

# Race and Gender biases in Student Evaluations of Teachers

Carolyn Chisadza\*

Nicky Nicholls<sup>†</sup>

Eleni Yitbarek <sup>‡</sup>

March 4, 2019

## Abstract

Student ratings of teaching (SETs) are vital for academic career trajectories of higher education lecturers. Although student bias against female lecturers is noted in previous studies, mostly in the developed world, the extent to which race affects such ratings has received limited attention. To better understand the role of race and gender bias in SETs, we conduct an experiment in South Africa, where racial bias is highly prevalent. Students are randomly assigned to follow video lectures with identical narrated slides and script but given by lecturers of different race and gender. We find that black lecturers receive lower ratings than white lecturers, particularly from black students.

**Key words:** student evaluations of teaching, gender and race bias, Africa

**JEL classification:** I23; J15; J16

## 1 Introduction

Lecturers often find themselves under pressure from two unforgiving forces: their institutions which expect them to balance teaching and research, and the students' expectations. To compound this situation, lecturers also have to contend with student evaluations of teaching (SETs) which are often subject to the students' biases ([Boring, 2017](#)). Given the importance of SETs in performance assessments of lecturers to ensure a better quality of education, understanding the potential biases in these assessments has been the focus of much research. However, while gender bias has been studied in SETs, mostly in developed countries, race has received limited attention ([Wagner et al., 2016](#)). This study aims to understand whether

---

\*University of Pretoria, Department of Economics

<sup>†</sup>Corresponding Author: Department of Economics, Tukkiewerf Building, University of Pretoria, Lynnwood Road, Pretoria, 0002, RSA, email: [nicky.nicholls@up.ac.za](mailto:nicky.nicholls@up.ac.za).

<sup>‡</sup>University of Pretoria, Department of Economics

student biases, particularly gender and race-based bias, influence the way students assess lecturers' performances. If students' assessments are biased against lecturers of a particular race and/or gender, this would call into question the fairness of using such metrics to evaluate lecturers' teaching abilities. Using biased SETs not only affects career progressions of lecturers, but also undermines the effort of higher education institutions to attract and maintain minority groups in an academic career.

To date, several experimental studies have focused on gender biases in SETs. An experiment conducted by [MacNell et al. \(2015\)](#) noted biases in favor of (male or female) lecturers perceived as male, which disappeared when comparing scores for the actual male versus female lecturer.<sup>1</sup> On the other hand, bias in favour of female lecturers is noted in [Bachen et al. \(1999\)](#) and [Rowden and Carlson \(1996\)](#).

This study contributes to the existing literature by looking at both race and gender biases in SETs using a randomized controlled experiment. Our study is conducted in South Africa, where the inclusion of race is particularly relevant given the racial divide that occurred during the Apartheid regime and which continues to plague South Africa today. In South African universities today less than one in ten university professors is black.<sup>2</sup> This is not encouraging given that 79% of South Africans are black. With regard to gender, more men are employed than women in academic roles, with the gap most prominent in senior positions.<sup>3</sup> Since the lower representation of black and female lecturers exists among junior academic staff too, this inequality is likely to persist, if not addressed.<sup>4</sup>

## 2 Data

### 2.1 Participants

We conducted an experiment with first year economics students at the University of Pretoria, South Africa. The experiment took the form of a quiz assignment as part of the first year economics course. In order not to bias responses, students were only notified at the end of the data gathering that their responses formed part of an experiment.<sup>5</sup> The quiz mark formed part of the students' assessment, giving students a strong incentive to follow the lecture. At the end of the online lecture (the experiment), students could either consent to

---

<sup>1</sup>See [Boring \(2017\)](#) and [MacNell et al. \(2015\)](#) for an overview of studies that find bias against female lecturers.

<sup>2</sup><http://www.thedailyvox.co.za/does-a-lecturers-race-matter-to-university-students>

<sup>3</sup>Females make up 52% of the total population (Census 2011, Statistics South Africa).

<sup>4</sup>[http://www.che.ac.za/focus\\_areas/higher\\_education\\_data/2012/staffing](http://www.che.ac.za/focus_areas/higher_education_data/2012/staffing)

<sup>5</sup>To comply with University of Pretoria ethics requirements, students had to give their informed consent for us to use their responses in our experiment.

have their responses used for the study, or request that only their quiz responses be saved for assessment purpose. Our final sample was composed of 1,599 first year economics students, out of 1,770 students. However, some students elected not to disclose their race. Of the 1,487 students who reported their race, 46%, 41%, and 96% were white, female, and South African, respectively.

## 2.2 Experiment

Students were required to watch a 12-minute video. Following the video, they answered a short multiple choice quiz on the video content and were then asked a series of SET-type questions to evaluate the quality of the lecturer and the content of the video. The assignment was accessed through the online Blackboard platform used by the University.

There were four different versions of the video, where only the lecturer's race and gender was varied. The videos, therefore, showed a lecturer who was either Black Male, Black Female, White Male or White Female. A number of measures were used to limit variation between videos to only lecturer's race and gender. First, we used an identical script and identical slides in the videos. Second, with the objective of controlling differences in body language and facial expression, the videos were presented as narrated slides rather than a live video. However, to make the lecturer's race and gender obvious to students, a photograph of the lecturer was included in the bottom corner of all slides. All lecturers were photographed with a neutral facial expression on a white background. Third, videos were re-recorded until all had a similar total duration so that all lecturers were speaking at a similar pace. Further, although qualification of the lecturers was not mentioned to students, all lecturers had identical qualifications, were of similar age and were not known to students. Finally, to ensure that English was not a confound, none of our four lecturers had English as their mother tongue.<sup>6</sup>

Selection bias was avoided by randomly assigning students to watch one of the four videos.<sup>7</sup> In order to minimize bias from students becoming aware of the different video versions, this assignment was done during class time. Since we have multiple first-year economics student groups, class times were selected where all groups had classes in two consecutive sessions. Students did the assignment in designated university computer labs. A supervisor was present to assist with technical issues and to ensure that each student watched the video on a separate device. In this way, we ensured that each student had

---

<sup>6</sup>Although most tertiary institutions in South Africa teach in English, English is not the mother tongue of most South Africans.

<sup>7</sup>To confirm the absence of student selection bias, we follow [Boring \(2017\)](#) and undertake Pearson correlation tests and a Chi-squared test. We find small and insignificant correlations between gender and race of students and lecturers.

watched the video randomly assigned individually to him/her. After watching the video, as is typical for SETs, students were asked to rate the lecturer on each dimension on a 5-point Likert-type scale. Table 1 gives the descriptive statistics of SET scores for overall lecture quality by lecturers’ race and gender. Overall, students give higher ratings to female and white lecturers. These differences are most pronounced for female and white students.

**Table 1** – Mean scores of overall quality by lecturer and gender & race

	<b>Lecturer</b>					
	Female	Male	Difference: Female-Male	White	Black	Difference: White-Black
<b>Student</b>						
Female	3.950	3.745	0.205***	3.969	3.722	0.247***
Male	3.861	3.705	0.157***	3.836	3.732	0.104*
Black	3.914	3.787	0.127**	3.940	3.764	0.176**
White	3.945	3.755	0.190***	3.939	3.753	0.186**
<b>Observations</b>						
Female	443	442		451	434	
Male	362	352		359	355	
Black	461	454		450	465	
White	344	340		360	324	

t-tests significance levels: \*: 10%\*\*: 5% \*\*\*: 1%

### 3 Analysis: student gender and race biases

Like other universities, University of Pretoria relies on overall lecture quality scores to measure teaching effectiveness. Accordingly, our analysis focuses on overall lecture quality SET scores. Let  $SETQ$  denote the vector of overall lecture quality scores. Our specification for the gender bias takes the following form:

$$SETQ = \alpha + \beta \mathbf{G} + \gamma \mathbf{x} + \varepsilon, \quad (1)$$

where  $\beta$ , and  $\gamma$  are vectors of parameters to be estimated ( $\alpha$  is the intercept),  $\mathbf{x}$  is a matrix of covariates of student characteristics and  $\mathbf{G}$  is a matrix, where the row represents the students’ gender and the column denotes the lecturers’ gender.<sup>8</sup>  $\varepsilon$  is the error term which is assumed to be i.i.d. We follow a similar specification for race only analysis. Here equation 1 includes a matrix for student and lecturer race instead of their gender.

<sup>8</sup> $g_{11}$ ,  $g_{12}$ ,  $g_{22}$  and  $g_{21}$  are dummy variables that equal one if the SET score is given by a female student to a female lecturer, a female student to a male lecturer, a male student to a male lecturer and a male student to a female lecturer, respectively.

When we interact the race and gender of students and lecturers, the estimated specification takes the following form:

$$SETQ = \alpha + \beta\mathbf{Z} + \gamma\mathbf{x} + \varepsilon, \quad (2)$$

where  $\mathbf{Z}$  stands for a matrix with the row representing the gender and race of students and the column denoting the gender and race of lectures.<sup>9</sup>

## 4 Results

We start by considering gender and race impacts on SET scores separately. Each panel in Table 2 reports 3 regressions: regression (1) includes only the gender (in panel A) or race (in panel B) interaction between students and lecturers. Regression (2) adds covariates: the student’s grade on the quiz and dummy variables for student department (first-year economics is a required module in many other degrees) and for student’s home language. Regression (3) adds the student’s rating of the lecturer’s accent.<sup>10</sup>

Our gender results indicate positive bias towards female lecturers from female students and to a lesser extent from male students, relative to the reference group: female students rating male lecturers. This result is somewhat surprising, since bias against female lecturers is common in the empirical literature. Perceived ease of understanding the lecturer’s accent is significant, and including this variable results in the gender interaction variables losing significance. The quiz grade has a small but positive correlation with ratings, consistent with other studies.<sup>11</sup>

Turning to the race results, we note negative bias against black lecturers from black students and to a lesser extent from white students, relative to the reference group: black students rating white lecturers. The bias against black lecturers from black suggests possible presence of “internalized racism”. The racial division of labor during Apartheid might have led black people to consciously or unconsciously accept racial hierarchy. Perceptions of the ease of understanding the lecturer’s accent is again highly significant, and race interaction variables lose significance after the lecturer’s accent is included in the regressions.

---

<sup>9</sup>For instance,  $z_{11}$ ,  $z_{22}$ ,  $z_{33}$  and  $z_{44}$  are dummy variables that equal one if the SET score is given by a white female student to a white female lecturer, a black female student to a black female lecturer, a white male student to a white male lecturer and a black male student to a black male lecturer, respectively.

<sup>10</sup>South Africans from different race and language groups speak English with slightly different accents.

<sup>11</sup>Regression analysis confirms that the quiz grade was not significantly impacted by lecturer race or gender.

**Table 2** – Determinants of overall quality scores, OLS regressions

Dependent variable:	Gender only (A)†		Race only (B)‡	
	(1)	(2)	(1)	(3)
Male student & male lecturer	-0.0409 (0.0622) [0.509, 0.509]	Overall quality of lecture -0.0468 (0.0623) [0.455, 0.455]	Black student & black lecturer -0.179*** (0.0556) [0.002, 0.004]	-0.173*** (0.0555) [0.073, 0.176]
Female student & female lecturer	0.205*** (0.0547) [0.000, 0.001]	0.199*** (0.0548) [0.000, 0.002]	White student & white lecturer 0.0521 (0.0569) [0.351, 0.351]	0.0603 (0.0570) [0.284, 0.284]
Male student & female lecturer	0.116* (0.0597) [0.056, 0.094]	0.107 (0.0598) [0.073, 0.126]	White student black & lecturer -0.134** (0.0602) [0.026, 0.048]	-0.127* (0.0601) [0.033, 0.063]
Quiz grade		0.0269** (0.0133)	Quiz grade 0.0261** (0.0140)	0.0259* (0.0662) [0.447, 0.662]
Economics Department		0.144 (0.110)	Economics Department 0.165 (0.110)	0.260** (0.129)
English home language		-0.0373 (0.0420)	English home language -0.0400 (0.0431)	-0.0121 (0.0430)
Accent is easy to understand			Accent is easy to understand 0.887*** (0.0746)	0.916*** (0.0783)
Constant	3.745*** (0.0419)	3.530*** (0.119)	Constant 3.887*** (0.118)	3.669*** (0.140)
N	1595	1595	N	1324
Adj. R-sq	0.012	0.014	Adj. R-sq	0.013

†Reference group = Female student & male lecturer

‡Reference group = Black student & white lecturer

Standard errors in parentheses

Raw and Romano-Wolf (Romano and Wolf, 2005; Romano and Wolf, 2016) step-down adjusted p-values in braces, respectively.

Significance levels based on Romano-Wolf step-down adjusted p-values: \*, 10% \*\*; 5% \*\*\*; 1%. We consider for each regression 3 hypotheses for student/teacher gender or race interaction.

Covariates for regressions 2 included quiz grade, whether the student is in the economics department and whether English is the student's home language;

with the ease of understanding lecturer accent added for regressions 3.

In Table 3, we consider the interaction between race and gender for all possible combinations of student and lecturer race and gender. Here we note positive bias towards white female lecturers from both black and white female students. As before, an easy to understand accent has a positive impact on SET score, but the race and gender bias loses significance after controlling for accent.

**Table 3** – Determinants of overall quality scores, OLS regressions

Dependent variable : Overall quality of lecture	Gender and race interaction †					
	(1)	(2)		(3)		
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Black-female student & black-female lecturer	0.000848 [0.991, 0.991]	(0.109)	0.00337 [0.976, 0.976]	(0.109)	-0.121 [0.301, 0.931]	(0.116)
Black-female student & white-female lecturer	0.297* [0.002, 0.052]	(0.109)	0.289* [0.004, 0.063]	(0.109)	0.201 [0.089, 0.525]	(0.116)
Black-female student & black-male lecturer	-0.157 [0.147, 0.714]	(0.108)	-0.156 [0.151, 0.727]	(0.108)	-0.0847 [0.467, 0.989]	(0.117)
Black-female student & white-male lecturer	0.0626 [0.596, 0.950]	(0.111)	0.0675 [0.571, 0.942]	(0.111)	0.0392 [0.779, 0.999]	(0.119)
White-female student & black-female lecturer	0.108 [0.309, 0.927]	(0.124)	0.112 [0.288, 0.929]	(0.124)	0.232 [0.061, 0.470]	(0.134)
White-female student & white-female lecturer	0.283* [0.003, 0.093]	(0.117)	0.289* [0.003, 0.090]	(0.117)	0.203 [0.099, 0.545]	(0.122)
White-female lecturer & black-male lecturer	-0.131 [0.329, 0.921]	(0.121)	-0.119 [0.387, 0.929]	(0.121)	-0.0127 [0.935, 0.999]	(0.131)
White-female student & white-male lecturer	0.111 [0.333, 0.927]	(0.116)	0.118 [0.313, 0.929]	(0.116)	0.0636 [0.607, 0.998]	(0.122)
Black-male student & white-female lecturer	0.0959 [0.423, 0.927]	(0.118)	0.0859 [0.471, 0.942]	(0.118)	-0.0271 [0.838, 0.999]	(0.124)
Black-male student & black-male lecturer	-0.115 [0.429, 0.927]	(0.118)	-0.113 [0.439, 0.929]	(0.118)	-0.0525 [0.742, 0.999]	(0.129)
Black-male student & white-male lecturer	-0.0719 [0.514, 0.950]	(0.116)	-0.0750 [0.495, 0.942]	(0.116)	-0.131 [0.292, 0.931]	(0.123)
White-male student & black-female lecturer	0.0675 [0.551, 0.950]	(0.122)	0.0643 [0.570, 0.942]	(0.122)	0.123 [0.341, 0.951]	(0.131)
White-male student & white-female lecturer	0.186 [0.126, 0.661]	(0.122)	0.185 [0.121, 0.659]	(0.122)	0.0466 0.724, 0.999	(0.128)
White-male student & black-male lecturer	-0.146 [0.255, 0.884]	(0.125)	-0.149 [0.250, 0.877]	(0.125)	-0.0159 [0.914, 0.999]	(0.139)
White-male student & white-male lecturer	0.0450 [0.688, 0.950]	(0.124)	0.0482 [0.669, 0.942]	(0.124)	-0.0597 [0.667, 0.999]	(0.129)
Quiz grade			0.0253*	(0.0130)	0.0274**	(0.0134)
Economics department			0.157	(0.112)	0.232*	(0.122)
English home language			-0.0365	(0.0419)	-0.00985	(0.0435)
Accent is easy to understand					0.893***	(0.0673)
Constant	3.778***	(0.0823)	3.570***	(0.138)	2.845***	(0.156)
N	1595		1595		1324	
Adj. R-sq	0.020		0.022		0.142	

†Reference group is black male student & black female lecturer

Raw and Romano-Wolf (Romano and Wolf, 2005; Romano and Wolf, 2016) step-down adjusted p-values in braces, respectively.

Significance levels based on Romano-Wolf step-down adjusted p-values: \*: 10% \*\*: 5% \*\*\*: 1%.

We consider 15 hypotheses for student/teacher race and gender interaction.

Covariates in the second regression included quiz grade, whether the student

is in the economics department and whether English is the student's home language;

with the ease of understanding the lecturer's accent added in the third regression.

To investigate the role of lecturer's accent further, we regress accent ratings on the same explanatory variables in Table 4. Ratings of black male lecturers from all four possible student race and gender combinations have significant negative coefficients; while the black female lecturer accent is rated negatively by all groups except black females.

It is particularly surprising that black students give lower ratings to black lecturers' accents, since any challenges in understanding accents would be expected between the different races i.e. when the lecturer and student races differ. The lower ratings of black lecturers' accents by black students, together with the lower ratings of lecture quality for black lecturers from black students in Table 2 supports the earlier argument for "internalized racism". Given that more than 85 percent of university lecturers are white, black students might think that black lecturers are not fit for the job. Accent ratings may also be used by students as a channel for expressing bias in a more acceptable way: expressing difficulty in understanding a lecturer's accent would be subject to less social sanction than complaining about race.



Table 4 – Determinants of easy to understand accent, OLS regressions

Gender only†		Race only§		Gender and race interaction‡				
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.		
Male student & male lecturer	-0.00352	(0.0742)	Black student & black lecturer	-0.223***	(0.0605)	Black-female student & black-female lecturer	-0.145	(0.115)
	[ 0.964, 0.964]			[0.000, 0.001]			[0.145 0.474]	
Female student & female lecturer	0.319***	(0.0653)	White student & white lecturer	0.106***	(0.0614)	Black-female student & white-female lecturer	-0.0382	(0.117)
	[0.000, 0.000]			[0.085, 0.085]			[0.705 0.885]	
Male student & female lecturer	0.257*	(0.0677)	White student & black lecturer	-0.532***	(0.0734)	Black-female student & black-male lecturer	-0.536***	(0.127)
	[0.000, 0.001]			[0.000, 0.000]			[0.000 0.000]	
						Black-female student & white-male lecturer	-0.276*	(0.119)
						White-female student & black-female lecturer	[0.010 0.057]	
							-0.536***	(0.142)
						White-female student & white-female lecturer	[0.000 0.001]	
							0.0433	(0.124)
						White-female lecturer & black-male lecturer	[0.695 0.885]	
							-0.935***	(0.158)
						White-female student & white-male lecturer	[0.000 0.000]	
							-0.170	(0.132)
						Black-male student & white-female lecturer	[0.174 0.474]	
							-0.0621	(0.115)
						Black-male student & black-male lecturer	[0.501 0.851]	
							-0.596***	(0.140)
						Black-male student & white-male lecturer	[0.000 0.000]	
							-0.297*	(0.130)
						White-male student & black-female lecturer	[0.015 0.057]	
							-0.432***	(0.140)
						Black-male student & black-female lecturer	[0.001 0.006]	
							-0.308**	(0.128)
						White-male student & black-male lecturer	[0.007 0.048]	
							-0.891***	(0.162)
						White-male student & white-male lecturer	[0.000 0.000]	
							-0.106	(0.131)
Quiz grade	0.00649	(0.0158)	Quiz grade	0.00741	(0.0156)	Quiz grade	[0.358 0.745]	(0.0155)
Department	-0.0728	(0.143)	department	-0.0589	(0.146)	Department	0.00278	(0.147)
English home language	0.0177	(0.0486)	English home language	0.0182	(0.0478)	English home language	-0.0637	(0.0474)
Constant	3.648***	(0.141)	Constant	3.934***	(0.142)	Constant	0.0178	(0.0474)
							4.139***	(0.161)
N	1597		N	1597		N	1597	
Adj. R-sq	0.020		Adj. R-sq	0.052		Adj. R-sq	0.072	

†Reference group = Female student & male lecturer

§Reference group = Black student & white lecturer

‡Reference group = white male student & white female lecturer

Standard errors in parentheses

Raw and Romano-Wolf (Romano and Wolf, 2005; Romano and Wolf, 2016) step-down adjusted p-values in braces, respectively.

Significance levels based on Romano-Wolf step-down adjusted p-values: \*, 10% \*\*; 5% \*\*\*; 1%. We consider, respectively for the 3 regressions, 3, 3 and 15 hypotheses for student/teacher race and gender interaction.

Covariates included quiz grade, whether the student is in the economics department and whether English is the student's home language.

## 5 Discussion

While several studies have noted biases in favor of male lecturers, we find bias towards female lecturers and bias against black lecturers. Accent also appears to play a role as a (possibly more socially acceptable) channel to express biases. Since 69% of school teachers in South Africa are female, students in their first year after high school might carry an expectation that people in a teaching role should be female.<sup>12</sup> The bias against black lecturers, while perhaps not surprising given South Africa’s apartheid history, highlights a real concern: If black lecturers receive lower SET scores based on their race, this could translate into slower career progression. Awareness of slower career progression among black academics, in general, might deter talented black students from considering an academic career. Further, a lack of black academic role models might negatively impact black students’ performance at university. It is particularly disconcerting that black students give lower ratings to black lecturers. Perhaps the relatively low representation of black people in academic careers in South Africa is impacting student expectations of lecturer race. A possible vicious cycle whereby the low representation of black academics helps to drive poor perceptions of black lecturers’ performance among students, resulting in fewer black people entering and succeeding in academia is a serious concern.

Our findings underline the importance of further research into the impact of race bias in SETs across South African universities as well as in other countries where race bias might exist. A better understanding of such biases may help address the delayed race transformation in higher education.

## References

- Bachen, C. M., McLoughlin, M. M., and Garcia, S. S. (1999). Assessing the role of gender in college students’ evaluations of faculty. *Communication Education*, 48(3):193–210.
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145:27–41.
- MacNell, L., Driscoll, A., and Hunt, A. N. (2015). What’s in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(2015):291–303.
- Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

---

<sup>12</sup>Recall that our sample is composed of first year undergraduate students.

- Romano, J. P. and Wolf, M. (2016). Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. *Statistics & Probability Letters*, 113:38–40.
- Rowden, G. V. and Carlson, R. E. (1996). Gender issues and students' perceptions of instructors' immediacy and evaluation of teaching and course. *Psychological Reports*, 78(3):835–839.
- Wagner, N., Rieger, M., and Voorvelt, K. (2016). Gender, ethnicity and teaching evaluations: Evidence from mixed teaching teams. *Economics of Education Review*, 54:79–94.