# Uneven Bars? Looking for Evidence of Environmental Microaggression Theory in NCAA Women's Gymnastics[1]

Tommy Morgan*
Brigham Young University

Seth Cannon
Brigham Young University

Click here for the most recent draft
Last updated 28 Dec 2023

**Abstract**

Research on racial microaggression theory suggests that environmental factors such as the name of a building can negatively impact Black individuals. This study utilizes a unique dataset of scores from NCAA women's gymnastics meets to examine whether Black gymnasts experience any negative effect on their performance when competing at Brigham Young University (BYU), a university with a history of racial controversy and a namesake who promoted racist policies during his life. We model gymnasts and their coaches as score maximizers and estimate a generalized differences-in-differences regression model with fixed effects to analyze the full regular seasons of every gymnastics team that visited BYU between the 2017 and 2022 seasons. We find no evidence that being Black has any significant effect on gymnasts performing at BYU in any of the four individual women's gymnastics events. When we introduce a triple difference by splitting our analysis to pre- and post-May 2020 (the death of George Floyd), we find similar non-results.

Keywords: women's gymnastics, racial gaps, racial microaggression theory

JEL Codes: J15, Z2, Z20

[1] The authors are grateful to Lizzie Mukai and Abbe McBride for their help in compiling much of the original data used in this study and also thank Joe Price and Jeff Denning for helpful direction. Any opinions expressed are those of the principal investigators alone and should not be construed as representing the opinions of BYU.

* Corresponding author: morganto@byu.edu

# Uneven Bars? Looking for Evidence of Intergroup Anxiety in NCAA Women's Gymnastics


## Abstract

Research on racial microaggression theory suggests that environmental factors such as the name of a building can negatively impact Black individuals. This study utilizes a unique dataset of scores from NCAA women's gymnastics meets to examine whether Black gymnasts experience any negative effect on their performance when competing at Brigham Young University (BYU), a university with a history of racial controversy and a namesake who promoted racist policies during his life. We model gymnasts and their coaches as score maximizers and estimate a generalized differences-in-differences regression model with fixed effects to analyze the full regular seasons of every gymnastics team that visited BYU between the 2017 and 2022 seasons. We find no evidence that being Black has any significant effect on gymnasts performing at BYU in any of the four individual women's gymnastics events. When we introduce a triple difference by splitting our analysis to pre- and post-May 2020 (the death of George Floyd), we find similar non-results.

**I. Introduction**

After the death of George Floyd in May 2020, the longtime call from social activists to rename places named in honor of controversial historical figures and symbols such as KKK members, eugenicists, racially charged terms or slurs, slave holders, Confederate soldiers, and the likes gained significant momentum. The argument to change these names usually focuses on their perceived negative impact on members of the groups disparaged by those namesakes, which is in line with prominent research on microaggression theory that suggests that environmental factors such as the name of a place can implicitly communicate to someone that "[They] don't belong / [They] won't succeed here" and that "There is only so far [they] can go" (Sue et. al. 2007). This paper aims to expand on this environmental microaggression theory and the general issue discussed in Pope & Schweitzer (2011) by investigating whether such a negative effect exists and is detectable in the face of "competition, large stakes, and experience."

Our study examines a scenario in which this potential negative effect would likely be present. Brigham Young University (BYU) has a history of racial controversy in its athletics programs due in large part to its affiliation with the Church of Jesus Christ of Latter-day Saints (Bergera 2013), and its namesake, Brigham Young, promoted racist policies as a religious and political leader in Utah throughout the mid-1800s. This history and association is relatively well known, so we contend that if a visiting athlete were to experience any negative effect from name-based environmental microaggression at any university, they would experience it at BYU.

We analyze the performance of NCAA women gymnastics teams over the entire regular season(s) in which they visited Brigham Young University at least once from 2017-2023. We include scores from the full span of the regular season in which they visited BYU. For example, we include every score from every University of Arizona gymnast at every one of their regular season meets from 2017 and 2022 because Arizona's women's gymnastics team competed at BYU during both of those years, but we include every score from every Illinois State University gymnast at every one of their regular season meets from just 2022 because this is the only season in which Illinois State competed at BYU over our 2017-2023 timespan. We choose to analyze women's gymnastics because the score a gymnast receives is given at the individual level. This is key, as any negative effect from environmental racial microaggression should present itself primarily at the individual level.

We model gymnasts as score maximizers and propose a general model for event scores in college-level gymnastics competition. We estimate a difference-in-differences model of those scores that uses an indicator variable for being a Black gymnast competing at BYU as the independent variable of greatest interest. We argue that our model controls for all relevant factors that could affect a gymnast's score, including innate ability, preparation, potential judge bias, and so on. In addition, we implement a triple difference design around the death of George Floyd and its aftermath and find similar non-results.

## II. Background & Related Literature

### II.1 Women's Artistic Gymnastics

In women's artistic gymnastics, a regular season meet is composed of four events: vault, floor, balance beam, and uneven bars. Gymnasts typically compete in only one or two of these events, but they are allowed to compete in all four at any given meet. Each performance is scored out of 10 by two to four judges whose independent scores are averaged to a final performance score. The typical regular season meet has four judges – two from in region and two from out of region – each judging two events, with two judges per event. When there are more than two teams at a given meet, at least eight individual judges judge one event each (still with two judges per event) with no rotation between events.

In each event, five or six gymnasts from each team perform, and if six gymnasts perform, the lowest of those six scores is dropped. At the NCAA level, scores are determined by two factors: the "start value" of the routine, which is the score a gymnast would receive by performing their prepared routine perfectly, and deductions taken from the start value that are related to technical or execution errors. Routines are required by rule to have at least a 9.4 start value, but gymnasts at the collegiate level are sufficiently skilled such that almost all routines begin at the maximum 10 point start value.

Though an individual score can theoretically range anywhere from zero to 10 in each event, they typically cluster within the 9.7 - 9.9 range. After all events are complete, the sum of the five highest scores in each event are summed to compute the team's meet score. Because the practical range of scores is so small, tiny differences in average scores separate elite teams from great and decent teams: elite gymnastics teams have the potential to hit a 198.00 meet score, which is obtainable only with an average score of 9.9 from every gymnast in every event across the entire meet; great teams can consistently hit a 197.00 meet score (a 9.85 average performance score); and good teams can consistently hit a 196.00 meet score (a 9.8 average performance score).

Choosing to analyze artistic gymnastics meets offers us several key advantages. One such advantage is that scores are assigned to gymnasts on an individual basis. Though research has shown that there may be an overall ordering bias in judging (Damisch, Mussweiler, & Plessner 2006; Morgan & Rotthoff 2014; Rotthoff 2015; Joustra, Koning, & Krumer 2020; Rotthoff 2020), that same research generally shows that a gymnasts' score is not affected by the performances immediately preceding it. Because scores are mostly independent across individual gymnasts, we can look for the presence of an environmental effect that would manifest at the individual level.

Each of the four gymnastics events is a high-intensity, high-focus exercise. We hypothesize that any unusual outside pressure could cause gymnasts to make mistakes they would not make absent such pressure, with potential pressures of this kind including a negative environmental microaggression effect as defined previously. We argue that if a gymnast were to experience such an effect at any venue, she would be likely to experience it at Brigham Young University, a university with a history of racial controversy in its athletics programs (Bergera 2013). Whether such an effect would be significant enough to affect competitive performance is the focus of the study.

Research that investigates behavioral effects using women's gymnastics has primarily focused on elite-level gymnasts. These papers most frequently deal with race-agnostic biases present in judges, such as difficulty bias (Rotthoff 2020) and the aforementioned ordering bias (Damisch, Mussweiler, & Plessner 2006; Morgan & Rotthoff 2014; Rotthoff 2015; Joustra, Koning, & Krumer 2020; Rotthoff 2020). However, other gymnastics-based research has focused on biases held by competitors, like the paper from Meissner, Rai, and Rotthoff (2021) on Simone Biles' superstar effect. These papers make use of the unique context of gymnastics competitions (individual scoring, tournament settings, judge panels, etc.) to investigate these biases in a convincing way. However, due to the infrequent nature of elite-level tournaments relative to other gymnastics competitions, these papers frequently rely on a single tournament's worth of data to draw their conclusions (with Meissner, Rai, & Rotthoff (2021) being a notable exception).

### II.2 Environmental Microaggression Theory

Recent research that investigates the effects of environmental microaggressions on Black individuals is largely based on the model of microaggression theory presented in Sue et al. (2007), which suggests that environmental factors such as the names of buildings or overall racial climates can constitute "[m]acro-level microaggressions, which are more apparent on systemic and environmental levels" than other types of interpersonal microaggressions. This research is often focused on qualitative interviews or surveys of Black students' experiences at predominantly White institutions (Mills 2020; Holliday & Squires 2020). This observation is also generally true of literature in this field historically, as evidenced by the many hundreds of papers based on interviewing Black students attending perennially White institutions (or PWIs) published from 1965-2013 that are summarized in Willie & Cunnigen (1981), Sedlacek (1987), and Holliday & Squires (2020).

In addition to this research on racial bias founded on microaggression theory, much research exists on racial biases within the world of sports. Generally, this research focuses on racial biases in referee/judge decisions (as in Price & Wolfers 2010; Parsons et al. 2011; Gallo, Grund, & Reade 2012; Rotthoff 2020; Eiserloh, Foreman, & Heintz 2020; and Pelechrinis 2023) or in fan/commentator preference outside of competition (as in Andersen & La Croix 1991; Preston & Szymanski 2008; Reilly & Witt 2011; Principe & van Ours 2022; and Quansah, Lang, & Frick 2023). These studies use data from professional sports leagues in many sports and around the world to generally show that racial biases can affect sports teams both in competitive outcomes and in perceived value.

Two more investigations warrant particular note in this paper. The first is Dix's body of work on sports programs at historically Black colleges and universities (or HBCUs) in which he shows teams from HBCUs experiencing negative effects in football (Dix 2017, 2021a), men's basketball (Dix 2022a, 2022b), women's basketball (Dix 2019, 2020a, 2022b), baseball (Dix 2020b), softball (Dix 2021b), and volleyball (Dix 2023). This line of research is helpful for bridging the gap between the above mentioned studies on professional sports to studies at the college level. Though it focuses only on averaged team results and not on individual level effects, it is nonetheless useful for contextualizing our work in this paper.

The second investigation of note is found in Caselli, Falco, & Mattera (2023). In this paper, the authors show that African players in a professional Italian soccer league improved their performance when COVID-19 prevented fans from attending their games. They argue that this effect stems from the absence of overtly racist fan behavior, which is reportedly common in that league. This research is relevant to our paper because it is to our knowledge the closest existing paper to this one. Like we will in this paper, Caselli, Falco, and Mattera evaluate individual-level performance scores (in this case, assigned algorithmically to individual soccer players) in a generalized fixed effects model that allows them to control for player- and match-based fixed effects. They also model the effects of a quasi-environmental removal of direct racial aggressions, which mirrors our empirical strategy in which we exploit the introduction of a potential environmental microaggression.

Our research contributes to literature that uses sports to research behavioral effects by trying to find evidence of an environmental microaggression in a competitive gymnastics setting. We contribute to racial microaggression research by trying to find evidence of it in a novel setting that may be externally valid to the experiences of students generally, especially in strenuous situations of high pressure such as difficult classes or exam season. Additionally, we aim to quantify this specific type of microaggression effect within a statistical framework that could uncover causality, which has not generally been done around environmental microaggression theory to this point. In regards to the two lines of research that most closely approach ours (the Dix papers and the Caselli, Falco, & Mattera paper), we differentiate ourselves from Dix by focusing on individual-level effects as opposed to team-level effects, and we differentiate ourselves from Caselli, Falco, and Mattera by focusing on the introduction of a racial microaggression to college-level athletes as opposed to the removal of an overt, quasi-environmental set of racial aggressions from professional athletes.

## III. Model & Data

In order to attribute causality to any estimate we produce, we must be reasonably sure that we have controlled for as many other factors that could influence a gymnast's score as possible. As such, we suggest a simple model of a gymnasts score in any given event that breaks down influential factors into three main categories:

$$(1) \qquad \text{score} = \text{ability} + \text{preparation} + \text{environment} + \varepsilon$$

In this model, ability-related factors could include, for example, the genetic makeup of a given gymnast, the age at which they began training, and the set of skills they have the physical capacity to perform. Preparation-related factors might include the quality of the team and coaches surrounding a gymnast, the number of years a gymnast has been competing, the types of skills a gymnast chooses to practice, and the number of meets that have already occurred in a season. Environment-related factors could include the relative competency or biases of judges at a given meet, the altitude, location, quality, and name of the venue, the time of the meet, and so on. There is also bound to be stochastic error in any model of human behavior.

It is important to note that the functional model used by judges to assign scores to gymnasts in a meet takes the following form:

$$(2) \qquad \text{score} = \text{start value} - \text{deductions}$$

where the skills a gymnast chooses to perform in their routine are assigned difficulty scores that contribute to the start value of that routine as discussed earlier. Those start values are then weakly deducted based on how well the gymnast performs those skills. We suggest that, generally, our model in Equation 1 can be viewed as an alternate interpretation of this official model. A gymnast's ability, preparation, and environment will generally inform the skills they choose to perform, with a large majority of gymnasts at the collegiate level having sufficient skill and preparation to maximize their routines to a 10 point start value. From there, the gymnast's ability, preparation, and environment will also contribute to the deductions they receive: a more flexible gymnast will be able to hit better splits, a gymnast who has been out of the gym due to injury may struggle "shaking off the rust", a certain judge or set of judges may be more lenient in one event than in another, the altitude change at an away venue may make some gymnasts light-headed, et cetera.

Having put forth this general framework, we further posit that gymnasts and their coaches behave as score-maximizers. We assume that these agents aim to maximize the score a gymnast will receive in a given event by selecting routines of appropriate difficulty and practicing them adequately; this implies that a gymnast has perfect information about and control over the ability- and preparation-related elements of the proposed scoring model. In our simple model, we assume that a gymnast has no control over the environment-related factors of her scoring. We discuss these assumptions further when we define the empirical model we use to estimate scores in a later section.

Our study uses hand-compiled data from official meet reports as published at RoadToNationals.com, the official statistical website of NCAA Gymnastics. Table 1 lists the teams and years included in our dataset. For each team-year, we gather every score in every event from every meet during their regular season, along with the gymnast who received the score. In total, our dataset covers the regular seasons of 17 different teams that each visited BYU at least once between 2017 and 2023.

Our dataset includes 9,007 scores across all four women's artistics gymnastics events; Table 2 shows a summary of these scores broken down by event. The scores in our dataset were received by 361 gymnasts (40 Black) over 323 meets (34 at BYU). Of those gymnasts and meets, each gymnast's race was coded as Black or not Black based on visual inspection of their official photographs as posted on each school's official gymnastics website, with Figure 1 showing an example of a typical web page from which this visual inspection was conducted.

## IV. Empirical Strategy

We use a fixed-effects regression design to isolate any heterogeneous effect of competing at BYU for Black and non-Black gymnasts. We begin with a baseline differences-in-differences model:

$$(3) \qquad \text{score}_{im} = \beta_0 + \beta_1 \text{ Black}_i + \beta_2 \text{ atBYU}_m + \beta_3 \text{ (Black x atBYU)}_{im} + u_{im}$$

where subscripts *i* and *m* refer to individual gymnasts and meets, respectively. The dependent variable is the score earned by a gymnast in a given event, and the interaction term (Black x atBYU)$_{im}$ takes a value of 1 if a score is received by a Black gymnast competing at BYU and 0 otherwise. Black$_i$ and atBYU$_m$ represent binary variables for a gymnast being Black and a meet being held at BYU. In this model, $\beta_0$ is the regression constant term, and the coefficient of interest is $\beta_1$, which we interpret as the differential impact of competing at BYU on Black gymnasts relative to non-Black gymnasts and non-BYU venues. The stochastic error term is represented by $u_{im}$.

This baseline model certainly suffers from omitted variable bias. To combat this, we introduce gymnast- and meet-level fixed effects:

$$(4) \qquad \text{score}_{im} = \beta_0 + \beta_1 * \text{(Black x atBYU)}_{im} + \beta_2 * \text{meet}_{im} + \text{[gymnast effects]}_i + \text{[meet effects]}_m + \varepsilon_{im}$$

where each of the square bracketed terms denotes a set of fixed effects and the error term is represented by $\varepsilon_{im}$. These fixed effects control for gymnast- and meet-specific characteristics that might influence the score a gymnast receives at a given meet, including (but not limited to) individual judge biases, venue effects, point-in-time effects (such as the start value a given skill is worth), gymnasts' skill and physical build, team effects (no gymnasts change teams in our sample), and so on. We argue that these fixed effects control for every possible relevant factor to the score a gymnast receives. Finally, to account for possible correlation between the scores received at a given meet, we report standard errors clustered at the meet level.

We are also interested in whether this effect presented itself differently in the aftermath of the death of George Floyd in May 2020. This turns our basic differences-in-differences model from Equation 3 into a triple difference, and we estimate the following model as described by Olden & Møen (2022):

$$\text{score}_{imt} = \beta_0 + \beta_1 \text{ Black}_i + \beta_2 \text{ atBYU}_m + \beta_3 \text{ postFloyd}_t + \beta_4 \text{ (Black x atBYU)}_{im} + \beta_5 \text{ (Black x postFloyd)}_{it} + \beta_6 \text{ (atBYU x postFloyd)}_{mt} + \beta_7 \text{ (Black x atBYU x postFloyd)}_{imt} + \varepsilon_{imt} \quad (5)$$

where the new subscript *t* is a time indicator and $\text{postFloyd}_t$ is a binary indicator for a score having been received in the 2021, 2022, or 2023 seasons (the 2020 season concluded in March 2020). All other terms are defined intuitively, and the coefficient of interest in this new model would be $\beta_7$, which we describe as the differential effect on scores of being Black at BYU in the George Floyd aftermath.

Due to a lack of fixed effects, this model likely shares its omitted variables with the model represented by Equation 3. As such, we introduce fixed effects to this model as well. However, as pointed out by Feigenberg, Ost, and Qureshi (2023), we need to consider whether to interact our fixed effects with the postFloyd indicator. This results in our estimating the following model:

$$\text{score}_{imt} = \beta_0 + \beta_1 \text{ (Black x atBYU)}_{im} + \beta_2 \text{ (Black x atBYU x postFloyd)}_{imt} + \text{[gymnast effects]}_i + \text{[meet effects]}_m + \text{[gymnast x postFloyd effects]}_{it} + \varepsilon_{imt} \quad (6)$$

where the $[\text{gymnast x postFloyd}]_{it}$ term is the interaction of the postFloyd binary indicator with the set of gymnast fixed effects and every other term is defined as above. We are able to omit the postFloyd indicator because the fixed effects for the meets that occurred post-2020 would be collinear with that indicator, and we are able to similarly omit the (atBYU x postFloyd) indicator as well. In this model, the coefficient of interest is $\beta_2$, but we are also interested in both 1) the joint significance of $\beta_1$ and $\beta_2$ ; and 2) whether $\beta_1$ and $\beta_2$ are statistically different from each other.

**V. Empirical Results**

Table 3 reports the results of estimating our initial differences-in-differences specifications. In the first row of Table 3, we report the coefficients on the interaction term obtained by estimating Equation 3 (the baseline difference-in-differences) across events. In the second row, we replace the binary Black variable with gymnast fixed effects; in the third row, we replace the binary atBYU variable with meet fixed effects; and in the final row, we estimate the full model in Equation 4 (i.e. with both sets of fixed effects).

We find that the coefficient on the interaction term (Black x atBYU) is not statistically different from zero in any individual event under any robust fixed-effect regression specification. This analysis reveals that even the most obvious looking racial gaps in event scoring from Figure 2 are not statistically significant. This implies that competing at BYU does not have an adverse effect on Black gymnasts' performance, meaning these gymnasts are either 1) fully unaffected by intergroup anxiety or other implicit social pressures at BYU; or 2) not sufficiently affected by such pressures to suffer a detectable drop in performance when at BYU.

Our difference-in-difference models appear to be hindered by the fact that the proportion of our sample made up of scores received by Black gymnasts while performing at BYU is so small. If one were to see the same distributions of scores reflected in Figure 2 in a better-balanced sample of similar size, one would likely see practically large and statistically significant effects on event scores produced by our models. However, a larger sample would also push those distributions closer to their asymptotic distributions, which may truly be equal to each other. As it stands, these results provide no evidence of an environmental microaggression effect on performance for Black gymnasts visiting BYU.

Table 4 reports the results of estimating the triple difference model described in Equation 6. It also reports p values and F values for tests of 1) the joint significance of the coefficients on (Black x atBYU) and (Black x atBYU x postFloyd); and 2) the equality of those coefficients. The only estimates that are even marginally significant are in the vault, and they do follow patterns that would be evident of an environmental microaggression effect (i.e. negative effect post-Floyd that is statistically different from the overall effect). However, these results are not significant at the conventionally acceptable 5% confidence level, and given the lack of statistically robust results in every other model, we interpret this result as likely spurious.

The results of our analysis should not be interpreted as confirming the non-existence of an environmental microaggression effect on the performance of Black gymnasts at BYU. The estimates are not precise enough around zero to support this interpretation. Rather, our results should be interpreted as failing to confirm the existence of such an effect. This result, though anti-climactic, is likely to be externally valid, in that it suggests neither the existence of a generalized environmental microaggression effect on Black people nor the non-existence of that effect. As a result, we can only call for further quantitative research into this intriguing niche.

## VI. Conclusion

We contribute to research that uses sports competition to research behavioral effects by using a unique dataset on collegiate women's gymnastics scores composed of teams that visit BYU to test whether Black gymnasts experience a change in performance when competing at BYU that their non-Black competitors do not experience. Our sample allows us to introduce many sets of fixed effects to control for relevant factors that influence scores, which allow us to isolate the interaction between a gymnast's race and the venue at which they perform. We estimate a difference-in-differences model with generalized fixed effects and then implement a triple differences model to investigate whether a hotbed racial controversy instigated the effect any further.

We find no significant difference in score distributions at BYU between Black and non-Black gymnasts in any individual event using either of our models, providing no evidence that the name of BYU affects Black gymnasts' performance to a notable degree. Our estimates are also very imprecise and therefore provide little evidence against the idea that this environmental microaggression effect could exist. However, we note that a more robust sample (i.e. larger) could yield different results under the same models if the score distributions we observe in our present sample persisted. As such, we call for further research into this effect, both in this context and others.

## References

Andersen, T., & La Croix, S. (1991). Customer Racial Discrimination in Major League Baseball. *Economic Inquiry*, 29(4), 665-677. https://doi.org/10.1111/j.1465-7295.1991.tb00853.x

Bergera, G. J. (2013). "This Time of Crisis": The Race-Based Anti-BYU Athletic Protests of 1968-1971. *Utah Historical Quarterly*, *81*(3), 204-229. https://issuu.com/utah10/docs/uhq_volume81_2013_number3/s/10418366

Caselli, M., Falco, P., & Mattera, G. (2023). When the Stadium Goes Silent: How Crowds Affect the Performance of Discriminated Groups. *Journal of Labor Economics*, 41(2), 431-451. https://doi.org/10.1086/719967

Damisch, L., Mussweiler, T., & Plessner, H. (2006). Olympic medals as fruits of comparison? Assimilation and contrast in sequential performance judgments. *Journal of Experimental Psychology: Applied*, 12(3), 166–178. https://doi.org/10.1037/1076-898X.12.3.166

Dix, A. C. (2017). A decade of referee bias against college football programs from historically black colleges and universities. *International Journal of Science Culture and Sport*, 5, 197-212. https://doi.org/10.14486/IntJSCS670

Dix, A. C. (2019). "And 1" more piece of evidence of discrimination against black basketball players. *Howard Journal of Communications*, 30, 211-229. https://doi.org/10.1080/10646175.2018.1491434

Dix, A. C. (2020a). And 10 more years of bias against HBCU female basketball players. *Texas Speech Communication Journal*, 44, 1-18.

Dix, A. C. (2020b). Critical race theory, the NCAA, and college baseball: Contradiction on the diamond. In M. Milford & L. R. Smith (Eds.), *Communication and contradiction in the NCAA: An unlevel playing field* (pp. 213-234). New York: Peter Lang.

Dix, A. C. (2021a). Referee judgments of communication in the field of play: A study on Historically Black Colleges and Universities in division II college football. International Journal of Sport Communication, 14, 554-573. https://doi.org/10.1123/ijsc.2021-0032

Dix, A. C. (2021b). Softball umpires call more walks than strikeouts when the pitcher plays for a Historically Black College and University. *International Journal of Sport Culture and Science*, 9(1), 91-103. https://web.archive.org/web/20231214222019/https://dergipark.org.tr/en/download/article-file/1412460

Dix, A. C. (2022a). The non-sweet sixteen: Referee bias against Historically Black Colleges and Universities in men's college basketball. *Sociology of Sport Journal*, 39, 118-124. https://doi.org/10.1123ssj.2020-0187

Dix, A. C. (2022b). Stay woke: An analysis of how referees evaluate the in-game communication of an HBCU that competes in a PWI conference. *Communication and Sport*, OnlineFirst. https://doi.org/10.1177/21674795221103407

Dix, A. C. (2023). Indications of referee bias in Division I women's college volleyball: Testing expectancy violations and examining nonverbal communication. *International Journal of Sport Communication*, 16(4), 414-422. https://doi.org/10.1123/ijsc.2023-0050

Eiserloh, D. G., Foreman, J. J., & Heintz, E. C. (2020). Racial Bias in National Football League Officiating. *Frontiers in Sociology*, 5(48). https://doi.org/10.3389/fsoc.2020.00048

Feigenberg, B., Ost, B., & Qureshi, J. A. (2023). Omitted Variable Bias in Interacted Models: A Cautionary Tale. *The Review of Economics and Statistics*, 1(47). https://doi.org/10.1162/rest_a_01361

Gallo, E., Grund, T., & Reade, J. (2012). Punishing the Foreigner: Implicit Discrimination in the Premier League Based on Oppositional Identity. *Oxford Bulletin of Economics and Statistics*, 75(1), 136-156. https://doi.org/10.1111/j.1468-0084.2012.00725.x

Holliday, N. R. & Squires, L. (2020). Sociolinguistic labor, linguistic climate, and race(ism) on campus: Black college students' experiences with language at predominantly white institutions. *Journal of Sociolinguistics*, 25(3), 418-437. https://doi.org/10.1111/josl.12438

Joustra, S. J., Koning, R. H., & Krumer, A. (2020). Order Effects in Elite Gymnastics. *De Economist*, 169, 21-35. https://doi.org/10.1007/s10645-020-09371-0

Meissner, L., Rai, A., & Rotthoff, K. W. (2021). The Superstar Effect in Gymnastics. *Applied Economics*, 53(24), 2791-2798. https://doi.org/10.1080/00036846.2020.1869170

Mills, K. J. (2020). "It's systemic": Environmental racial microaggressions experienced by Black undergraduates at a predominantly White institution. *Journal of Diversity in Higher Education*, 13(1), 44–55. https://doi.org/10.1037/dhe0000121

Morgan, H. N., & Rotthoff, K. W. (2014). The Harder the Task, the Higher the Score: Findings of a Difficulty Bias. *Economic Inquiry*, 52(3), 1014-1026. https://doi.org/10.1111/ecin.12074

Olden, A., & Møen, J. (2022). The triple difference estimator. *The Econometrics Journal*, 25(3), 531-553. https://doi.org/10.1093/utac010

Parsons, C. A., Sulaeman, J., Yates, M. C., & Hamermesh, D. S. (2011). Strike Three: Discrimination, Incentives, and Evaluation. *American Economic Review*, 101(4), 1410-1435. https://doi.org/10.1257/aer.101.4.1410

Pelechrinis, K. (2023). Quantifying implicit biases in refereeing using NBA referees as a testbed. *Scientific Reports*, 13. https://doi.org/10.1038/s41598-023-31799-y

Pope, D. G., & Schweitzer, M. E. (2011). Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes. *The American Economic Review*, *101*(1), 129–157. http://www.jstor.org/stable/41038785

Preston, I., & Szymanski, S. (2008). Racial Discrimination in English football. *Scottish Journal of Political Economy*, 47(4), 342-363. https://doi.org/10.1111/1467-9485.00168

Price, J. & Wolfers, J. (2010). Racial discrimination among NBA referees. *The Quarterly Journal of Economics*, 125(4), 1859-1887. https://doi.org/10.1162/qjec.2010.125.4.1859

Principe, F., & van Ours, J. C. (2022). Racial bias in newspaper ratings of professional football players. *European Economic Review*, 141, https://doi.org/10.1016/j.euroecorev.2021.103980

Quansah, T. K., Lang, M., & Frick, B. (2023). Color blind - Investigating customer-based discrimination in European soccer. *Current Issues in Sport Science*, 8(2). https://doi.org/10.36950/2023.2ciss007

Reilly, B., & Witt, R. (2011). Disciplinary sanctions in English Premiership Football: Is there a racial dimension?. *Labour Economics*, 18(3), 360-370. https://doi.org/10.1016/j.labeco.2010.12.006

Rotthoff, K. W. (2015). (Not Finding a) Sequential Order Bias in Elite Level Gymnastics. *Southern Economic Journal*, 81(3), 724-741. https://doi.org/10.4284/0038-4038-2013.052

Rotthoff, K. W. (2020). Revisiting Difficulty Bias, and Other Forms of Bias, in Elite Level Gymnastics. *Journal of Sports Analytics*, 6(1), 1-11. https://doi.org/10.3233/JSA-200272

Sedlacek, W. E. (1987). Black students on White campuses: 20 years of research. *Journal of College Student Personnel*, 28(6), 484–495. https://psycnet.apa.org/record/1988-37333-001

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A. M., Nadal, K. L., & Esquilin, M. (2007) Racial microaggressions in everyday life: implications for clinical practice. *American Psychologist*, 62(4), 271-286. https://doi.org/10.1037/0003-066X.62.4.271

Willie, C. V. & Cunnigen, D. (1981). Black students in higher education: A review of studies, 1965-1980. Annual Review of Sociology, 7, 177-198. https://www.jstor.org/stable/2946027

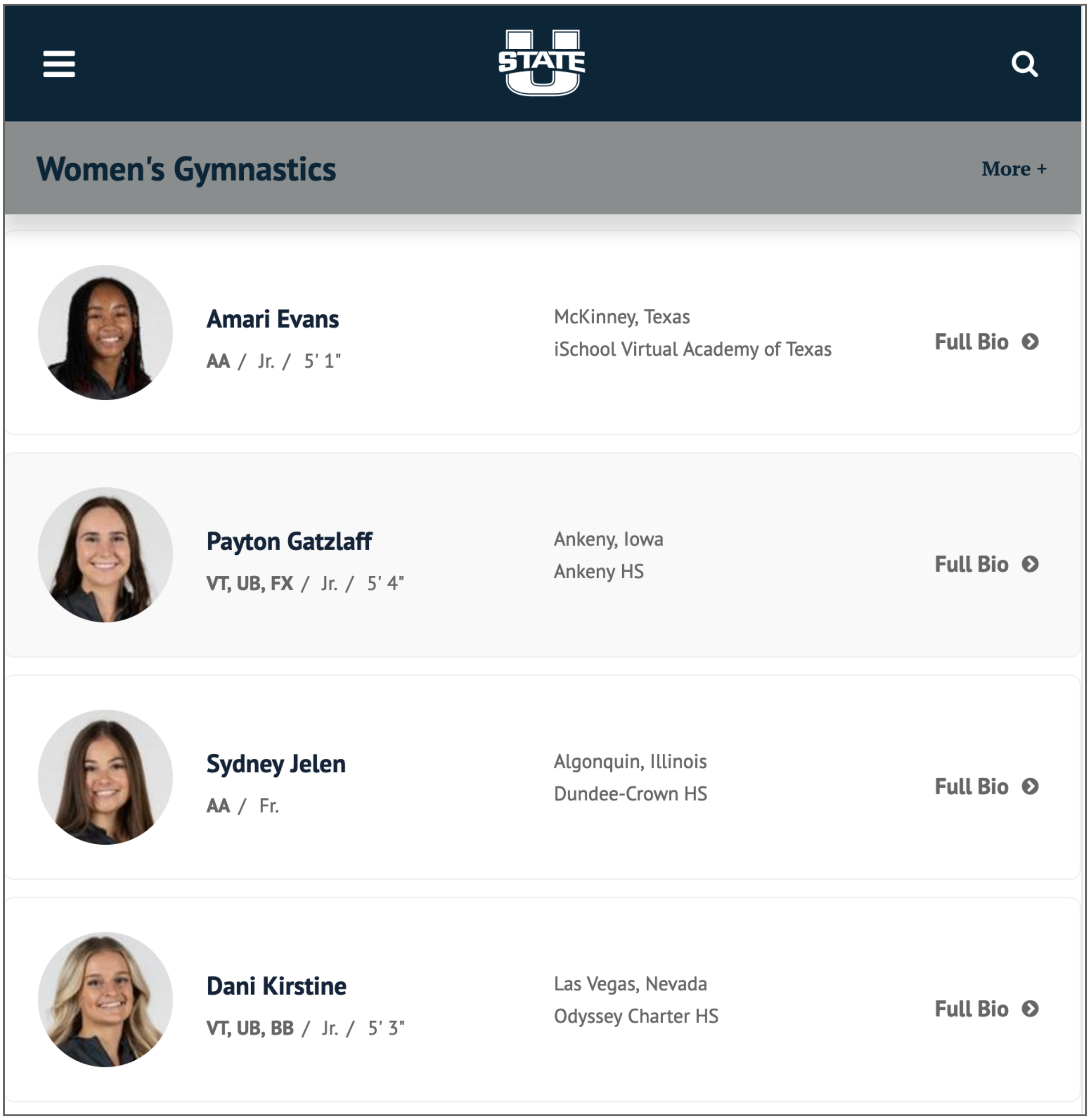
# Figures and Tables

**Table 1:** Team-Years Included in Sample

| Team | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|---|---|
| Air Force | | ✔ | | | ✔ | | |
| Alaska | | | | | | | ✔ |
| Arizona | ✔ | | | | | ✔ | |
| Boise State | | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ |
| Denver | | ✔ | | | ✔ | | |
| Illinois State | | | | | | ✔ | |
| Iowa | ✔ | | | | | | |
| Nebraska | | | | ✔ | | | |
| Penn State | | ✔ | | | | | |
| Southern Utah | ✔ | ✔ | ✔ | | ✔ | ✔ | ✔ |
| Sacramento State | | | | ✔ | | | |
| Texas Women's | | ✔ | | | | | |
| UC Berkeley | | ✔ | | | | | |
| UC Los Angeles | | | | ✔ | | | |
| Utah | ✔ | | ✔ | | | | |
| Utah State | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ | ✔ |
| Washington | | | | | | ✔ | |

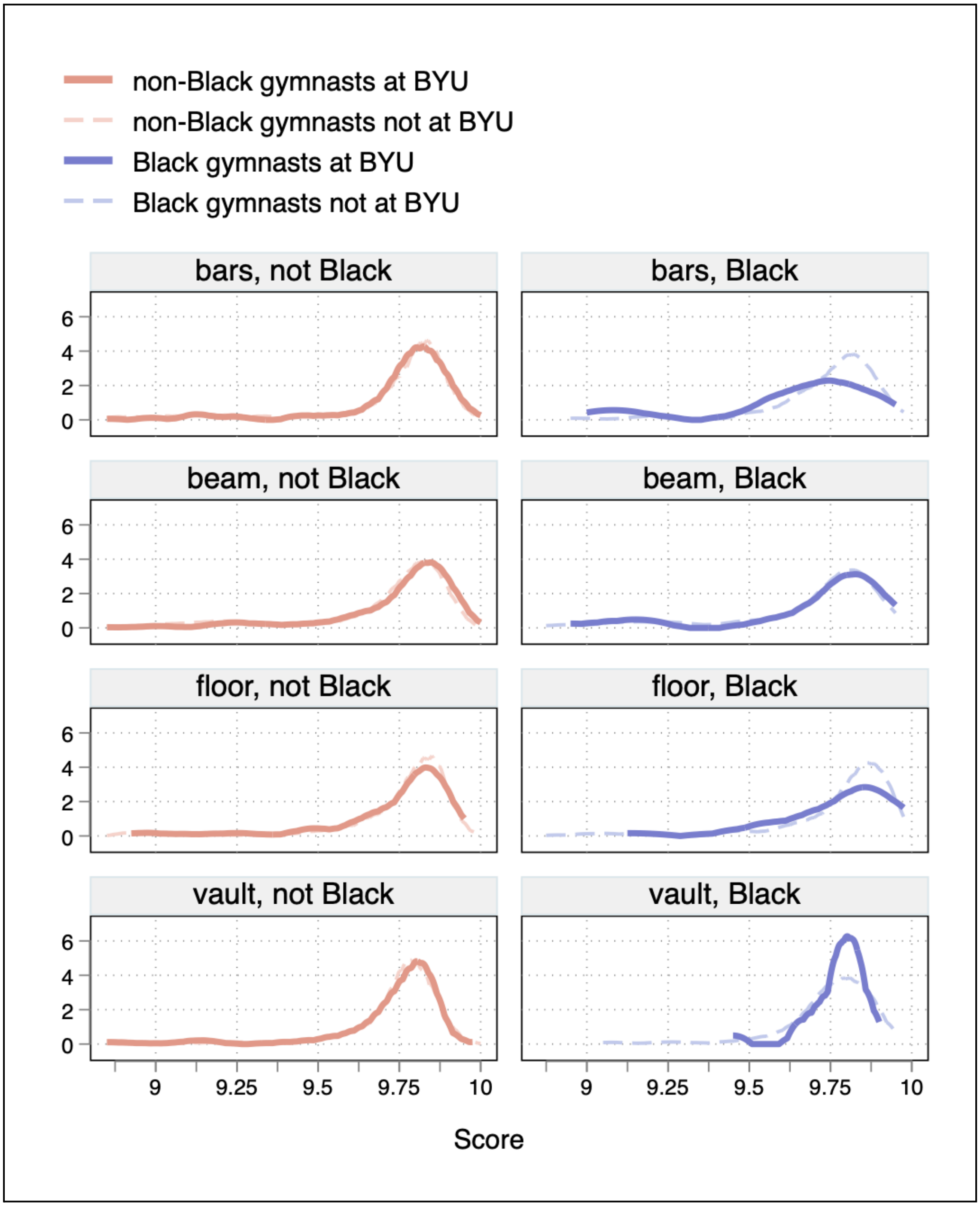
*Notes*. Each check represents a year in which the team on a given row competed at BYU in a regular season meet. If a team has a check for a given year, every regular season meet in which that team competed in that year is included in our dataset. All data and code can be accessed at github.com/tmorg46/uneven_bars.

**Figure 1:** Example Screenshot of a Women's Gymnastics Roster




*Notes*. This screenshot captures four gymnasts from the 2022 Utah State University women's gymnastics roster. The screenshot was taken from https://utahstateaggies.com/sports/womens-gymnastics/roster/2022 on 28 December 2023. More photos of each gymnast were available upon clicking their names. Amari Evans was coded as Black, while the other three gymnasts pictured were not.

**Figure 2:** Kernel Density Estimations of Scores at BYU & Not at BYU




*Note:* These kernel density estimates use an Epanechnikov kernel with the bandwidth selected by Silverman's h. All data and code can be accessed at github.com/tmorg46/uneven_bars.

**Table 2:** Score Summary Statistics

| VARIABLES | | (1) Vault | (2) Bars | (3) Beam | (4) Floor |
|---|---|---|---|---|---|
| Score | Mean | 9.738 | 9.673 | 9.682 | 9.735 |
| | (SD) | (0.170) | (0.415) | (0.286) | (0.244) |
| Black | | 0.143 | 0.127 | 0.105 | 0.129 |
| | | (0.350) | (0.333) | (0.306) | (0.336) |
| at BYU | | 0.100 | 0.101 | 0.102 | 0.100 |
| | | (0.300) | (0.301) | (0.303) | (0.300) |
| Black and at BYU | | 0.012 | 0.012 | 0.010 | 0.012 |
| | | (0.107) | (0.109) | (0.098) | (0.111) |
| Observations | | 2,240 | 2,256 | 2,267 | 2,244 |

*Notes*. The non-score variables are binary variables, so they represent the fraction of scores in a given event that fall in a given category. All data and code can be accessed at github.com/tmorg46/uneven_bars.

**Table 3**: Fixed Effects Regression Estimates of Interaction Term on Scores by Event

| | | Coefficient on (Black x atBYU) | | | |
|---|---|---|---|---|---|
| MODELS | | (1) Vault | (2) Bars | (3) Beam | (4) Floor |
| Baseline | Coefficient | 0.045** | -0.085 | -0.061 | 0.003 |
| | (Std. Error) | (0.022) | (0.076) | (0.067) | (0.039) |
| With Gymnast effects | | 0.036* | -0.049 | -0.071 | -0.017 |
| | | (0.019) | (0.063) | (0.055) | (0.031) |
| With Meet effects | | 0.092** | -0.009 | -0.035 | 0.025 |
| | | (0.042) | (0.075) | (0.073) | (0.042) |
| With Gymnast and Meet effects | | 0.043 | 0.025 | -0.064 | 0.001 |
| | | (0.033) | (0.069) | (0.080) | (0.041) |
| Observations | | 2,240 | 2,256 | 2,267 | 2,244 |
| R-squared, by Model | | 0.012 | 0.006 | 0.007 | 0.023 |
| | | 0.426 | 0.249 | 0.316 | 0.386 |
| | | 0.339 | 0.198 | 0.220 | 0.257 |
| | | 0.533 | 0.365 | 0.450 | 0.501 |

** p<0.05, * p<0.1
*Notes:* Standard errors in this table are clustered at the meet level. The first row estimates Equation 3, and the fourth row estimates Equation 4. All data and code can be accessed at github.com/tmorg46/uneven_bars.

**Table 4**: Triple Difference Estimates

| VARIABLES | | (1) Vault | (2) Bars | (3) Beam | (4) Floor |
|---|---|---|---|---|---|
| (Black x atBYU) | Coefficient | 0.090* | 0.052 | -0.010 | 0.040 |
| | (Std. Error) | (0.050) | (0.127) | (0.099) | (0.047) |
| (Black x atBYU x postFloyd) | | -0.095* | -0.083 | -0.089 | -0.071 |
| | | (0.057) | (0.140) | (0.139) | (0.076) |
| Constant | | 9.535** | 10.205** | 9.514** | 9.870** |
| | | (0.073) | (0.248) | (0.164) | (0.129) |
| Observations | | 2,240 | 2,256 | 2,267 | 2,244 |
| R-squared | | 0.541 | 0.371 | 0.462 | 0.516 |
| <u>F Tests</u> | | | | | |
| Row 1 + Row 2 = 0 | P Value | 0.866 | 0.579 | 0.311 | 0.607 |
| | (F Stat) | (0.03) | (0.31) | (1.03) | (0.26) |
| Row 1 = Row 2 | | 0.075* | 0.605 | 0.722 | 0.315 |
| | | (3.17) | (0.61) | (0.13) | (1.01) |

** p<0.05, * p<0.1

*Notes:* Each model also controls for meet number, in which the first regular season meet is marked as meet 1, the fifth as meet 5, and so on. All data and code can be accessed at github.com/tmorg46/uneven_bars.