white noise testing (null hypothesis: the residuals are white noise). Table 5 above presents information on the Q statistic of the model (specifically, the Ljung-Box Q test statistic), including the statistic value and p-value. Typically, if the corresponding p-value is greater than 0.1, it indicates that the white noise test is satisfied (otherwise, it suggests that it is not white noise). Q6 is used to test whether the autocorrelation coefficients of the residuals up to lag 6 satisfy the white noise assumption. In common cases, analyzing Q6 directly is sufficient. From the results of the Q statistic, since the p-value of Q6 is greater than 0.1, the null hypothesis cannot be rejected at a significance level of 0.1. This indicates that the residuals of the model are white noise, and the model meets the requirements.

**Table 5.** Predicted value (12 period ahead)

| Time | 2022/ 2023 | 2023/ 2024 | 2024/ 2025 | 2025/ 2026 | 2026/ 2027 | 2027/ 2028 | 2028/ 2029 | 2029/ 2030 | 2030/ 2031 | 2031/ 2032 | 2032/ 2033 | 2033/ 2034 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| value | 0.536 | 0.526 | 0.515 | 0.505 | 0.496 | 0.489 | 0.484 | 0.482 | 0.482 | 0.484 | 0.488 | 0.493 |

Subsequently, the prediction of ratio in coming years is presented. Table 5 above shows the precise predicted values of the ratio from 2022/2023 to 2033/2034. Prior to analyzing the predicted trend, this study conducted a series of tests on various indicators to ensure the accuracy and reliability of the ARIMA model's forecasting results. The small values of the root mean square error (RMSE) and mean square error (MSE) -- 0.0144 and 0.0002, respectively -- highlight the model's ability to predict with minimal differences from the actual observations, indicating a high overall predictive accuracy of the model. Furthermore, the mean absolute error (MAE) and mean absolute percentage error (MAPE) also exhibit relatively small values of 0.0092 and 0.0189 respectively. This suggests that the model has low average prediction errors and minimal relative deviation in the predicted results. In conclusion, the ARIMA model employed in this study is relatively reliable and accurate for prediction purposes.

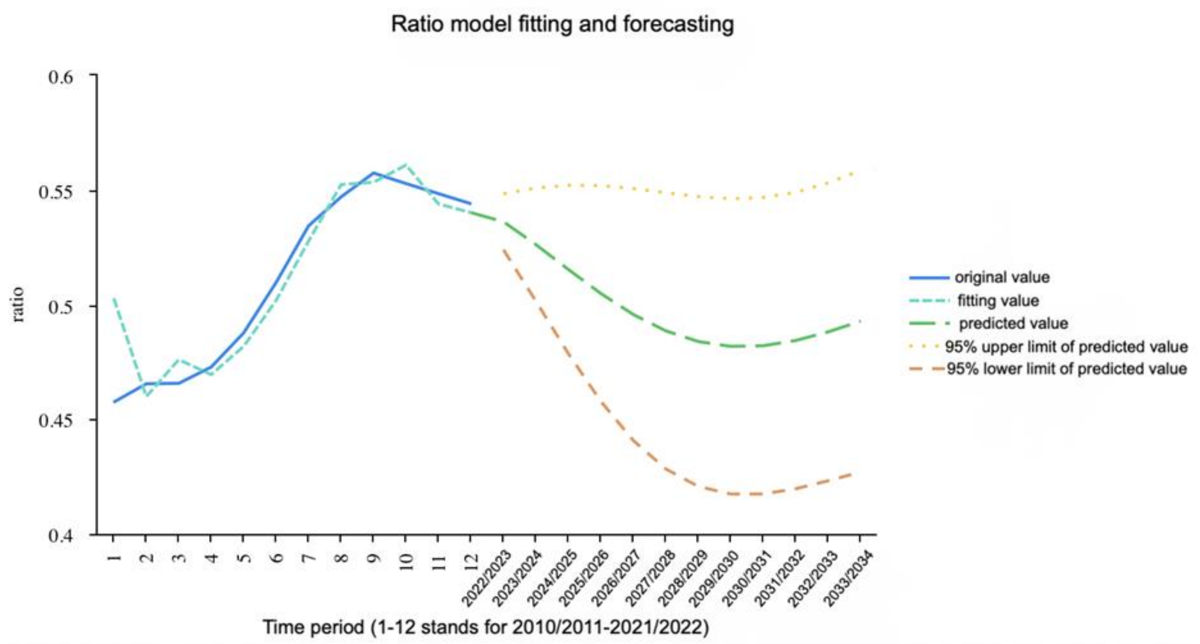


**Fig. 4** Ratio model fitting and forecasting

Additionally, a time series fitting graph illustrating the trend of the data in the next 12 periods and the model's fitting performance for the ratio of female to male golf participants in American high schools is provided (Fig 4). According to the figures, the overall trend of the ratio indicates a decline in the next 12 time periods. Specifically, there is a significant decline from 2022/2023 to 2030/2031, followed by a very slight increase from 2030/2031 to 2033/2034. The decreasing ratio of female to male high school golf participants from 2022/2023 to 2033/2034 reflects an increasing severity of the gender imbalance issue in American high school golf over the next twelve years. This suggests that the issue of gender equality in high school sports is much more complex than anticipated. The current efforts in education and institutional systems to promote gender equality in sports are insufficient to achieve sustained and stable progress in female participation compared to male counterparts. The downward trend in the ratio is likely a result of lingering issues in high school sports. Therefore, this forecast should be considered as a warning sign for the future of golf in American high schools. It is crucial to implement more effective measures to prevent further exacerbation of gender imbalance.

Schools should encourage students to eliminate negative stereotypes and misconceptions about women's golfing abilities, such as lack of skill and strength. Schools should also provide equal resources for both female and male golf enthusiasts by offering both women's and men's teams, ensuring equal practice and competition opportunities. Governments should investigate discriminatory policies in golf courses, such as shorter opening hours for women, and enforce stricter regulations to promote gender equality in sports. Coaches and players should not normalize inequality, as accepting gender inequality as the norm may perpetuate the disparity in treatment and opportunities for female golfers. Sports media should enhance the visibility of female golfers, and the coverage of women golfers should not be confined to their appearance, stature, and personal matters. Instead, it should focus on their accomplishments and highlight exceptional female players in the public domain. This approach will serve as an inspiration for more young girls to engage in the sport of golf. In conclusion, based on the unfavorable predicted outcome, establishing true gender equality in sports is a long-term battle that requires collective and sustained efforts of multiple groups.

## 4. Conclusion

This study presents a time series analysis of female-male American high school golf participants from 2010/2011 to 2021/2022, revealing an overall upward trend in the participation ratio of female high school golfers over the past 12 years. This indicates some progress in addressing gender equality in high school golf. However, the issue remains unresolved as girls' golf participation still lags significantly behind that of boys. Additionally, using the ARIMA model, the study forecasts a downward trend in the participation ratio of female high school golfers in the United States for the next 12 years (2022/2023 to 2033/2034). This trend highlights the persistent gender inequality issues in sports and the potential threat they pose to future female sports participation. It also underscores the shortcomings of current policies across different domains.

To prevent the predicted downward trend and mitigate the exacerbation of gender inequality in high school sports, this paper provides several recommendations. Firstly, the government should eliminate discriminatory rules within golf clubs and establish policies that promote equality. Secondly, schools should ensure equal opportunities and resources for male and female golfers while educating students on gender equality issues. Lastly, coaches and golfers should actively reject and challenge any normalization of inequality. It is crucial that multiple relevant groups of people actively work together to create a fair and inclusive sports environment.

It is important to note that this study had certain limitations. The research conducted solely focused on the gender ratio issue in American high school golf. However, further research is necessary to fill the gaps and expand the study to a broader population. Future articles could also delve into potential factors that may contribute to the increasing gender imbalance in golf in the coming years.

## References

[1] Kosofsky S. Toward gender equality in professional sports. Hastings Women's LJ, 1993, 4: 209.

[2] Senne J A. Examination of gender equity and female participation in sport. Sport J, 2016, 19: 1-9.

[3] Reis H, Correia A. Gender inequalities in golf: a consented exclusion. International Journal of Culture, Tourism and Hospitality Research, 2013, 7(4): 324-339.

[4] Kitching N, Grix J, Phillpotts L. Shifting hegemony in 'a man's world': Incremental change for female golf professional employment. Sport in Society, 2017, 20(11): 1530-1547.

[5] Haig-Muir M. Many a slip'twixt cup and the lip: equal opportunity and Victorian golf clubs. Sporting Traditions, 2000, 17(1): 19-38.

[6] Kitching N. Women in golf: A critical reflection. Routledge international handbook of golf science, 2017, 404-413.

[7] Golf E. Golf club membership questionnaire 2014 results book. Woodhall Spa, Lincolnshire: England Golf, 2014.

[8] Xu Jia, Xie Baoguo. The relationship between career self-concept clarity and depression among college students: the moderating effect of gender and gender equality. Psychological Development and Education, 2021, 37 (4): 8.

[9] Chen Hengming, Yang Guanqiang. From Gender Equality to Domain Transformation: The Evolution of Socialist Women's Sports Concept. Sports and Science, 2021, 42 (3): 6.

[10] Tian Fen. MIT School of Engineering's Approach to Promoting Gender Equality in Academic Careers. World Education Information, 2021, 34 (7): 6.