

2014

Fair and Balanced? Experimental Evidence on Partisan Bias in Grading

Paul Musgrave
Dickinson College

Mark Rom

Follow this and additional works at: http://scholar.dickinson.edu/faculty_publications



Part of the [Political Science Commons](#)

Recommended Citation

Musgrave, Paul and Rom, Mark, "Fair and Balanced? Experimental Evidence on Partisan Bias in Grading" (2014). *Dickinson College Faculty Publications*. Paper 78.
http://scholar.dickinson.edu/faculty_publications/78

This article is brought to you for free and open access by Dickinson Scholar. It has been accepted for inclusion by an authorized administrator. For more information, please contact scholar@dickinson.edu.

Fair and Balanced?

Experimental Evidence On Partisan Bias In Grading

Abstract: Is grading polarized in political science classrooms? We offer experimental evidence that suggests it is not. Many have argued that instructors' grading in political science classrooms is skewed by the political characteristics of the instructor, the student, or an interaction between the two. Yet the evaluations of whether such biases exist has been asserted and denied with little evidence—even though prominent theories in political science suggest that the charge is not entirely implausible. Using a set of anonymous essays by undergraduates graded by teaching assistants at a variety of institutions, we test for the presence of bias in a framework that avoids the usual selection bias issues that confound attempts at inference. After evaluating the evidence carefully, we find that the evidence for bias is much weaker than activists claim.

Keywords: polarization, bias, undergraduate education, pedagogy

Acknowledgements The authors thank participants at the 2012 APSA Teaching and Learning Conference, the 2012 Midwest Political Science Association annual meeting, Hans Noel, Jonathan Ladd, Patrick Carr, and three anonymous reviewers for their comments. The authors also thank Ashley Branca and Rose Tintera-Baldauf for their research assistance, and Rachel Blum for proofreading assistance.

Introduction

Political bias in the academy is a subject of substantial popular controversy (for example, Mariani & Hewitt 2008; Maranto & Hess 2009; Tierney 2011; Yancy 2011; Wood 2011; Fosse, Freese & Gross 2011; Schmidt 2011; Saad 2012; Savlov 2012; Chiaramante 2012; Gross 2013 among many others.). Conservatives often argue that liberals use their dominance in college classrooms to indoctrinate students (e.g. Horowitz & Laksin 2009; Gross 2013) or to punish conservative students for their beliefs by lowering their grades (Elder 2012; Boyle 2011). Former Republican senator Rick Santorum, for instance, said during the 2012 presidential primary that he was “docked for my conservative views” during his college years (Sonmez 2012). Academics’ responses to such charges often come as denunciations of studies that purport to demonstrate liberal bias as propaganda or as attacks on academic freedom (e.g. Lee 2006; Giroux 2006; Gitlin 2006; Gross 2006). But in a time when political science faces greater partisan scrutiny, it seems ill-advised for the scholarly community to argue (or insist) that it is immune from bias without providing evidence to support that claim.

Scholarly claims of neutrality notwithstanding, it is at least plausible that instructors do consider students’ partisanship in assigning scores despite the strong scholarly norm against it. The political science literature on the nature of partisan perceptions suggests bias could operate even if instructors are unaware of their students’ partisan predilections and even if they grade students’ work anonymously. If partisans see the world through distinctive filters, then their interpretations of even objective facts might be distinctive as well (e.g. Converse et al, 1960, and Gaines et al, 2007).

We label the Horowitz-Santorum claims as the “political affinity bias” conjecture. To make it amenable to testing, we formalize it in two hypotheses. The first, more general, claim is

that instructors tend to respond positively to arguments and claims made by co-partisans and negatively to those from other party's adherents, giving higher scores than deserved in the former case and lower scores in the latter. The second, a subset of the first, is specifically that Democratic instructors tend to reward students who share their partisan affiliations and to discriminate against those who do not.¹

This paper seeks to test these claims. We thus make three contributions. First, we apply insights from the theoretical and empirical literature on partisanship and perception to a new area of analysis, partisan bias in the classroom, that is important to political scientists in their role as instructors as well as scholars (and as the targets of activism). We therefore arrive at a way to assess whether conservative activists' charges are grounded in reality. Second, we present results from an experiment designed to assess whether affinity bias is present. We find no evidence that there is partisan affinity bias in grading. Finally, we provide a framework for further, systematic inquiry into this topic.

Theory

The idea that partisans "see the world in a manner that is consistent with their political views...dates as far back as *The American Voter* (Campbell *et al.* 1960)" (Jerit & Barabas 2012: 673). As Achen & Bartels (2006) note, "very few politically consequential facts are subject to direct, personal verification" and so assessing "the plausibility of alternative states of the world" based on "how well they square with [voters'] partisan predispositions" is rational in a world of costly information. But rational, they caution, is not equal to optimal. A long literature supports the view that partisan filtering yields divergent assessments even of apparently objective facts. In examining individual responses to 'objective' political news, Bartels concludes "partisan bias in

political perceptions plays a crucial role in perpetuating and reinforcing sharp differences between Republicans and Democrats” (Bartels 2002: 117). Gaines, *et al.* (2007) similarly argue that “Democrats and Republicans could accurately perceive the same fact and yet make different judgments about its meaning”. Achen and Bartels (2006) find that partisanship matters differently for high- and low-information voters in assessing questions about the deficit and economic growth, but that overall “the average citizen’s perception of the federal budget deficit is constructed of four parts folk wisdom, one part partisan inference, and a trace element of reality.” Most important, “*attitudinally congruent arguments are evaluated as stronger than attitudinally incongruent arguments*” (Taber and Lodge, 2006: 755; emphasis added) and partisans “denigrate and counter argue claims that run contrary to their existing opinions” (Jerit & Barabas, 2012: 672). Considered together, these studies provide compelling evidence that partisans filter information through a partisan lens, influencing their interpretations of political facts, causes, and consequences. Partisans tend to hold favorable views on matters that accord with their preconceptions and less favorable opinions on those that do not.

Not all scholars agree with all of these claims. Ansolabehere *et al* (2013) finds that survey respondents are more accurate in assessing trends in *quantifiable* data than in *qualitative* data (e.g. the price of gasoline compared to voters’ responses about whether the economy is doing “well” or “poorly”), which they hold reduces partisan perception bias (or partisan cheerleading). Similarly, Bullock *et al.* (2013) find that the use of incentives in online experiments can reduce partisan bias in assessing objective conditions by 45 to 80 percent and conclude that much of the effect of partisan filters is due to expressive behavior instead of sincere representations of reality. However, if partisan filters exist, as the balance of the literature suggests, then there are clear implications for grading when an instructor’s subjective assessment is part of evaluation.

A substantial literature has examined the sources and incidence of grading bias. There are many mechanisms that can produce bias, but two are especially relevant for our purposes. The first, most heavily studied set of factors involves attributes of the student like race, ethnicity, nationality, gender, and physical attractiveness (for a summary of the literature, see Fleming 1999: 85-86). Spear (1984) finds that male and female teachers both tended to evaluate the work of male students more highly in subjects perceived to be ‘masculine’ (such as physics) and female students more highly in ‘feminine’ subjects. The second mechanism is that preconceptions about the quality of the student can also bias grading through ‘halo’ and ‘horn’ effects. Prior beliefs about a student’s competence can lead an instructor to give higher (‘halo’) or lower (‘horn’) grades than the student’s work merits (Fleming 1999: 86).

Some researchers have focused on how the political beliefs of faculty affect students. Mariani & Hewitt (2008) find that there is little evidence that faculty ideological preferences cause students’ shifts in ideological affiliations during college, although this of course does not refute the conservative allegation that faculty *attempt* to do so. One study finds that liberal and Democratic professors are perceived to grade more fairly but that the greater the perceived ideological difference between a student and a faculty member the lower the students’ evaluation of that professor (Kelly-Woessner & Woessner 2006).

Yet surprisingly little research directly examines the possibility of political bias in grading (Bar & Zussman 2012; see also Jacobson 2006 & Jaschik 2006). A single major study has assessed the relationship between student ideology and grading patterns at a single large public university; it concluded that conservative students received grades that were equal to or higher than their liberal counterparts (Kimmelmeier, Danielson & Baston 2005). It has been suggested that this study does not disprove politically biased grading because it does not take

into account the possibility that students ‘write to the professor’ by modifying their work to match their professors’ biases (Leef 2006; Balch 2006). Barr & Zussman (2012) finds that Republican professors favor a more unequal distribution of grades than their Democratic colleagues, which the authors analogize to Republican and Democratic preferences over taxation and redistribution.

Hypotheses and Methodology

We seek to test the two hypotheses about affinity bias that we lay out in the introduction. The general claim is that instructors will assign higher grades when their perceptual filters about politics align with students’ beliefs. The subsidiary claim is that Democratic professors in general will be more apt to use their influence in the classroom to reward liberal students and punish conservatives. Accordingly, *affinity bias*—our chief theoretical concept—is best tested through examining not only the partisan characteristics of students and instructors but also what happens when their partisan affiliations match.

Identifying causation is difficult in observational studies. The partisan-perceptual and grading-bias literatures are clear that there are many potential sources for bias in the classroom, ranging from perceptions about students’ gender or race (often their most obvious categorical traits) to their beliefs about individual students’ capacity and competence. Furthermore, students can choose which courses and which instructors to take. If they assume that the instructor’s partisan affiliation, either in isolation or in interaction with their own, will affect their grade, then they may self-select out of certain instructors’ classes—a clear source of endogeneity. Even those students who cannot avoid taking classes from instructors whose political beliefs they believe will not match their own may nevertheless write to please the professor, concealing their

own beliefs in order to earn a higher grade. Finally, the incentive for students to perform well in classes means that strict assessments of political knowledge—entirely objective questions—may obscure partisan filters, along the lines that Bullock et al and Ansolabehere et al suggest.

The best way to overcome such challenges is to avoid observational studies in favor of experiments. Accordingly, as we explain in more detail below, we use a double-blind trial in which neither students nor instructors know each other. Doing so allows us to account for the most severe challenges to studies on grading bias and perceptual filtering within our study population. In particular, we avoid ‘halo’ and ‘horn’ effects and most racial and gender bias because the instructors have no knowledge of the students besides any cues the students themselves choose to disclose. Perhaps most important, we mitigate the problems caused by students’ writing to please an individual instructor by subjecting their work to multiple instructors’ assessments. Indeed, if the write-to-please explanation is correct, then showing essays written for one instructor to instructors with an array of ideological and partisan affiliations is likely to reveal such traits.

Of course, we cannot directly measure bias, since the ‘true’ grade that given assignment should receive can never be directly observed (and in some sense may not even exist). Thus, we test whether observable characteristics—students’ and TAs’ reported partisan affiliations—in general are linked to predictable variations in reported grades across a diverse population of instructors.

Study Design²

We recruited a panel of undergraduate students to write the essays to be scored and a panel of graduate teaching assistants (TA) to score the essays.

All undergraduates enrolled in an introductory course on U.S. politics at a private, selective, East Coast university during the fall semester of 2011 were directed to write two essays on the first day of class as a regular assignment. The first essay prompt was:

What does the Democratic party stand for? What are its major goals (regarding economic, social and foreign policy)? Who are its major leaders? Who are its major constituencies/supporters?

An identical prompt was given regarding the Republican party. Students had 30 minutes to complete each essay during that class session. Later, students were solicited to participate in the “Political Science Essay Grading” study. Informed consent to allow their essays to be scored by TAs from other universities was obtained from 110 students out of the approximately 240 enrolled. Those who consented completed a brief survey indicating their sex, partisanship, and ideology.³ To select the essays most likely to elicit affinity bias, we removed essays by independents and weak partisans and chose 30 students (ten “strongly Republican” students and 20 “strongly Democratic” students selected at random) for a total of 60 essays.⁴ These essays, along with the survey information, were identified only by randomly-generated student ID number. Descriptive information on the participating students (and TAs) is provided in Table 1.

[Insert Table 1 about here.]

Separately, we solicited TAs with experience teaching courses in American politics to participate in this study through three electronic discussion boards hosted by the American Political Science Association.⁵ We provided prospective TA participants with background material on the “Political Science Essay Grading” study. The only information TAs were given about its purpose was that the study would “assess some of the factors that affect the scores that

teaching assistants assign undergraduate political science essays.” Fifty-two TAs expressed interest; 39 gave informed consent. We assigned each TA participant a randomized code number and deleted the link between the TAs’ names and their code numbers to preserve their anonymity. TAs who successfully completed all tasks were offered a \$50 honorarium.

We sent the consenting TAs two packets at different times. The first contained a brief survey concerning background characteristics, partisanship, and ideology as well as the 30 ‘Democratic’ essays they were to score.⁶ The TAs were to fill out the survey, score the essays and record the scores, and return the material to us before we sent them the second packet. The second packet contained the 30 ‘Republican’ essays to be scored and returned. We sent the packets separately to reduce the possibility that the TAs would compare the scores they assigned to the two sets of essays. We instructed the TAs to:

[Q]uickly grade these essays to assign [each] a single score between 0-100. You should complete this first batch in approximately 90 minutes [spending about three minutes on each essay]. We have no prior expectations regarding the mean or the distribution. You should grade them as if they were first year, first day students in one of your university’s introductory courses. Your score can be a ‘holistic’ score considering a) factual content; b) analytic skills; c) grammar and style.

To conceal the purpose of our study, we further instructed the TAs to guess the sex of the essay writer and to record their guesses on the scoring sheet along with the essay scores.

Thirty TAs completed all the tasks. Notably, only four TAs identified themselves as Republicans, with another five TAs identified themselves as independent but leaning Republican. However, since the main criticism that has been directed toward the academy focuses on the

perceived liberal bias, our panel seems to be reflective of the preponderance of Democratic instructors in the classroom.

Results

We begin by visualizing the scores that students earned in Figures 1 and 2, which employ kernel density estimation procedures to show non-parametric estimates of the probability distribution function of students' scores. Figure 1 demonstrates that Republican, Democratic, and Independent TAs assigned roughly similar grades to students on the Democratic essay prompt. However, the distribution of scores on the Republican essay prompt suggests that TAs of all partisan affiliations found that Republican students performed worse on the Republican essay prompt. Figure 2 focuses on Democratic and Republican TAs and divides students in each pane into those whose party affiliations match the TAs' and those who do not. We again find that Republican and Democratic TAs agree that Republican students underperformed their Democratic peers.

[Insert Figure 1 and Figure 2 about here.]

Kernel density functions are useful for describing data, but they are less well adapted to hypothesis testing. Accordingly, we use least-squares regression to test for partisan influence on assigning grades. Our core model is:

$$\begin{aligned} Student\ Score_{i,j} = & \beta_0 + \beta_1(TA\ School_i) + \beta_2(TA\ Courses_i) + \\ & \beta_3(TA\ Sex_i) + \beta_4(Student\ Sex_j) + \\ & \beta_5(TA-Student\ Sex\ Match_{i,j}) + \beta_6(TA\ Partisanship_i) + \\ & \beta_7(Student\ Partisanship_j) + \beta_8(Partisan\ Match_{i,j}) + \varepsilon \end{aligned}$$

The dependent variable, *Student Score*, is each TA's score for each student's essay.

The key independent variables are *TA Partisanship*, *Student Partisanship*, and *Partisan Match*. *TA Partisanship* is a binary variable that takes a value of 1 if the TA reports that she or he is a Democrat and 0 if Republican. We measured this in two ways. First, we used a ‘strict’ partisan measure that included only those TAs who were self-identified partisans. Second, we used a ‘loose’ partisan measure that also included TAs who, although independent, identified themselves as leaning to one party or another. This variable would take into account any differential grading patterns of Republican and Democrat TAs. *Student Partisanship* takes a value of 1 if the student is a Democrat and 0 if a Republican. Since all students in the sample were either self-identified strong Democrats or strong Republicans, this does not require a ‘strict’ or ‘loose’ variant.

Our key independent variable, *Partisan Match* takes a value of 1 if the TA and the student share a partisan affiliation and 0 otherwise. Again, we have a ‘strict’ (self-identified partisans only) and a ‘loose’ measure (including both self-identified partisans and independents leaning to one party). The partisan filtering literature suggests that, if scholarly norms are too weak to prevent such problems, co-partisan TA-student pairs should have higher scores than discordant pairs.

We include five variables to account for the impact of TA background (experience and type of schooling) as well as TA and student sex. *TA School* is 1 if the TA reported currently attending a public school and 0 if a private school. *TA Courses* is a factor variable that takes a value of 1 if the TA had taught or TA’d for one course, 2 for two courses, 3 for three courses, and 4 for more than three. *TA Sex* and *Student Sex* are binary variables coded 1 if the participant is female and 0 if male.⁷ *Sex Match* is a binary variable coded 1 if the TA and the student are of the same sex and 0 otherwise.

Table 2, columns 1-3, presents our regressions. Table 2, column 1, gives the results for all essays pooled together; columns 2 and 3, respectively, present the results when we segregate by essay topic. The models we present use heteroskedastic-robust standard errors calculated via the default package in Stata. Note that the R^2 for these tests is consistently low. The model explains only a relatively small fraction of the variance in the data (between 3.2 percent and 6.3 percent). This is unsurprising, as ideally most variation should come from students' performance, not from the observable characteristics of the students or TAs. Nevertheless, the F -tests continually demonstrate that the models perform better at explaining variance than chance alone.

[Insert Table 2 about here.]

After accounting for student and TA characteristics, we find no evidence for affinity bias: the coefficients for *Partisan Match* are small and do not nearly approach statistical significance.⁸ The estimated coefficient for *Student Partisanship* is positive and statistically significant in both the overall sample and the Republican essays sample. In these models, Democratic students received scores 2.7 to 3.7 percentage points higher on the 100-point scale than did Republican students. Conversely, the coefficient for *TA Partisanship* is negatively signed and significant only in the Democratic essays sample. Students graded by a Democratic TA on this assignment received scores averaging two points lower than students graded by Republican TAs, after accounting for other factors.

What do these models tell us? From the student's point of view, the instructor's nonpolitical characteristics (i.e., sex, experience, and type of school) matter more than the instructor's political characteristics in determining grades. For instance, having an instructor who is very experienced—having graded four courses instead of one—is worth a bonus of between just over 2 and just under 3 percentage points on a 100-point scale. Similarly, female TAs appear

to be more generous in their grading overall. By contrast, a TA's partisanship does not appear to affect grading outcomes except in one specification. Even in that specification, since all students generally received a lower grade on the Democratic essays from Democratic TAs, the implication is that the Democratic TAs were tougher graders, not that they were singling out Republican students.

Students' political characteristics, by contrast, have more obvious effects on the grades they received. Democratic students received higher grades on this assignment than their Republican peers, even after accounting for other relevant factors. That bonus is most obvious in Table 2, column 3, where the Democratic students outscored their Republican peers by an average of 3.7 points when writing essays about the Republican party. That finding that calls for further exploration. Recall first that the higher scores received by Democratic students accrue from Republican TAs as well as Democratic TAs. Over a variety of specifications, we see that Republican students are receiving lower marks from Republican and Democratic TAs alike. Concluding that there is bias in favor of Democratic students based on these results might be misleading. Throughout these models, however, *Partisan Match*—the key variable for assessing the affinity bias hypothesis—never reaches statistical significance and the estimates of its effects are less than 1 point on a 100-point scale.

We test the robustness of our findings in a number of ways. The most important alternative specification we present is derived from a procedure we implemented to determine if there were influential outliers that skewed our estimate of the coefficients. The usual diagnostic statistics (such as *dffits* and *dfbeta*) did not detect influential points, but these tests are insufficient. Since we know that the data-generating process depends on many fewer TAs and students than the overall N of student essays in our model, we are worried not about outliers

caused by any individual essays but instead due to outlying students or TAs. Consequently, we looked for influential outliers by sequentially re-estimating our models after dropping each student and TA in turn. We did not find any significant difference in the estimated coefficients by dropping each TA and re-estimating the models, but we did find that the exclusion of one student—student 11, a Republican male—did affect the estimated significance of the variable indicating the students’ partisanship (as the supplemental information explains in more detail). After dropping student 11, the control variables continue to perform much as they do in Table 2, but the partisan variables perform much differently. *Student Partisanship* is negatively signed and significant only for Democratic essays, while it is positively signed and significant for Republican essays only.

This re-estimation suggests that, dropping one significantly low-performing student, we find that *Republican* students performed better on Democratic essays while *Democratic* students did better on Republican essays. We view this as evidence against an interpretation that conservative students engaged in political typing (Gross 2013) altered their behavior in writing strategically to please presumed liberal graders. It may be that students instead provided more careful explanations of the “other side’s” objective characteristics but were blinded by partisan filters in providing objective considerations, although that would flip most accounts of filtering logic on its head. Significantly, our coefficient estimate of *Partisan Match* remained far from significance at conventional levels throughout all of these procedures. In other words, the phenomenon does not relate to the indicator that should best measure the presence of the conventional bias theory.

The supplemental information (available on request) includes many more specifications than we present here, including both additional least-squares regression model specifications and

quantile regressions. Across a wide range of specifications and examinations of subsets of the data (only Republican or only Democratic essays, only Republican or only Democratic students, and so on) we find the same basic pattern: that partisan variables are usually significant in directions that are opposite those that the affinity bias hypothesis would predict, that the key affinity bias variable, *Partisan Match*, rarely if ever rises to conventional levels of significance, and that most or nearly all of the gap between Republican and Democratic scores (a gap which, itself, is inconsistently significant) comes not from Democratic TAs unfairly punishing Republican students but from Republican and Democratic TAs alike marking down Republican students.

Discussion

Our findings are not congruent with most accounts of partisan bias. However, we do not intend for our research to be the final word on the subject. Our findings come with several caveats. Most obvious, we examine affinity bias among TAs, not professors, for primarily logistical reasons. Obtaining consent from college professors would pose greater challenges than recruiting TAs. Even though we provided the TAs with strong confidentiality protections, it is plausible that professors might be wary of participating in a study that examines their grading patterns for bias. Our findings thus cannot necessarily be generalized to professors, but we believe that our findings have a reasonable amount of external validity. All of the TAs were experienced graders, with fully half of our TA panel having taught three or more courses. Moreover, it seems unlikely that affinity bias would change substantially as the TAs become instructors themselves. And, of course, in many universities TAs perform an immense amount of grading.

Second, the partisan cues that appear in the students' essays – written on the first day of class during what was, for many students, the first week of their college careers – might be different from those that more advanced students would present. Affinity-grading bias could become more pronounced as students encounter more sophisticated arguments for (or against) the political opinions they hold when they arrive at college. More advanced students might write essays with higher, or lower, levels of partisan signaling that might have prompted the TAs to respond differently. Recruitment of upperclassmen, however, would pose greater difficulties, as we discuss more fully below.

Third, a different essay prompt might have altered the possibilities for affinity bias in grading. A more controversial prompt (e.g., “Are Democrats the party of special interests?” “Are Republicans the party of the rich?”) might have elicited responses that more strongly engage partisan filtering. But we should also note that such experiments might lack external validity, since (in our experience) it is rare to find professors who would consider assigning such prompts.

Fourth, readers might argue that the grading exercise was unrealistic. After all, we asked the TAs to grade the papers quite quickly, without offering justifications for their grades. Our reasoning again was driven by a desire to see whether we could induce biased responses under “real-world” conditions. It is not unusual for instructors to grade rapidly, especially on summative rather than formative assignments. However, perhaps partisan filtering engages more strongly when instructors have more time to grade. Different experimental designs could begin to unpack these issues.

Fifth, we are cautious about drawing conclusions from a sample that included comparatively few Republican TAs. On the other hand, although it is logically possible that Republican TAs exhibit positive or negative affinity biases, the more frequent (and more loudly

expressed) version of the bias claim is that Democratic TAs discriminate against Republican students. We do not find this pattern even among our relatively larger sample of Democratic TAs after dropping scores from one student who happened to be both a negative outlier and a Republican—a removal that does not affect the behavior of most other covariates. In this regard, we note that both Republican and Democratic students fared worse on essays about the Republican Party. It is logically possible that this reflects a bias against the Republican Party by Democratic and Republican TAs as well, who might be negatively reacting to the GOP. Our study cannot disprove this assertion, but it does suggest a potential vector for bias that future researchers may find fruitful. (Alternatively, of course, it may be that students are simply better informed about Democratic leaders and spokespeople because President Obama is a Democrat and more visible than Republican leaders. Running a similar test under a Republican president might offer a way to test this alternative explanation; unfortunately, we will probably have to wait until at least 2017 to conduct such an experiment.)

Sixth, readers might wonder about measuring the ideological content of the essays as a way to refine our claims. Our study design intended to test the most common (and potentially most dangerous) form of the political bias allegation. We therefore chose the most objective question we could think of in the study of American politics: a description of the objective characteristics of the two major parties. Had we found evidence supporting academia's critics, this would have been a warning sign to academics (and other constituencies) requiring action. This study has not exonerated academia, however, but it has shown that the extent of bias can at least be delimited. Future studies should proceed beyond our relatively straightforward prompt to incorporate testing materials with greater potential for bias (as a suggestion, abortion, income redistribution, or support for religious instruction in the schools). Future studies might also

benefit from taking steps to measure both students and graders in the same ideological space using a separate battery and an ideal-point estimation technique.

Finally, we are unable to test fully the effects of self-selection into disciplines based on students' ideological and political commitments.⁹ If students' political leanings incline them away from choosing a certain major, then our design would not be able to find these effects. Consequently, our estimations of the effect of partisan filters would be skewed by underlying endogeneity: there might be an unobserved factor governing selection into treatment that means that "real" conservatives do not show up in political science classrooms. However, although it would be objectionable on other grounds, such a selection bias would also refute the Santorum-Horowitz conjecture: how can students be indoctrinated or punished for conservative views by liberal professors if they never arrive in their classrooms?

Conclusion

Questions about partisan bias often generate more heat than light. Although professors – especially in the humanities, but in the social sciences as well – are predominantly Democratic, the impact of their partisanship on what they teach and how they grade, as well on how they treat their students, remains fertile grounds for empirical research. We may live in a world in which empirical evidence is often politicized, but we nevertheless hope that further research in this area can shift the terms of the debate.

Table 1: Descriptive Statistics of Participants

		<i>Students</i>	<i>TAs</i>
<i>Sex</i>	Male	16	15
	Female	14	15
<i>Party ID</i>	Democrat	20	15
	Republican	10	4
	Independent		9
	Other/NA		2
<i>If Democratic</i>	Strong	20	9
	Weak		6
<i>If Republican</i>	Strong	10	3
	Weak		1
<i>If Independent</i>	Lean R		5
	Lean D		4
<i>Courses Taught/ TA'd Previously</i>	One		9
	Two		5
	Three		4
	Four or More		11
<i>University Type</i>	Private	30	5
	Public		25

Table 2. OLS Models of Student Scores, By Essay Topic and With and Without Student 11.

	(1)	(2)	(3)	(4)	(5)	(6)
	Full	Dem Essays Only	Rep Essays Only	Full, No Student 11	Dem Essays, No Student 11	Rep Essays, No Student 11
<i>TA Sex</i> (<i>M=0, F=1</i>)	1.391* (0.498)	0.991 (0.661)	1.723* (0.749)	1.211* (0.473)	0.822 (0.613)	1.546* (0.726)
<i>Student Sex</i> (<i>M=0, F=1</i>)	-2.032* (0.498)	-0.251 (0.652)	-4.012* (0.753)	-2.884* (0.483)	-1.232* (0.626)	-4.721* (0.736)
<i>Sex Match</i> (<i>No=0, Yes=1</i>)	-0.173 (0.498)	-0.257 (0.658)	0.0919 (0.748)	-0.0256 (0.472)	-0.0986 (0.606)	0.211 (0.726)
<i>TA Attends Public College</i>	2.756* (0.809)	1.664+ (0.997)	3.898* (1.275)	2.555* (0.765)	1.586+ (0.906)	3.567* (1.234)
<i>TA Courses</i>	0.821* (0.193)	0.963* (0.247)	0.657* (0.304)	0.828* (0.183)	0.997* (0.223)	0.645* (0.297)
<i>TA Partisanship</i> (<i>GOP=0, Dem=1</i>)	-0.770 (0.810)	-2.008+ (1.197)	0.589 (1.087)	-0.713 (0.592)	-2.109* (0.726)	0.778 (0.935)
<i>Student is a Democrat</i>	2.655* (0.700)	1.675 (1.018)	3.723* (0.958)	-0.153 (0.549)	-1.659* (0.677)	1.476+ (0.857)
<i>Partisan Match</i> (<i>No=0, Yes=1</i>)	0.547 (0.861)	0.724 (1.280)	0.379 (1.146)	0.556 (0.633)	0.924 (0.787)	0.221 (0.986)
<i>Constant</i>	78.20* (1.171)	79.46* (1.508)	76.89* (1.806)	81.62* (1.051)	83.28* (1.259)	79.86* (1.702)
Observations	1770	930	840	1711	899	812
R^2	0.036	0.032	0.063	0.041	0.040	0.068
F	9.277	4.128	7.961	9.549	5.396	8.058

Robust standard errors in parentheses. Strict partisan definitions displayed. Loose partisan definitions are substantively similar.

+ $p < 0.10$, * $p < 0.05$

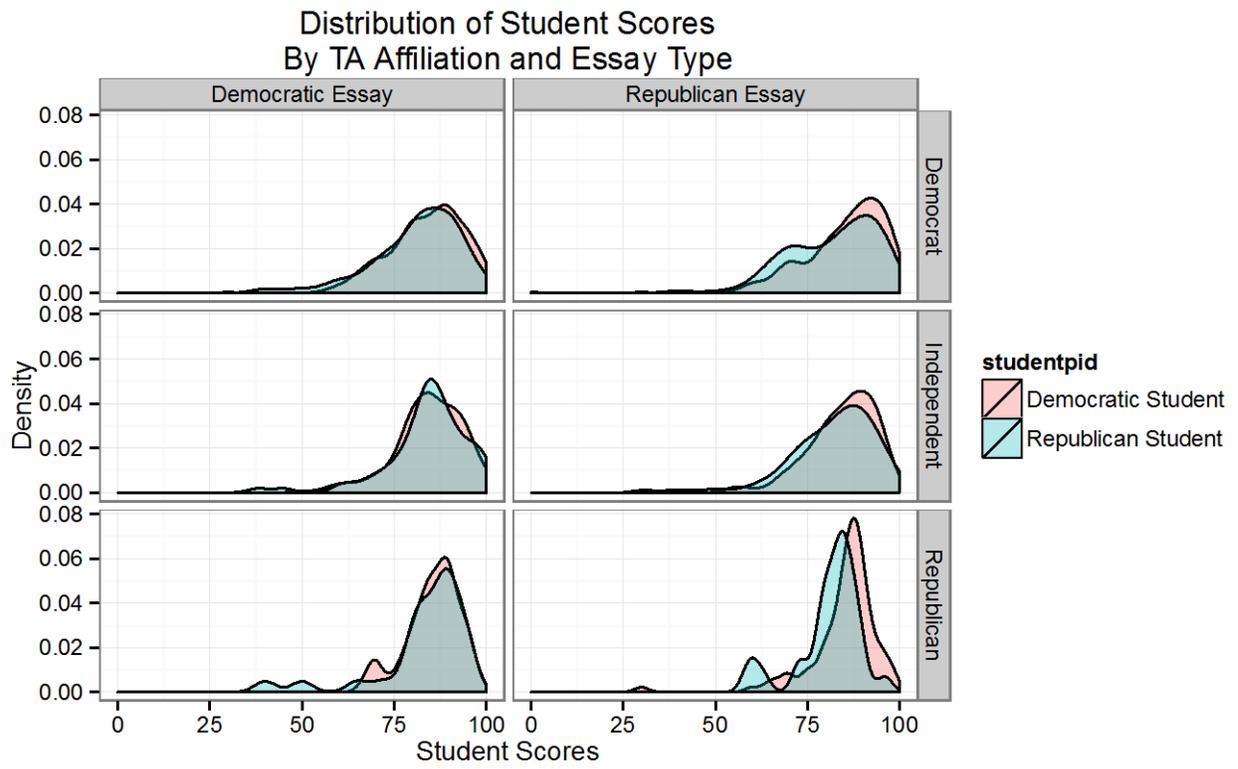


Figure 1. Kernel density plots for Republican and Democratic students by essay type and by TA partisan affiliation.

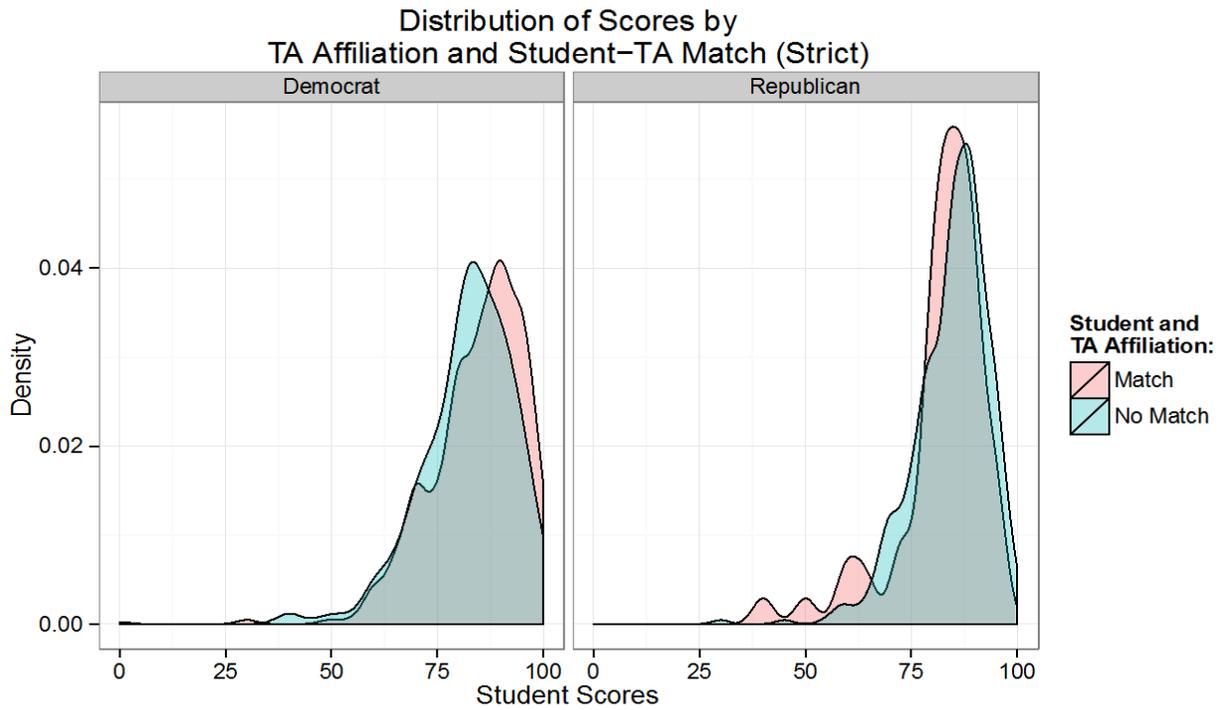


Figure 2. Kernel density plots by TA party affiliation and student-TA partisan match. This kernel density plot uses the “strict” partisan definition, including only those who have declared a positive affiliation with one party or another and excluding independents, including independents who “lean” toward one party or another. However, changing the partisan definition does not substantively affect the shape of the estimated p.d.f.s.

References

- Achen, C. H. & Bartels, L. M. (2006). It feels like we're thinking: The rationalizing voter and electoral democracy. Prepared for presentation at the annual meeting of the American Political Science Association, August 2006. Philadelphia, Pennsylvania.
- Balch, S. (2006, March 30). The study needs study. [Blog post]. *National Review Online: Phi Beta Cons*. Retrieved from <http://www.nationalreview.com/phi-beta-cons/46829/study-needs-study>.
- Barr, T. & Zussman, A. (2012). Partisan grading. *American Economic Journal: Applied Economics*, 4(1), 30-48.
- Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, 24, 117-150.
- Boyle, M. (2011). Conservative college students battle liberal academia with new website. *The Daily Caller*. Retrieved from <http://dailycaller.com/2011/11/27/conservative-college-students-battle-liberal-academia-with-new-website/> (Accessed September 16, 2013).
- Bullock, J G., Gerber, A., Hill, S., & Huber, G. (2013). *Partisan Bias in Factual Beliefs About Politics*. (No. w19080). National Bureau of Economic Research.
- Campbell, A., Converse, P., Miller, W. & Stokes, D. (1960). *The American Voter*. New York: John Wiley & Sons.
- Chiarante, P. (2012). Conservative student group at University of North Carolina claims political bias robbed it of funding. *Fox News*. Retrieved from <http://www.foxnews.com/us/2013/09/04/conservative-student-group-at-unc-claims-funding-lowered-because-political-bias/>
- Conservapedia* (N.d.). Liberal bias in academia. Retrieved from http://conservapedia.com/Liberal_bias_in_academia.

- Daly, J.A. & Dickson-Markman, F. (1982). Contrast effects in evaluating essays. *Journal of Educational Measurement*, 19, 309-16.
- Elder, L. (2012, April 12). Leftist California professors "corrupt" higher education. *Townhall.com*. April 12. Retrieved from http://townhall.com/columnists/larryelder/2012/04/12/leftist_california_professors_corrupt_higher_education/page/full.
- Ekins, E. (2012). 53 percent say California public universities teach political bias in the classroom. *Reason.com*. Retrieved from <http://reason.com/blog/2012/12/07/53-percent-say-california-public-univers>.
- Fleming, N. D. (1999). Biases in marking students' written work: quality? In Sally Brown and Angela Glasner (Eds.), *Assessment matters in higher education: choosing and using diverse approaches* (83-92). Buckingham: Society for Research into Higher Education and Open University Press.
- Fosse, E., Freese, J. & Gross, N. (2011). Political liberalism and graduate student attendance: a longitudinal analysis. Working paper. Retrieved from https://www10.arts.ubc.ca/fileadmin/template/main/images/departments/soci/faculty/gross/fosse_freese_gross_2_25.pdf.
- Gaines, B., Kuklinski, J, Quirk, P., Peyton, B. & Verkuilen, J. (2007). Same facts, different interpretations: partisan motivation and opinions on Iraq. *Journal of Politics*, 69, 957-94.
- Giroux, H A. (2006). "Academic freedom under fire: the case for critical pedagogy." *College Literature*, 33(4), 1-42.

- Gitlin, T. (2006). Professor's post: Todd Gitlin on Horowitz' "Dangerous Professors". Retrieved from <http://www.studentsforacademicfreedom.org/dangprofs//2006/03/professors-post-todd-gitlin-on.html>.
- Gross, N. (2006, February 26). Right, left, and wrong: David Horowitz's latest attack on America's left-leaning college professors doesn't add up. *Boston Globe*. Retrieved from http://www.boston.com/news/globe/ideas/articles/2006/02/26/neil_gross/
- Gross, N. (2013). *Why are professors liberal and why do conservatives care?* Cambridge, MA: Harvard University Press.
- Horowitz, D. & Laksin, J. (2009). *One party classroom: how radical professors at America's top colleges indoctrinate students and undermine democracy*. New York, NY: Crown Forum.
- Jacobson, J. (2006, March 30). Study casts doubt on claims that conservative students face discrimination in classes. *Chronicle of Higher Education*. Retrieved from <http://www.freerepublic.com/focus/f-news/1606178/posts> .
- Jaschik, S. (2006, March 20). The real bias in the classroom. *Inside Higher Ed*. Retrieved from <http://www.insidehighered.com/news/2006/03/20/politics>.
- Jaschik, S. (2011, March 21). (Liberal) academic self-selection. *Inside Higher Ed*. Retrieved from http://www.insidehighered.com/news/2011/03/21/new_studies_back_theory_that_the_professoriate_is_liberal_because_of_self_selection .
- Jerit, J. & Barabas, J. (2012). Partisan perceptual bias and the information environment. *Journal of Politics*, 74, 672-84.
- Kelly-Woessner, A. & Woessner, M.C. (2006). My professor is a partisan hack: how perceptions of a professor's political views affect student course evaluations. *PS: Political Science and Politics*, 39, 495-501.

- Kemmelmeier, M., Danielson, C., & Basten, J. (2005). What's in a grade? Academic success and political orientation. *Personality and Social Psychology Bulletin*, 31, 1386-1399.
- Lee, J. (2006). The 'faculty bias' studies: science or propaganda? American Federation of Teachers, Washington, D.C.
- Leef, G. (2006, March 30). Re: classroom bias. [Blog post.] *National Review Online: Phi Beta Cons*. Retrieved from <http://drupal6.nationalreview.com/phi-beta-cons/46831/re-classroom-bias>
- Mariani, M. D. & Hewitt, G. (2008). Indoctrination U.? Faculty ideology and changes in student political orientation. *PS: Political Science & Politics*, 41, 773-783.
- Maranto, R. & Hess, F. (2009). *The politically correct university: problems, scope and reforms*. Washington, DC: AEI Press.
- Saad, G. (2012, February 12). Is there a liberal bias among American professors? *Psychology Today*. Retrieved from <http://www.psychologytoday.com/blog/homocousmericus/201202/is-there-liberal-bias-among-american-professors>
- Schmidt, P. (2011, March 21). Experiment tricked graduate-program officials in fruitless search for political bias. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/Experiment-Tricked/126845/?sid=at&utm_source=at&utm_medium=en
- Slavov, S. (2012, October 3). Surviving academe's liberal bias. *The Chronicle of Higher Education*. Retrieved from <http://chronicle.com/blogs/conversation/2012/10/03/surviving-academes-liberal-bias/>.

Sonmez, F. (2012, February 26). Santorum stands by 'snob' comment, says conservatives 'singled out' and 'ridiculed' at colleges. *The Washington*

Post. http://www.washingtonpost.com/blogs/election-2012/post/santorum-stands-by-snob-comment-says-conservatives-singled-out-and-ridiculed-at-colleges/2012/02/26/gIQA7InobR_blog.html.

Spear, M-G. (1984). The biasing influence of pupil sex in a science marking exercise. *Research in Science and Technological Education*, 2, 55-60.

Taber, C. & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50, 755-769.

Tierney, J. (2011, July 22). The left-leaning tower. *The New York Times*. Retrieved from http://www.nytimes.com/2011/07/24/education/edl-24notebook-t.html?pagewanted=all&_r=2&.

Wood, P. (2011, March 22). Unnatural selection. *The Chronicle of Higher Education*. Retrieved from <http://chronicle.com/blogs/innovations/unnatural-selection/28946>.

Yancy, G. (2011). *Compromising scholarship: education*. Waco, TX: Baylor University Press.

Religious and political bias in

Footnotes

¹ This claim can take two forms. One is that liberal/Democratic instructors are more likely to be politically biased than are conservative/Republican instructors. The second is that, even if both types of instructors are equally biased, the predominance of liberal/Democratic instructors in higher education implies that overall the bias will primarily affect conservative/Republican students.

² This study was approved by the Georgetown University Institutional Review Board. Study protocols available from the authors on request.

³ The survey form is in Appendix One.

⁴ We selected the thirty students to simulate a tractable section size for the graders while retaining as much partisan variation as the population of students who agreed to participate allowed. We included all the Republicans and most of the Democrats who agreed; the Democrats were selected arbitrarily.

⁵ The discussion lists were: 1) Simulations and Role Play; 2) American Politics and Institutions and 3) Political Science Education.

⁶ The survey form is in Appendix Two.

⁷ No TAs or students identified as being transgendered.

⁸ Table 2 uses our strict partisan definition, which codes Republican and Democratic TAs as belonging to a party only if they identify as being members of that party and not as independents. However, there is no substantive difference if we use loose partisan definitions, which code Republican and Democratic TAs as belonging to a party if they are either members of that party or independents who say they lean to one side or another.

⁹ The class used to generate the essays is a required course for government majors, although most students in the class had not yet declared their majors. The course also fulfilled a general election requirement, so some substantial proportion of the students are likely to major in other disciplines.