Replicating and Extending Effects of "Achievement Gap" Discourse

David M. Quinn
University of Southern California

Tara-Marie Desruisseaux
University of Southern California

Scholars argue the "racial achievement gap" frame perpetuates deficit mindsets. Previously, we found teachers gave lower priority to racial equity when disparities were framed as "achievement gaps" versus "inequality in educational outcomes." In this brief, we analyze data from two survey experiments using a teacher sample and an MTurk sample. We find: (1) the effect of "achievement gap" (AG) language on equity prioritization is moderated by implicit bias, with larger negative effects among teachers holding stronger anti-Black/pro-White stereotypes, (2) the negative effect of AG language replicates with non-teachers, and (3) AG language causes respondents to express more negative racial stereotypes.

Replicating and Extending Effects of "Achievement Gap" Discourse

David M. Quinn
Tara-Marie Desruisseaux

Rossier School of Education
University of Southern California

Forthcoming in *Educational Researcher*

Abstract

Scholars argue the "racial achievement gap" frame perpetuates deficit mindsets. Previously, we found teachers gave lower priority to racial equity when disparities were framed as "achievement gaps" versus "inequality in educational outcomes." In this brief, we analyze data from two survey experiments using a teacher sample and an MTurk sample. We find: (1) the effect of "achievement gap" (AG) language on equity prioritization is moderated by implicit bias, with larger negative effects among teachers holding stronger anti-Black/pro-White stereotypes, (2) the negative effect of AG language replicates with non-teachers, and (3) AG language causes respondents to express more negative racial stereotypes.

Replicating and Extending Effects of "Achievement Gap" Discourse

Scholars argue the "racial achievement gap" frame in education discourse is rooted in a deficit paradigm (e.g., Carey, 2014; Ladson-Billings, 2007). Rather than framing the problem around the structural injustices that lead to unequal learning opportunities by race, the "achievement gap" frame focuses attention on students, as if they are the ones who need "fixing" (e.g., Ladson-Billings, 2006; Milner, 2012). As such, the frame plays into, and may perpetuate, racist stereotypes.

In line with these critiques, we found in a recent survey experiment the "racial achievement gap" frame led teachers to place lower priority on racial equity, compared to a "racial inequality" frame (Quinn et al., 2019). We randomly assigned teachers to one of two versions of a survey item, asking some to rate how much of a priority they felt it was to "close the (Black/White) achievement gap," and others to rate the conceptually synonymous "ending (Black/White) inequality in educational outcomes." Teachers gave lower priority when disparities were framed as an "achievement gap" versus as "inequality" (ES = -0.11 SD). Furthermore, this result was driven by White teachers (ES = -0.18 SD). (However, item language did not affect the explanations teachers gave for why racial gaps/inequalities exist.)

If the negative effect of "achievement gap" language occurs because the term primes deficit mindsets, we would expect the effect to be larger among people who already hold stronger anti-Black stereotypes. We would also expect the term to elicit stronger expressions of anti-Black stereotypes. In the present study, we show evidence for both of these hypotheses. We also improve the external validity of Quinn et al. (2019) by replicating the finding in a new sample of non-teachers.

Methods

We analyze data from two separate survey experiments: (1) the teacher sample from Quinn et al. (2019) and (2) a second U.S. sample drawn from Amazon MTurk (n=500). (See Appendix A for sample descriptive statistics with comparisons to nationally-representative data, randomization balance, and disaggregated results)[1]. In each experiment, respondents were randomly assigned to one of two versions of our main survey item. One version used the term "racial achievement gap" while the other used "racial inequality in educational outcomes."[2] The main item read:

> *As you may know, there is* ***[a racial achievement gap/racial inequality in educational outcomes]*** *between Black and White students in the US. Thinking about all of the important issues facing the country today, how much of a priority do you think it is to* ***[close the racial achievement gap/end racial inequality in educational outcomes]*** *between Black and White students?*

Response options were "not a priority," "low priority," "medium priority," "high priority," or "essential" (adapted from Valant & Newark [2016]). For the MTurk sample, we used this item with 5 additional items to create an index (Cronbach's $\alpha = .97$; see Appendix B).

The teacher sample also completed an implicit association test (IAT; Greenwald et al., 1998) measuring respondents' automatic associations between race (Black/White) and academic competence (see Appendix C). We use the IAT as a moderator to test whether the effect of "achievement gap" language on priority ratings differs depending on teachers' implicit stereotypes.

After answering the gap/inequality item, respondents in the MTurk sample answered 10 stereotype items (based on the General Social Survey) in which they rated racialized groups (Black/White) on five bipolar traits (hardworking/lazy; intelligent/unintelligent;

competent/incompetent; capable/incapable; confident/unconfident). The difference in

respondents' average trait ratings of racialized groups comprises our stereotype index

(Cronbach's $\alpha = .78$; see Table 1 note and Appendix D).

Results

In Figure 1, we present results from a fitted logistic regression model predicting whether

teachers rated closing the gap/ending inequality as "high priority" or "essential." We find

teachers' implicit racial stereotypes moderate the effect of "achievement gap" (AG) language on

priority ratings (see Appendix E for estimates and robustness checks). On the x-axis, positively-

signed IAT scores represent teachers' automatic association of White students as being more

competent than Black students, negatively-signed scores represent the reverse, and zero

represents neutrality. As seen by the vertical distance between the two fitted curves, the negative

effect of AG language is largest among teachers holding strong implicit pro-White/anti-Black

stereotypes. Teachers with automatic associations that Black students are more competent than

White students (negatively-signed IAT scores) give high priority regardless of framing condition.

The negative effect of AG language on priority levels is statistically significant for respondents

with IAT scores above 0.45. Said differently, higher levels of implicit racial bias predict lower

prioritization of racial equity when AG language is used, but "inequality" language neutralizes

that negative relationship (IAT does not significantly predict priority level in the "inequality"

condition).

In Table 1, we show effects of AG language in the MTurk sample. First, we replicate the

finding from the teacher sample in Quinn et al. (2019): AG language lowers the extent to which

respondents prioritize racial equity (as measured by the original priority item, as well as the

priority index). Descriptively, this negative main effect is larger in the MTurk sample (ES = -0.26 SD vs. -0.11 SD for teachers).

Consistent with the theory that AG-framing activates deficit mindsets, we also find AG language increased explicit anti-Black/pro-White stereotypes (see Table 1 and table note for detail). On the stereotype index, positively-signed values indicate anti-Black/pro-White stereotypes (0=neutrality). As seen, respondents in both conditions expressed significant anti-Black/pro-White stereotypes. The AG language, however, increased the magnitude of stereotyping, by 0.19 SD (see Appendix F for robustness checks).

Discussion

We present further evidence that the language used to discuss racial equity in education matters. First, we improve the external validity of Quinn et al. (2019) by replicating the negative effect of "achievement gap" language in a sample of non-teachers. Importantly, we show evidence consistent with the hypothesis that AG language primes deficit thinking. We show the effect from Quinn et al. (2019) was driven by teachers who held stronger implicit stereotypes that Black students are less competent than White students. Furthermore, in the new non-teacher sample, AG language increased explicit anti-Black/pro-White stereotypes, compared to "inequality" language.

Scholars and advocates sometimes use language associated with deficit models in service of advancing an anti-deficit agenda; as such, it is important to take account of context, and to recognize that the use of a term does not automatically ground a discourse in a deficit framework (Patton Davis & Museus, 2019). At the same time, phrases can call to mind specific frames with which they are often associated (Lakoff, 2004). The findings in the present study suggest that in the absence of an explicitly anti-deficit framework, the language of "racial achievement gaps"

6

may carry more negative connotations compared to the language of "racial inequality." Combined with recent experimental evidence that a TV news story reporting on racial achievement gaps magnified viewers' racial stereotypes (Quinn, 2020), the present findings indicate that care should be taken in how racial disparities are framed in efforts to advance educational equity.

References

Carey, R.L. (2014). A cultural analysis of the achievement gap Discourse: Challenging the language and labels used in the work of school reform. *Urban Education, 49,* 440-468.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, *74*(6), 1464.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher*, *35*(7), 3–12. doi:10.3102/0013189x035007003

Ladson-Billings, G. (2007). Pushing past the achievement gap: An essay on the language of deficit. *The Journal of Negro Education, 76,* 316-323.

Lakoff, G. (2004). *Don't think of an elephant: Know your values and frame the debate.* Vermont: Chelsea Green Publishing.

Milner IV, H. R. (2012). Beyond a test score: Explaining opportunity gaps in educational practice. *Journal of Black Studies*, 43(6), 693-718.

Patton Davis, L., & Museus, S.D. (2019). What is deficit thinking? An analysis of conceptualizations of deficit thinking and implications for scholarly research. *Currents, 1(1),* 117-130.

Quinn, D. M., Desruisseaux, T. M., & Nkansah-Amankra, A. (2019). "Achievement gap" language affects teachers' issue prioritization. *Educational Researcher*, *48*(7), 484-487.

Quinn, D. M. (2020). Experimental effects of "achievement gap" news reporting on viewers' racial stereotypes, inequality explanations, and inequality prioritization. *Educational Researcher*, *49*(7), 482-492.

Valant, J., & Newark, D. A. (2016). The politics of achievement gaps: U.S. public opinion on race-based and wealth-based differences in test scores. *Educational Researcher*, *45*(6), 331–346.

Notes

[1] The sample size for the teacher sample in Quinn et al. (2019) differs from the sample size for the moderation analysis in this research brief due to the large number of participants who did not complete the implicit association test, or for whom valid IAT scores could not be calculated (see Appendix C for detail). As seen in Table A3 of Appendix A, the negative effect of AG language on priority in the MTurk sample is larger for males than females and this difference in effects is statistically significant. While we consider this result exploratory (it did not replicate in the teacher sample), we note potential generalizability implications.

[2] One concern is that respondents might interpret "achievement gaps" and "inequalities in educational outcomes" as referring to different academic outcomes, and such a difference in interpretation may drive differences in the priority ratings they give. As reported in Appendix A of Quinn et al. (2019), we found no evidence that respondents interpreted these terms as referring to different academic outcomes. However, we cannot rule out the possibility that the phrases differentially prime social desirability bias.

Figure 1.
Teachers' Implicit Racial Stereotypes (IAT score) Moderate the Effect of "Achievement Gap" Language on the Priority Ratings Teachers Give to Racial Disparities in Education (n=675).
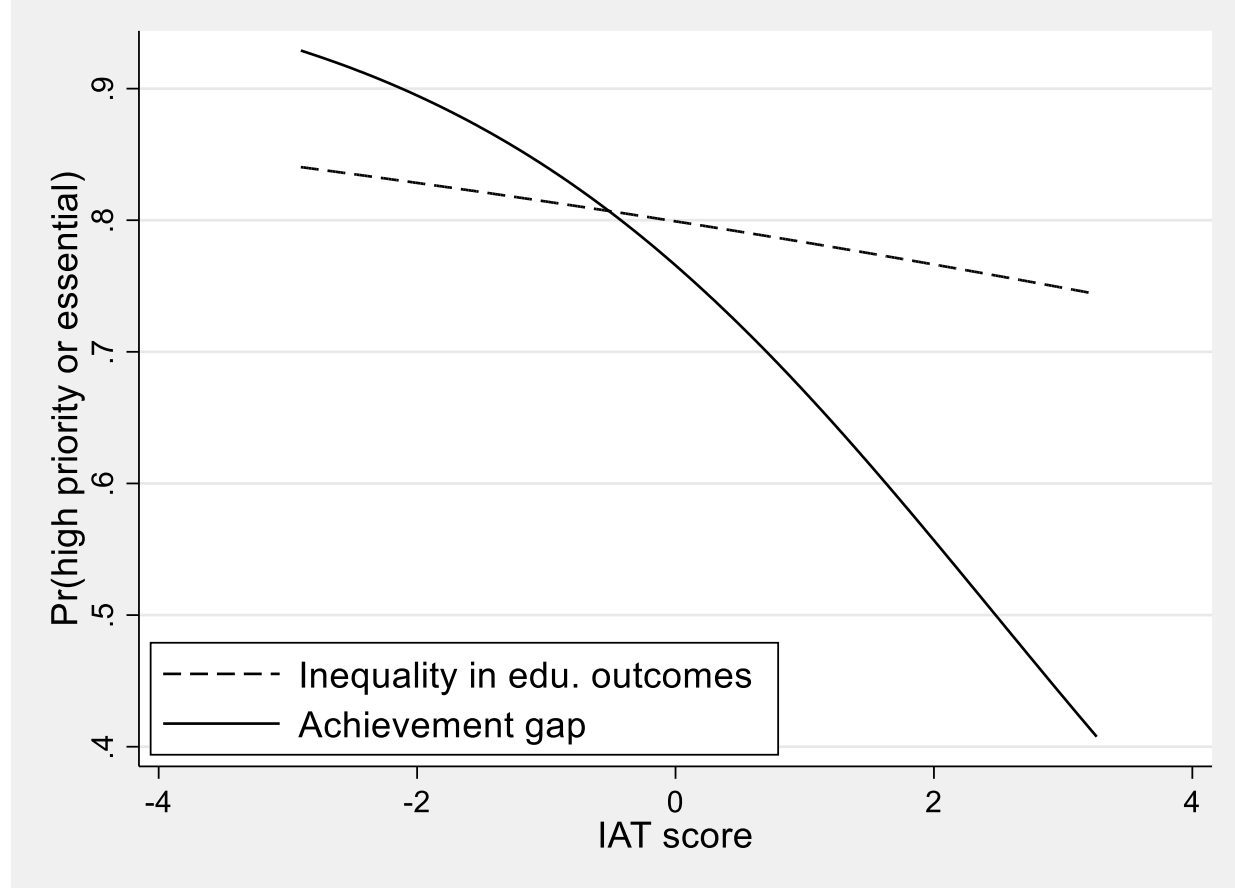
Note. Fitted logistic regression model predicting whether teacher responded "high priority" or "essential" to question, "*As you know, there is [a racial achievement gap/racial inequality in educational outcomes] between Black and White students in the US. Thinking about all of the important issues facing the country today, how much of a priority do you think it is to [close the racial achievement gap/end racial inequality in educational outcomes] between Black and White students? not a priority; low priority; medium priority; high priority; essential*" "Inequality in edu. outcomes" = teacher was randomly assigned to "inequality" version of the survey item; "Achievement gap" = teacher was randomly assigned to AG version of item. Model includes interaction between randomized survey condition and respondent's score on an implicit association test (IAT) measuring automatic association between race (Black/White) and academic competence (p=0.046 on likelihood-ratio test comparing model with versus without the interaction term; post-hoc tests indicate predicted probabilities by condition are statistically different when IAT score ≥ 0.45). On the IAT scale, positively-signed scores indicate the respondent holds automatic associations that White students are more academically competent than Black students, negative scores indicate automatic association that Black students are more academically competent than White students, and zero indicates neutrality. See Appendix C for detail on the IAT; see Appendix E for model estimates and alternative modelling strategies (results are robust).

Table 1.
Effects of "Achievement Gap" Language on Priority Ratings for Racial Disparities and on Explicit Racial Stereotypes (MTurk Sample).

	Ach. Gap	Inequality		
	Mean (SD)	Mean (SD)	Diff. (SE)	P
Priority outcome				
Priority index	3.30	3.62	-0.32**	0.003
	(1.18)	(1.22)	(0.11)	
High Priority/Essential (1=Y, 0=N)	0.44	0.57	-0.13**	0.003
	(0.50)	(0.50)	(0.04)	
Priority (single item)	3.26	3.59	-0.33**	0.003
	(1.20)	(1.26)	(0.11)	
Stereotype outcome				
Stereotype index (White mean index-Black mean index)	0.48	0.30	0.18*	0.039
	(1.09)	(0.87)	(0.09)	
Black mean index	4.77	4.92	-0.15	0.14
	(1.12)	(1.08)	(0.10)	
White mean index	5.25	5.22	0.04	0.628
	(0.86)	(0.87)	(0.08)	
N	*250*	*250*		

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note. "Ach. Gap" = respondents randomly assigned to "achievement gap" version of priority items; "Inequality" = assigned to the "inequality in educational outcomes" version of items. "Diff." = Ach Gap - Inequality mean difference; standard error of the mean difference is in parentheses in "Diff" column. p=p-value for t-test of equal means across the two conditions. Priority (single item): "*As you may know, there is [a racial achievement gap/racial inequality in educational outcomes] between Black and White students in the US. Thinking about all of the important issues facing the country today, how much of a priority do you think it is to [close the racial achievement gap/end racial inequality in educational outcomes] between Black and White students?*" Responses on 5-point scale, 1=not a priority; 2=low priority; 3= medium priority; 4=high priority; 5=essential. "High priority/essential" = 0/1 variable indicating respondent answered "high priority" or "essential" on priority item. Priority index = mean on priority item and 5 similar items (see Appendix B for scale detail and Appendix F for additional robustness checks). Stereotype items are comprised of 10 items: respondents rated Black Americans and White Americans on 5 bipolar traits (hardworking/lazy; intelligent/unintelligent; competent/incompetent; capable/incapable; confident/unconfident), each with a 7-point scale on which 7 = the respondent believes that "almost all" of the given racialized group tends to exhibit the positive pole of the trait, 1= "almost all" exhibit the negative pole. Black (White) mean index = mean score respondents gave across 5 traits for Black (White) Americans; Stereotype index (White mean-Black mean) = White-Black difference in mean score across 5 traits. See Appendix D for scale detail and Appendix F for robustness checks.

Appendix A. Descriptive Statistics.

In Table A1, we present descriptive statistics for the teacher sample, with comparisons to national estimates of the US teacher population. At the start of our survey, respondents were asked to provide the grade-level(s) and subject area(s) of their current teaching assignment. In addition to grade-level options, respondents were given an option "I am not a teacher" (in case any non-teachers somehow ended up in the Qualtrics teacher panel) and "other teaching position." A total of 87 respondents were dropped from the sample because they did not identify as a current full-time preK-12 teacher (e.g., retired teachers, substitute teachers, teachers' aides, parents home-schooling their children). However, results are unchanged if all respondents are included in the analysis.

As seen in Table A1, compared with the population of US teachers, the teachers in our sample are slightly less likely to identify as White, more likely to identify as multi-racial, and less likely to identify as female. This raises potential concerns about external validity. If IAT scores moderate the effect of AG language on teachers' priority levels differentially by race or gender, the interaction we find in our sample may not generalize to the overall teacher population. We therefore fit versions of our moderation models (the logit model, but also the sensitivity analyses using the linear probability model and the OLS model with 5-point survey scale outcome) that included the three-way interactions between IAT scores, survey condition, and each of the race/ethnicity indicator variables (along with all of the required pairwise interactions and main effects). In none of these models was the set of three-way interactions between the race indicators, IAT, and survey condition jointly statistically significant. We also fit analogous models testing the three-way interaction between IAT, survey condition, and gender; this three-way interaction was not significant in any of the models either.

Table A1.
Descriptive Statistics for Teacher Sample by Condition with Comparisons to Nationally Representative Data.

	Study Sample			National Estimates
	Ach. Gap	Inequality		
	Mean	Mean	p	Mean
Asian	0.01	0.03	0.17	0.02
Black	0.09	0.07	0.30	0.07
Hispanic or Latino	0.06	0.08	0.47	0.09
Multi-racial	0.08	0.11	0.23	0.01
Another race	0.00	0.00	0.33	0.01
White	0.76	0.72	0.34	0.80
Female	0.62	0.62	0.88	0.77
Non-binary	0.01	0.02	0.24	
Pre-K teacher	0.12	0.12	0.78	
K-2 teacher	0.15	0.16	0.94	
3-5 teacher	0.14	0.18	0.15	0.476 (elementary)
6-8 teacher	0.19	0.17	0.49	0.178 (middle)
9-12 teacher	0.39	0.36	0.44	0.287 (high)
School not primarily any race/ethnicity	0.17	0.20	0.27	0.449 (school is <50% White)
School primarily Asian/Pac. Is.	0.01	0.02	0.29	
School primarily Black	0.13	0.12	0.76	
School primarily Latinx	0.09	0.10	0.69	
School primarily Nat. Am.	0.01	0.00	0.58	
School primarily White	0.60	0.56	0.29	
<1 year in field of education	0.02	0.02	0.63	
1–3 years in field of education	0.16	0.13	0.21	0.099 (<3 years)
4–6 years in field of education	0.16	0.20	0.11	
7–10 years in field of education	0.22	0.21	0.95	0.283 (3–9 years)
11–15 years in field of education	0.22	0.19	0.23	
16–20 years in field of education	0.09	0.11	0.43	0.393 (10–20 years)
>20 years in field of education	0.13	0.14	0.61	0.225 (>20 years)
IAT (std.)	.40 (.97)	.42 (1.01)	0.87	
N	343	332		

Note. "Ach. Gap" = teachers assigned to "achievement gap" version of item; "Inequality" = teachers assigned to the "inequality in educational outcomes" version of items. All variables are 0/1 indicator variables, except IAT score (SD in parentheses). *p* value is for test of mean difference between conditions. "Another race" category includes Pacific Islander, Native American, and "another race." National estimates for teachers are for 2015-16 school year from Snyder, de Brey, & Dillow (2019), tables 209.1 (race/ethnicity, gender, years in edu field) and 209.24 (grade-level, school demographics). Pre-K teachers are not included in sample for national estimates but are included in our sample.

In Table A2, we present descriptive statistics for the MTurk sample, with comparisons to national estimates from the 2020 US census. (We refer to this sample as the "non-teacher" sample in this brief because unlike the Qualtrics sample, we did not target teachers for the MTurk sample. However, we did not query MTurk respondents on their occupation, and some may in fact be teachers). As can be seen, the racial make-up of our sample differs from that of the US as a whole, and a smaller percentage of our sample identifies as female compared to the general population. Again, this raises questions of external validity. If the effect of AG language differs by race or gender, the main effects we find in our sample may not generalize to the overall population. We therefore refit our models including interactions between survey condition and each of the race/ethnicity indicator variables. The set of interactions between the race indicators and survey condition was not jointly statistically significant for any of our outcome variables (though we note potential Type 2 errors due to small sample sizes for some racialized groups). However, we find a significant interaction between gender and survey condition when predicting priority level, such that the negative effect of gap language is stronger for males than for females (for these analyses, we removed the three respondents who identified as non-binary). Consequently, it is possible that our estimate of the main effect of AG language on priority level is biased away from zero in the MTurk sample. This finding requires replication, however, given the exploratory nature of the analysis.

In Table A3, we present our estimates of the main effects of survey condition on each of the outcomes in the MTurk sample, disaggregated by race/ethnicity and gender. Given the small sample sizes of many of these subgroups, we urge extreme caution when interpreting the results.

Table A2.
Descriptive Statistics for MTurk Sample by Condition with Comparisons to Census Data.

	Study Sample			2020 Census
	Ach. Gap	Inequality		
	Mean	Mean	p	Mean
Asian	0.08	0.10	0.34	0.06
Black	0.06	0.06	1.00	0.12
Hispanic or Latino	0.03	0.06	0.05	0.19
Multi-racial	0.07	0.06	0.72	0.10
Native American	0.01	0.00	0.32	0.01
Another race	0.00	0.01	0.56	0.08
White	0.75	0.70	0.23	0.62
Female	0.44	0.42	0.65	0.51
Non-binary	0.01	0.00	0.56	
HS degree	0.07	0.09	0.32	
Some college	0.21	0.18	0.36	
Associate's degree	0.14	0.14	1.00	
Bachelor's degree	0.42	0.44	0.53	
Master's degree	0.15	0.14	0.70	
Doctorate	0.02	0.02	0.52	
Age: 18-29	0.24	0.29	0.19	
Age: 30-39	0.40	0.37	0.46	
Age: 40-49	0.20	0.19	0.74	
Age: 50-59	0.11	0.09	0.55	
Age: 60-69	0.04	0.05	0.83	
Age: 70-79	0.01	0.01	1.00	
Age: 80+	0.00	0.00	0.32	
N	250	250		

Note. "Ach. Gap" = respondents assigned to "achievement gap" version of item; "Inequality" = respondents assigned to the "inequality in educational outcomes" version of items. All variables are 0/1 indicator variables. *p* value is for test of mean difference between conditions. National comparisons for race come from 2020 census (US Census, 2021); gender comes from Quick Facts (U.S. Census Bureau, 2020).

Table A3.
Effects of AG Language, Disaggregated by Race/ethnicity and Gender (MTurk Sample; note small sample sizes for some subgroups).

	Ach. Gap	Inequality		
	Mean	Mean	Diff.	
Full sample	(SD)	(SD)	(SE)	*p*
Priority outcome				
Priority mean index	3.30	3.62	-0.32**	0.003
	(1.18)	(1.22)	(0.11)	
High Priority/ Essential	0.44	0.57	-0.13**	0.003
	(0.50)	(0.50)	(0.04)	
Priority	3.26	3.59	-0.33**	0.003
	(1.20)	(1.26)	(0.11)	
Stereotype outcome				
White mean-Black mean index	0.48	0.30	0.18*	0.039
	(1.09)	(0.87)	(0.09)	
Black mean index	4.77	4.92	-0.15	0.14
	(1.12)	(1.08)	(0.10)	
White mean index	5.25	5.22	0.04	0.628
	(0.86)	(0.87)	(0.08)	
N	*250*	*250*		
Asian				
Priority outcome				
Priority mean index	3.19	3.77	-0.57~	0.083
	(0.98)	(1.12)	(0.32)	
High Priority/ Essential	0.32	0.56	-0.24	0.112
	(0.48)	(0.51)	(0.15)	
Priority	3.26	3.68	-0.42	0.203
	(0.87)	(1.18)	(0.32)	
Stereotype outcome				
White mean-Black mean index	0.93	0.42	0.51~	0.081
	(1.08)	(0.81)	(0.29)	
Black mean index	4.02	4.70	-0.68*	0.023
	(0.90)	(0.99)	(0.29)	
White mean index	4.95	5.12	-0.17	0.459
	(0.62)	(0.84)	(0.23)	
N	*19*	*25*		
Black				
Priority outcome				
Priority mean index	4.14	4.05	0.08	0.816

	(0.95)	(1.06)	(0.36)	
High Priority/ Essential	0.75	0.63	0.13	0.462
	(0.45)	(0.50)	(0.17)	
Priority	4.06	3.81	0.25	0.561
	(1.06)	(1.33)	(0.43)	
Stereotype outcome				
White mean-Black mean index	0.08	0.10	-0.02	0.926
	(0.72)	(0.80)	(0.27)	
Black mean index	5.24	5.14	0.10	0.772
	(0.98)	(0.96)	(0.34)	
White mean index	5.31	5.24	0.07	0.796
	(0.85)	(0.77)	(0.29)	
N	*16*	*16*		
Latino/a/x/Hispanic				
Priority outcome				
Priority mean index	3.52	3.76	-0.24	0.681
	(1.09)	(1.31)	(0.57)	
High Priority/ Essential	0.29	0.56	-0.28	0.24
	(0.49)	(0.51)	(0.23)	
Priority	3.57	3.81	-0.24	0.662
	(0.98)	(1.28)	(0.54)	
Stereotype outcome				
White mean-Black mean index	0.34	0.53	-0.18	0.716
	(0.85)	(1.18)	(0.49)	
Black mean index	4.74	4.64	0.11	0.857
	(0.98)	(1.37)	(0.58)	
White mean index	5.09	5.16	-0.08	0.825
	(0.76)	(0.76)	(0.34)	
N	*7*	*16*		
Multi-racial				
Priority outcome				
Priority mean index	3.75	3.79	-0.04	0.915
	(1.12)	(1.17)	(0.40)	
High Priority/ Essential	0.59	0.73	-0.15	0.405
	(0.51)	(0.46)	(0.17)	
Priority	3.76	3.87	-0.10	0.802
	(1.15)	(1.13)	(0.40)	
Stereotype outcome				
White mean-Black mean index	0.15	0.35	-0.19	0.42
	(0.55)	(0.79)	(0.24)	
Black mean index	4.98	5.07	-0.09	0.8

	(1.07)	(0.91)	(0.35)	
White mean index	5.13	5.41	-0.28	0.406
	(1.09)	(0.76)	(0.34)	
N	17	15		
White				
Priority outcome				
Priority mean index	3.20	3.52	-0.31*	0.015
	(1.20)	(1.24)	(0.13)	
High Priority/ Essential	0.41	0.54	-0.13*	0.012
	(0.49)	(0.50)	(0.05)	
Priority	3.14	3.50	-0.35**	0.008
	(1.23)	(1.29)	(0.13)	
Stereotype outcome				
White mean-Black mean index	0.52	0.28	0.24*	0.028
	(1.15)	(0.87)	(0.11)	
Black mean index	4.78	4.93	-0.16	0.187
	(1.14)	(1.09)	(0.12)	
White mean index	5.30	5.21	0.08	0.369
	(0.87)	(0.90)	(0.09)	
N	187	175		
Female				
Priority outcome				
Priority mean index	3.70	3.78	-0.08	0.607
	(1.13)	(1.18)	(0.16)	
High Priority/ Essential	0.59	0.60	-0.01	0.894
	(0.49)	(0.49)	(0.07)	
Priority	3.60	3.70	-0.11	0.539
	(1.20)	(1.30)	(0.17)	
Stereotype outcome				
White mean-Black mean index	0.47	0.22	0.25*	0.021
	(0.92)	(0.63)	(0.11)	
Black mean index	4.99	5.15	-0.17	0.225
	(1.04)	(0.98)	(0.14)	
White mean index	5.45	5.37	0.08	0.435
	(0.73)	(0.84)	(0.11)	
N	109	104		
Male				
Priority outcome				
Priority mean index	2.97	3.51	-0.53***	<0.001
	(1.12)	(1.25)	(0.14)	
High Priority/ Essential	0.31	0.54	-0.24***	<0.001

	Ach. Gap	Inequality	Diff.	p
	(0.46)	(0.50)	(0.06)	
Priority	2.98	3.51	-0.53***	<0.001
	(1.13)	(1.24)	(0.14)	
Stereotype outcome				
White mean-Black mean index	0.52	0.36	0.16	0.219
	(1.19)	(1.01)	(0.13)	
Black mean index	4.59	4.74	-0.16	0.244
	(1.15)	(1.13)	(0.14)	
White mean index	5.11	5.10	0.00	0.977
	(0.90)	(0.88)	(0.11)	
N	*139*	*145*		

Note. "Ach. Gap" = respondents assigned to "achievement gap" version of item priority items; "Inequality" = assigned to the "inequality in educational outcomes" version of items. Diff.= Ach Gap – Inequality mean difference; Standard error of diff. in parentheses in "Diff" column. *p*=p-value for t-test of equal means across conditions. Priority item: "*As you may know, there is [a racial achievement gap/racial inequality in educational outcomes] between Black and White students in the US. Thinking about all of the important issues facing the country today, how much of a priority do you think it is to [close the racial achievement gap/end racial inequality in educational outcomes] between Black and White students?"* Responses on 5-point scale, 1=not a priority; 2=low priority;3= medium priority; 4=high priority; 5=essential. Stereotype outcomes are comprised of 10 items: respondents rated Black Americans and White Americans on 5 bipolar traits (hardworking/lazy; intelligent/unintelligent; competent/incompetent; capable/incapable; confident/unconfident), each with a 7-point scale on which 7 = the respondent believes that "almost all" of the given racialized group tends to exhibit the positive pole of the trait, 1= "almost all" exhibit the negative pole. Black (White) mean index = mean score respondents gave across 5 traits for Black (White) Americans; White mean-Black mean = White-Black difference in mean score across 5 traits.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix B. Priority Index

We created the priority index by taking the mean response on 6 items (see below for the "achievement gap" version of the items). An exploratory factor model for the 6 items showed one factor with an eigenvalue above one (5.06), with all factor loadings above .90.

EFFECTS OF "ACHIEVEMENT GAP" DISCOURSE

As you may know, there is a racial achievement gap between Black and White students in the US.

Thinking about all of the important issues facing the country today, how much of a priority do you think it is to close the racial achievement gap between Black and White students?

○ Not a priority (1)	○ Low priority (2)	○ Medium priority (3)	○ High priority (4)	○ Essential (5)

How important is closing the Black/White achievement gap as a social justice issue?

○ Not important (1)	○ A little important (2)	○ Somewhat important (3)	○ Quite important (4)	○ Extremely important (5)

How important is closing the Black/White achievement gap to the future of the United States?

○ Not important (1)	○ A little important (2)	○ Somewhat important (3)	○ Quite important (4)	○ Extremely important (5)

How important is it that our national political leaders are committed to closing the Black/White achievement gap?

○ Not important (1)	○ A little important (2)	○ Somewhat important (3)	○ Quite important (4)	○ Extremely important (5)

How important is it that local education leaders are committed to closing the Black/White achievement gap?

○ Not important (1)	○ A little important (2)	○ Somewhat important (3)	○ Quite important (4)	○ Extremely important (5)

How urgent is it that we close the Black/White achievement gap?

○ Not urgent (1)	○ A little urgent (2)	○ Somewhat urgent (3)	○ Quite urgent (4)	○ Extremely urgent (5)

Appendix C. Race/Competence Implicit Association Test

To measure implicit racial stereotypes, we use an adapted version of the Black/White race implicit association test (IAT; Greenwald et al., 1998). The original Black/White race IAT is a computerized timed classification test designed to measure the respondent's automatic association between (a) a positive versus negative valence with (b) White people compared to Black people. The IAT does this by comparing how quickly and accurately respondents can classify stimuli representing White people (e.g., photographs of faces) when the race category is paired with a good vs. bad valence term (e.g., "joy" vs. "hurt") to how quickly and accurately they can classify stimuli representing Black people when paired with good vs. bad valence terms.

Like the traditional race IAT, our adapted competence IAT uses the categories "African American" and "European American." The stimuli were photographs of Black and White adolescents (4 male, 4 female for each racialized group), with competence target words "intelligent," "confident," "capable," and "efficient" (from Fiske et al. [2002]) and incompetence target words "disorganized," "unqualified," "stupid," and "unskilled" (inspired by Vitriol, Ksiazkiewicz, & Farhart [2018]). The adapted IAT was built using the iatgen online software (Carpenter et al., 2018). The adapted IAT showed an internal consistency reliability (based on split-half with Spearman-Brown correction) of .86. See Quinn (2020a) and Quinn (2020b) for additional development detail and validity evidence, including detail on the selection and validation of the student photos used in the IAT.

In the survey sequencing, the IAT appeared later in the survey than the gap/inequality priority item. As seen in Table A1, IAT scores were nearly identical across survey conditions, suggesting scores were not impacted by the gap/inequality item, enabling us to use IAT scores as a moderator.

The survey saw substantial respondent drop-off at the IAT. Of the analytic sample of 1,549 teachers in Quinn et al. (2019), 675 completed a valid IAT. Although respondents with valid IAT scores differ on some observables compared with respondents without valid IAT scores, F-tests regressing a survey condition indicator variable on the collection of demographic variables show no significant differences on observables between conditions in either the full sample or the IAT analytic sample. Compared with respondents with IAT scores, respondents missing IAT scores were more likely to identify as Asian (5.4% vs. 2.2%), more likely to identify as Black (10.6% vs. 7.7%), less likely to identify as White (66% vs. 74%,) more likely to teach Pre-K (18% vs. 12%), more likely to teach in a school with predominantly Native American students (1.4% vs. 0.4%), and less likely to teach in a school with predominantly White students (50% vs. 58%) (all differences significant at $\alpha = .05$).

Table C1 provides descriptive statistics for the IAT, by survey condition and respondent demographics.

Table C1.

Descriptive statistics for race/competence implicit association test by survey condition and respondent race and gender identity

	Ach. Gap			Inequality				
	Mean	SD	N	Mean	SD	N	Diff.	p
Full sample	0.40	0.97	343	0.42	1.01	332	-0.01	0.87
Asian	1.43	1.09	5	0.54	0.77	10	0.89	0.09
Black	-0.25	1.06	30	-0.27	0.91	22	0.02	0.93
Latinx	0.18	0.71	21	0.31	0.77	25	-0.13	0.56
Multi-racial	0.37	1.12	27	0.29	1.06	35	0.08	0.76
White	0.48	0.92	259	0.51	1.02	240	-0.02	0.81
Female	0.40	0.98	212	0.44	1.07	207	-0.04	0.72
Male	0.38	0.94	129	0.38	0.89	120	0.00	0.98
Non-binary	1.81	0.10	2	0.25	1.35	5	1.57	0.18

Note. IAT scores are divided by full sample SD to render scores interpretable in SD units. Positively-signed values = pro-White/anti-Black stereotypes; negatively-signed values are the reverse; 0 = neutrality. *P*-values are for test of the null hypothesis of mean equality across conditions.

Appendix D. Stereotype Indices.

Our stereotype indices use items adapted from the General Social Survey (GSS; Smith, Marsden, & Hout, 2015). The question item stems mirror those used in the GSS, but the response scales are adapted to incorporate best practices in scale development (Gehlbach & Brinkworth, 2011). The GSS stereotype items include bipolar scales for hard-working/lazy and unintelligent/intelligent; we add items for incompetent/competent, incapable/capable, and unconfident/confident to match the IAT target words drawn from Fiske et al. (2002). The actual survey items are shown below (the order of items within each set was determined randomly for survey-takers).

To create the Black (White) mean index, we calculate the mean score (on the 7-point scale) across the 5 traits that respondents gave for Black (White) Americans. To create the White mean – Black mean index, we take the White – Black mean difference in ratings across the 5 traits.

STEREOTYPE ITEMS

In what follows, you will see a scale on which the characteristics of people from different groups can be rated.

Do people in this group tend to be **hard-working or** tend to be **lazy**?
White Americans

Almost all are lazy (1)	Many are lazy (2)	Slight majority are lazy (3)	No tendency to one or another (4)	Slight majority are hardworking (5)	Many are hardworking (6)	Almost all are hardworking (7)

Do people in this group tend to be **hard-working or** tend to be **lazy**?

Black Americans

Almost all are lazy (1)	Many are lazy (2)	Slight majority are lazy (3)	No tendency to one or another (4)	Slight majority are hardworking (5)	Many are hardworking (6)	Almost all are hardworking (7)

Do people in this group tend to be **unintelligent or** tend to be **intelligent**?
White Americans

Almost all are unintelligent (1)	Many are unintelligent (2)	Slight majority are unintelligent (3)	No tendency to one or another (4)	Slight majority are intelligent (5)	Many are intelligent (6)	Almost all are intelligent (7)

Do people in this group tend to be **unintelligent or** tend to be **intelligent**?
Black Americans

Almost all are unintelligent (1)	Many are unintelligent (2)	Slight majority are unintelligent (3)	No tendency to one or another (4)	Slight majority are intelligent (5)	Many are intelligent (6)	Almost all are intelligent (7)

Do people in this group tend to be **incompetent or** tend to be **competent**?
White Americans

Almost all are incompetent (1)	Many are incompetent (2)	Slight majority are incompetent (3)	No tendency to one or another (4)	Slight majority are competent (5)	Many are competent (6)	Almost all are competent (7)

Do people in this group tend to be **incompetent or** tend to be **competent**?
Black Americans

Almost all are incompetent (1)	Many are incompetent (2)	Slight majority are incompetent (3)	No tendency to one or another (4)	Slight majority are competent (5)	Many are competent (6)	Almost all are competent (7)

Do people in this group tend to be **incapable or** tend to be **capable**?
White Americans

Almost all are incapable (1)	Many are incapable (2)	Slight majority are incapable (3)	No tendency to one or another (4)	Slight majority are capable (5)	Many are capable (6)	Almost all are capable (7)

Do people in this group tend to be **incapable or** tend to be **capable**?
Black Americans

Almost all are incapable (1)	Many are incapable (2)	Slight majority are incapable (3)	No tendency to one or another (4)	Slight majority are capable (5)	Many are capable (6)	Almost all are capable (7)

Do people in this group tend to be **unconfident or** tend to be **confident**?
White Americans

Almost all are unconfident (1)	Many are unconfident (2)	Slight majority are unconfident (3)	No tendency to one or another (4)	Slight majority are confident (5)	Many are confident (6)	Almost all are confident (7)

Do people in this group tend to be **unconfident or** tend to be **confident**?
Black Americans

EFFECTS OF "ACHIEVEMENT GAP" DISCOURSE

Almost all are unconfident (1)	Many are unconfident (2)	Slight majority are unconfident (3)	No tendency to one or another (4)	Slight majority are confident (5)	Many are confident (6)	Almost all are confident (7)

Appendix E. Estimates and Robustness Checks: Moderation Model for Teacher Sample

In Table E1, we present the estimates from logistic regression models predicting dichotomized responses to the priority item (1=high priority or essential; 0=otherwise) in the teacher sample. Column 1 shows the main effect of "achievement gap" language in the full teacher sample. Columns 2-4 show estimates from models fit using the analytic sample of teachers with valid IAT scores. The fitted model in Column 4 corresponds to Figure 1 in the main text.

As robustness checks, we also fit the following models: (1) alternative versions of the logistic regression models, fit using linear probability models (Table E2), (2) OLS regression models using the full 5-point priority scale outcome (Table E3), (3) ordered logit models using the full 5-point priority scale outcome (Table E4), and (4) ordered probit models using the full 5-point priority scale outcome (Table E5).

As seen in the tables, the interaction between IAT scores and survey condition is statistically significant in each model. As relevant for the nonlinear models, we include the p-values for the likelihood ratio tests of the significance of the interaction terms (for the logit, ordered logit, and ordered probit models); all of these tests are statistically significant. While this significance test is sufficient when one is interested in interpreting the exponentiated interaction coefficient in the logistic model as a ratio of odds ratios (Buis, 2010), we also test significance of the interaction term using the entire cross-derivative with the "inteff" command in Stata, as suggested by Norton et al. (2004).

Table E1.
Logistic Regression

	(1) High Priority/Essential	(2) High Priority/Essential	(3) High Priority/Essential	(4) High Priority/Essential
Ach. Gap	-0.414*** (0.117)	-0.393* (0.181)	-0.404* (0.182)	-0.197 (0.210)
IAT			-0.293** (0.0950)	-0.0965 (0.135)
Ach. Gap*IAT				-0.381* (0.192)
Constant	1.270*** (0.0871)	1.338*** (0.135)	1.485*** (0.147)	1.381*** (0.150)
N	1549	675	675	675

Standard errors in parentheses. p=0.046 on likelihood-ratio test comparing model with versus without the interaction term. P=.023 for significance test on the cross-derivative for the interaction via the inteff command.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E2.
Linear Probability Models

	(1) High Priority/Essential	(2) High Priority/Essential	(3) High Priority/Essential	(4) High Priority/Essential
Ach. Gap	-0.0790*** (0.0222)	-0.0721* (0.0330)	-0.0727* (0.0328)	-0.0424 (0.0337)
IAT			-0.0518** (0.0164)	-0.0158 (0.0222)
Ach. Gap*IAT				-0.0739* (0.0322)
Constant	0.781*** (0.0149)	0.792*** (0.0223)	0.814*** (0.0232)	0.799*** (0.0237)
N	1549	675	675	675
R^2	0.008	0.007	0.021	0.028

heteroskedasticity-robust standard errors in parentheses
$^{\sim}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E3.
OLS Regression

	(1) Priority	(2) Priority	(3) Priority	(4) Priority
Ach. Gap	-0.110*	-0.107	-0.109	-0.0308
	(0.0492)	(0.0727)	(0.0721)	(0.0765)
IAT			-0.144***	-0.0512
			(0.0391)	(0.0544)
Ach. Gap*IAT				-0.191*
				(0.0760)
Constant	4.097***	4.160***	4.220***	4.181***
	(0.0348)	(0.0507)	(0.0547)	(0.0577)
N	1549	675	675	675
R^2	0.003	0.003	0.026	0.036

heteroskedasticity-robust standard errors in parentheses
$^\sim p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

Table E4.
Ordered Logit Models

	(1) Priority	(2) Priority	(3) Priority	(4) Priority
Ach. Gap	-0.246**	-0.207	-0.229	-0.0962
	(0.0935)	(0.142)	(0.143)	(0.158)
IAT			-0.305***	-0.162
			(0.0745)	(0.102)
Ach. Gap*IAT				-0.300*
				(0.149)
/				
cut1	-4.126***	-4.669***	-4.851***	-4.799***
	(0.197)	(0.388)	(0.391)	(0.392)
cut2	-2.753***	-2.823***	-2.998***	-2.942***
	(0.113)	(0.177)	(0.184)	(0.186)
cut3	-1.179***	-1.235***	-1.393***	-1.329***
	(0.0760)	(0.116)	(0.124)	(0.127)
cut4	0.341***	0.211*	0.0813	0.151
	(0.0694)	(0.105)	(0.111)	(0.115)
N	1549	675	675	675

Standard errors in parentheses. p=0.043 on likelihood-ratio test comparing model with versus without the interaction term.
$\tilde{}\ p < 0.10$, $^*\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Table E5.
Ordered Probit Models

	(1) Priority	(2) Priority	(3) Priority	(4) Priority
Ach. Gap	-0.128[*]	-0.117	-0.120	-0.0266
	(0.0554)	(0.0847)	(0.0849)	(0.0935)
IAT			-0.175[***]	-0.0742
			(0.0436)	(0.0605)
Ach. Gap*IAT				-0.208[*]
				(0.0874)
/				
cut1	-2.161[***]	-2.375[***]	-2.476[***]	-2.455[***]
	(0.0813)	(0.149)	(0.151)	(0.154)
cut2	-1.560[***]	-1.600[***]	-1.696[***]	-1.658[***]
	(0.0570)	(0.0887)	(0.0926)	(0.0941)
cut3	-0.713[***]	-0.753[***]	-0.837[***]	-0.792[***]
	(0.0449)	(0.0687)	(0.0722)	(0.0744)
cut4	0.225[***]	0.137[*]	0.0671	0.115
	(0.0424)	(0.0646)	(0.0673)	(0.0699)
N	1549	675	675	675

Standard errors in parentheses. p=0.017 on likelihood-ratio test comparing model with versus without the interaction term.
$\tilde{}\ p < 0.10$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Appendix F. Robustness Checks: Priority and Stereotype Outcomes for MTurk Sample

In Table F1, we present the effects on each item that comprises the "priority index." For reference, we also include the results reported in the main text for the index and the single priority item, as well as the dichotomized single priority item (1 = "high priority" or "essential") (all results are also robust to ordered logit and ordered probit models, not shown).

Table F1.
Effects of "Achievement Gap" Language on Full Set of Issue Priority Items (MTurk Sample).

	Ach Gap		Inequality			
	Mean	SD	Mean	SD	Diff.	*p*
Priority mean index	3.30	1.18	3.62	1.22	-0.32	0.003
High Priority/Essential	0.44		0.57		-0.13	0.003
Priority	3.26	1.20	3.59	1.26	-0.33	0.003
Social justice	3.36	1.36	3.68	1.30	-0.32	0.008
Future of US	3.36	1.24	3.72	1.29	-0.36	0.001
National politicians	3.28	1.30	3.58	1.30	-0.30	0.009
Local educators	3.46	1.27	3.67	1.30	-0.21	0.071
Urgent	3.10	1.28	3.48	1.32	-0.38	0.001
N	250		250			

Note. "Ach. Gap" = respondents assigned to "achievement gap" version of item priority items; "Inequality" = assigned to the "inequality in educational outcomes" version of items. Diff.= Ach Gap – Inequality mean difference; p=p-value for t-test of equal means across conditions. Priority item: "*As you may know, there is [a racial achievement gap/racial inequality in educational outcomes] between Black and White students in the US. Thinking about all of the important issues facing the country today, how much of a priority do you think it is to [close the racial achievement gap/end racial inequality in educational outcomes] between Black and White students?"* Responses on 5-point scale, 1=not a priority; 2=low priority;3= medium priority; 4=high priority; 5=essential. "High priority/essential" = 0/1 variable indicating respondent answered "high priority" or "essential" on priority item. Priority index = mean on priority item and the 5 other items included in the table. See Appendix B for full survey items.

In Table F2, we present robustness checks for the stereotype items. We draw attention to the bottom section of the table (section header "*Indicator, 1=rated White higher than Black*") where we show robustness of the results to the interval scale assumption on which the analyses using the mean indices rely. Here, we create a set of indicator variables for each item pair, indicating whether the respondent rated White Americans more positively than Black Americans on a given dimension (1=rated White higher, 0=otherwise). As seen, in both conditions respondents gave higher average ratings to White Americans than Black Americans on each dimension. Importantly, we see that the "achievement gap" language increased the proportion of respondents who rated White Americans more positively than Black Americans for each dimension, and this increase in pro-White/anti-Black stereotyping was statistically significant for all but one dimension (lazy/hardworking). On average, when respondents received the "inequality" version of the main survey item, they rated White Americans more positively on 26% of the item pairs. When assigned to the "achievement gap" version of the main item, this increased to 35% ($p=.001$).

Second, recall that the White mean – Black mean stereotype index in the main analysis is the difference in average ratings that respondents gave to White Americans and Black Americans across the five bipolar scales (lazy/hardworking; unintelligent/intelligent; incompetent/competent; incapable/capable, unconfident/confident). An exploratory factor analysis on these White-Black differences for each item pair showed one factor with an eigenvalue above 1 (2.39); all items showed factor loadings above .69, except for the White-Black unconfident/confident item, which had a loading of .17. In Table F2, following the results from the main text (included for reference), we include stereotype index results using indices excluding the unconfident/confident item pair. As seen in the table, the effect of "achievement

gap" language using the White-Black mean index without the unconfident/confident items has nearly identical magnitude compared with the effect on the full index (.19 vs. .18), with $p=.06$ for the test of mean difference.

Table F2.
Effects of "Achievement Gap" Language on Full Set of Stereotype Items (MTurk Sample).

	Ach Gap		Inequality			
	Mean	SD	Mean	SD	Diff	p
Mean indices						
White-Black mean index	0.48	1.09	0.30	0.87	0.18	0.039
Black mean index	4.77	1.12	4.92	1.08	-0.15	0.140
White mean index	5.25	0.86	5.22	0.87	0.04	0.628
White-Black mean index (w/o confidence item)	0.50	1.24	0.31	1.02	0.19	0.060
Black mean index (w/o confidence item)	4.72	1.26	4.90	1.21	-0.18	0.098
White mean index (w/o confidence item)	5.22	0.95	5.21	0.91	0.01	0.923
White-Black, indiv. Items						
Lazy/hardworking	0.44	1.66	0.30	1.38	0.14	0.320
Unintelligent/intelligent	0.53	1.34	0.35	1.15	0.18	0.107
Incompetent/competent	0.54	1.43	0.30	1.26	0.24	0.051
Incapable/capable	0.48	1.36	0.27	1.18	0.21	0.064
Unconfident/confident	0.42	1.47	0.27	1.34	0.15	0.227
Black, individual items						
Lazy/hardworking	4.56	1.45	4.71	1.43	-0.15	0.251
Unintelligent/intelligent	4.66	1.33	4.79	1.34	-0.14	0.256
Incompetent/competent	4.76	1.39	4.98	1.30	-0.22	0.073
Incapable/capable	4.90	1.36	5.14	1.33	-0.23	0.055
Unconfident/confident	4.98	1.30	4.97	1.26	0.00	0.972
White, individual items						
Lazy/hardworking	5.00	1.22	5.01	1.16	-0.01	0.910
Unintelligent/intelligent	5.19	1.12	5.14	1.05	0.04	0.650
Incompetent/competent	5.30	1.09	5.28	1.06	0.02	0.835
Incapable/capable	5.38	1.06	5.40	1.00	-0.02	0.828
Unconfident/confident	5.40	1.02	5.24	1.05	0.16	0.093
Indicator, 1=rated White higher than Black						
Lazy/hardworking	0.31		0.29		0.02	0.559
Unintelligent/intelligent	0.36		0.24		0.12	0.003
Incompetent/competent	0.34		0.25		0.09	0.031
Incapable/capable	0.33		0.21		0.12	0.003
Unconfident/confident	0.42		0.29		0.13	0.003
Prop. rated White > Black across all items	0.35		0.26		0.10	0.001
Prop. rated White > Black across all items, w/o confident item	0.33		0.25		0.09	0.007
N	250		250			

Note. For stereotype items, respondents rated Black Americans and White Americans on the 5 bipolar traits shown (hardworking/lazy; intelligent/unintelligent; competent/incompetent; capable/incapable; confident/unconfident),

each with a 7-point scale on which 7 = the respondent believes that "almost all" of the given racialized group tends to exhibit the positive pole of the trait, 1= "almost all" exhibit the negative pole. Black (White) mean index = mean score respondents gave across 5 traits for Black (White) Americans; White -Black mean index = White-Black difference in mean score across 5 traits. See Appendix D for item detail.

References (Appendices)

Buis, M.L. (2010). Stata tip 87: Interpretation of interactions in nonlinear models. *Stata Journal, 10,* 305-308.

Carpenter, T., Pogacar, R., Pullig, C., Kouril, M., LaBouff, J., Aguilar, S. J., … Chakroff, A. (2018, April 3). Conducting IAT Research within Online Surveys: A Procedure, Validation, and Open Source Tool. https://doi.org/10.31234/osf.io/6xdyj

Fiske, S.T., Cuddy, A.J.C., Glick, P., & Xu, J. (2002) A model of (often mixed) stereotype content: Competence and warmth respectively flow from perceived status and competition. *Journal of Personality and Social Psychology, 82,* 878-902.

Gehlbach, H., & Brinkworth, M.E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology, 15,* 380-387.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, *74*(6), 1464.

Norton, E.C., Wang, H., & Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal, 4,* 154-167.

Quinn, D. M. (2020a). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis*, *42*(3), 375-392

Quinn, D. M. (2020b). Experimental effects of "achievement gap" news reporting on viewers' racial stereotypes, inequality explanations, and inequality prioritization. *Educational Researcher*, *49*(7), 482-492.

Quinn, D. M., Desruisseaux, T. M., & Nkansah-Amankra, A. (2019). "Achievement gap" language affects teachers' issue prioritization. *Educational Researcher*, *48*(7), 484-487.

Smith, T.W.; Marsden, P.V., & Hout, M. General Social Surveys, 1972-2014. [machine-readable data file]. Principal Investigator, Tom W. Smith; Co-Principal Investigators, Peter V. Marsden and Michael Hout, NORC ed. Chicago: National Opinion Research Center, 2015. 1 data file (59,599 logical records) and 1 codebook (3,485 pp)

Snyder, T. D., de Brey, C., & Dillow, S. A. (2019). *Digest of Education Statistics 2017* (NCES Publication No. 2018-070). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

U.S. Census Bureau (2020). Quick Facts United States. Retrieved from: https://www.census.gov/quickfacts/fact/table/US/PST045219.

U.S. Census Bureau. (2021). Race and ethnicity in the United States: 2010 Census and 2020

Census. Retrieved from: https://www.census.gov/library/visualizations/interactive/race-and-ethnicity-in-the-united-state-2010-and-2020-census.html

Vitriol, J.A., Ksiazkiewicz, A., & Farhart, C.E. (2018). Implicit candidate traits in the 2016 U.S. presidential election: Replicating a dual-process model of candidate evaluations. *Electoral Studies, 54,* 261-268.