UCLA School of Law

Public Law & Legal Theory Research Paper No. 23-26

A Response to Thaxton

By

Richard H. Sander

Jesse Dukeminier Professorship in Law

A Response to Thaxton
By Richard Sander
October 2023

In his critique of "Mismatch and Bar Passage," Sherrod Thaxton argues that there are six "problems" with the analysis. I examine each of these in turn.

Critique one. "First," observes Thaxton, "is the improper specification of the effect of race upon bar passage. S&S do not show, as mismatch theory hypothesizes, that race has no independent impact on law school GPA after accounting for students' degree of mismatch….[but] a proper test of any race effect must either disentangle the direct and indirect effects (through law grades) of race on bar passage or adjust for confounding variables before an analyst can claim that racial disparities in bar passage are explained by mismatch."

Thaxton is wrong. As we will see with his other "critiques," he seems to understand neither the hypotheses we are testing, nor the detailed empirical analyses we present.

The question we are examining in "Mismatch and Bar Passage," is whether a student who attends School A with a large preference is hurting her chances of passing the bar, relative to a student with the same credentials who attends School B with a smaller preference or without any preference at all.

It is helpful to provide concrete examples. Suppose a student has an LSAT score of 155 and an undergraduate GPA of 3.4. For simplicity's sake, we combine LSAT and UGPA credentials into a single number, the academic index, which in this case would be 690. A student with a 690 index who attend a school where her median classmate has an index of 790 has, in our analysis, a bar passage rate of 40%. But if they attend a school where their median classmate has an index of 690 – in other words, where our student has a level of academic preparation that, so far as we can tell from her index, puts her in the middle of her class – she will have an expected bar passage rate of 75%. At School A, the student is "mismatched" by 100 points, which is equivalent to a very substantial admissions preference. At School B, the student is not mismatched at all. Large mismatch, we find across a large number of students, is associated with bar failure at a much higher rate than the absence of mismatch. Moreoer, we find that the harm to one's chances of passing the bar goes up steadily, on average, as mismatch increases.

We also find that these effects of mismatch upon bar passage are essentially the same for students of all races. When we control for a student's pre-law "index" level, race has no statistically detectable effect upon bar passage whatsoever.

However, since most large admissions preferences at law schools in recent decades have been "racial" preferences, this means that Blacks are vastly more likely than whites to receive a large admissions preference and thus to be subject to the mismatch effect.[1] Our findings, as elaborated in a follow-up article, indicate that mismatch – i.e., the use of excessively large admissions preferences by law schools – accounts for more than three-quarters of the Black-white gap in bar passage. The rest is accounted for by the lower distribution of entering credentials among Black students relative to whites. Race, per se, explains none of the bar passage gap.

Thaxton seems to believe that our models are "improper" because we do not examine the mediating effect of law school grades upon bar passage. But why should this be so? None of the three scholars who peer-reviewed our paper for the *Journal of Legal Education* even raised this as an issue. Nor have any of the several other economists who read drafts of the paper and provided feedback. None of the mismatch critics who have offered alternative models have included law school GPA in the way Thaxton seems to favor. Daniel Ho has specifically argued that a model of law school mismatch should *not* include law school GPA; he argues that in studying mismatch, the admissions decision to grant or not grant preferences is the "treatment" and bar passage is the "effect"; therefore, introducing law school grades as a predictor of bar passage introduces "post-treatment bias" since grades are assigned after the "treatment" has occurred. While I disagree with Ho on some points, I think his point here is a reasonable one, and I have always thought that if one could develop an accurate measure of mismatch at the point of admissions, it would be important and valuable to observe mismatch's effect on bar passage.

Thaxton does not explain exactly how he would incorporate law school GPA into a model of bar passage. Since mismatch hurts law school GPA, then if both a measure of mismatch and a measure of GPA are included in a model, how does one sort out the direct effects of GPA from the indirect effects of mismatch through GPA? There may be a good answer, but Thaxton certainly hasn't provided it.

In short, Thaxton's first critique falls somewhere between "off the mark" and "incoherent."

<u>Critique two</u>. Thaxton contends that we make "improper" comparison of the effect of race across "nested" models. In Tables 3 and 4 of our paper, we show that, without any other controls, Blacks and Hispanics are more likely to fail the bar than are other groups, but that, as one adds additional and better measures to control for entering credentials, mismatch, and school, these effects disappear.

The critique is bizarre, for several reasons. First, it is a standard and common technique in social science research, including careful research on educational outcomes, to show how disparities across groups change when we add additional controls. Second, the comparisons across models are useful but by no means necessary for our conclusions. The models that best

[1] Hispanics are also disproportionately likely to receive law school admissions preferences, but these average about half the size of Black preferences.

describe the operation of mismatch are, of course, the most complete models – Models 5 and 6 in Table 3, and, particularly, Models 8 and 10 in Table 4.

Thaxton cites to a paper by Clogg, Petkova, and Haritou to buttress his argument, but the citation is inapt. Clogg et al are making the valid point that if one comparing two models, one of which keeps all the variables in the first model but adds additional variables, then one has to be cautious in interpreting small changes in the coefficients across models. They outline a test for assessing whether the changes are real, or are incidental artifacts produced by the addition of variables. But we are ***not*** comparing small changes – in our models, the race effects are very large in our simplest model, and then essentially disappear when we add controls. If one runs the Clogg test on our results (a standard feature in Stata), it confirms that the differences are highly statistically significant ($p<.001$).

Critique three. "Third," says Thaxton, "is the poor predictive ability of the estimated regression models. There are two related issues here. One issue is the use of the model fit statistic Somers' D, which is inadvisable because it inflates the predictive capacity of the regression models. The other issue is the lack of incremental predictive improvement for mismatch….models that control for mismatch are no better at predicting bar passage than models that simply include [LSAT and index]."

Thaxton is again attempting to manufacture issues that do not exist. None of the conclusions or findings of our paper depend on the Somers' D results. We provide the Somers' D merely as an additional bit of information about our results that some readers may find useful or interesting. Many (and probably most) high-quality papers reporting logistic regression results do not include any measure of "explanatory power" for the logistic regression, partly because there is really no satisfactory counterpart in a logistic context (where the outcome is either "yes" or "no") to the R^2 usually reported for OLS results (where the outcome is a continuous variable).

That said, the Somers' D results we report are not weak; they are impressive. [Do models as a whole satisfy something analogous to an F-test?]. It is useful to keep in mind that for a large share of the students at each school, the probability of bar passage is very high. Failure is concentrated among a relatively small share of students; consequently, there is a limited amount of "work" any regression predicting bar outcomes can do. The reader need only examine Table 2 to see that (a) mismatch explains rather dramatic differences for students who are highly mismatched, but that (b) the share of the thousands of students in our data who are highly mismatched is relatively small.

Note that, here again, Thaxton cites an authority for his argument, but misunderstands (or deliberately mischaracterizes) what that authority says. In this case, he cites [first name] Harrell for the proposition that [the Somers' D can produce misleading estimates of model fit]. In fact, however, Harrell suggests that Somers' D is superior to other measures of model fit when the outcomes across the data as a whole tilt heavily in one direction (i.e., towards bar passage). If anything, Harrell's article should be taken as supportive of our use of Somers' D, not a critique of it. But the larger point is that Thaxton's criticism, even if it were valid (which it is not) would be tangential to the arguments of the paper.

Critiques Four and Five. Thaxton contends that the "race effect" we observe in our models is so sensitive to particular model specifications that we can draw no reliable conclusions about whether race is a crucial determinant of bar passage. So far as I can tell, Thaxton has simply not studied our tables carefully. It is a strength of our paper, not a weakness, that we went out of our way to not simply present our main result (Model 10 in Table 4), but that we ran many variations on our model to test the robustness of our findings and to let the reader see for herself the work done by various inputs into the model. It is a strength of our paper, not a weakness, that our results were so strikingly *consistent* across these different formulations.

In the table below, I have summarized the race coefficients from all sixteen of our models, organized in terms of decreasing completeness. The models varied in many ways – some included school fixed effects and some did not; some used "categorical" measures of mismatch (to test for possible non-linearities in the mismatch effect) while some used a continuous, linear measure. Some analyses included only two of the schools rather than all three, because UCLA's data did not include undergraduate GPA, and because we excluded Bowen to test the consistency of our results when looking at two schools whose students took the same bar exam. We had no prediction about whether these sorts of variations would affect the role of race in the model. But we did expect the race coefficient to be affected by how well we controlled for a student's entering credentials, and whether we controlled for mismatch. (1) We expected that (consistent with prior research), if we fully controlled for student credentials and mismatch, there would be no race effect – i.e., the coefficients on "Black" and "Hispanic" would be close to 1.0 and not statistically significant. (2) We expected that if we lacked UGPA data, and thus could only control for LSAT and could only measure mismatch in terms of LSAT, then there would be small race effects in the models. This is because (as we explain at length in the paper, and document in Table 5) if a school is generally using large admissions preferences based on race, then enrolled Blacks and Hispanics with high LSATs will also tend to have particularly low UGPAs; thus, the omission of UGPA will somewhat bias our estimates with respect to those groups; their LSAT scores alone will overstate their chances of passing the bar, and thus the race coefficients will capture this residual underperformance. (3) If we control for LSAT and do not control for mismatch, the race effect will be somewhat more pronounced (i.e., the race coefficients will be further from 1.0 and more statistically significant), because, of course, we are arguing that mismatch helps explain the low bar performance of students receiving large preferences. (4) If we omit all academic controls, then the race coefficients will be even further from 1.0 and even more statistically significant, because in such a model we are simply measuring the large, well-known racial disparities in bar passage.

We also expect that for effects (2), (3), and (4), the race effects will be more apparent for Blacks than for Hispanics, because Blacks at all three of our schools received larger preferences than Hispanics.

As inspection of our table shows, the actual race coefficients follow these predictions with remarkable fidelity. Data is *of course* intrinsically noisy, and each of our models involves slightly different controls or slightly different combinations of schools, so we expect some variation in coefficients – indeed, it would be very suspicious if the coefficients did not vary. But the pattern is unmistakable – and as we noted earlier, if we use [ ]'s method for testing whether the differences are real….[wording]

Model	Key controls	Schools	Expectation	Logistic coefficient and significance	
				Blacks	Hispanics
10	LSAT & UGPA (index), categorical mismatch, school fixed effects	Bowen Davis	No race effect	1.11	.96
8	LSAT & UGPA (index), continuous mismatch	Bowen Davis	No race effect	1.02	.93
6	LSAT, categorical mismatch, school fixed effects	All Three	Small race effect, a little larger for blacks	.77	.81
5	LSAT, categorical mismatch	All Three	Small race effect, a little larger for blacks	.71*	.77*
4	LSAT, continuous mismatch, school fixed effects	All Three	Small race effect, a little larger for blacks	.71*	.79
3	LSAT, continuous mismatch	All Three	Small race effect, a little larger for blacks	.71*	.76*
9	LSAT, categorical mismatch measured by LSAT	Bowen Davis	Small race effect, a little larger for blacks	.63**	.79
7	LSAT, continuous mismatch measured by LSAT	Bowen Davis	Small race effect, a little larger for blacks	.64**	.77
A6	LSAT, categorical mismatch measured by LSAT, school fixed effects	UCLA Davis	Small race effect, a little larger for blacks	.59**	.77
A5	LSAT, categorical mismatch measured by LSAT	UCLA Davis	Small race effect, a little larger for blacks	.58**	.75*
A4	LSAT, continuous mismatch, school fixed effects	UCLA Davis	Small race effect, a little larger for blacks	.60**	.76*

A3	LSAT, continuous mismatch	UCLA Davis	Small race effect, a little larger for blacks	.58**	.74*
2	LSAT	All Three	Moderate race effect, a little larger for blacks	.56**	.53**
A2	LSAT	UCLA Davis	Moderate race effect, a little larger for blacks	.52**	.67**
A1	(race only)	UCLA Davis	Largest race effects, largest for blacks	.25***	.37***
1	(race only)	All Three	Largest race effects, largest for blacks	.28***	.41***

It's true that not all of the four "classes" of race effect are statistically distinguishable from one another, but that is not our claim. Our claims are that (a) the race effect diminishes as one adds better controls for entering credentials and mismatch, and that (b) when we control for LSAT, UGPA, and mismatch, the race effect on bar passage becomes statistically insignificant.

Thus, there's simply no basis for Thaxton's claim that the racial coefficients in our models are "sensitive" to particular model specifications and therefore unreliable.

Critique Six. Thaxton contends that the results in our appendix – i.e, the regressions that include only UCLA and Davis – are inconsistent with our other results. Here again, the problem lies not in our analysis, but in Thaxton's failure to read or understand our findings. Thaxton apparently thinks that the race coefficients in our Appendix Model 6 (A6 in the table above) are inconsistent with the models in the body of the paper. They are not. "Race" is weakly significant in A-6 because our only control for entering credentials, in the models that include UCLA, is LSAT. But the race coefficients in A-6 are very similar to the race coefficients in all of the other models that control for LSAT and mismatch. ***As we observe many times in this note, and in the paper itself, race effects disappear when one controls for LSAT and undergraduate UPGA, and our models predict, correctly, that there will still be some race effect when one controls only for LSAT.***

Conclusion. To recap, every one of Thaxton's six criticisms is not just wrong, but goofy. Thaxton mischaracterizes what we argue, incorrectly reads our tables, and misapplies the statistical literature. This reply is six pages long not because Thaxton says anything substantive, but only to make clear why he is not saying anything that is even a little bit right.