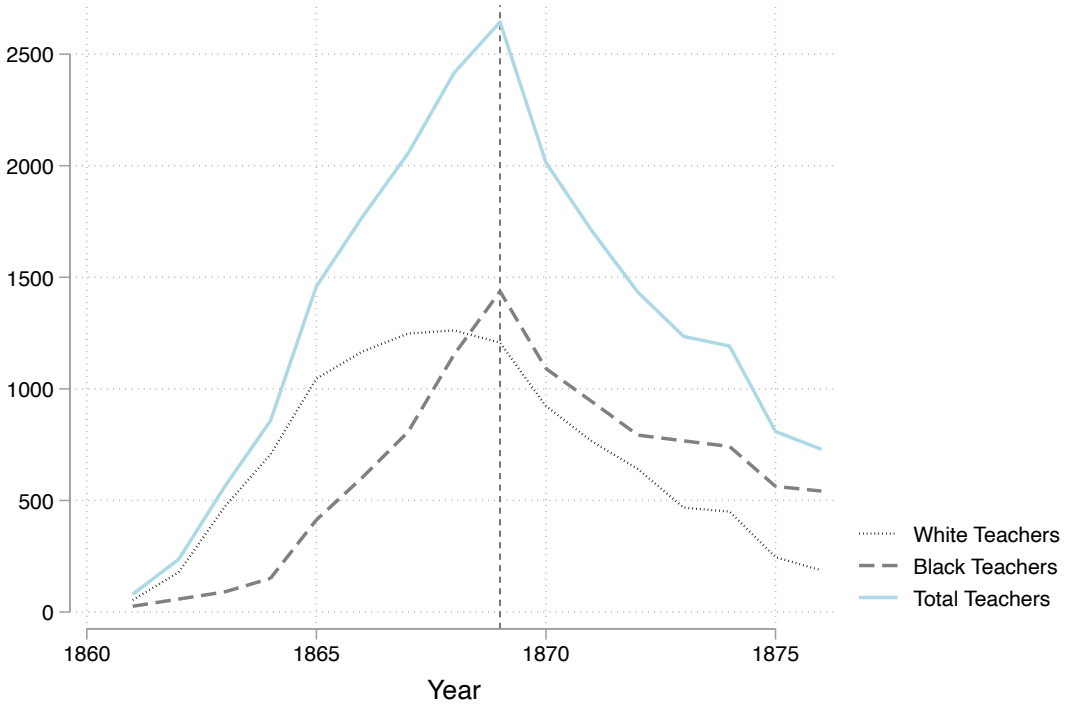
Online Appendix

Reconstruction-Era Education and Long-Run Black-White Inequality

By Daniel B. Jones and Ethan Schmick

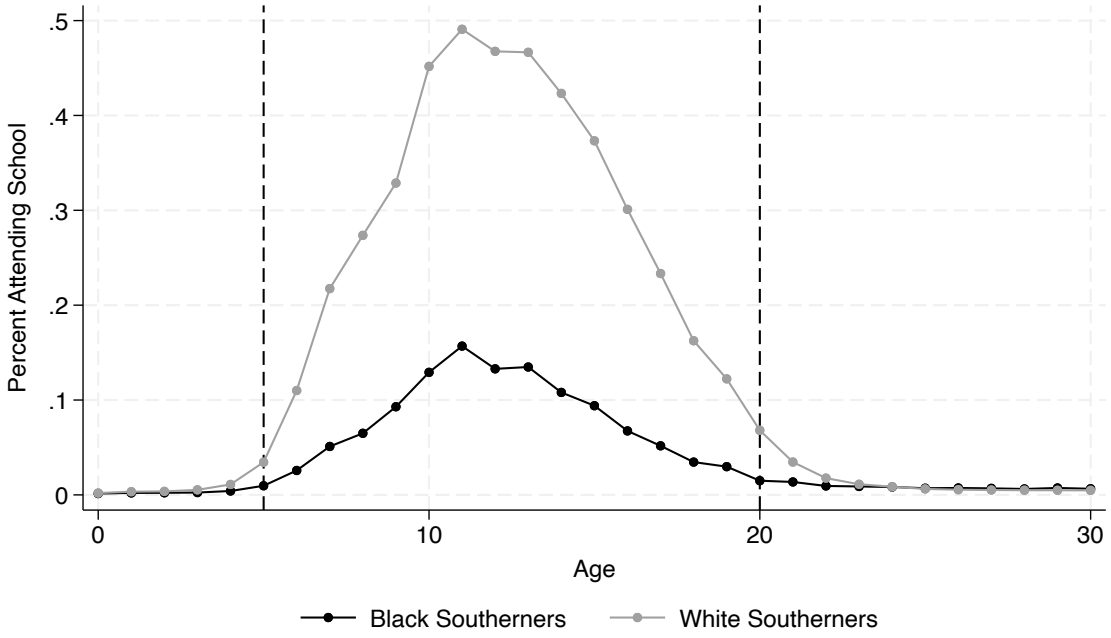
Appendix A: Supplemental Figures and Tables

Figure A.1: Counts of Teachers by Year in Former Confederate States



Notes: Data are most complete prior to 1870; vertical dashed line marks that point.
Sources: Butchart (2016).

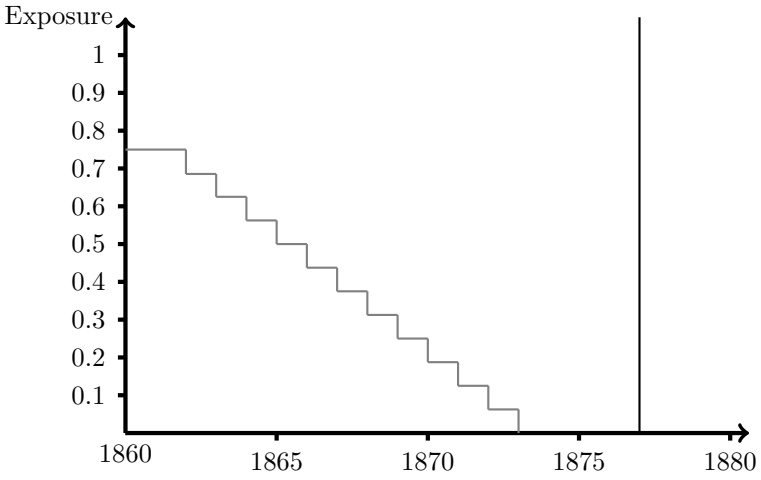
Figure A.2: Southern School Attendance by Age and Race in 1870



Notes: This figure displays the percentage of Black and white Southerners that reported attending school in the 1870 Census by age. Enumerators in the 1870 Census were to code an individual as attending school if they had attended school at all during the previous 12 months.
Sources: Ruggles et al. (2021).

Figure A.3: Exposure to Reconstruction Functions

(a) Louisiana



(b) Texas

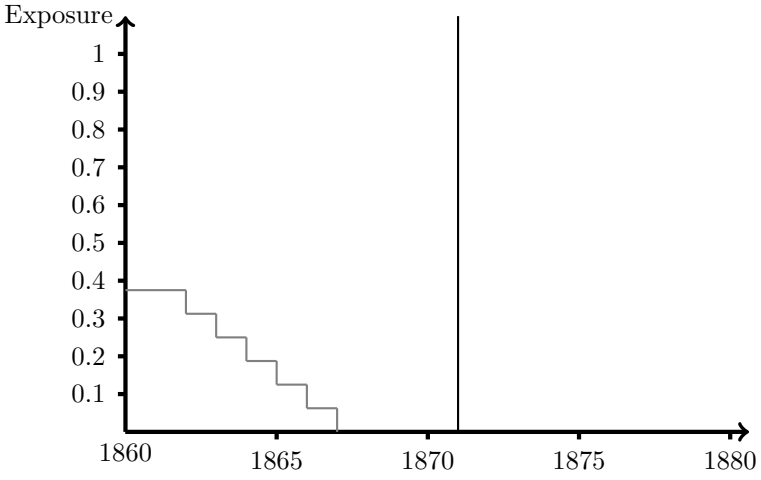
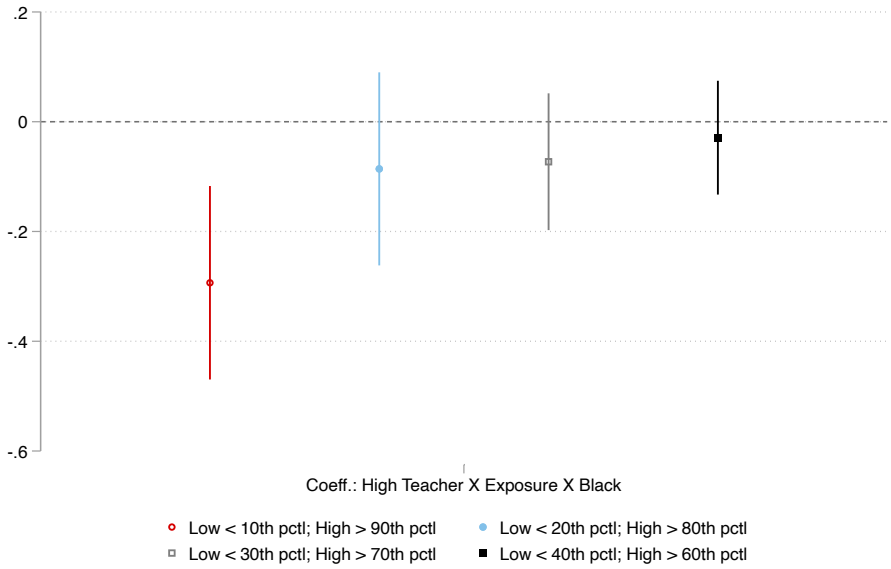
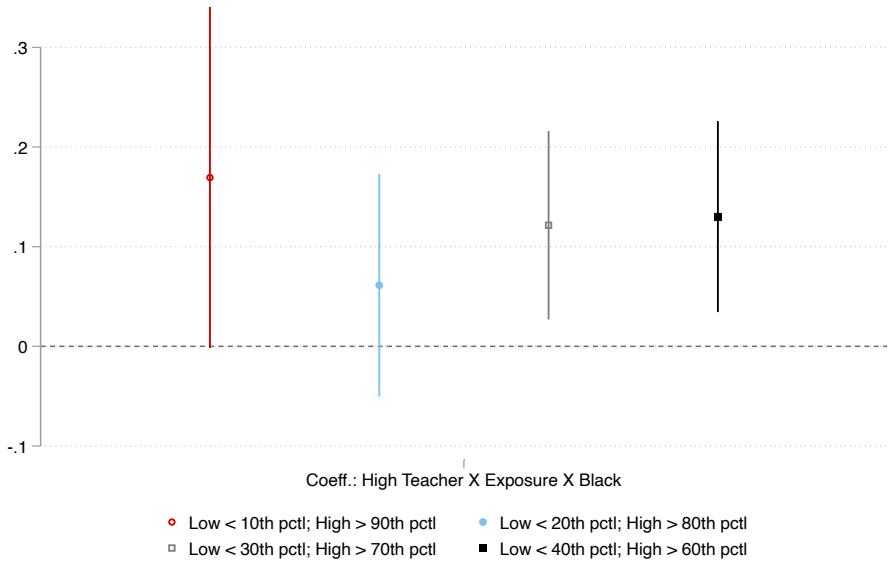


Figure A.4: Triple Differences Estimates, adopting Different Cutoffs for Low, Medium, and High Teacher Intensity



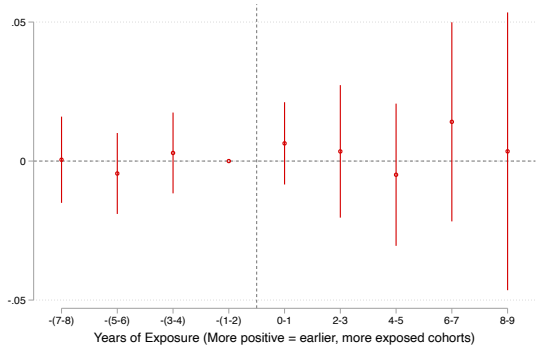
(a) Outcome: Literacy



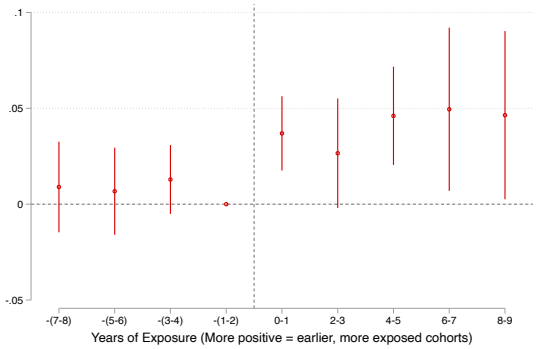
(b) Outcome: log(Occscore)

Notes: Each figure depicts the main triple difference coefficient for four different specifications, each adopting different thresholds for high, medium, and low teacher intensity. Whereas in main analysis, high, medium, and low are defined as terciles, here, from left to right, we instead take: below 10th percentile as low and above 90th percentile as high, below 20th percentile as low and above 80th percentile as high, below 30th percentile and above 70th percentile as high, and below 40th percentile as low and above 60th percentile as high.

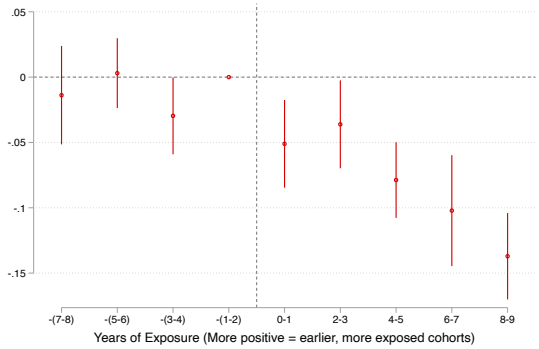
Figure A.5: Event studies - Broad Occupation Categories



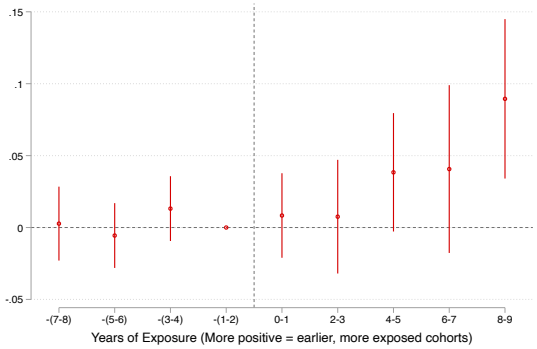
(a) White collar



(b) Blue collar



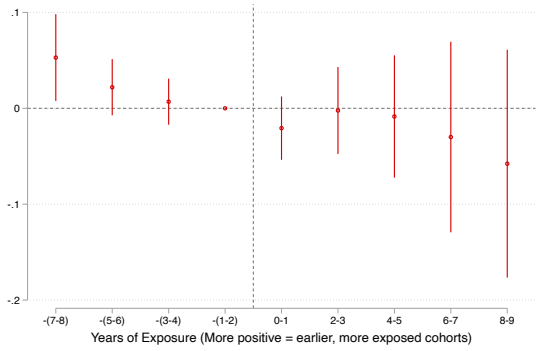
(c) Agriculture



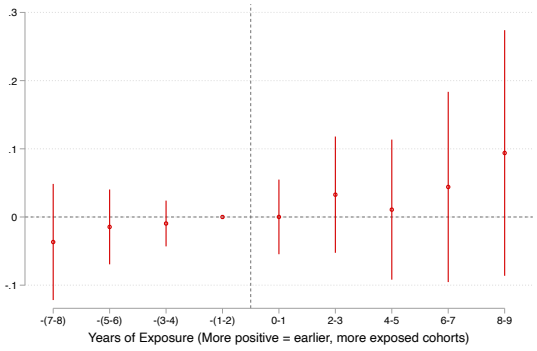
(d) Labor or service

Notes: The unit of observation is an individual linked from either 1870 to 1900 or from 1880 to 1900. We bin individuals who turned 5 more than 8 years after Democratic takeover of their state's legislature into a "<-8" bin, which is not reported. We bin individuals who turned 5 more than 9 years prior to legislative takeover into a ">9" bin, which is not reported. We control for: initial county-of-residence (i.e. in 1870 or 1880) fixed effects and its interaction with an indicator if an individual is Black, birth-year fixed effects and its interaction with an indicator if an individual is Black, census enumeration year fixed effects and its interaction with an indicator if an individual is Black, birth-year fixed effects interacted with the high teacher intensity variable, the Black indicator interacted with the high teacher intensity variable, and the triple interaction of birth-year fixed effects, the Black indicator, the high teacher intensity variable, and a continuous control for the fraction of a county-cohort's school-age years that Union Army troops occupied a county in their urban area and its interaction with an indicator if an individual is Black. Panel A additionally controls for an indicator if the individual is female and its interaction with an indicator if an individual is Black. Standard errors are clustered at the urban area level. 95% confidence intervals are displayed.

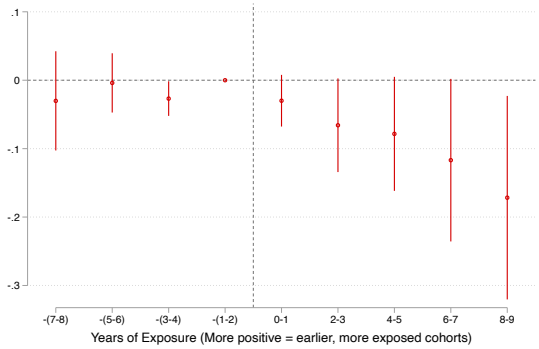
Figure A.6: Event studies - Agricultural Occupations



(a) Farm owners



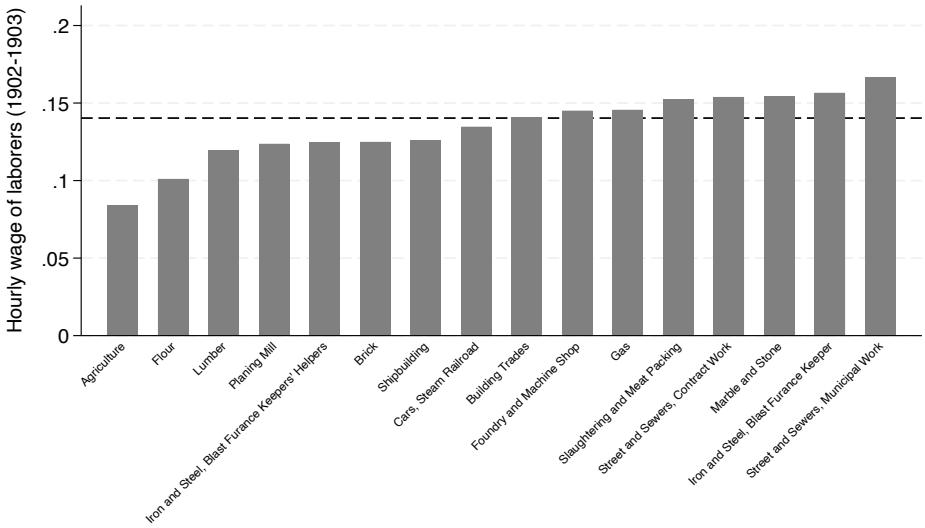
(b) Tenants/sharecroppers



(c) Farm laborers

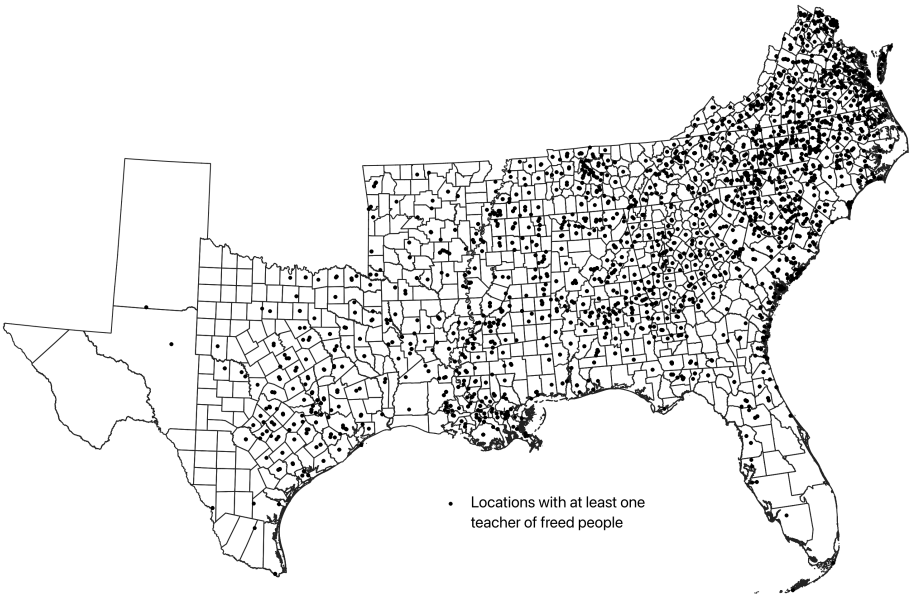
Notes: The unit of observation is an individual linked from either 1870 to 1900 or from 1880 to 1900. We bin individuals who turned 5 more than 8 years after Democratic takeover of their state's legislature into a "<-8" bin, which is not reported. We bin individuals who turned 5 more than 9 years prior to legislative takeover into a ">9" bin, which is not reported. We control for: initial county-of-residence (i.e. in 1870 or 1880) fixed effects and its interaction with an indicator if an individual is Black, birth-year fixed effects and its interaction with an indicator if an individual is Black, census enumeration year fixed effects and its interaction with an indicator if an individual is Black, birth-year fixed effects interacted with the high teacher intensity variable, the Black indicator interacted with the high teacher intensity variable, and the triple interaction of birth-year fixed effects, the Black indicator, the high teacher intensity variable, and a continuous control for the fraction of a county-cohort's school-age years that Union Army troops occupied a county in their urban area and its interaction with an indicator if an individual is Black. Panel A additionally controls for an indicator if the individual is female and its interaction with an indicator if an individual is Black. Standard errors are clustered at the urban area level. 95% confidence intervals are displayed.

Figure A.7: Average hourly wages of laborers in the South



Notes: Data are from U.S. Bureau of Labor Statistics, Bulletin nos. 59 (1905) and 604 (1932) and from *Crops and Markets* published by the U.S. Department of Agriculture Vol. 19, No. 6 pp. 150-151 (1942). Table I of Bulletin no. 59 provides information on the average hourly wages of occupations within manufacturing industries in 1903 in the South (e.g. laborers within the brick making industry). These averages come from a survey of manufacturing establishments and the number of establishments surveyed is also reported. The above figure graphs the weighted average hourly wage for the occupation "laborer" in manufacturing industries in the South Atlantic and South Central regions. Averages are weighted by the number of establishments surveyed. The South Atlantic region includes: Delaware, Washington, D.C., Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, and West Virginia. The South Central region includes: Alabama, Arkansas, Kentucky, Louisiana, Mississippi, Tennessee, and Texas. A table on pages 150-152 of *Crops and Markets*, 19(6), provides a time series of the average wage *per day* for farm laborers in most states. Data is available for 1902 and the average in the figure above is for the South Atlantic and South Central states (i.e. the average farm laborer's daily wage in each of these states is equally weighted to obtain the overall average). According to Table D.1 in the Bureau of Labor Statistics bulletin 604, the average number of hours worked per week for farm laborers around 1900 was 60, equating to six, 10-hour days (this is almost identical to the number of hours worked by laborers in manufacturing industries). Thus, the average wage *per day* for farm laborers was divided by 10 to obtain the average hourly wage shown above. Data for wages for farm laborers are from contracts that did not include board. Daily wages were lower if the contract included board.
Sources: Bureau of Labor Statistics (1905), Bureau of Labor Statistics (1934), and Department of Agriculture (1942).

Figure A.8: Locations of teachers of freed people



Notes: This map displays all locations with at least one teacher of freed people during one year.
Sources: The data on the locations of teachers of freed people come from Butchart et al. (2022).

Table A.1: Summary statistics - 1860 to 1880 birth cohorts

	Black		White	
	Mean	N	Mean	N
	Panel A: Men born 1860 to 1880			
Literate	0.610	166433	0.876	587771
Occscore	14.300	146439	17.725	508334
Agriculture	0.701	146439	0.654	508334
White collar	0.020	146439	0.146	508334
Blue collar	0.064	146439	0.113	508334
Laborer or service	0.214	146439	0.085	508334
Farm owner	0.078	146439	0.241	508334
Tenant farmer	0.258	146439	0.162	508334
Farm laborer	0.366	146439	0.251	508334
No occupation	0.120	166433	0.135	587771
	Panel B: Women born 1860 to 1880			
Literate	0.555	122771	0.862	350140

Notes: This table reports summary statistics for our main dependent variables for both men and women.

Table A.2: Teacher Intensity and School Attendance in 1870 - Women

	Pr(School Attendance=1)	
	(1)	(2)
Medium teacher intensity	-0.034	
	(0.023)	
High teacher intensity	-0.006	
	(0.025)	
Black	-0.191	-0.181
	(0.011)	(0.009)
Medium teacher intensity * Black	0.050	0.037
	(0.017)	(0.014)
High teacher intensity * Black	0.052	0.050
	(0.019)	(0.015)
Observations	1912248	1912248
Birth Year FE	X	X
County FE		X

Notes: The unit of observation is a female in the 1870 census. The estimation is run on all individuals aged 5 through 20 living in a former Confederate state. Standard errors are clustered at the county-of-residence level.

Table A.3: Exposure to Reconstruction and Literacy - Women

	DiD| Black	DiD| White	DiD	DiD| Low teacher	DiD| High teacher	DDD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Exposure	0.085	0.021	0.021	-0.060	0.060	-0.057	-0.066
	(0.038)	(0.019)	(0.019)	(0.041)	(0.030)	(0.041)	(0.038)
Exposure * Black			0.064	0.170	0.013	0.166	0.160
			(0.035)	(0.092)	(0.033)	(0.089)	(0.089)
Exposure * High teacher intensity						0.115	0.138
						(0.050)	(0.048)
Exposure * Black * High teacher intensity						-0.151	-0.133
						(0.095)	(0.092)
Observations	122771	350140	472911	147421	187937	472911	472911

Notes: Estimates are for Specifications (1), (2), and (3) in the text. The unit of observation is an individual linked from either 1870 to 1900 or from 1880 to 1900. Columns 1 and 2 control for: initial county-of-residence (i.e. in 1870 or 1880) fixed effects, birth-year fixed effects, and census enumeration year fixed effects. Columns 3, 4, and 5 control for everything in columns 1 and 2 and their interaction with an indicator variable if an individual is Black. Columns 6 and 7 control for everything included in columns 3, 4, and 5, and additionally control for birth-year fixed effects and its interaction with the high teacher intensity variable, the Black indicator interacted with the high teacher intensity variable, and the triple interaction of birth-year fixed effects, the Black indicator, and the high teacher intensity variable. Column 7 additionally controls for the fraction of a county-cohort's school-age years that Union Army troops occupied a county in their urban area and its interaction with an indicator if an individual is Black. Standard errors are clustered at the urban area level.

Additional Motivation for Specifications in Table A.4

As discussed in the text, there is conflicting evidence on the impacts of the Freedman's Savings Bank on outcomes related to those studied in our paper (Celerier and Tak, 2021; Fu, 2021; Stein and Yannelis, 2020); for that reason, we do not control for exposure to nearby Banks in our main specifications, but we do document that results are very similar when such a control is included (Column 2).

Next, our main analysis includes individuals born from 1860 to 1880, such that the earliest cohorts we study became school-age at the start of Reconstruction. However, we recognize that Black individuals born between 1860 and 1865 might have been born into slavery and, therefore, had very different early life experiences than Black individuals born in 1865 and later (of course, we include race-by-birth year fixed effects throughout all specifications, which partially addresses this issue) – reported in Column 3.

Next, while we take school-age to be 5-20 in our main sample, this choice was largely aimed at adopting the most conservative definition possible. It is therefore important to show that results are similar when trimming to a subset of those years (7-17).[1] See Column 4.

Our linked sample is not representative of the entire population, especially with regards to the share of individuals that are Black. Accordingly, we generate weights for each individual in our linked sample; these weights result in our linked sample appearing more representative of the population on observable characteristics. This process is described in Appendix C. Our point estimates are very similar when using these weights. See Column 5.

Finally, since we measure educational opportunity at the urban area level we use urban area fixed effects rather than county fixed effects (Column 6).

[1] We again refer readers to Appendix Figure A.2 for a descriptive account of the ages at which Black and white southerners are observed attending school in 1870.

Table A.4: Robustness checks - alternative controls, birth cohorts, school-age, and weighting

	Baseline	Control: Freedman's Savings Bank	Birth cohorts: 1865-1880	School-age: 7-17	Weighted	Urban area fixed effects
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Literacy			
Exposure	-0.012	-0.015	-0.008	-0.012	-0.012	-0.010
	(0.015)	(0.013)	(0.013)	(0.013)	(0.027)	(0.014)
Exposure * High teacher intensity	0.046	0.051	0.039	0.037	0.046	0.061
	(0.016)	(0.015)	(0.019)	(0.014)	(0.029)	(0.021)
Exposure * Black	0.142	0.159	0.119	0.104	0.132	0.151
	(0.054)	(0.040)	(0.036)	(0.045)	(0.071)	(0.048)
Exposure * Black * High teacher intensity	-0.014	-0.039	0.007	-0.005	0.006	-0.021
	(0.057)	(0.046)	(0.044)	(0.047)	(0.075)	(0.052)
Observations	754204	754204	650408	754204	754204	754232
			Panel B: Log(Occscore)			
Exposure	-0.066	-0.068	-0.078	-0.049	-0.059	-0.061
	(0.029)	(0.029)	(0.033)	(0.022)	(0.028)	(0.034)
Exposure * High teacher intensity	0.041	0.044	0.033	0.030	0.028	0.060
	(0.051)	(0.053)	(0.066)	(0.035)	(0.050)	(0.049)
Exposure * Black	0.004	0.010	0.015	0.007	0.007	0.006
	(0.036)	(0.040)	(0.047)	(0.025)	(0.033)	(0.042)
Exposure * Black * High teacher intensity	0.126	0.118	0.174	0.080	0.121	0.113
	(0.047)	(0.053)	(0.055)	(0.035)	(0.048)	(0.055)
Observations	654768	654768	562558	654768	654768	654798

Notes: Estimates are for Specification (3) in the text. The unit of observation is an individual linked from either 1870 to 1900 or from 1880 to 1900. All columns in all panels control for: birth-year fixed effects and its interaction with an indicator if an individual is Black, census enumeration year fixed effects and its interaction with an indicator if an individual is Black, a continuous control for the fraction of a county-cohort's school-age years that Union Army troops occupied a county in their urban area and its interaction with an indicator if an individual is Black, birth-year fixed effects interacted with the high teacher intensity variable, the Black indicator interacted with the high teacher intensity variable, and the triple interaction of birth-year fixed effects, the Black indicator, and the high teacher intensity variable. Columns 1-5 control for initial county-of-residence (i.e. in 1870 or 1880) fixed effects and its interaction with an indicator if an individual is Black. Column 6 controls for initial urban-area-of-residence (i.e. in 1870 or 1880) fixed effects and its interaction with an indicator if an individual is Black. Column 2 additionally controls for: the fraction of a county-cohort's school-age years that a branch of the Freedman's Savings Bank was open in a county in their urban area and its interaction with an indicator if an individual is Black. Standard errors are clustered at the urban area level.

Table A.5: Exposure to Reconstruction and other measures of occupational standing Score

	Log(Occ-score)	Log(Imputed income)	Log(LIDO score)
Exposure	-0.064	0.146	-0.037
	(0.031)	(0.102)	(0.026)
Exposure * High teacher intensity	0.029	-0.240	-0.010
	(0.053)	(0.118)	(0.035)
Exposure * Black	-0.021	-0.153	-0.017
	(0.030)	(0.090)	(0.031)
Exposure * Black * High teacher intensity	0.101	0.230	0.076
	(0.042)	(0.106)	(0.046)
Observations	579363	579363	579363

Notes: Estimates are for Specification (3) in the text. The unit of observation is an individual linked from either 1870 to 1900 or from 1880 to 1900. All columns in all panels control for: initial county-of-residence (i.e. in 1870 or 1880) fixed effects and its interaction with an indicator if an individual is Black, birth-year fixed effects and its interaction with an indicator if an individual is Black, census enumeration year fixed effects and its interaction with an indicator if an individual is Black, a continuous control for the fraction of a county-cohort's school-age years that Union Army troops occupied a county in their urban area and its interaction with an indicator if an individual is Black, birth-year fixed effects interacted with the high teacher intensity variable, the Black indicator interacted with the high teacher intensity variable, and the triple interaction of birth-year fixed effects, the Black indicator, and the high teacher intensity variable. Standard errors are clustered at the urban area level.

Table A.6: Robustness checks - alternative treatment measures

	Baseline	1861-1877 teacher intensity	Continuous teacher intensity	County treatment	Any exposure
	(1)	(2)	(3)	(4)	(5)
			Panel A: Literacy		
Exposure	-0.012	0.004	-0.077	0.013	0.005
	(0.015)	(0.015)	(0.025)	(0.011)	(0.003)
Exposure * High teacher intensity	0.046	-0.011	0.046	0.024	-0.003
	(0.016)	(0.021)	(0.011)	(0.011)	(0.005)
Exposure * Black	0.142	0.112	0.203	0.091	0.020
	(0.054)	(0.059)	(0.105)	(0.033)	(0.006)
Exposure * Black * High teacher intensity	-0.014	-0.026	-0.039	0.041	0.012
	(0.057)	(0.071)	(0.048)	(0.037)	(0.012)
Observations	754204	754204	754204	730381	754204
			Panel B: Log(Occscore)		
Exposure	-0.066	-0.049	-0.149	-0.041	-0.006
	(0.029)	(0.032)	(0.086)	(0.019)	(0.006)
Exposure * High teacher intensity	0.041	0.021	0.054	0.001	0.003
	(0.051)	(0.038)	(0.049)	(0.043)	(0.014)
Exposure * Black	0.004	0.006	-0.079	0.054	0.003
	(0.036)	(0.036)	(0.107)	(0.027)	(0.009)
Exposure * Black * High teacher intensity	0.126	0.125	0.079	0.053	0.032
	(0.047)	(0.055)	(0.055)	(0.042)	(0.011)
Observations	654768	654768	654768	638584	654768

Notes: Estimates are for Specification (3) in the text. The unit of observation is an individual linked from either 1870 to 1900 or from 1880 to 1900. All columns in all panels control for: initial county-of-residence (i.e. in 1870 or 1880) fixed effects and its interaction with an indicator if an individual is Black, birth-year fixed effects and its interaction with an indicator if an individual is Black, census enumeration year fixed effects and its interaction with an indicator if an individual is Black, a continuous control for the fraction of a county-cohort's school-age years that Union Army troops occupied a county in their urban area and its interaction with an indicator if an individual is Black, birth-year fixed effects interacted with the high teacher intensity variable, the Black indicator interacted with the high teacher intensity variable, and the triple interaction of birth-year fixed effects, the Black indicator, and the high teacher intensity variable. Panel A controls for an indicator if the individual is female and its interaction with an indicator if an individual is Black. Standard errors are clustered at the urban area level.

Table A.7: Exposure to Reconstruction, Literacy, and Income - Alternative Measure of Exposure

	DiD| Black	DiD| White	DiD	DiD| Low teacher	DiD| High teacher	DDD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: Literacy				
Exposure * Years Republican	0.131	0.017	0.017	-0.010	0.038	-0.008	-0.008
	(0.025)	(0.009)	(0.009)	(0.019)	(0.008)	(0.019)	(0.019)
Exposure * Years Republican * Black			0.114	0.188	0.117	0.186	0.176
			(0.020)	(0.069)	(0.022)	(0.067)	(0.068)
Exposure * Years Republican * High teacher intensity						0.044	0.045
						(0.020)	(0.021)
Exposure * Years Republican * Black * High teacher intensity						-0.065	-0.038
						(0.071)	(0.071)
Observations	166433	587771	754204	249755	290029	754204	754204
			Panel B: Log(Occscore)				
Exposure * Years Republican	0.096	-0.010	-0.010	-0.080	-0.003	-0.079	-0.075
	(0.039)	(0.027)	(0.027)	(0.038)	(0.053)	(0.036)	(0.035)
Exposure * Years Republican * Black			0.106	0.017	0.122	0.016	0.005
			(0.023)	(0.051)	(0.040)	(0.048)	(0.043)
Exposure * Years Republican * High teacher intensity						0.075	0.064
						(0.063)	(0.058)
Exposure * Years Republican * Black * High teacher intensity						0.108	0.138
						(0.062)	(0.057)
Observations	146434	508334	654768	217792	250788	654768	654768
Control for Union troop presence							X

Notes: Estimates are for Specifications (1), (2), and (3) in the text. The unit of observation is an individual linked from either 1870 to 1900 or from 1880 to 1900. The "Exposure" measure in this table is not vary by state. Rather it is simply the share of school-age years (5-20) between 1866 and 1877. This is interacted with the share of years during Reconstruction (1866-1877) that an individual's state had a majority Republican legislature. Columns 1 and 2 control for: initial county-of-residence (i.e. in 1870 or 1880) fixed effects, birth-year fixed effects, and census enumeration year fixed effects. Columns 3, 4, and 5 control for everything in columns 1 and 2 and their interaction with an indicator variable if an individual is Black. Columns 6 and 7 control for everything included in columns 3, 4, and 5, and additionally control for birth-year fixed effects and its interaction with the high teacher intensity variable, the Black indicator interacted with the high teacher intensity variable, and the triple interaction of birth-year fixed effects, the Black indicator, and the high teacher intensity variable. Column 7 additionally controls for the fraction of a county-cohort's school-age years that Union Army troops occupied a county in their urban area and its interaction with an indicator if an individual is Black. Panel A controls for an indicator if an individual is female in columns 1 and 2 and the female indicator interacted with an indicator if an individual is Black in columns 3-7. Standard errors are clustered at the urban area level.

Table A.8: Exposure to Reconstruction and broad occupational categories

	Ag.	White collar	Blue collar	Service or laborer	Service or laborer modified	No occ.
	(1)	(2)	(3)	(4)	(5)	(6)
Exposure	0.061	-0.009	-0.011	-0.040	-0.028	-0.030
	(0.045)	(0.027)	(0.012)	(0.023)	(0.014)	(0.013)
Exposure * High teacher intensity	-0.083	0.011	0.006	0.064	0.041	0.011
	(0.052)	(0.031)	(0.039)	(0.028)	(0.017)	(0.028)
Exposure * Black	-0.003	-0.003	-0.003	0.008	0.022	-0.047
	(0.033)	(0.032)	(0.018)	(0.027)	(0.017)	(0.023)
Exposure * Black * High teacher intensity	-0.094	-0.015	0.005	0.106	0.072	0.039
	(0.040)	(0.039)	(0.046)	(0.038)	(0.030)	(0.034)
Observations	654768	654768	654768	654768	611741	754204

Notes: Estimates are for Specification (3) in the text. The unit of observation is an individual linked from either 1870 to 1900 or from 1880 to 1900. All columns in all panels control for: initial county-of-residence (i.e. in 1870 or 1880) fixed effects and its interaction with an indicator if an individual is Black, birth-year fixed effects and its interaction with an indicator if an individual is Black, census enumeration year fixed effects and its interaction with an indicator if an individual is Black, a continuous control for the fraction of a county-cohort's school-age years that Union Army troops occupied a county in their urban area and its interaction with an indicator if an individual is Black, birth-year fixed effects interacted with the high teacher intensity variable, the Black indicator interacted with the high teacher intensity variable, and the triple interaction of birth-year fixed effects, the Black indicator, and the high teacher intensity variable. Column 5, which examines "Service or laborer modified" excludes laborers for which we have no reported industry and individuals coded as a general "laborer", but whose industry is reported as agriculture. Standard errors are clustered at the urban area level.

Table A.9: Exposure to Reconstruction and the agricultural ladder

	Farm laborer	Tenant	Farm owner
	(1)	(2)	(3)
Exposure	0.048	-0.132	0.146
	(0.022)	(0.025)	(0.034)
Exposure * High teacher intensity	0.079	0.103	-0.266
	(0.112)	(0.044)	(0.094)
Exposure * Black	0.003	0.042	-0.050
	(0.055)	(0.068)	(0.084)
Exposure * Black * High	-0.337	0.206	0.038
teacher intensity	(0.087)	(0.102)	(0.122)
Observations	654768	654768	654768

Notes: Estimates are for Specification (3) in the text. The unit of observation is an individual linked from either 1870 to 1900 or from 1880 to 1900. All columns in all panels control for: initial county-of-residence (i.e. in 1870 or 1880) fixed effects and its interaction with an indicator if an individual is Black, birth-year fixed effects and its interaction with an indicator if an individual is Black, census enumeration year fixed effects and its interaction with an indicator if an individual is Black, a continuous control for the fraction of a county-cohort's school-age years that Union Army troops occupied a county in their urban area and its interaction with an indicator if an individual is Black, birth-year fixed effects interacted with the high teacher intensity variable, the Black indicator interacted with the high teacher intensity variable, and the triple interaction of birth-year fixed effects, the Black indicator, and the high teacher intensity variable. Standard errors are clustered at the urban area level.

Table A.10: Intergenerational Effects on Occupational Standing (with weighting)

	(1) Ref. parent: Fathers	(2) Ref. parent: Mothers
Parent Expo.	0.031	-0.040
	(0.087)	(0.086)
Parent Expo. X Black	-0.138	-0.044
	(0.099)	(0.100)
Parent Expo. X High tch.	-0.065	0.100
	(0.090)	(0.099)
Parent Expo. X Black X High tch.	0.228	0.174
	(0.127)	(0.158)
Observations	85,438	50,210

Notes: Estimates are for Specification (3) in the text. The unit of observation is a son of a father or mother that was successfully linked from the 1870 to 1900 or from 1880 to 1900. These sons themselves had to be successfully linked from 1900 to 1920. In Panel A the reference parent is the father. In Panel B the reference parent is the mother. All columns in all panels control for: parents' initial county-of-residence (i.e. in 1870 or 1880) fixed effects and its interaction with an indicator if an individual is Black, parents' birth-year fixed effects and its interaction with an indicator if an individual is Black, parents' census enumeration year fixed effects and its interaction with an indicator if an individual is Black, parents' birth-year fixed effects interacted with the high teacher intensity variable for the parents' urban area, the Black indicator interacted with the high teacher intensity variable for the parents' urban area, and the triple interaction of parents' birth-year fixed effects, the Black indicator, and the high teacher intensity variable for the parents' urban area. Standard errors are clustered at the parent's urban area level.

Appendix B: Construction of Urban Areas and Migration

Construction of Urban Areas

Our aim was to first identify counties that served as urban population centers within their states. We did so through a process of eliminating counties within states that clearly did not fit this definition. Our process was as follows. We first sorted counties by population within states based on the 1870 Census and constructed a cumulative within-state population share. This allowed us to identify the smallest set of counties, starting from the highest population counties, that collectively formed at least half of the state's population. We dropped any counties that met two conditions: (1) they were outside of this set and (2) the share of the county population residing in an urban area was zero. (Requiring the second condition was meant to ensure that we did not drop any highly urban counties that are otherwise low in population.) Next, we followed the same procedure, but sorting counties by their urban population and constructing a cumulative within-state urban population share. Ordering counties by their state share of urban population, we identify the smallest set of counties that contains three-fourths of the urban population. We dropped any counties that met two conditions: (1) they were outside of this second set and (2) they were not in the top three counties with regards to overall population. (The latter restriction was meant to allow for a more conservative definition of urban population centers.) Finally, we dropped any counties with zero urban population. These counties constituted our "urban center" counties.

We then created within-state groupings of counties around the "urban center" counties. We did so simply based on (centroid-to-centroid) distance. For every non-urban center county, we measured the distance to the urban center counties within the same state and assigned a county to the urban center county with the minimum distance. We refer to the resulting groupings of counties as "urban areas".

This process resulted in thirty unique "urban areas" across the eleven former-Confederate states in our sample, ranging from one urban area in two states to five urban areas in one state. The modal number of urban areas within a state is three.

Migration

Our main reason for constructing urban areas was to account for the fact that families may move and may do so in particular to seek out the educational opportunity that forms the basis of our "treatment". If families (or at least some families) moved to a nearby urban center to obtain greater educational opportunity, then educational opportunity is better thought of at an urban area level than at a county level.

Moves from rural to urban areas were not uncommon during the period (Foner, 1988). Whether our method of constructing urban areas captures such moves is an empirical question. We provide some evidence on that question here.

If families seeking greater educational opportunity moved to the *nearest* county with such opportunity, then most of the increase in teacher intensity amongst movers should come from moves within, and not across, urban areas. We test this by linking families across the 1870 and 1880 Census and restrict the sample to individuals who moved. We then explore how *county level* teacher intensity (measured as teachers per 1,000 Black school-aged children, as in the paper) differs in 1880 relative to 1870 for these moving families. Recall that teacher intensity is a static measure, so the changes in teacher intensity depicted here are based solely on locational change and not local changes in teacher intensity. (Put differently, we test whether families in 1880 are seen in counties with higher 1870 teacher intensity than the 1870 teacher intensity of their 1870 county.)

As the plot depicts, amongst all movers ("any move"), there is a general increase in teacher intensity – families are moving towards higher teacher intensity counties. However, we find that this is entirely driven by within-area moves and not across-area moves.

Second, we are not only concerned that people are moving to higher teacher intensity counties, but also that there may be selection in doing so. We tested whether there was a relationship between the occupational score of head-of-household and the likelihood of moving. Results from simple bivariate regressions on this are presented below. In short, there is a (weak and - surprisingly - negative) relationship, but not for those moving *across* areas.

The two pieces of evidence presented here can be summarized as: families moving within "urban areas" moved to higher teacher intensity and there is some relationship between

Figure B.1: Migration and change in teacher intensity

head-of-household occupational income score and likelihood of making such moves. None of that is true for across-area movers. As such, we take "urban area" rather than county as our unit of treatment – the level at which we measure teacher intensity. Leaving our measure of teacher intensity at the county-level would ignore that there was movement across counties and some selection on said movement.

Table B.1: Relation between head-of-household occscore and moves

VARIABLES	(1) Moved	(2) Moved across zones	(3) Moved within zone
Occ. Inc. Score (std. norm.)	-0.003*	-0.001	-0.002*
	(0.001)	(0.001)	(0.001)
Observations	266,553	266,553	266,553
R-squared	0.060	0.099	0.079

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix C: Linking, Linking Representativeness, and Weighting

Linked sample - main analysis

Our main analysis uses linked samples from the Multigenerational Longitudinal Panel (MLP; Helgertz et al. (2020)) provided by IPUMS (Ruggles et al., 2021). The MLP provides links between decadal censuses (e.g. individuals in the 1870 census linked to the 1880 census) and between 1880 and 1900 since the 1890 census manuscripts were destroyed in a fire. We use two samples in our main analysis: a sample linked from 1870 to 1900 and a sample linked from 1880 to 1900. The 1870 to 1900 sample simply consists of the set of individuals that appear in both the 1870 to 1880 sample and the 1880 to 1900 sample.

Our final sample consists of the pooled 1870 to 1900 and 1880 to 1900 linked samples with a few restrictions. First, if an individual appears in both samples, we only keep their observation from the 1870 to 1900 sample. Second, we consider only children who were born in 1860 or later, were living in a Confederate state in the initial census (either 1870 or 1880), and were living in the state they were born in when first observed (either in 1870 or 1880). Finally, in our main analysis we restrict our attention to men, although we do examine literacy for women in the appendix. This results in 166,686 men linked from 1870 to 1900 and 587,518 men linked from 1880 to 1900, for a pooled sample of 754,204 men.[2] Our link rate of children from 1870 to 1900 is 11%; our link rate of children from 1880 to 1900 is 17%.

The MLP establishes links at a higher rate and with greater accuracy than competing methods. To construct the MLP, Helgertz et al. (2020) proceed in two steps. In Step 1, they train data to identify links based on *both* immutable individual characteristics (name, birth year, and place of birth) and mutable household characteristics. For example, the MLP uses information on parental/spousal names, birth year, and place of birth to help identify links. It also uses information on other household members and residential characteristics (such as street name) to help identify links. This approach contrasts with other linking methods,

[2]For women, we have 50,465 linked from 1870 to 1900 and 422,446 linked from 1880 to 1900 for a total of 472,911 women.

which rely exclusively on immutable individual characteristics. After training the data, a machine learning algorithm is applied to link between complete count censuses. All of this is performed only for men.

Step 2 attempts to link other individuals from linked mens' households. Data is once again trained and a machine learning algorithm applied, but potential matches are restricted to the household in which they are expected to appear. This step allows women to be linked. Helgertz et al. (2022) provide more details on the entire procedure and compare the results to other linking methods (Abramitzky, Boustan and Eriksson, 2014; Feigenbaum, 2016).

A concern with any linked sample is whether the individuals that are linked resemble the entire population. If certain types of individuals are linked at higher rates they can generate biased estimates. This is particularly true when studying intergenerational mobility. To use an example provided by Ward (2023), "if children from low socioeconomic status families who remain poor in adulthood are less likely to be linked" than one would overestimate the amount of intergenerational mobility (Ward, 2023, p. 3222). Our analysis is not concerned with rank-rank correlations of intergenerational mobility, but instead uses a triple differences specification that exploits variation in: year of birth, teacher intensity in the local area, and race. Accordingly, our estimates would only be biased if, for example, we were more likely to link literate Black children educated during Reconstruction in high teacher-intensity areas than literate white children educated during Reconstruction in the same areas.

To make our linked sample appear representative of the population, we re-weight using inverse propensity weighting, as described in Appendix B of Ward (2023). We perform the following steps:

1. We pool the linked sample with the sample of all children from the complete count census who satisfy the same restrictions (i.e. were born in 1860 or later, were living in a Confederate state, and were living in the state they were born in). Our re-weighting is based on the population in the initial (i.e. 1870 or 1880), rather than final census (i.e. 1900), because we restrict linked children to be living in the same state they were born in during the initial census. It is not possible to make this same restriction for unlinked individuals in 1900.

2. We estimate a probit model to predict who will be in the linked sample. We use the following variables to predict who will be in the linked sample: a Black indicator, age (in 10-year bins) and its interaction with the Black indicator, sex and its interaction with the Black indicator, state of residence and its interaction with the Black indicator, and farm status (living on a farm) and its interaction with the Black indicator.

3. Using the estimates from the probit model, we calculate $\hat{p}$, the probability that an individual is linked. Panel A of Figure C.1 shows the distribution of link probabilities for males linked from 1870 to 1900 and for unlinked males in the 1870 census; Panel B shows the same, but for 1880. Both figures show a large amount of overlap in the probability of being linked meaning we are not just linking certain types of individuals.

4. We re-weight the sample using an inverse propensity weight: $(\frac{1-\hat{p}}{\hat{p}}) * (\frac{1}{1-q})$. $\hat{p}$ is the predicated probability that an individual is linked and q is the share of the population that is linked, which was discussed above.

Table C.1 shows that our linked sample is unrepresentative of the population, but much of this is corrected by re-weighting the linked sample using the inverse propensity weight. Columns 1 and 2 show that we are much less likely to link Black and female children. We are also more likely to link children living on farms. Column 3 re-weights column 2, which now appears more similar to column 1.

Column 5 of Table A.4 displays estimates from our baseline specification weighted by the inverse propensity of being linked. The estimates are extremely similar to our baseline estimates without weighting. Thus, it does not appear that the unrepresentativeness of our linked sample is biasing our estimates.

Linked sample - intergenerational analysis

For our intergenerational analysis, we first use the 1900 census to locate *sons* of the linked individuals used in our main analysis. We then use the MLP to link these sons to the 1920 census to obtain adult outcomes. We generate one linked dataset of fathers and their sons and a second linked dataset of mothers and their sons and estimate the intergenerational results separately to see if both fathers' and mothers' educational opportunities during Re-

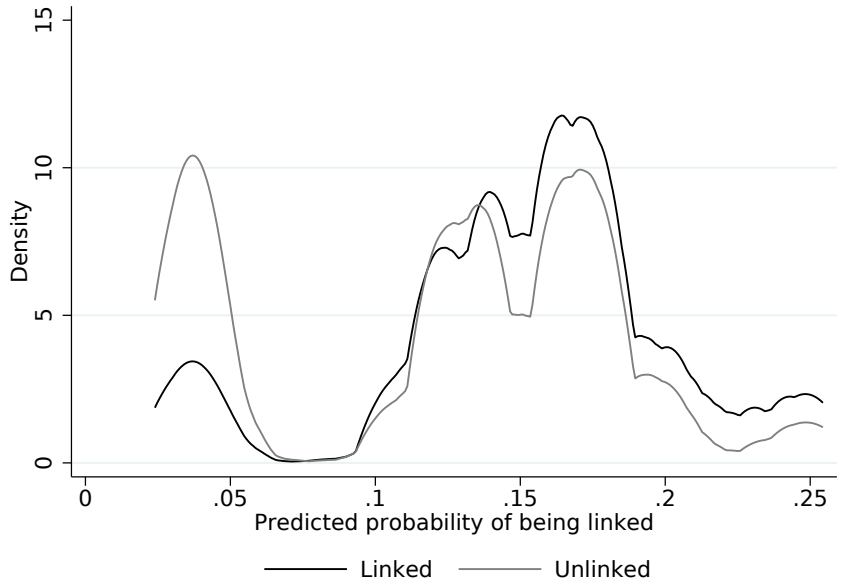
construction matter for the next generation. 85,310 sons for whom we have information about their father were linked and 50,259 sons for whom we have information about their mother were linked.

Table C.2 compares the representativeness of sons in our linked intergenerational samples to the relevant population. We consider the relevant population to be males in the 1900 census who were under the age of 20, living with their father (or mother depending on the dataset), and whose father (mother) was born in a former Confederate state. We acknowledge that this definition is not perfect. In particular, it would be preferable to define the relevant population using the same restrictions we use in our main analysis: sons of individuals who were (1) born in a former Confederate state, (2) were living in that same state in 1870 or 1880, and (3) were born in 1860 or later. However, the census does not contain information on parental birth year or the location the parent was living in 1870 or 1880 (it does contain information on parental birthplace). Despite these imperfections, the above definition of the relevant population results in a link rate of 2% for sons for whom have information about their father and 1% for sons for whom we have information about their mother. Very low link rates are expected here; we first have to link an individual's parent from either 1870 or 1880 to 1900 and then link that individual themselves from 1900 to 1920. At a minimum this requires links between three censuses (1880 to 1900; 1900 to 1910; 1910 to 1920) and possibly four (1870 to 1880; 1880 to 1900; 1900 to 1910; 1910 to 1920). As shown in the first two columns of Table C.2, there are fewer Black sons in our linked sample than in the population.

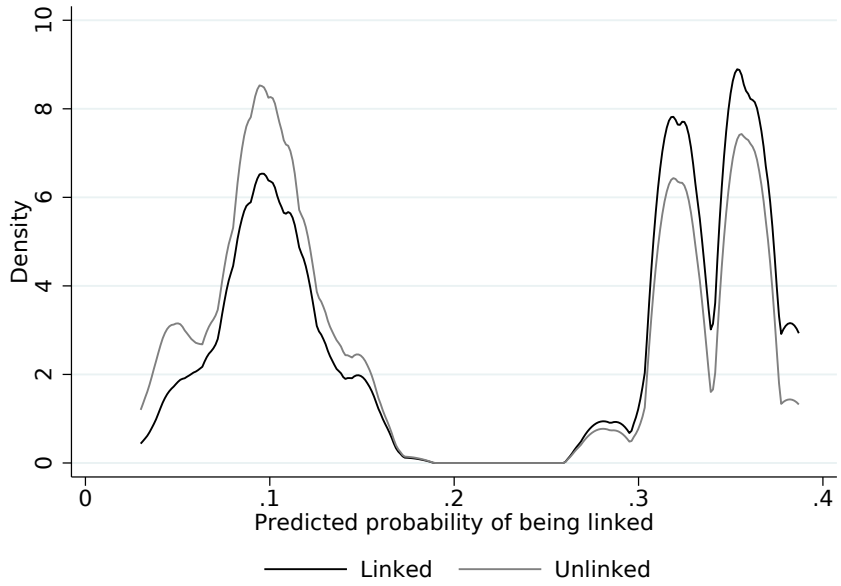
We, again, calculate inverse propensity weights for each of these sons using the steps outlined in the previous subsection. The probit model used to generate these weights includes a Black indicator, age (in 10-year bins) and its interaction with the Black indicator, and division of residence in 1900 and its interaction with the Black indicator. Once we re-weight our linked sample, it is more representative of the population (column 3 of Table C.2). Figure C.2 shows the distribution of link probabilities for linked and unlinked sons. Panel A is sons for whom we have information about their father and Panel B is sons for whom we have information about their mother. There is overlap in the distributions.

Figure C.1: Predicted probability of being in the linked sample - main sample

(a) 1870-1900



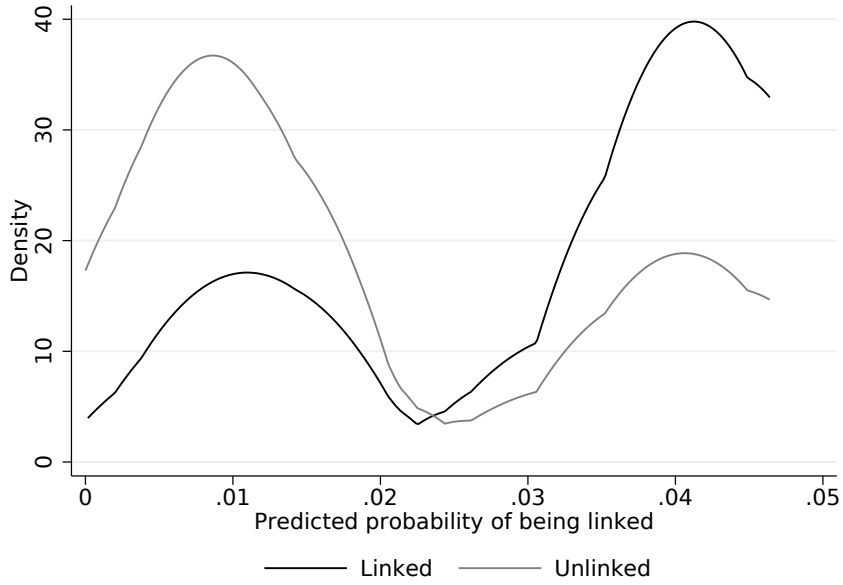
(b) 1880-1900



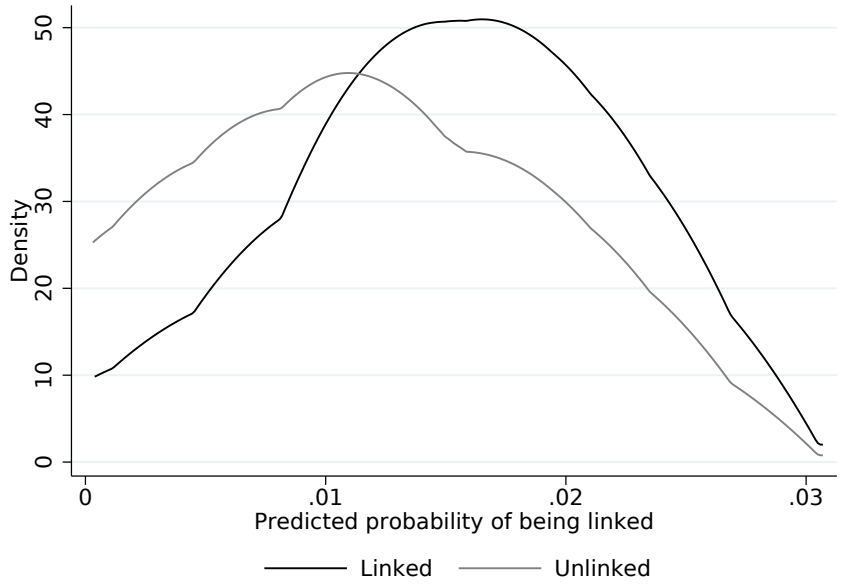
Notes: These figures present kernel density estimates of the predicated probability of being linked for both linked and unlinked individuals. The predicated probability of being linked was obtained from a probit regression where the dependent variable was a dummy if the individual was linked and the independent variables were: a Black indicator, age (in 10-year bins) and its interaction with the Black indicator, sex and its interaction with the Black indicator, state of residence and its interaction with the Black indicator, and farm status (living on a farm) and its interaction with the Black indicator. In panel A, the linked sample of individuals from 1870-1900 that are in our baseline specification in Panel A of Table 3 are pooled with all individuals in the 1870 census who were living in a former Confederate state and born after 1860 (i.e. the same restrictions we use in our baseline specification). In panel B, the linked sample of individuals from 1880-1900 that are in our baseline specification in Panel A of Table 3 are pooled with all individuals in the 1880 census who were living in a former Confederate state and born after 1860 (i.e. the same restrictions we use in our baseline specification). The kernel density estimates use an Epanechnikov kernel function with a bandwidth of 0.008.

Figure C.2: Predicted probability of being in the linked sample - intergenerational sample

(a) Sons with information on fathers



(b) Sons with information on mothers



Notes: These figures present kernel density estimates of the predicated probability of being linked for both linked and unlinked individuals. The predicated probability of being linked was obtained from a probit regression where the dependent variable was a dummy if the individual was linked and the independent variables were: a Black indicator, and region of residence in 1900 and its interaction with the Black indicator. In panel A, the linked sample of sons for whom we have information about their fathers are pooled with all individuals in the 1900 census who were under the age of 20, living with their father, and whose father was born in a former Confederate state. In panel B, the linked sample of sons for whom we have information about their mothers are pooled with all individuals in the 1900 census who were under the age of 20, living with their mother, and whose mother was born in a former Confederate state. The kernel density estimates use an Epanechnikov kernel function with a bandwidth of 0.005.

Table C.1: Comparison of linked to unlinked individuals - main sample

	Population	Linked unweighted	Linked weighted
	Panel A: 1870 Sample		
Age	4.801	4.664	4.686
Black	0.442	0.147	0.481
Live on farm	0.447	0.613	0.425
State of residence:			
Alabama	0.108	0.090	0.111
Arkansas	0.045	0.049	0.045
Florida	0.019	0.012	0.019
Georgia	0.132	0.122	0.131
Louisiana	0.073	0.061	0.075
Mississippi	0.088	0.058	0.092
North Carolina	0.115	0.153	0.108
South Carolina	0.078	0.060	0.081
Tennessee	0.133	0.161	0.130
Texas	0.083	0.079	0.085
Virginia	0.127	0.155	0.123
Observations	1459325	166686	166686
	Panel B: 1880 Sample		
Age	8.418	5.616	7.834
Black	0.444	0.241	0.492
Live on farm	0.580	0.657	0.575
State of residence:			
Alabama	0.101	0.100	0.103
Arkansas	0.052	0.059	0.058
Florida	0.019	0.016	0.020
Georgia	0.126	0.109	0.121
Louisiana	0.070	0.070	0.074
Mississippi	0.090	0.076	0.099
North Carolina	0.114	0.126	0.100
South Carolina	0.082	0.073	0.085
Tennessee	0.121	0.138	0.117
Texas	0.107	0.115	0.115
Virginia	0.118	0.118	0.106
Observations	3374969	587518	587518

Notes: This table presents a comparison of means for individuals that were not linked and individuals that were linked in our main sample. Panel A compares men linked from 1870 to 1900 that are in our baseline specification (Panel A of Table 3 with the population of men in the 1870 census who were living in a former Confederate state and born after 1860 (i.e. the same restrictions we use in our baseline specification). Panel B compares men linked from 1880 to 1900 that are in our baseline specification (Panel A of Table 3 with the population of men in the 1880 census who were living in a former Confederate state and born after 1860 (i.e. the same restrictions we use in our baseline specification). Column (1) shows the mean for the population, column (2) shows the mean for linked individuals, and column (3) weights the mean for linked individuals so they appear more representative of the population. The weight applied to individuals is given by the formula: $(\frac{1-\hat{p}}{\hat{p}}) * (\frac{1}{1-q})$, where $\hat{p}$ is the predicated probability that an individual is linked and q is the share of the population that is linked. In Panel A, 11.4% of the population is linked. In Panel B, 17.4% of the population is linked.

Table C.2: Comparison of linked to unlinked individuals - intergenerational sample

	Population	Linked unweighted	Linked weighted
	Panel A: Sons with information on fathers		
Age	8.641	5.975	8.200
Black	0.336	0.110	0.340
Division of residence in 1900:			
New England	0.001	0.000	0.001
Mid-Atlantic	0.007	0.001	0.007
East North Central	0.014	0.002	0.013
West Noth Central	0.023	0.005	0.023
South Atlantic	0.428	0.448	0.430
East South Central	0.269	0.310	0.265
West South Central	0.250	0.235	0.252
Mountain	0.004	0.000	0.005
Pacific	0.004	0.000	0.005
Observations	4267150	85450	85439
	Panel B: Sons with information on mothers		
Age	8.743	8.903	9.080
Black	0.373	0.141	0.376
Division of residence in 1900:			
New England	0.001	0.000	0.001
Mid-Atlantic	0.005	0.001	0.005
East North Central	0.009	0.001	0.009
West Noth Central	0.017	0.003	0.017
South Atlantic	0.412	0.456	0.410
East South Central	0.282	0.319	0.282
West South Central	0.267	0.219	0.269
Mountain	0.003	0.001	0.003
Pacific	0.004	0.000	0.004
Observations	4479573	50259	50210

Notes: This table presents a comparison of means for individuals that were not linked and individuals that were linked in our intergenerational sample. Panel A compares sons linked from 1900 to 1920 for whom we have information about their father (Panel A of Table 4) with the population of men in the 1900 census who were under the age of 20, living with their father, and whose father was born in a former Confederate state. Panel B compares sons linked from 1900 to 1920 for whom we have information about their mother (Panel B of Table 4) with the population of men in the 1900 census who were under the age of 20, living with their mother mother, and whose mother was born in a former Confederate state. Column (1) shows the mean for the population, column (2) shows the mean for linked individuals, and column (3) weights the mean for linked individuals so they appear more representative of the population. The weight applied to individuals is given by the formula: $(\frac{1-\hat{p}}{\hat{p}}) * (\frac{1}{1-q})$, where $\hat{p}$ is the predicated probability that an individual is linked and q is the share of the population that is linked. In Panel A, 2% of the population is linked. In Panel B, 1.12% of the population is linked.

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2014. "A nation of immigrants: Assimilation and economic outcomes in the age of mass migration." *Journal of Political Economy*, 122(3): 467–506.

Bureau of Labor Statistics, United States. 1905. "Bulletin of the United States Bureau of Labor, No. 59, Volume XI."

Bureau of Labor Statistics, United States. 1934. "History of Wages in the United States From Colonial Times to 1928 : Bulletin of the United States Bureau of Labor Statistics."

Butchart, Ronald E. 2016. "Freedmen's education during reconstruction." *New Georgia Encyclopedia*, 13: 4–13.

Butchart, Ronald E., Melanie Pavich, Mary Ella Engel, Christina Davis, and Amy F. Rolleri. 2022. "The Freedmen's Teacher Project: Teachers among the Freed People in the U.S. South, 1861-1877." Harvard Dataverse.

Celerier, Claire, and Purnoor Tak. 2021. "The Impact of Financial Inclusion on Minorities: Evidence from the Freedman's Savings Bank." *Available at SSRN 3825143*.

Department of Agriculture, United States. 1942. "Crops and Markets."

Feigenbaum, James J. 2016. "Automated census record linking: A machine learning approach."

Foner, Eric. 1988. *Reconstruction: America's Unfinished Revolution.* HarperCollins.

Fu, Xuanyu. 2021. *Life Cycle and Intergenerational Effects of Income and Wealth.* University of California, Los Angeles.

Helgertz, Jonas, Joseph Price, Jacob Wellington, Kelly J Thompson, Steven Ruggles, and Catherine A Fitch. 2022. "A new strategy for linking US historical censuses: A case study for the IPUMS multigenerational longitudinal panel." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 55(1): 12–29.

Helgertz, Jonas, Steven Ruggles, John Robert Warren, Catherine A. Fitch, Ronald Goeken, J. David Hack, Matt A. Nelson, Joseph P. Price, Evan Roberts, and Matthew Sobek. 2020. "IPUMS Multigenerational Longitudinal Panel: Version 1.0 [dataset]." IPUMS.

Ruggles, Steven, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Evan Roberts, Megan Schouweiler, and Matthew Sobek. 2021. "IPUMS Ancestry Full Count Data: Version 3.0 [dataset]." IPUMS.

Stein, Luke CD, and Constantine Yannelis. 2020. "Financial inclusion, human capital, and wealth accumulation: Evidence from the freedman's savings bank." *The Review of Financial Studies*, 33(11): 5333–5377.

Ward, Zachary. 2023. "Intergenerational Mobility in American History: Accounting for Race and Measurement Error." *American Economic Review*, 113(12): 3213–48.