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 7 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 fixed-effects regression model 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. Notwithstanding, when we split our analysis to pre- and post-May 2020 (the death of George Floyd), we find that coefficient estimates for being a Black gymnast at BYU become more negative generally, with two of the four event estimates showing statistically significant negative differences. We speculate on the implications of these results for people interested in the effects of environmental microaggressions on affected populations.

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

JEL Codes: J15, Z2, Z20

---

* 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 fixed-effects regression model 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. Notwithstanding, when we split our analysis to pre- and post-May 2020 (the death of George Floyd), we find that coefficient estimates for being a Black gymnast at BYU become more negative generally, with two of the four event estimates showing statistically significant negative differences. We speculate on the implications of these results for people interested in the effects of environmental microaggressions on affected populations.

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 the general issue discussed in Pope & Schweitzer (2011) by investigating whether this negative effect 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 place, they would experience it at BYU.

We analyze the performance of NCAA women gymnasts of all races who visited Brigham Young University at least once in a given season from 2017-2022. We include scores from the full span of the regular season in which they visited BYU. For example, we include every score from every Utah State University gymnast at every one of their regular season meets from 2017-2022 because USU visited BYU every year in that timespan, 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 visited BYU over that timespan. We choose to analyze women's gymnastics because the score a gymnast receives in a given event is given at the individual level. This is key, as any negative effect from environmental racial microaggression should also vary at the individual level.

We model gymnasts and their coaches as score maximizers and propose a general model for event scores in college-level gymnastics competition. We then estimate a fixed-effects regression 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. Though our model does not produce any statistically significant effect of being Black at BYU on scores in any of the four events, it does return some results that would be concerning if replicated on a larger dataset with greater statistical power.

II. Background & Related Literature

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, eight individual judges judge one event each with no rotation between events.

In each event, six gymnasts from each team perform, and the lowest of those six scores is dropped. Though an individual score can theoretically range anywhere from zero to 10 in each event, they typically cluster within the 9.5 - 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).

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 that uncomfortable experience is significant enough to translate to a performance effect 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 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 affecting decisions made by individual competitors, like Meissner, Rai, and Rotthoff's 2021 paper on the 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.)

Recent research that investigates the effects of environmental microaggressions on Black individuals is largely based on the model of microaggression theory 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 microaggression. This research is often focused on qualitative interviews or surveys of Black students' experiences at predominantly White institutions (or PWIs) (Mills 2020; Holliday & Squires 2021). This is also true of literature in this field historically, as evidenced by the many hundreds of papers on or related to the subject of Black students in PWI contexts published from 1965-2013 that are summarized in Willie & Cunnigen (1981), Sedlacek (1987), and Holliday & Squires (2020).

Our research contributes to the literature using 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 to this point.

III. Model & Data

This study uses hand-compiled data from official meet reports. These report sheets contain information on each gymnasts' score(s) at a meet, the judges for each event, and the venue of the meet. Most of those sheets were found at RoadToNationals.com, the official statistical website of

NCAA Gymnastics. Any sheets not found on RoadToNationals were pulled from individual team websites. 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. The data is well-suited for our study because it allows for analysis of scores across race groups and across meet venues. Because variation in scores is still small and the "perfect 10" upper bound can give undue weight to exceptionally low scores, we drop the top and bottom first percentiles of scores for each event prior to estimation.

Kernel density estimations of score distributions by event (Figure 1) show that scoring in vault and floor appears consistent between race groups and venues, while scoring distributions in beam and bars appear to be less consistent across the same groups. This apparent gap motivates our primary analysis, in which we use a fixed-effects regression model to see if those gaps are statistically significant. However, tracking the actual score a gymnast receives is not the only way to determine whether they were under any unexpected outside stress.

In NCAA scoring, every gymnast's routine is worth 9.5 points as a baseline value with additional points available for meeting certain requirements up to a maximum starting value of 10 points. At the skill level represented in our dataset, almost every routine any gymnast runs will be worth 10 points if performed perfectly; as such, a score below 9.5 on any given routine would represent a (potential series of) major error(s) that could be interpreted as being independent of any individual judge's decision or bias. A binary indicator marking such scores serves as the dependent variable in a conditional logit model that we run to supplement the primary analysis.

Our analysis is hindered by the number of scores received by Black gymnasts at BYU making up less than two percent of all scores in the dataset (see Table 1 for full summary statistics). For example, it makes running conditional logit on the large deduction binary variable impossible for the vault, as the dataset does not contain a single instance of a Black gymnast scoring that low on

vault when performing at BYU. Indeed, running any model with such a small sample in the "treated group" likely gives undue weight to outlier scores and obfuscates its results; we discuss these implications further in Section IV.

IV. Empirical Strategy

Our analysis begins by using a fixed-effects regression design to isolate any heterogeneous effect of competing at BYU for Black and non-Black gymnasts. Our set of fixed effects control for gymnast, venue, year, and judge-specific characteristics that might influence the score a gymnast receives at a given meet. In Table 2, we report the results from estimating the following linear model:

$$(1) \quad \text{Score}_{imvjy} = \alpha + \beta(\text{Black x atBYU})_{imv} + \text{meet}_m + [\text{gymnast effects}]_i$$
$$+ [\text{venue effects}]_v + [\text{judge effects}]_j + [\text{year effects}]_y$$
$$+ [\text{gymnast x year effects}]_{iy} + \varepsilon_{imvjy}$$

where subscripts *i, m, v, j,* and *y* refer to individual gymnasts, meet weeks, venues, judges, and years, respectively. The dependent variable is the score earned by a gymnast in a given event. Each of the square brackets denote a set of fixed effects, meet_m denotes the match week (i.e. $\text{meet}_m = 1$ for the first meet of the season), and the interaction term $(\text{Black x atBYU})_{iv}$ takes a value of 1 if a score is received by a Black gymnast competing at BYU and 0 otherwise. α is a regression constant, and the coefficient of interest is β, which we interpret as the differential impact of competing at BYU on Black gymnasts relative to non-Black gymnasts. To account for possible correlation between the scores received by a given gymnast, we report standard errors clustered at the gymnast level.

In the first row of Table 2, we report coefficients on the interaction term from regressions of event score on binary variables for being Black, being at BYU, and being Black at BYU. In the second row, we replace the binary Black and at BYU variables with gymnast and venue fixed effects,

respectively. In the third row, we add judge and year fixed effects, and in the final row, we add

gymnast by year linear time trends.

We then use a fixed-effects logistic regression to estimate odds ratios for getting a score

below 9.5 on any given individual routine. We control for match week number in all models, and add

various sets of fixed effects as above. In Table 3, we report odds ratios from estimating the following

logit model:

$$(2) \quad Logit(P_{low}) = ln(P_{low} / 1\text{-}P_{low}) = \tau_0 + \tau_1(\text{Black x atBYU}) + \tau_2(\text{meet number})$$
$$+ [\text{gymnast effects}] + [\text{venue effects}]$$
$$+ [\text{year effects}] + u_{low}$$

where P_{low} is the probability of getting a score below 9.5, $(P_{low} / 1\text{-}P_{low})$ is the reported odds ratio, the

set of τ_k represent marginal effects, and the square bracketed terms represent sets of fixed effects as

defined above. As in Equation (1), standard errors are clustered at the gymnast level.

In the first row of Table 3, we report odds ratios for the interaction term from the above

logit of event score on binary variables for being Black, being at BYU, and being Black at BYU. In

the second row, we replace the binary Black and at BYU variables with gymnast and venue fixed

effects, respectively, and in the third row, we add year fixed effects. Unlike in Table 2, we do not

include a specification with judge fixed effects or gymnast-year fixed effects, for two reasons. First,

including those sets of fixed effects made the log-likelihood function non-concave, precluding us

from including them. Second, it could be argued that these sets of fixed effects are not relevant

when considering such large deductions, as they seem more likely to occur if a gymnast is under

unusual pressure than as a result of any particular judge's influence or gymnast's experience level.

V. Empirical Results

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. The odds ratios and corresponding marginal effects obtained via logit are similarly insignificant. This analysis reveals that even the most obvious looking racial gaps in event scoring from Figure 1 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 models appear to be hindered by 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 1 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, our analysis provides no evidence to support the presence of intergroup anxiety or other unique social pressures for Black gymnasts visiting BYU.

VI. Conclusion

Using a unique dataset on collegiate women's gymnastics scores composed of teams that visit BYU, we test whether Black gymnasts experience a change in performance when competing at BYU that their non-Black competitors do not experience. Our data 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 find no significant difference in score distributions at BYU between Black and non-Black gymnasts in any individual

event and fail to detect a similar type of difference via logit odds ratio estimation, indicating no

evidence of intergroup anxiety or any similar bias affecting gymnasts' performance to a notable

degree. However, we note that a more robust sample could yield different results under the same

models if the score distributions we observe in our present sample persisted.

FIGURE 1

KERNEL DENSITY ESTIMATIONS OF SCORE DISTRIBUTIONS
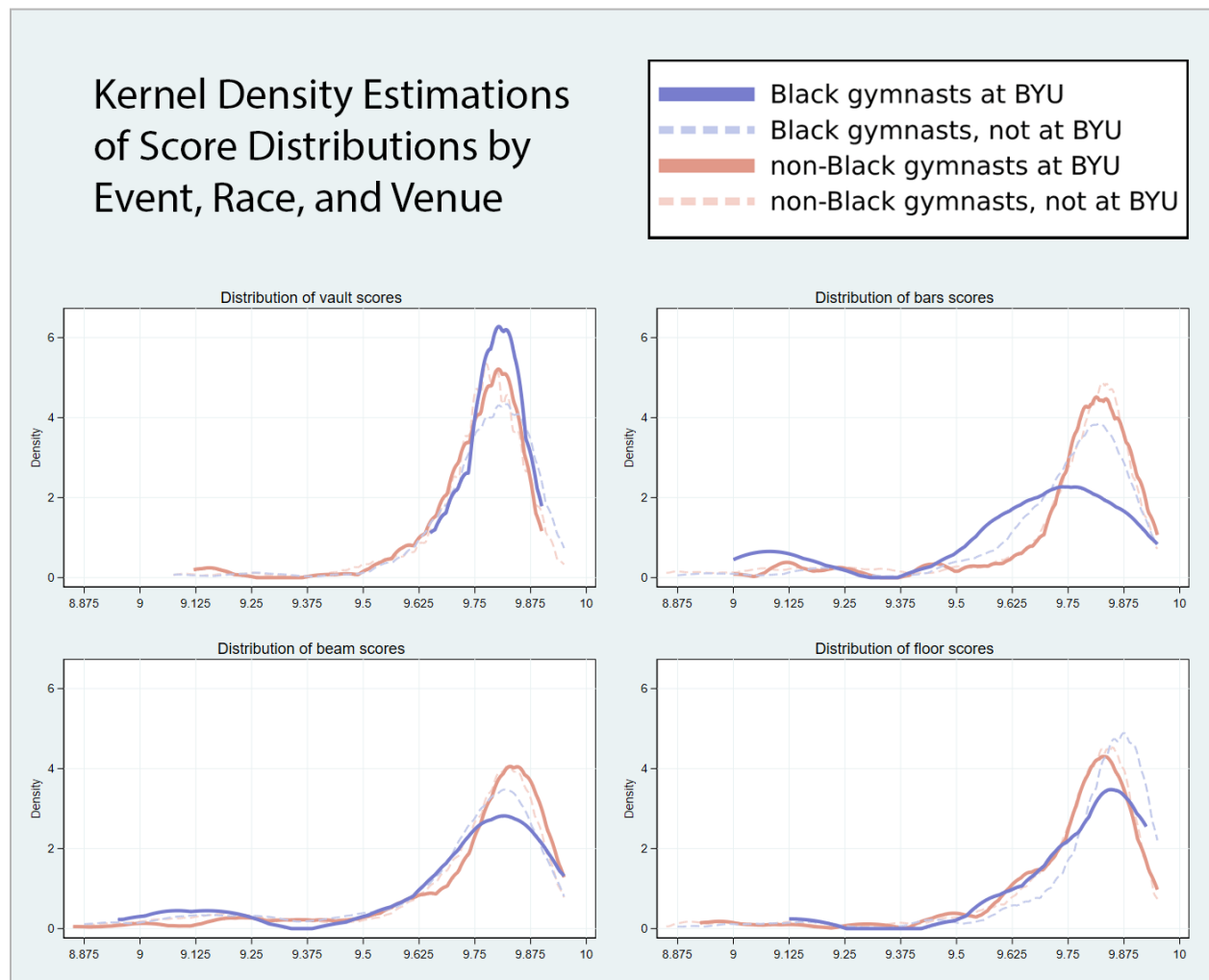
Note: It uses an Epanechnikov kernel with the bandwidth selected by Silverman's h. All data and code can be accessed at github.com/tmorg46/uneven_bars.

TABLE 1. SUMMARY STATISTICS

VARIABLES		(1) Vault	(2) Bars	(3) Beam	(4) Floor
Score	Mean	9.747	9.686	9.691	9.739
	(SD)	(0.160)	(0.409)	(0.277)	(0.241)
Deducted		0.033	0.120	0.133	0.067
		(0.180)	(0.325)	(0.340)	(0.250)
Black		0.170	0.131	0.113	0.131
		(0.375)	(0.337)	(0.316)	(0.337)
at BYU		0.093	0.093	0.095	0.093
		(0.290)	(0.291)	(0.294)	(0.290)
Black and at BYU		0.014	0.012	0.010	0.012
		(0.117)	(0.110)	(0.097)	(0.110)
Observations		2,035	2,046	2,057	2,030

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 2. Fixed Effects Regression Estimates of Interaction Term on Scores by Event

MODELS		Coefficient on (Black x atBYU)			
		(1) Vault	(2) Bars	(3) Beam	(4) Floor
OLS on Black, atBYU, meet week, and interaction	Coefficient (Std. Error)	0.027** (0.014)	-0.117** (0.050)	-0.027 (0.064)	-0.008 (0.040)
Adding Gymnast and Venue effects		0.016 (0.016)	-0.079* (0.047)	-0.041 (0.070)	-0.014 (0.040)
Adding Year and Judge effects		0.021 (0.018)	-0.040 (0.050)	-0.024 (0.075)	-0.008 (0.048)
Adding Gymnast-by-Year linear time trends		0.020 (0.019)	-0.035 (0.051)	-0.024 (0.078)	-0.008 (0.049)
Observations		2,009	2,009	2,057	1,995
R-squared, by Model		0.015	0.005	0.006	0.024
		0.424	0.309	0.322	0.387
		0.523	0.396	0.426	0.482
		0.567	0.437	0.462	0.533

** Significant at the 5% confidence level

 * Significant at the 10% confidence level

TABLE 3
FIXED EFFECTS LOGIT ODDS RATIOS OF BEING DEDUCTED (SCORE<9.5)

MODELS		(1) Bars	(2) Beam	(3) Floor
Logit on Black, atBYU, meet week, and interaction	Odds Ratio (Std. Error)	2.763 (1.801)	1.264 (0.921)	0.526 (0.605)
Introduce Gymnast and Venue effects		2.058 (1.596)	1.494 (1.270)	0.512 (0.626)
Adding Year fixed effects		2.035 (1.585)	1.491 (1.265)	0.597 (0.727)
Observations, by Model		2,009 1,476 1,476	2,030 1,467 1,467	1,995 996 996

** Significant at the 5% confidence level

 * Significant at the 10% confidence level

TABLE 1A. Fixed Effects Regression Estimates on Vault Scores

VARIABLES		(1)	(2)	(3)	(4)
Black and at BYU	Coefficient	0.027**	-0.117**	-0.027	-0.008
	(Std. Error)	(0.014)	(0.050)	(0.064)	(0.040)
Black		0.016	-0.079*	-0.041	-0.014
		(0.016)	(0.047)	(0.070)	(0.040)
At BYU		0.021	-0.040	-0.024	-0.008
		(0.018)	(0.050)	(0.075)	(0.048)
Constant		0.020	-0.035	-0.024	-0.008
		(0.019)	(0.051)	(0.078)	(0.049)
Observations		2,057	2,057	2,057	2,057
R-squared		0.007	0.328	0.426	0.462
Gymnast Effects			X	X	X
Venue Effects			X	X	X
Year Effects				X	X
Judge Effects				X	X
Gymnast-Year Effects					X

** Significant at the 5% confidence level * Significant at the 10% confidence level

References

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

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

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

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

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

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