Stanford cepa | Center for Education Policy Analysis

CEPA Working Paper No. 17-04A

The Global Increase in the Socioeconomic Achievement Gap, 1964-2015: Online Appendices

AUTHORS

Anna K. Chmielewski
University of Toronto

ABSTRACT

These appendices report additional details of the results from the main text of the paper, as well as supplementary analyses undertaken to test the sensitivity of results to a number of different limitations of the data. The finding of global increases in SES achievement gaps is generally robust to differences across test instruments, changes in the distribution of achievement and of SES, and changes in the measurement error of achievement and of SES. The multivariate findings predicting changing country achievement gaps from changing country characteristics and policies are generally robust across a variety of model specifications.

VERSION

May 2017

The Global Increase in Socioeconomic Achievement Gaps, 1964-2015

Anna K. Chmielewski

May 22, 2017

Online Appendices

These appendices report additional details of the results from the main text of the paper, as well as supplementary analyses undertaken to test the sensitivity of results to a number of different limitations of the data. The finding of global increases in SES achievement gaps is generally robust to differences across test instruments, changes in the distribution of achievement and of SES, and changes in the measurement error of achievement and of SES. The multivariate findings predicting changing country achievement gaps from changing country characteristics and policies are generally robust across a variety of model specifications.

The contents of the appendices are as follows:

L. Specification of trend model

M. Specification of multivariate model

A1. Table A1. List of Included Countries and Datasets

code	country	region	fims1964	fiss1970gr4	fiss1970gr8	fircs1970gr4	fircs1970gr8	sims1980	siss1984gr4	siss1984gr8	rls1991gr4	rls1991gr8	timss1995gr4	timss2003gr4	timss2007gr4	timss2011gr4	timss2015gr4	timss1995gr8	timss1999	timss2003gr8	timss2007gr8	timss2011gr8	timss2015gr8	pirls2001	pirls2006	pirls2011	pisa2000	pisa2003	pisa2006	pisa2009	pisa2012	pisa2015
ALB	Albania	4																									x			x	(x)	(x)
DZA	Algeria	3													x						x											x
ARG	Argentina	5																						x			x		x	x	x	
ARM	Armenia	4												x	(x)	x				x	x	x										
AUS	Australia	6	x		x				x	x			x	x	x	x	x	x	x	x	x	x	x			x	x	x	x	x	x	x
AUT	Austria	6											x		x	x		x							x	x	x	x	x	x	x	x
AZE	Azerbaijan	4														x										x			x	x		
BHR	Bahrain	3														(x)	x			x	x	x	x									
BEL	Belgium	6	x																													
BFL	Belgium-Flemish	6		x	x	x	x	x						x		x	x	x	x	x					x		x	x	x	x	x	x
BFR	Belgium-French	6		x	x	x	x	x			x	x						x							x	x	x	x	x	x	x	x
BLZ	Belize	5																						x								
BIH	Bosnia & Herzegovina	4																			x											
BWA	Botswana	1										x				x				x	x	x	x			x						
BRA	Brazil	5																									x	x	x	x	x	x
BGR	Bulgaria	4															x	(m)	x	x	x			x	x	x	x		x	x	x	x
CAN	Canada	6							(m)	(m)			x		x		x	x	x		x		x		x	x	x	x	x	x	x	x
CHL	Chile	5		(x)	x	x	x									x	x		x	x		x	x				x		x	x	x	x
CHN	China	2								x																						
TWN	Chinese Taipei	2												x	x	x	x		x	x	x	x	x		x	x			x	x	x	x
COL	Colombia	5													x			x			x			x		x			x	x	x	x
CRI	Costa Rica	5																												x	x	x
HRV	Croatia	4														x	x									x			x	x	x	x
CYP	Cyprus	6									(x)	x	x	x			x	x	x	x	x			x								
CZE	Czech Rep.	4											x		x	x	x	x	x		x			x		x	x	x	x	x	x	x
DNK	Denmark	6									x	x			x	x	x	x							x	x	x	x	x	x	x	x
DOM	Dominican Republic	5																														x
EGY	Egypt	3																		x	x		x									
SLV	El Salvador	5													x						x											
ENG	England	6	x	x	x	x	x	x	x	x			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
EST	Estonia	4																		x									x	x	x	x
FIN	Finland	6	x	x	x	x	x	x	x	x	x	x				x	x		x			x				x	x	x	x	x	x	x
FRA	France	6	x					x			x	x					x	x						x	x	x	x	x	x	x	x	x
GEO	Georgia	4													x	x	x				x	x	x		x	x				x		x
DEU	Germany	6													x	x	x	x						x	x	x	x	x	x	x	x	x
GDR	Germany-East	6									x	x																				
FRG	Germany-West	6	x	x	x						x	x																				
GHA	Ghana	1								x										x	x	x										
GRC	Greece	6									x	x	x					x						x			x	x	x	x	x	x
HND	Honduras	5														x						x				x						
HKG	Hong Kong	2						x	x	x	(x)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
HUN	Hungary	4		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
ISL	Iceland	6									x	x	x					x						x	x		x	x	x	x	x	x
IND	India	2		x	x	x	x																									
IDN	Indonesia	2									x		(m)				x		x	x	x	x			x	x	x	x	x	x	x	x
IRN	Iran	3		x	x	x	x						x	x	x	x	x	x	x	x	x	x	x	x	x	x						
IRL	Ireland	6									x	x	x			x	x	x					x			x	x	x	x	x	x	x
ISR	Israel	6	x			x	x	x	x	(m)			x					x	x	x	x	x	x	x	x	x	x		x	x	x	x
ITA	Italy	6		x	x	x	x		x	x	(x)	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
JPN	Japan	2	x	x	x			x	x	x			(m)	x	x	x	x	x	(m)	x	x	x	x				x	x	x	x	x	x
JOR	Jordan	3																	x	x	x	x	x						x	x	x	x
KAZ	Kazakhstan	4													x	x	x					x	x							x	x	
KOR	Korea, Rep.	2							x	x			x			x	x	x	x	x	x	x	x				x	x	x	x	x	x
KSV	Kosovo	4																														x
KWT	Kuwait	3											x		x	x	x	x			x		x	x	x	x						
KGZ	Kyrgyzstan	4																											x	x		
LVA	Latvia	4											x	x	x			x	x	x				x	x		x	x	x	x	x	x

Notes: Shaded countries were excluded because they participated in only one test. (m) denotes missing SES data; (x) denotes that gaps could not be computed, usually because of low-quality SES data. Regions include: (1) sub-Saharan Africa, (2) east and southeast Asian and Pacific countries, (3) Middle Eastern and North African countries, (4) Eastern Europe and the Commonwealth of Independent States, (5) Latin America and the Caribbean, and (6) Western countries (Western Europe and Anglophone countries).

Table A1 (cont.)

code	country	region	fims1964	fiss1970gr4	fiss1970gr8	fircs1970gr4	fircs1970gr8	sims1980	siss1984gr4	siss1984gr8	rls1991gr4	rls1991gr8	timss1995gr4	timss2003gr4	timss2007gr4	timss2011gr4	timss2015gr4	timss1995gr8	timss1999	timss2003gr8	timss2007gr8	timss2011gr8	timss2015gr8	pirls2001	pirls2006	pirls2011	pisa2000	pisa2003	pisa2006	pisa2009	pisa2012	pisa2015
LBN	Lebanon	3																		x	x	x	x									x
LIE	Liechtenstein	6																									x	x	x	x	x	
LTU	Lithuania	4												x	x	x	x	x	x	x	x	x	x	x	x	x	(m)		x	x	x	x
LUX	Luxembourg	6						x																	x		x	x	x	x	x	x
MAC	Macao-China	2																										x	x	x	x	x
MKD	Macedonia	4																	x	x		x		x	x		x					x
MYS	Malaysia	2																	x	x	x	x	x							x	x	
MLT	Malta	6														x					x		x			x				x		x
MUS	Mauritius	1																												x		
MEX	Mexico	5											(m)														x	x	x	x	x	x
MDA	Moldova	4												x					x	x				x	x					x		x
MNG	Mongolia	2													x						x											
MNE	Montenegro	4																											x	x	x	x
MAR	Morocco	3												x	x	x	x		x	x	x	x	x	(x)	x	x						
NLD	Netherlands	6	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x				x	x	x	x	x	x	x	x	x
NZL	New Zealand	6			x		x	x			x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x
NGA	Nigeria	1						(x)	x	x		x																				
NIR	Northern Ireland	6														x	x									x						
NOR	Norway	6							x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x
OMN	Oman	3														x	x				x	x	x			x						
PAN	Panama	5																												x		
PNG	Papua New Guinea	2								x																						
PER	Peru	5																									x			x	x	x
PHL	Philippines	2							x	x		x		x				x	x	x												
POL	Poland	4							x	x						x	x								x	x	x	x	x	x	x	x
PRT	Portugal	6									x	x	x			x	x	x								x	x	x	x	x	x	x
QAT	Qatar	3													x	x	x				x	x	x		x	x			x	x	x	x
ROM	Romania	4														x		x	x	x	x	x		x	x	x	x		x	x	x	x
RUS	Russian Fed.	4												x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
SAU	Saudi Arabia	3														x	x			x	x	x	x			x						
SCO	Scotland	6	x	x	x	x	x	x					x	x	x			x		x	x			x	x		x	x	x	x	x	x
SRB	Serbia	4														x	x			x	x							x	x	x	x	
QCN	Shanghai-China	2																												x	x	
SGP	Singapore	2							x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				x	x	x
SVK	Slovak Rep.	4													x	x	x	x	x	x				x	x	x		x	x	x	x	x
SVN	Slovenia	4									x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			x	x	x	x
ZAF	South Africa	1																x	x	x		x	x		x	x						
ESP	Spain	6									x	x				x	x	x							x	x	x	x	x	x	x	x
SWZ	Swaziland	1						x																								
SWE	Sweden	6	(m)	x	x	x	x	x	x	x	x	x			x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x	x
CHE	Switzerland	6									x	x						x									x	x	x	x	x	x
SYR	Syria	3																		x	x	x										
THA	Thailand	2		x	x			x		x		x	x			x		x	x		x	x	x				x	x	x	x	x	x
TTO	Trinidad and Tobago	5									x	x													x	x				x		x
TUN	Tunisia	3												x	x	x			x	x	x	x						x	x	x	x	x
TUR	Turkey	6														x	x		x		x	x	x	x				x	x	x	x	x
UKR	Ukraine	4													x						x	x										
ARE	United Arab Emirates	3														x	x					x	x			x				x	x	x
USA	United States	6	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	(x)	x	x	x	x	x	x	x
URY	Uruguay	5																										x	x	x	x	x
VEN	Venezuela	5									x	x																				
VNM	Vietnam	2																												(x)	x	x
PSE	West Bank and Gaza	3																		x	x	x										
YEM	Yemen	3												(x)	x	x																
ZWE	Zimbabwe	1								x		x																				

Notes: Shaded countries were excluded because they participated in only one test. (m) denotes missing SES data; (x) denotes that gaps could not be computed, usually because of low-quality SES data. Regions include: (1) sub-Saharan Africa, (2) east and southeast Asian and Pacific countries, (3) Middle Eastern and North African countries, (4) Eastern Europe and the Commonwealth of Independent States, (5) Latin America and the Caribbean, and (6) Western countries (Western Europe and Anglophone countries).

B. Combining different test instruments

In order to estimate international trends in SES achievement gaps over a 50-year period, this paper combines data from a variety of different international assessments of math, science, and reading. However, among tests of the same subject, a comparison of skills frameworks from the official study reports reveals differences. For example, the IEA math and science tests—FIMS, FISS, SIMS, SISS, and TIMSS—are curriculum-based, while the PISA math and science tests, as well as all reading tests—PISA, FIRCS, RLS, and PIRLS—are literacy-based. Though the early IEA tests contained anchor items to enable studying trends in achievement, the scores were not placed on common scales, and they did not have the advantage of improvements to testing methodology in the 1990s; thus, early and recent IEA tests are not strictly comparable. The analyses in the main text of the paper deal with this issue by standardizing achievement within each study and each country and assuming only that each test is interval-scaled and that different tests rank students similarly.

However, there are six recent studies that repeat the same test instrument to enable measuring trends over time: TIMSS 4^{th} and 8^{th} grade math and science; PIRLS; and PISA reading, math, and science. These trend studies allow us to investigate the sensitivity of gap trend results to differences across test instruments—but only over the recent 9-20 years that the studies have been conducted. TIMSS trends (for both grades and subjects) can be estimated from 1995 to 2015; PIRLS trends from 2001 to 2011; PISA reading trends from 2000 to 2015; PISA math trends from 2003 to 2015; and PISA science trends from 2006 to 2015. In addition, because each instrument remains the same over time, it is not necessary to standardize achievement within studies or countries, meaning we can examine changes in SES achievement gaps in light of possible changes in the variance of skills (which will be addressed in Appendix C).

Table B1 computes trends in SES achievement gaps separately for each test instrument, as well as for each SES variable and each reporter, where applicable (student or parent). Each cell in the table is the coefficient for cohort birth year from a separate hierarchical growth model with no controls (as the subject and age at testing dummies from the models in the main text do not vary across country-years within each test instrument). The first three columns report trends in gaps estimated without standardizing achievement within each country-year. The last three columns report gaps estimated when achievement is standardized within each county-year, as in the main text of the paper.[1,2] It can be seen from the reported country sample sizes that the number of participating countries varies widely across the different test instruments. It is not possible to reliably estimate trends across all test instruments for a core group of countries that has participated in every test, as there are too few countries that have done so. Thus, the reported trends for each test instrument should be interpreted only as a rough indication of the sensitivity of the general finding of increasing SES achievement gaps over time. The size of coefficients can be compared across different test instruments only for the models using standardized achievement, not for those using unstandardized achievement, as they are in different metrics and 1 point in PISA, for example, is not the same as 1 point in PIRLS or TIMSS. Significance levels should be interpreted with caution because of changing sample sizes and the large number of significance tests conducted; significance is reported only as a general indication of the strength of association. Overall, the estimated gap trends are positive for nearly every test

[1] The Standardized Achievement models adjust each gap estimate for the reliabilities of test instruments and SES report; the Unstandardized Achievement models do not adjust for either type of reliability. It is not necessary to adjust for test reliability in these models as gap estimates are not attenuated since they have not been divided by the test score variance. The Unstandardized Achievement gaps have been left unadjusted for SES reliability as well because estimated trends are reported separately for student-reported and parent-reported data.

[2] When computing these gaps, rather than using all available categories of each SES variable as in the main text of the paper, each SES variable was recoded to ensure that the SES instrument remained the same across test years. There were six categories for parent education and parent occupation and five categories for household books.

instrument, with one notable exception: Gap trends for all three PISA subject tests based on student-reported parent education are negative. Further analysis shows that this pattern may be due to problems in the measurement of parent education in the PISA student survey (see Appendix H). In contrast, nearly all other PISA gaps show increasing trends: those for parent-reported parent education, those for student- and parent-reported parent occupation, and those for student-reported household books (parent-reported household books are unavailable). Gaps are also consistently increasing for all three SES variables in PIRLS and for both available SES variables in both subjects of TIMSS at both the 4th and the 8th grades. Therefore, positive increases in SES achievement gaps over time are quite robust across the different test instruments that are combined in the main text of the paper.

Table B1. Coefficients for Cohort Birth Year from Models Run Separately by Test Instrument

				Unstandardized Achievement						Standardized Achievement					
Study	Subject	Years	SES report	Education	(N)	Occupation	(N)	Books	(N)	Education	(N)	Occupation	(N)	Books	(N)
PISA	Math	2006-2012	Parent	1.514 +	(7)	1.395 *	(7)			-0.004	(7)	0.012	(7)		
PISA	Math	2003-2015	Student	-0.115	(41)	-0.128	(41)	0.177	(41)	0.003	(41)	0.003 +	(41)	0.010 ***	(41)
PISA	Reading	2006-2012	Parent	0.805	(7)	0.362	(6)			0.007	(7)	0.013 +	(6)		
PISA	Reading	2000-2015	Student	-0.311 +	(41)	0.109	(42)	0.561 *	(42)	-0.003	(41)	0.003	(42)	0.009 **	(42)
PISA	Science	2006-2012	Parent	-0.345	(7)	0.203	(6)			-0.009 *	(7)	0.007	(6)		
PISA	Science	2006-2015	Student	-0.810 **	(54)	0.035	(54)	-0.034	(54)	-0.008 *	(54)	0.002	(54)	0.000	(54)
PIRLS	Reading	2001-2011	Parent	0.180	(19)	0.337	(20)	0.379	(21)	0.008 *	(19)	0.007 +	(20)	0.009 *	(21)
PIRLS	Reading	2001-2011	Student					0.664 *	(23)					0.017 **	(23)
TIMSS Grade 4	Math	1995-2015	Student					0.563 +	(18)					0.010 *	(18)
TIMSS Grade 4	Science	1995-2015	Student					0.429	(18)					0.011 *	(18)
TIMSS Grade 8	Math	1995-2015	Student	0.862 **	(18)			1.138 ***	(21)	0.013 ***	(18)			0.019 ***	(21)
TIMSS Grade 8	Science	1995-2015	Student	0.812 ***	(19)			1.311 ***	(21)	0.012 ***	(19)			0.021 ***	(21)

+ p<.1, * p<.05, ** p<.01, *** p < .001
Note: N refers to sample size of countries.

C. Changing distribution of achievement

Table B1 in the previous section reports trends estimated using both unstandardized and standardized achievement in order to address a further concern: that the variance of student achievement may be changing over time. Changes in achievement variance are not captured in models where achievement is standardized within each country-year, including the models using standardized achievement in Table B1 and results in the main text of the paper. As a result, estimated trends in standardized achievement measure changes in the relative *strength* of the SES-achievement association (e.g., the correlation) rather than the absolute *size* of the association. This decision was necessary as the models in the main text combine achievement from different studies. However, it is not immediately clear whether we should prefer to know the strength or the absolute size of the SES-achievement association. On the one hand, the strength may be preferred because it is not confounded with changes in the variance of student achievement (all else equal, if the variance of an outcome variable—achievement—increases, then its unstandardized association with an independent variable—SES—will increase; if the variance of the outcome decreases, then its unstandardized association with the independent variable will decrease). On the other hand, the absolute size of the association may be preferred because it captures whether the results are meaningful in terms of the skills that students have.

In contrast to the main analyses that pool different studies, in separate analyses of trend studies (PISA, TIMSS, PIRLS), it is possible to estimate whether the variance of student achievement has changed over time. Figures C1-C8 display score variance at the student, school/classroom, and country levels for each year of each trend study, estimated from separate three-level hierarchical models, as follows:

$$\hat{T}_{ij} = \gamma_{00} + v_j + u_{ij} + \epsilon_{ij},$$

$$v_j \sim N(0, \tau_{000}); u_{ij} \sim N(0, \tau_{00}); \;\; \epsilon_{ij} \sim N(0, \sigma^2),$$

[C1]

where $\hat{T}_{ij}$ is the estimated test score of student *i* in school or classroom *j* in country k, τ_{000} is the between-country variance of scores, τ_{00} is the between-school variance of scores, and σ^2 is the within-school student-level variance of scores. Total student weights are applied at the student level, meaning all students are weighted in proportion to their probability of selection, and all countries are weighted in proportion to the size of their target population (i.e., more populous countries receive greater weight). Models are estimated once for each plausible value and averaged. Only countries that participated in all years of a given trend study are included. Samples of included countries vary depending on the study. After estimating the student-, school/classroom-, and country-level variances, all three are adjusted for estimated test reliability (α) for the relevant set of countries in each year, as follows:

$$\sigma^2_{True} = \alpha * \sigma^2_{Total}$$

[C2]

Figure C1

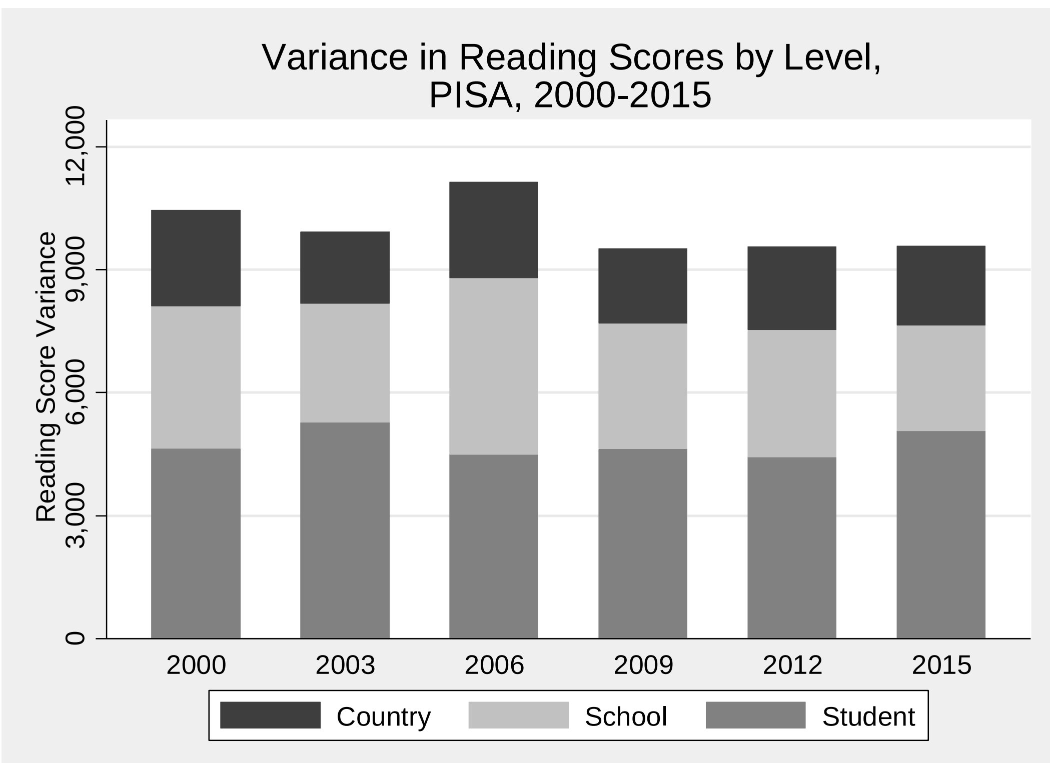


Figure C2

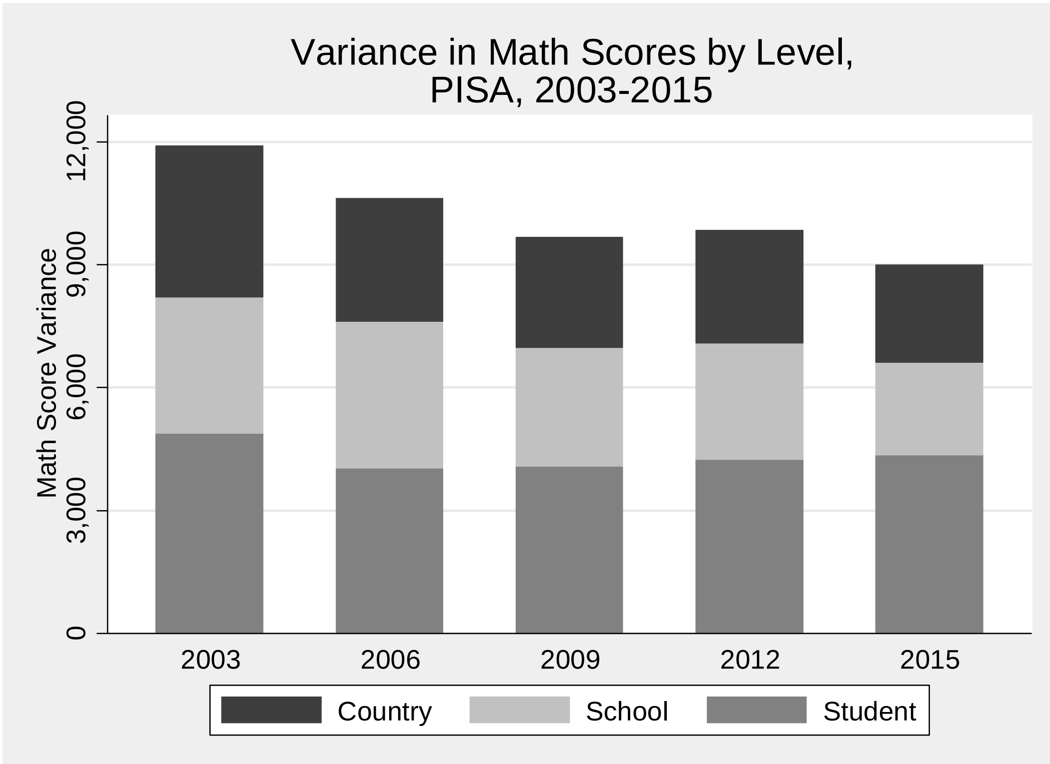


Figure C3

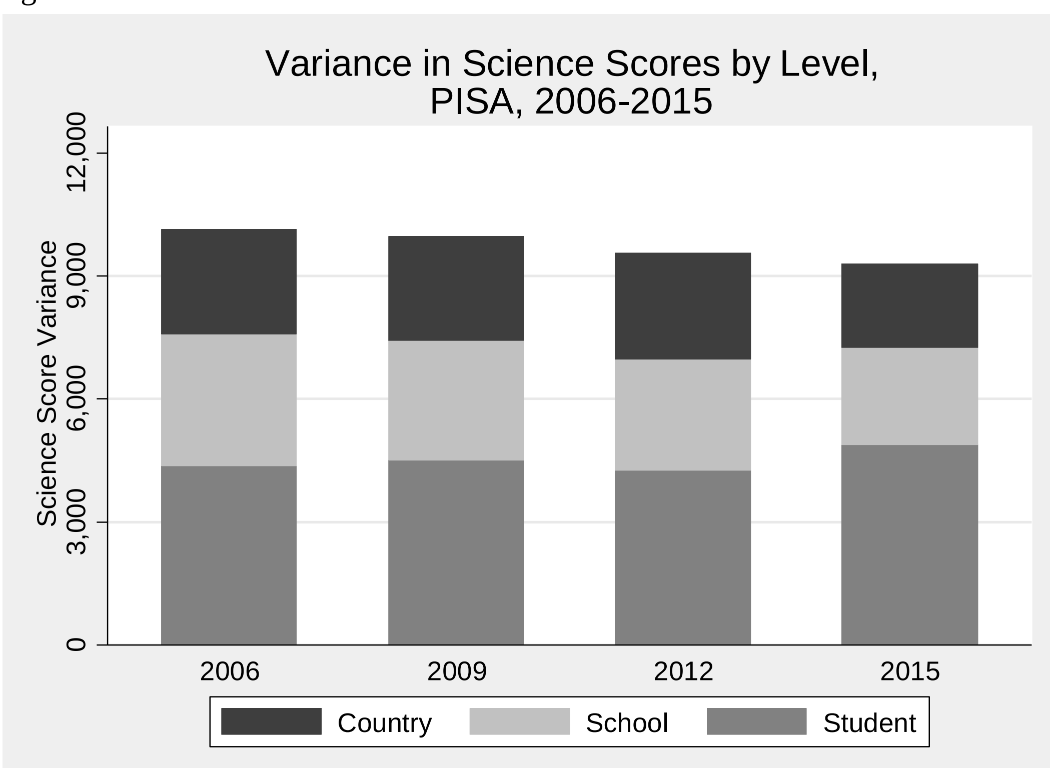


Figure C4

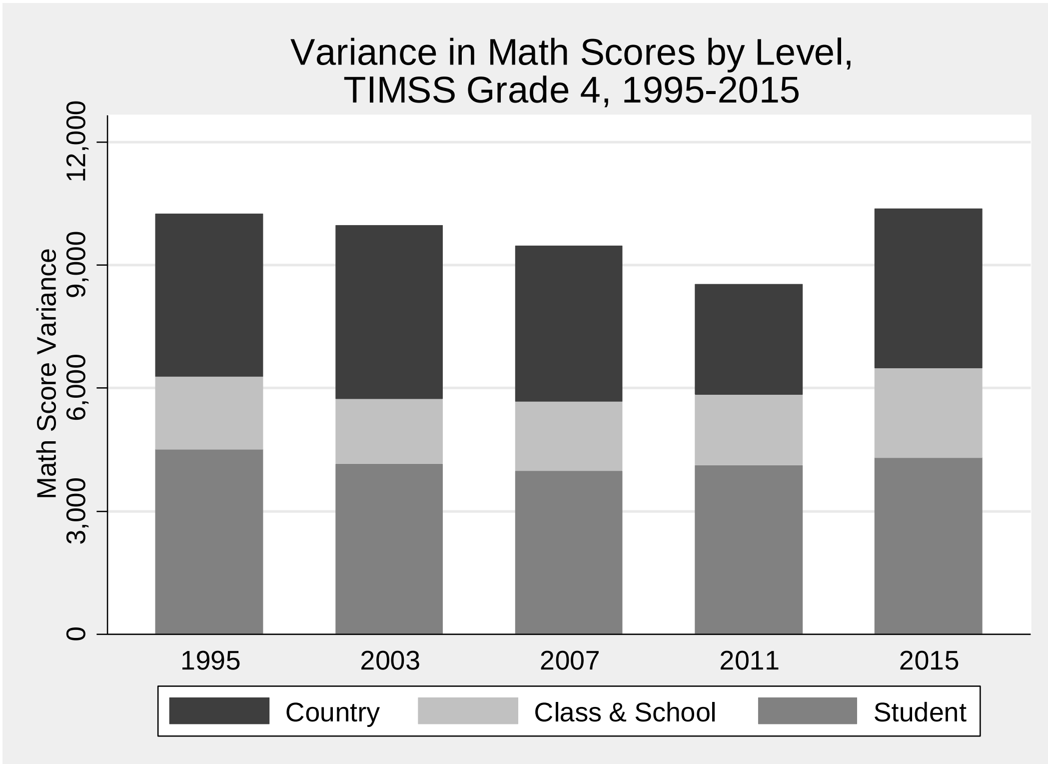


Figure C5

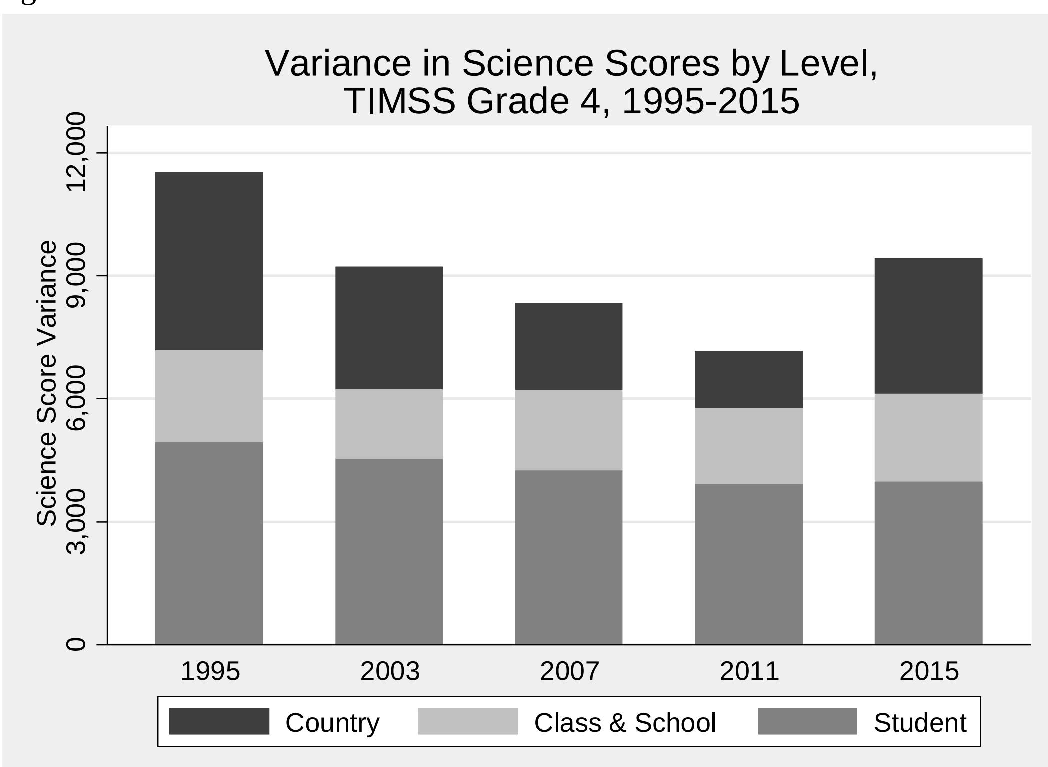


Figure C6

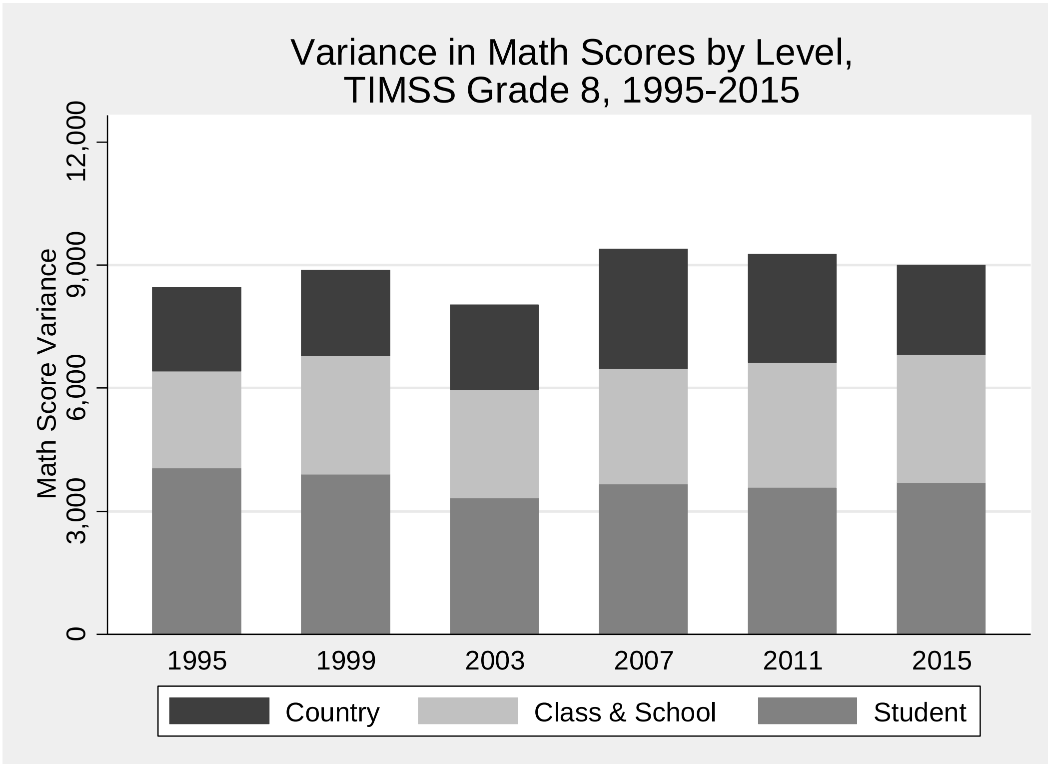


Figure C7

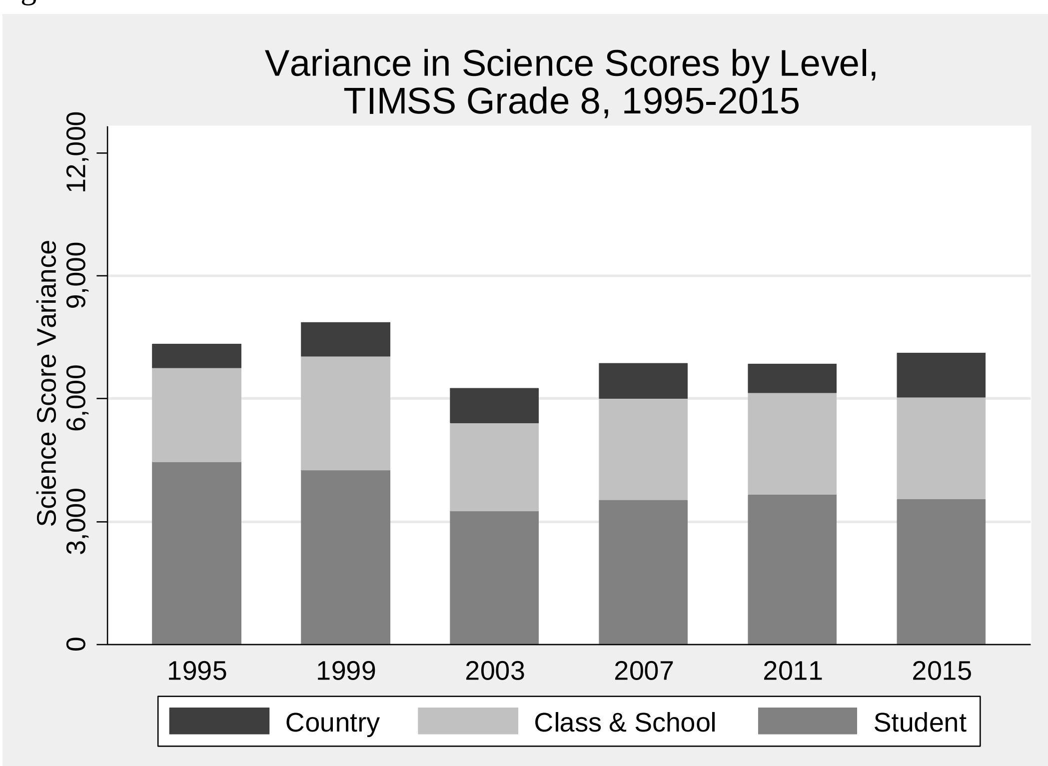
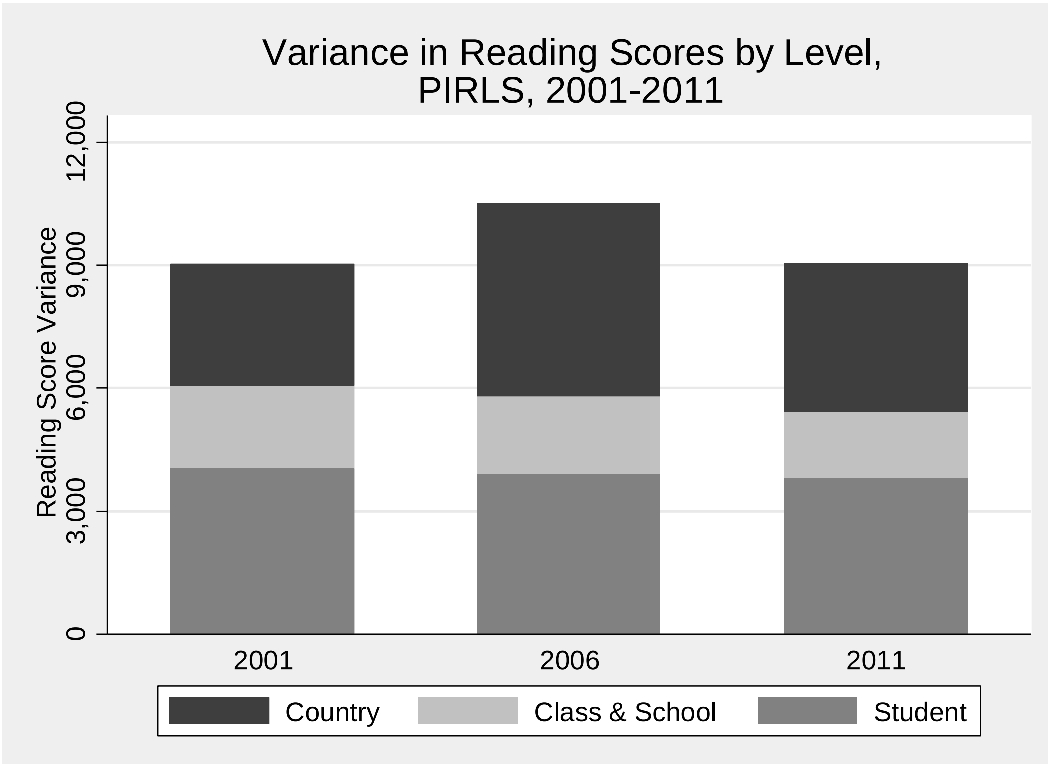


Figure C8



The hierarchal models reveal that within-country test score variance (the sum of the student- and school/classroom-level variances) has declined for all test instruments except TIMSS math (4^{th} and 8^{th} grades). These decreases in variance might lead us to question whether the absolute *size* of SES achievement gaps in terms of skills is declining even as their *strength* increases. However, the results for unstandardized achievement in Table B1 show positive trends for nearly all test instruments. This indicates that, for the specific sets of math, science, and reading skills tested by the trend studies, differences in the degree to which high- and low-SES students have mastered those skills have indeed grown over the past 9-20 years.

The trend studies also allow us to examine not only whether achievement gaps have truly grown in size but also the changing *levels* of achievement for students of different SES. In other words, we can ask: Do SES achievement gaps increase because low-SES students' achievement is declining or because it is not rising as quickly as that of high-SES students? A series of models addressing this question, separately for each test instrument, are estimated as follows:

$$\hat{A}_{ij} = \gamma_{00} + \gamma_{10}Y_{ij} + \gamma_{20}P10_{ij} + \gamma_{30}\left(Y_{ij} * P10_{ij}\right) + v_j + r_jY_{ij} + s_j(Y_{ij} * P10_{ij}) + u_{ij} + \epsilon_{ij},$$

$$v_j \sim N(0, \tau_{00}); r_j \sim N(0, \tau_{11}); s_j \sim N(0, \tau_{22}); \; u_{ij} \sim N(0, \sigma^2); \; \epsilon_{ij} \sim N\left(0, \omega_{ij}\right),$$

[C3]

where $\hat{A}_{ij}$ is the estimated mean achievement of the 90th or the 10th SES percentile in country j in country-year-SES level i, γ_{10} is the coefficient for cohort birth year Y_{ij}, γ_{20} is the coefficient for the dummy variable $P10_{ij}$ indicating whether the mean achievement of country-year-SES level i was estimated for the 10th (1) or the 90th (0) SES percentile, γ_{30} is the coefficient for the interaction $Y_{ij} * P10_{ij}$ between cohort birth year and the 10th percentile dummy, τ_{00} is the between-country variance of the true skills means, τ_{11} is the between-country variance of true slopes of cohort birth year, τ_{22} is the between-country variance of interactions between cohort birth year and the 10th percentile dummy, σ^2 is the true within-country variance of the gaps, and $\omega_{ij} = \left[s.e.\left(\hat{G}_{ij}\right)\right]^2$ is the sampling variance of $\hat{G}_{ij}$.

Table C1 reports selected coefficients from these models for each test instrument. Each vertical pair of cells reports coefficients for cohort birth year (the trend for the reference category, the 90th SES percentile) and the interaction between cohort birth year and the 10th percentile dummy.[3] Adding the coefficients allows one to compute the trend in achievement level for 10th percentile students. Overall, the achievement of high-SES students has increased in most of the trend studies, with the exceptions of the PISA math and science tests (when the SES

[3] Note that if the models in Tables B1 and C1 were simple OLS regression models, the coefficients for the Cohort birth year*p10 interactions in Table C1 would be equal to the Cohort birth year trends in the gaps in Table B1, but opposite in sign. This is not true in the reported models due to the precision weighting and random cohort slopes in the hierarchical growth models. However, estimated coefficients are generally similar in size and opposite in sign across Tables B1 and C1.

variables are student-reported only). In the primary school studies (PIRLS and TIMSS 4th grade) and in PISA reading, the achievement of low-SES students has also increased, but generally by a smaller amount than that of high-SES students. In the secondary school math and science studies, the achievement of low-SES students has generally declined. The declining math and science achievement of low-SES secondary school students may indicate that low-SES students truly have fewer learning opportunities in the 2010s than 9-20 years prior and/or may reflect a changing population of secondary school students due to increased access in some countries.

Table C1. Models Predicting Achievement Levels for 90th and 10th SES Percentiles (Coefficients for Cohort Birth Year and Interaction between Cohort and 10th Percentile), Run Separately by Test Instrument

					Unstandardized Achievement					
Study	Subject	Years	SES report		Education	(N)	Occupation	(N)	Books	(N)
PISA	Math	2006-2012	Parent	Cohort birth year	2.725 ***	(7)	2.906 ***	(7)		
				Cohort birth year*p10	-0.729		-1.542 ***			
PISA	Math	2003-2015	Student	Cohort birth year	-0.537 *	(41)	-0.418 +	(41)	-0.403 +	(41)
				Cohort birth year*p10	0.338		0.062		-0.028	
PISA	Reading	2006-2012	Parent	Cohort birth year	1.923 **	(7)	1.658 *	(7)		
				Cohort birth year*p10	-0.413		-0.485			
PISA	Reading	2000-2015	Student	Cohort birth year	0.295	(41)	0.422 +	(42)	0.785 ***	(42)
				Cohort birth year*p10	0.333		-0.171		-0.701 **	
PISA	Science	2006-2012	Parent	Cohort birth year	1.552 ***	(7)	1.996 ***	(7)		
				Cohort birth year*p10	0.783		-0.322			
PISA	Science	2006-2015	Student	Cohort birth year	-0.416	(54)	-0.072	(54)	-0.072	(54)
				Cohort birth year*p10	0.812 ***		-0.1		-0.146	
PIRLS	Reading	2001-2011	Parent	Cohort birth year	1.449 **	(20)	1.394 **	(21)	1.348 **	(21)
				Cohort birth year*p10	-0.681 +		-0.568		-0.465	
PIRLS	Reading	2001-2011	Student	Cohort birth year					1.183 **	(24)
				Cohort birth year*p10					-0.571 +	
TIMSS Grade 4	Math	1995-2015	Student	Cohort birth year					1.278 ***	(18)
				Cohort birth year*p10					-0.451 +	
TIMSS Grade 4	Science	1995-2015	Student	Cohort birth year					0.806 *	(18)
				Cohort birth year*p10					-0.311	
TIMSS Grade 8	Math	1995-2015	Student	Cohort birth year	0.862 +	(20)			0.835 *	(21)
				Cohort birth year*p10	-0.944 ***				-1.037 ***	
TIMSS Grade 8	Science	1995-2015	Student	Cohort birth year	0.972 *	(20)			0.941 ***	(21)
				Cohort birth year*p10	-0.959 ***				-1.205 ***	

+ p<.1, * p<.05, ** p<.01, *** p < .001

Note: Each observation in the model is the estimated achievement at the 90th or 10th SES percentile. Models include a 10th percentile dummy (not shown), cohort birth year, and the interaction between cohort birth year and the 10th percentile dummy. Thus, the cohort birth year coefficient is the estimated achievement trend for 90th SES percentile students, and the interaction is the difference between the achievement trends for 90th and 10th percentile students. N refers to sample size of countries.

D. Declining measurement error of achievement

Even assuming the true variance of achievement had remained constant over time, if measurement error of achievement declines over time (e.g., because of improvements in testing methodology), the SES achievement gap estimates in the main text of the paper will artificially appear to increase because they were attenuated in early years where measurement error was higher. This is because the method of standardizing achievement in each country-year involves dividing by the standard deviation of achievement, which will be inflated due to measurement error. That SES achievement gaps are also increasing for most of the unstandardized scores reported in Appendix B is evidence that findings are robust, even when not standardizing achievement.

Tables D1-D3 report median, minimum, and maximum test reliabilities by age group for math, reading, and science tests. Median test reliabilities have not consistently increased over time for all test subjects and age groups. Reliabilities have increased for 4th grade tests and for secondary science tests, but appear to have declined somewhat for secondary math and reading tests. However, it should be kept in mind that the sample of countries participating in international assessments has become more diverse over time, and countries at a lower level of development often have lower test reliabilities.

Table D1. Median, Minimum, and Maximum Test Reliability for Math Tests

Study	Year	Median	Minimum	Maximum
Grade 4 Math				
TIMSS	1995	0.84	0.74	0.88
TIMSS	2003	0.87	0.76	0.91
TIMSS	2007	0.83	0.55	0.88
TIMSS	2011	0.82	0.57	0.89
TIMSS	2015	0.88	0.78	0.92
Grade 8 Math				
FIMS	1964	0.92	0.87	0.95
SIMS[a]	1980	0.85[a]	0.81[a]	0.85[a]
TIMSS	1995	0.89	0.73	0.92
TIMSS	1999	0.89	0.69	0.94
TIMSS	2003	0.89	0.51	0.94
TIMSS	2007	0.88	0.62	0.93
TIMSS	2011	0.87	0.66	0.94
TIMSS	2015	0.91	0.81	0.94
Age 15 Math				
PISA	2000	0.88	0.82	0.92
PISA	2003	0.90	0.83	0.93
PISA	2006	0.88	0.83	0.93
PISA	2009	0.88	0.77	0.92
PISA	2012	0.92	0.84	0.94
PISA	2015	0.85	0.67	0.89

[a] SIMS test reliability was not reported in the available documentation and was estimated for each country using a model that included age, subject, year, and countries' level of development.

Table D2. Median, Minimum, and Maximum Test Reliability for Reading Tests

Study	Year	Median	Minimum	Maximum
Grade 4 Reading				
FIRCS	1970	0.85	0.74	0.89
RLS	1991	0.93	0.89	0.97
PIRLS	2001	0.88	0.83	0.91
PIRLS	2006	0.87	0.81	0.92
PIRLS	2011	0.88	0.79	0.93
Grade 8 Reading				
FIRCS	1970	0.85	0.64	0.90
RLS	1991	0.92	0.77	0.95
Age 15 Reading				
PISA	2000	0.92	0.87	0.94
PISA	2003	0.83	0.70	0.88
PISA	2006	0.88	0.80	0.93
PISA	2009	0.92	0.86	0.94
PISA	2012	0.89	0.81	0.93
PISA	2015	0.86	0.72	0.89

Table D3. Median, Minimum, and Maximum Test Reliability for Science Tests

Study	Year	Median	Minimum	Maximum
Grade 4 Science				
FISS	1970	0.82	0.68	0.87
SISS	1984	0.74	0.70	0.79
TIMSS	1995	0.77	0.70	0.83
TIMSS	2003	0.84	0.74	0.87
TIMSS	2007	0.80	0.69	0.88
TIMSS	2011	0.78	0.62	0.85
TIMSS	2015	0.85	0.81	0.90
Grade 8 Science				
FISS	1970	0.83	0.57	0.89
SISS	1984	0.75	0.60	0.80
TIMSS	1995	0.78	0.69	0.84
TIMSS	1999	0.80	0.62	0.86
TIMSS	2003	0.84	0.63	0.91
TIMSS	2007	0.84	0.65	0.91
TIMSS	2011	0.83	0.67	0.89
TIMSS	2015	0.89	0.81	0.92
Age 15 Science				
PISA	2000	0.87	0.75	0.92
PISA	2003	0.82	0.68	0.88
PISA	2006	0.91	0.84	0.94
PISA	2009	0.89	0.79	0.93
PISA	2012	0.89	0.80	0.93
PISA	2015	0.91	0.77	0.93

For the models in the main text of the paper, which pool different tests with different scales and must standardize achievement, all SES achievement gaps (and their standard errors) are adjusted according to each country's test reliability for each study, as published in the corresponding technical reports (as well as for the estimated reliability of SES reports, which is explained in more detail in Appendix H). The adjustment is computed as follows:

$$\widehat{Gap}_{adj} = \widehat{Gap}_{raw} * \frac{1}{\sqrt{\alpha_{ach}} * \sqrt{\alpha_{SES}}}$$

[D1]

Table D4 reports estimated trends in SES achievement gaps without adjusting those gaps for differences in test reliability (and also without adjusting for the reliability of SES report, which is discussed in more detail in Appendix H). Gap trends are positive and significant, and are very similar to those reported in the models in the main text. Without adjusting for reliability, increases in parent education and parent occupation gaps are slightly larger and increases in books gaps are slightly smaller than those reported in the main text. Gaps adjusted for reliability are preferred, as they are likely more accurate and also more conservative estimates of increases in SES achievement gaps, since earlier studies tend to have lower test reliabilities (and may also have lower reliabilities of SES reports; see Appendix H).

Table D4. Estimated trends in 90/10 SES achievement gaps, without disattenuating by test or SES reliability

	Education		Occupation		Books	
	Adj.	No adj.	Adj.	No adj.	Adj.	No adj.
	(1)	(2)	(1)	(2)	(1)	(2)
Within countries						
Age 10 at testing	-0.026	0.100 ***	0.021	-0.009	-0.234 ***	-0.175 ***
	(0.031)	(0.025)	(0.020)	(0.016)	(0.027)	(0.017)
Age 15 at testing	-0.166 ***	-0.104 ***			0.085 ***	0.062 ***
	(0.022)	(0.017)			(0.025)	(0.017)
Math	0.029 **	0.015 +	0.019 *	0.013 +	-0.021 +	-0.024 **
	(0.011)	(0.009)	(0.008)	(0.007)	(0.013)	(0.009)
Science	0.023 *	-0.005	0.014 *	0.003	0.032 **	-0.003
	(0.010)	(0.008)	(0.006)	(0.005)	(0.012)	(0.009)
Cohort birth year	0.005 ***	0.005 ***	0.004 **	0.004 ***	0.008 ***	0.007 ***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Between countries						
Intercept	1.053 ***	0.774 ***	0.919 ***	0.771 ***	1.230 ***	0.809 ***
	(0.032)	(0.024)	(0.025)	(0.021)	(0.040)	(0.027)
Residual variance (within countries)	0.01287	0.00930	0.00709	0.00523	0.01821	0.00901
Residual variance (country intercepts)	0.05820	0.03355	0.04476	0.03230	0.12190	0.05829
Residual variance (cohort slopes)	0.00010	0.00006	0.00008	0.00005	0.00014	0.00007
N (observations)	1915	1915	1405	1405	2190	2190
N (countries)	94	94	82	82	100	100

+ p<.1, * p<.05, ** p<.01, *** p < .001

Aside from test reliability, measurement error in achievement could also decline if tests are administered in a more standardized way over time. Standardization of test administration could explain the reductions in between-country achievement variance reported in Appendix C above. The hierarchical models also found declining between-school and between-classroom achievement variance (for all trend studies except TIMSS 4^{th} and 8^{th} grade math), which may also be partially explained by standardization of test administration. However, achievement variance has also declined within schools and classrooms for most trend studies, and it is less clear how greater standardization of test administration could explain this change.

E. Changing distribution of SES

Educational attainment and the occupational structure have shifted dramatically in most countries since 1964. The trends reported in the main text of the paper refer to changes in the achievement gap between the 90^{th} and the 10^{th} percentiles of each SES variable in each country-year, even though the meaning of the 90^{th} and 10^{th} percentiles has changed over time. In FIMS 1964, the 90^{th} percentile of parent education was 1-2 years of postsecondary in most countries, while in the same set of countries in PISA 2015, it was a university BA or more. The 10^{th} percentile of parent education in 1964 was lower secondary school, while in 2015 it was a vocational high school diploma. The 90^{th} percentile of parent occupation was semi-professional or technical occupations in 1964 and was professional work in 2015; the 10^{th} percentile of parent occupation was semi-skilled or skilled blue-collar work in most countries in 1964 and was skilled blue-collar in 2015. The 90/10 percentile method (Reardon 2011b) was chosen to avoid changes in the selectivity of different SES categories as their frequencies changed over time. However, treating these historical and contemporary percentiles as equivalent also makes a theoretical assumption that these SES characteristics confer mainly positional advantages to children. Alternatively, it may be that having a parent with a university degree always confers the same absolute advantage, regardless of whether that parent was among the elite few who earned a degree in the mid-20th century or the larger share who earned a degree at the turn of the 21st century.

One piece of evidence that the increasing SES achievement gaps reported here are not merely an artifact of the general upgrading of SES is that increases are found not only for parental education and occupation, whose levels have increased over time, but also for household books, whose levels have *declined*. In SISS 1984, 45 percent of 8th grade students reported having more than 100 books at home, while in the same set of countries in PISA 2015, only 40 percent of students reported more than 100 books. (It should be noted that this decline does not appear to be entirely attributable to the recent popularity of electronic reading devices, as it occurred gradually over the entire period.)

In addition to changing levels of the three SES variables, the dispersion of SES has also changed. The variance of parent education and occupation has declined somewhat in most countries over time, while the variance of household books has remained relatively constant. (The changing variances of these ordinal SES variables were computed after recoding into the same categories in every study—6 categories for parent education and occupation and 5 categories for books.) All else equal, if the variance of an independent variable—SES—decreases, then its unstandardized association with an outcome variable—achievement—will increase. The models in the main text of the paper avoid this problem, as converting the SES variables into percentiles is a form of standardization. However, these changes in variance should be kept in mind for the next set of models, where SES is unstandardized.

An additional piece of evidence that increasing SES achievement gaps are not an artifact of changing SES distributions comes from the models reported in Tables E1 and E2, which compute achievement gaps between three consistently-defined categories of each SES variable rather than percentiles. These analyses examine the robustness of the finding of increasing SES achievement gaps to treating SES as an absolute rather than a positional good. Each SES variable is coded into three categories for all studies: parent education is coded into (1) less than secondary [less than ISCED 3], (2) secondary or non-degree vocational postsecondary [ISCED 3 or 4], and (3) an academically- or vocationally-oriented higher educational degree [ISCED 5A or 5B or more]. Parent occupation is coded into (1) working class [unskilled, semiskilled, or agricultural labor], (2) intermediate class [skilled trades, service, clerical, or small business], (3) salariat class [semi-professional, managerial, or professional]. Household books are coded into (1) 0-10 books, (2) 11-100 books, (3) 101 books or more. In addition to checking the robustness of the results from the 90/10 gaps models by reporting trends in the gap between the top and bottom categories of each variable (Table E1), the models also allow us to see whether gaps have increased primarily between the top and middle (the top panel of Table E2) or the middle and bottom categories of each variable (the bottom panel of Table E2). The results in the table show that gaps have increased between all three categories of all three variables. However, for all three variables, increases between the top and middle categories are substantially larger than increases between the middle and bottom categories. For all three variables, the achievement gap between the middle and bottom categories has increased by about 0.001 SD per year, a change which is positive but not significantly different from 0 for any variable. This is true both for parent education and occupation, where the top category has become a larger share of students, and for household books, where the top category has become a smaller share of students. Thus, the achievement advantage of students with college-educated or professional parents or many books at home has increased, even as the share of students with college-educated or professional parents has increased (and the share with many books has declined). These findings are consistent with Reardon's (2011b) finding for the US that 90/50 income achievement gaps grow more than 50/10 gaps. However, it should be noted that the three categories used for each SES variable here do not correspond to the 90^{th}, 50^{th}, and 10^{th} percentiles. Broadly, the results from these models indicate that SES achievement gaps still appear to increase, whether we think of SES as a positional an absolute good.

Table E1. Estimated Trends in Achievement Gaps between High and Low SES Categories

	Education	Occupation	Books
High-low			
Age 10 at testing	-0.038	0.099 ***	-0.221 ***
	(0.032)	(0.020)	(0.028)
Age 15 at testing	-0.083 **		-0.001
	(0.030)		(0.025)
Math	0.03 **	0.002	-0.02 +
	(0.010)	(0.008)	(0.012)
Science	0.032 ***	0.001	0.044 ***
	(0.009)	(0.007)	(0.013)
Cohort birth year	0.007 ***	0.002 *	0.003 +
	(0.001)	(0.001)	(0.002)
Intercept	0.976 ***	0.825 ***	1.318 ***
	(0.029)	(0.021)	(0.044)
Residual variance (within countries)	0.03481	0.00851	0.02560
Residual variance (country intercepts)	0.05793	0.02703	0.16224
Residual variance (cohort slopes)	0.00009	0.00007	0.00015
N (observations)	1889	1334	2086
N (countries)	93	80	95

+ p<.1, * p<.05, ** p<.01, *** p < .001. Note: See text for definitions of high, middle, and low for each SES variable.

Table E2. Estimated Trends in Achievement Gaps between High-Middle and Middle-Low SES Categories

	Education	Occupation	Books
High-middle			
Age 10 at testing	-0.029	-0.036 **	-0.169 ***
	(0.019)	(0.012)	(0.015)
Age 15 at testing	-0.088 ***		0.011
	(0.016)		(0.013)
Math	0.031 ***	0.012 *	0.001
	(0.006)	(0.005)	(0.008)
Science	0.019 ***	0.012 ***	0.0270 ***
	(0.005)	(0.004)	(0.007)
Cohort birth year	0.005 ***	0.002 +	0.002 *
	(0.001)	(0.001)	(0.001)
Intercept	0.496 ***	0.548 ***	0.62 ***
	(0.014)	(0.016)	(0.022)
Residual variance (within countries)	0.01243	0.00319	0.00907
Residual variance (country intercepts)	0.01691	0.01686	0.04203
Residual variance (cohort slopes)	0.00003	0.00005	0.00007
N (observations)	1889	1334	2086
N (countries)	93	80	95
Middle-low			
Age 10 at testing	-0.004	0.137 ***	-0.05 *
	(0.021)	(0.015)	(0.024)
Age 15 at testing	0.004		-0.004
	(0.023)		(0.018)
Math	-0.001	-0.009 +	-0.024 ***
	(0.007)	(0.005)	(0.007)
Science	0.012 +	-0.013 **	0.015 *
	(0.007)	(0.004)	(0.007)
Cohort birth year	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)
Intercept	0.484 ***	0.273 ***	0.698 ***
	(0.020)	(0.010)	(0.026)
Residual variance (within countries)	0.01781	0.00297	0.01407
Residual variance (country intercepts)	0.02717	0.00471	0.05037
Residual variance (cohort slopes)	0.00004	0.00004	0.00005
N (observations)	1889	1334	2086
N (countries)	93	80	95

+ p<.1, * p<.05, ** p<.01, *** p < .001. Note: See text for definitions of high, middle, and low for each SES variable.

F. Achievement gaps by mother's and father's SES characteristics

One possible explanation for increasing SES achievement gaps is the increasing educational attainment and occupational status of mothers. If children's achievement tends to be more strongly associated with their mother's than with their father's education and occupation (because mothers perform the majority of childcare), then the dramatic rise in women's status since the mid-20th century could result in greater variability in the SES characteristics of mothers, and therefore increasing SES achievement gaps. The dramatic global increase in the educational attainment and occupational status of women is easily visible in the international assessment data. In early international assessments, most students reported a higher level of educational attainment for their father than their mother. The mother's occupation was not collected at all in the earliest three datasets (FIMS 1964, FISS 1970, and FIRCS 1970). In the most recent assessments, mothers and fathers are about equally educated, and mothers have somewhat higher occupational status than fathers (reflecting women's greater likelihood of working in white collar jobs).

Since the main analyses in study use only the *highest* of the two parents' education and occupation as a measure of the child's SES, the increasing education and occupational status of mothers means that the highest parent education and occupation are increasingly likely to come from the mother. This could explain why SES achievement gaps are increasing, if mothers' SES characteristics are more strongly associated with children's achievement. (Children and/or parents were asked to report the education and occupation of both parents, whether or not both were present in the home, and most international assessments did not collect data on whether each parent lived in the home. For parents not currently working, the most recent occupation was reported. For parents who had never worked—who were very likely to be mothers performing home duties, particularly in earlier years—this study treats that parent's occupation as missing and imputes an occupation as part of the multiple imputation model described in the Methods section of the main text of the paper.)[4]

To check the robustness of the main results to changes in the relative status of mothers and fathers, Table F1 reports trends in SES achievement gaps based on fathers' and mothers' education and occupation separately. The results show that trends in gaps based on both father's and mother's education are positive and significant, while the trend in the gap based on mother's occupation is positive but not significant, and the trend in the gap based on father's occupation is slightly negative and not significant. As expected, increases in gaps based on mothers' education and occupation are larger than those based on fathers' characteristics. Between the 1950 and the 2005 birth cohorts, the father's education achievement gap grew from about 0.87 SDs to 1.09 SDs (about a 25% increase), while the mother's education achievement gap grew from 0.78 SDs to 1.11 SDs (about a 43% increase). Between the 1966 and 2005 birth cohorts (1966 is the birth cohort corresponding to SIMS 1980, the first cohort for which both mothers' and fathers' occupation are available), the father's occupation achievement gap declined slightly from 0.87 to 0.86 SDs (a 2% decrease), and the mother's occupation achievement gap grew from 0.83 to 0.91 SDs (a 9% increase). That achievement gaps have increased not only for mothers' but also fathers' education suggests that the global increase in parent education achievement gaps is not fully explained by the increasing educational attainment of mothers. In contrast, the stable father's occupation achievement gap suggests that the global increase in parent occupation achievement gaps may indeed be explained by the increasing occupational status of mothers.

[4] See Appendix L for models computing gap trends with listwise deletion of missing data rather than multiple imputation. In these models, for students with one missing and one nonmissing parent education or occupation, the nonmissing value was used as the "highest" parent education or occupation. Results are very similar to those from the models with imputed data.

Table F1. Trends in Achievement Gaps Based on Fathers' and Mothers' SES Characteristics

	Education		Occupation	
	Father	Mother	Father	Mother
Within countries				
Age 10 at testing	0.007	0.031	0.028	-0.018
	(0.028)	(0.031)	(0.018)	(0.022)
Age 15 at testing	-0.172 ***	-0.150 ***		
	(0.023)	(0.026)		
Math	0.029 **	0.031 **	0.030 ***	0.027 ***
	(0.011)	(0.012)	(0.008)	(0.008)
Science	0.019 *	0.028 **	0.010 +	0.020 ***
	(0.009)	(0.010)	(0.005)	(0.006)
Cohort birth year	0.004 *	0.006 ***	-0.0003	0.002
	(0.002)	(0.002)	(0.002)	(0.001)
Between countries				
Intercept	1.025 ***	1.010 ***	0.862 ***	0.876 ***
	(0.029)	(0.032)	(0.021)	(0.026)
Residual variance (within countries)	0.01036	0.01360	0.00036	0.00205
Residual variance (country intercepts)	0.05740	0.06020	0.03418	0.04851
Residual variance (cohort slopes)	0.00011	0.00011	0.00013	0.00009
N (observations)	1902	1877	1337	1333
N (countries)	95	94	82	82

+ p<.1, * p<.05, ** p<.01, *** p < .001

Another possible explanation for increasing SES achievement gaps is increasing homogamy among the parents of participating students. That is, the SES characteristics (education and occupation) of mothers and fathers are likely growing more correlated over time. Students with two highly-educated or high-occupational-status parents may be more advantaged than students with only one highly-educated or high-occupational-status parent. Table F2 reports hierarchical growth models estimating trends in the correlation between mothers' and fathers' education and occupation. The results show that the average correlation between mothers' and fathers' education has increased only moderately (from about 0.55 in the 1950 birth cohort to about 0.61 in the 2005 cohort), and the average correlation between mothers' and fathers' occupation has remained relatively constant (declining slightly from about 0.40 in the 1966 birth cohort to about 0.39 in the 2005 cohort). These results demonstrate a less pronounced increase in homogamy than expected, which may be because this analysis treats parent education and occupation as continuous positional goods, converted into percentiles within each country-year, as in the main text of this paper. While it is true that an increasing number of children have two parents with higher education degrees or professional occupations, the associations between the *relative* positions of mothers and fathers within their own gender distributions have not increased dramatically over time.

Table F2. Trends in Correlations between Mothers' and Fathers' Education and Occupation

	Education	Occupation
Within countries		
Age 10 at testing	-0.025 ***	0.090 ***
	(0.007)	(0.008)
Age 15 at testing	-0.054 ***	
	(0.006)	
Cohort birth year	0.0010 *	-0.0003
	(0.0003)	(0.0010)
Between countries		
Intercept	0.589 ***	0.392 ***
	(0.006)	(0.009)
Residual variance (within countries)	0.00234	0.00364
Residual variance (country intercepts)	0.00341	0.00524
Residual variance (cohort slopes)	0.00000	0.00001
N (observations)	865	549
N (countries)	95	82

+ p<.1, * p<.05, ** p<.01, *** p < .001

G. Achievement gaps conditional on other SES variables

Results in the main paper text model trends in achievement gaps separately for each SES variable (parent education, parent occupation, and household books) rather than constructing an SES index to avoid loss of information because not all SES variables are available in every dataset. However, it could be the case that only one of the three SES variables is growing more strongly associated with achievement over time, while the other two SES variables only appear to be growing more strongly associated with achievement due to their correlations with this one most salient SES variable. Or it may even be the case that the independent associations between each SES variable and achievement have remained constant over time, but correlations between the three SES variables are growing stronger over time. This would create the appearance of increasing SES achievement gaps for all three variables because an increasing share of students would experience “double-” or “triple disadvantage.” That is, students with university-educated parents would be more likely also to have parents with professional occupations and to have a large number of books at home. Conversely, students whose parents have not completed secondary education would be more likely to have parents with working-class occupations and very few books at home. Thus, there may be a pattern of increasing polarization of socioeconomic advantage and disadvantage among schoolchildren, which may completely explain away increasing SES achievement gaps for all three SES variables.

Table G1 reports hierarchical growth models estimating trends in the correlation between pairs of SES variables. The results show that the average correlation between parent education and parent occupation has increased quite substantially (from about 0.40 in the 1950 birth cohort to about 0.51 in the 2005 cohort). This finding is consistent with international research showing an increasing association between education and occupation across most countries (Kreidl, Ganzeboom and Treiman 2014).[5] In contrast, the average correlations between household books

[5] Note the literature shows an increasing association between education and occupation only when both variables are treated as linear (as here); the true pattern of change may be more complex. For example, research on over-education shows a declining relationship between attainment of tertiary education and a professional occupation when treating both variables as categorical rather than linear—that is, assuming education and occupational status are absolute rather than positional goods (the opposite of the assumption made in this paper).

and each of parent education and occupation, respectively, have increased more moderately (the correlation between household books and parent education increased from about 0.31 in the 1956 birth cohort to about 0.35 in the 2005 cohort; the correlation between household books and parent occupation increased from about 0.27 in the 1956 birth cohort to about 0.32 in the 2005 birth cohort).

Table G1. Trends in Correlations between Three SES Variables

	Education & Occupation	Education & Books	Occupation & Books
Within countries			
Age 10 at testing	0.105 ***	0.064 ***	0.046 ***
	(0.011)	(0.007)	(0.008)
Age 15 at testing		-0.03 ***	
		(0.005)	
Cohort birth year	0.0020 ***	0.0010 +	0.0010 **
	(0.0005)	(0.0004)	(0.0004)
Between countries			
Intercept	0.48 ***	0.338 ***	0.307 ***
	(0.009)	(0.007)	(0.008)
Residual variance (within countries)	0.00539	0.00195	0.00194
Residual variance (country intercepts)	0.00431	0.00338	0.00523
Residual variance (cohort slopes)	0.00001	0.00001	0.00001
N (observations)	575	834	566
N (countries)	82	95	83

+ p<.1, * p<.05, ** p<.01, *** p < .001

But do these increasing correlations fully explain increasing SES achievement gaps for all three variables? And even if increasing correlations do not fully explain increasing gaps, could it be the case that only one or two SES variables are growing more strongly associated with achievement, while the other SES variable(s) only appear to be growing more strongly associated with achievement due to their correlation(s) with the most salient SES variable(s)? One way to address both of these questions is by computing SES achievement gaps for each variable conditional on one or both of the other SES variables. The results of these models are presented in Table G2. Conditional gaps can be estimated only from studies that collected more than one SES variable, meaning sample sizes are reduced. To obtain an accurate comparison to cohort trends in unconditional SES achievement gaps, trends in unconditional gaps are also estimated using the same reduced set of studies. In Table G2, each row reports two models estimating gaps based on a particular SES variable. The "Conditional gaps" columns report the intercept and cohort birth year coefficients for a model whose gaps are conditional on one or both of the other SES variables. The "Unconditional gaps" columns report the coefficients for a model based on unconditional gaps using the same sample of studies. In each model, the intercept is the estimated gap for the variable in question (conditional or unconditional) for the 1989 birth cohort, while the cohort trend is the estimated annual change in the gap. All models control for subject and age at testing. Since there is no established method to adjust conditional associations for attenuation due to measurement error, conditional gaps are not adjusted for test or SES reliability. For an accurate comparison, the unconditional gaps are also not adjusted for reliability.

Table G2. Intercept and Cohort Birth Year Coefficients from Models Predicting 90/10 SES Achievement Gaps Conditional on Other SES Variables

		Conditional gaps			Unconditional gaps		
Variable	Conditional on	Intercept	Cohort trend	(N)	Intercept	Cohort trend	(N)
Education	Occupation	0.344 ***	-0.001	(81)	0.688 ***	0.003 *	(81)
Education	Books	0.383 ***	0.002 +	(95)	0.735 ***	0.005 ***	(95)
Education	Occupation & Books	0.200 ***	-0.002	(81)	0.691 ***	0.001	(81)
Occupation	Education	0.541 ***	0.005 ***	(81)	0.763 ***	0.006 ***	(81)
Occupation	Books	0.484 ***	0.002 +	(82)	0.765 ***	0.004 ***	(82)
Occupation	Education & Books	0.422 ***	0.001	(81)	0.777 ***	0.002 +	(81)
Books	Education	0.488 ***	0.005 ***	(95)	0.812 ***	0.007 ***	(95)
Books	Occupation	0.674 ***	0.005 ***	(82)	0.912 ***	0.007 ***	(82)
Books	Education & Occupation	0.636 ***	0.007 ***	(81)	0.929 ***	0.007 ***	(81)

+ p<.1, * p<.05, ** p<.01, *** p < .001

By computing the predictions of the model with all three SES variables (rows 3, 6, and 9) for the 1956 birth cohort (the first cohort with all three variables available) and the 2005 cohort, it can be seen that in both years, books had the strongest independent association with achievement, followed by occupation and then education. However, over this time period, this ranking became even more pronounced. The independent 90/10 books gap increased markedly from 0.41 to 0.75, while the independent 90/10 parent occupation gap increased more modestly from 0.39 to 0.44, and the independent parent education gap *decreased* from 0.27 to 0.17. (Note this does not mean parent education is not significantly related to student achievement after controlling for parent occupation and household books, only that the relationship between parent education and achievement is not growing, after accounting for the other two SES variables). Thus, it appears that the independent associations of each SES variable with achievement have changed at different rates over time.

That the conditional 90/10 parent occupation and household books gaps show increases suggests that the global increases in unconditional SES achievement gaps for these variables are not fully explained by their correlations with other SES characteristics. In contrast, the declining conditional 90/10 parent education gap suggests that the growing parent education gap may be fully explained by the correlations between parent education and other SES characteristics. However, it should be noted that the model with all three SES variables included is based on a substantially reduced sample of only those datasets that collected all three SES variables, and thus may not be representative. Most notably, all TIMSS 8th grade data is absent, as the dataset does not collect parent occupation. As we saw in Appendix B, TIMSS shows large increases in parent education achievement gaps, while PISA shows no increase or even declines in parent education achievement gaps (a discrepancy which may be due to problems in the measurement of parent education in PISA, which is addressed in Appendix H). In fact, for this limited sample, the trend in unconditional gaps is very small and non-significant (see row 3, "Unconditional gaps", "Cohort trend" column), in contrast with the large and significant increase in the parent education achievement gap seen in the main analyses. Possibly a better comparison is the model including only parent education and books (row 2), where the unconditional parent education gap shows a large and significant increase that is much closer to that estimated in the main analyses. In this sample, the conditional 90/10 parent education gap (controlling only for books) increased from 0.32 to 0.42 between the 1956 and 2005 birth cohorts, an increase that was marginally significant. This suggests that the increase in the parent education achievement gap is not fully explained by the correlation between parent education and books. From the data available, it is not possible to discern whether the independent 90/10 parent education gap is truly increasing or not.

Regardless of the relative importance of parent education, parent occupation, and books, the results in Table G2 clearly indicate that increasing correlations between SES variables do not fully explain increasing SES achievement gaps. Nearly all trends in conditional SES achievement gaps are positive and significant or marginally significant, with the exceptions noted above, which may be the result of sample reductions. These results suggest that, even after accounting for the growing number of children with "double-" or "triple-disadvantage" due to increasing correlations among SES variables, each SES variable (or at least parent occupation and books) has become more consequential for students' academic achievement.

One further piece of evidence that increasing SES achievement gaps for each SES variable are not entirely due to increasing correlations between variables is that the R^2 of the models used to compute conditional gaps with all three variables has increased over time. Table

G3 shows results from hierarchical growth models (country-subject-years within countries) predicting R^2 from cohort birth year (with controls for age and subject). (R^2 is adjusted for test reliability before running these models.) Results indicate that the R^2 of the model including all three SES variables nearly doubled from about 0.11 in the 1956 birth cohort to about 0.20 in the 2005 birth cohort. Thus, it appears that the overall predictive power of SES on achievement has grown substantially stronger over this 49-year time period.

Table G3. Trends in R^2 from Models with Two or Three SES Variables

	R^2 (Education & Occupation)	R^2 (Education & Books)	R^2 (Occupation & Books)	R^2 (Education, Occupation & Books)
Within countries				
Age 10 at testing	0.026 ***	-0.018 **	-0.041 ***	-0.027 ***
	(0.005)	(0.006)	(0.005)	(0.006)
Age 15 at testing		-0.004		
		(0.005)		
Math	0.007 ***	0.001	0.001	0.004
	(0.002)	(0.002)	(0.002)	(0.002)
Science	0.007 ***	0.009 ***	0.007 ***	0.009 ***
	(0.002)	(0.002)	(0.002)	(0.002)
Cohort birth year	0.0008 **	0.0021 ***	0.0017 ***	0.0018 ***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Between countries				
Intercept	0.107 ***	0.140 ***	0.161 ***	0.171 ***
	(0.005)	(0.006)	(0.006)	(0.006)
Residual variance (within countries)	0.00082	0.00120	0.00088	0.00098
Residual variance (country intercepts)	0.00153	0.00294	0.00303	0.00288
Residual variance (cohort slopes)	0.00000	0.00001	0.00000	0.00001
N (observations)	1379	1907	1386	1348
N (countries)	82	95	83	82

+ p<.1, * p<.05, ** p<.01, *** p < .001

H. Declining measurement error of SES

Another reason why SES achievement gaps may artificially appear to be increasing over time is that the reliability with which SES is measured may be increasing over time. Lower reliability of SES in early years could cause the estimated association between SES and achievement in early years to be attenuated, creating the appearance of increasing SES achievement gaps over time. Reliability is defined as the ratio of the variance in true SES to the total variance in SES (including both true variance and the variance of errors of measurement). As reported in Appendix E, the total variance of SES has declined for two SES variables—parent education and occupation—while the total variance has remained relatively constant for household books. It is likely that much of the decline in total variance in parent education and occupation reflects a decline in true variance, due to the large increase in the average *levels* of both variables as more parents attain higher education and professional occupations.

However, it is also likely that some of the reduction in the total variance of parent education and occupation is due to declining measurement error in these variables. Measurement error could potentially decline, for example, because current students have more accurate knowledge of their families' SES characteristics than in the past, because of improvements in survey wording, or because more recent SES data are more likely to be reported by parents rather than students. As stated in the main paper text, some international assessments have added parent questionnaires in recent years (PIRLS 2001-2011, TIMSS 4th grade 2011-2015, and PISA 2006-2012). The main models use parent-reported SES variables when available and student-reported SES otherwise. However, it is expected that parents report SES variables more reliably than their children. Some other patterns reported in previous appendix sections are consistent with increasing reliability of SES due to declining measurement error, while other patterns are inconsistent with this story. The large increase in the correlation between parent education and occupation (reported in Appendix G) is consistent with increasing reliability in these variables. However, the increase in the correlations between household books and each of the other two SES variables have been more modest. An increasing correlation between mothers' and fathers' education and occupation (reported in Appendix F) could also be evidence for increasing reliability of these variables. However, the correlation between mothers' and fathers' education has increased only moderately, and the correlation between mothers' and fathers' occupations has *declined* slightly.

As described in the Methods section of the main paper text, all SES achievement gaps have been adjusted for estimated reporting reliability of SES variables. Following Reardon (2011a), this adjustment consists of by multiplying each gap estimate by $\frac{1}{\sqrt{r}}$, where r is the reliability of the SES measure. In order to estimate the reliability r of each student- or parent-reported SES measure, we can take advantage of having two measures of the same variable reported by different sources (i.e., students and parents) (Jerrim and Micklewright 2014). The reliability of students' and parents' SES reports can be computed from the following formulas (Reardon 2011a):

$$r_s = corr(s,p) \cdot \frac{corr(s,y)}{corr(p,y)}$$

$$r_p = corr(s,p) \cdot \frac{corr(p,y)}{corr(s,y)}$$

where r_s is the reliability of student-reported SES variable s, r_p is the reliability of parent-reported SES variable p, and y is a third variable that would have a particular correlation with true SES, were SES measured without error. In reliability calculations from PISA and TIMSS, math achievement is used for this third variable y; in reliability calculations from PIRLS, reading achievement is used for *y*. Parent education is reported by both 15-year-old students and parents in PISA 2006, 2009, and 2012. Parent occupation is reported by both 15-year-old students and parents in PISA 2006 and 2012. Household books are reported by both 4th grade students and parents in PIRLS 2001, 2006, and 2011 and in TIMSS 2011 and 2015. I estimate the average reliability of parents' reports of their own educational attainment (across PISA 2006-2012) at 0.84 and students' reports of their parents' education at 0.62. I estimate the average reliability of parents' reports of their own occupational category (across PISA 2006 and 2012) at 0.81 and students' reports of their parents' occupation at 0.79. I estimate the average reliability of parents' reports of the number of household books (across PIRLS 2001-2011 and TIMSS 2011-2015) at 0.52 and students' reports of household books at 0.46. The higher accuracy of parent occupation reports and low accuracy of household books reports is consistent with findings by Jerrim and Micklewright (2014) using some of the same international datasets. In order to estimate reliabilities for other age groups, I assume that 8th grade students report all SES variables with the same reliability as 15-year-old students, but 4th grade students report parent education 80% as reliably and parent occupation and household books 90% as reliably as 15-year-old students. Finally, in order to estimate reliabilities for other years where parent reports are unavailable, I use the average reliabilities for each of these age groups. Since the reliabilities applied to all years are derived from parent reports in recent years, this procedure adjusts only for differences in reliability between parents and children, but cannot account for possible changes in reliability over time.

Table H1 reports estimated trends in 90/10 SES achievement gaps, using only student-reported data for all three SES variables (gaps are not adjusted for SES or test reliability). Sample sizes are reduced, mainly due to the omission of 4th grade assessments from recent years with parent-reported education and occupation. In these models, the estimated positive trends in 90/10 SES achievement gaps are reduced compared to those reported in the main paper text but are still positive and highly significant for all three SES variables. Using these results, it is possible to estimate the sensitivity of estimated gap trends to potential increases in the accuracy of students' reports of their parents' SES characteristics. For the increases in SES achievement gaps reported in Table H1 to be fully accounted for by measurement error alone, the reliability of students' reports of parental education, estimated at 0.62 for recent cohorts, would have to be only 0.41 for the 1950 cohort. The reliability of students' reports of parental occupation, estimated at 0.79 for recent cohorts, would have to be 0.61 for the 1950 cohort; and the reliability of students' reports of household books, estimated at 0.46 for recent cohorts, would have to be 0.32 for the 1956 cohort (the first cohort for which the household books variable was collected). Without parental reports, it is impossible to know from these data whether the reliability of students' reports could have increased by 25-50% over this 50 year period. However, a thorough literature search did not reveal published evidence that survey reporting of SES characteristics by either adults or children has become more accurate over time.

Table H1. Estimated trends in 90/10 SES achievement gaps, student-reported SES data only

	Education	Occupation	Books
Within countries			
Age 10 at testing			-0.179 ***
			(0.018)
Age 15 at testing	-0.111 ***		0.057 ***
	(0.016)		(0.017)
Math	0.037 ***	0.025 ***	-0.014 +
	(0.009)	(0.006)	(0.008)
Science	0.017 *	0.008 *	0.007
	(0.007)	(0.004)	(0.008)
Cohort birth year	0.004 ***	0.003 **	0.006 ***
	(0.001)	(0.001)	(0.001)
Between countries			
Intercept	0.752 ***	0.786 ***	0.807 ***
	(0.024)	(0.021)	(0.027)
Residual variance (within countries)	0.00623	0.00363	0.00840
Residual variance (country intercepts)	0.03600	0.03073	0.05967
Residual variance (cohort slopes)	0.00005	0.00006	0.00007
N (observations)	1627	1092	2182
N (countries)	92	72	100

+ p<.1, * p<.05, ** p<.01, *** p < .001

The estimates of the reliability of students' SES reports from recent years (for countries that collected SES from both students and parents) do vary substantially. Reliability estimates of students' reports of parent education from PISA 2012 range from 0.43 in Croatia to 0.91 in Portugal; of parent occupation range from 0.66 in Mexico to 0.92 for Croatia; and of household books from PIRLS 2011 range from 0.03 in Kuwait to 0.67 in Bulgaria. Yet for parent education and occupation, even the countries with the lowest estimated reliabilities do not quite reach what would need to be the *average* level of reliability for early cohorts in order to fully explain the global increase in SES achievement gaps. For household books, in contrast, there is more variability in the estimated accuracy of students' reports, and it is conceivable that average reliability could have increased by 40% in 49 years.

Estimated SES reliability could potentially also be compared across multiple waves of each study that collects a parent questionnaire (PIRLS 2001-2011 and PISA 2006-2012). Computed SES reliabilities for students and parents do appear to vary somewhat across waves, although it is not clear that these differences represent meaningful trends, given the small number of years and countries represented. It is also possible to compare trends in SES achievement gaps estimated from students' versus parents' reports across waves of PIRLS and PISA. These estimates are reported in Table H2. The estimated trends do not differ systematically depending on whether they are estimated from parent or student reports (this comparison was done before adjusting for computed SES reliability, as gaps based on parent- and student-reported data will be nearly identical after adjustment by construction).

Table H2. Trends in Gaps by Student-Reported and Parent-Reported SES, by Test

Study	Subject	Years	SES report	Education		(N)	Occupation		(N)	Books		(N)
PISA	Math	2006-2012	Parent	0.017	**	(13)	0.013	*	(7)			
PISA	Math	2006-2012	Student	0.013	***	(13)	0.015	*	(7)			
PISA	Reading	2006-2012	Parent	0.030	***	(13)	0.013	*	(7)			
PISA	Reading	2006-2012	Student	0.014	**	(13)	0.013	+	(7)			
PISA	Science	2006-2012	Parent	0.013	*	(13)	0.008		(7)			
PISA	Science	2006-2012	Student	-0.001		(13)	0.009	+	(7)			
PIRLS	Reading	2001-2011	Parent							0.004		(40)
PIRLS	Reading	2001-2011	Student							0.009	**	(40)

+ p<.1, * p<.05, ** p<.01, *** p < .001

Increased accuracy of the SES variables could result not only from improvements in students' knowledge of their family SES characteristics but also from improved questionnaire wording over time. The wording of the background questionnaires differs across the older IEA studies, the new IEA studies (TIMSS and PIRLS), and PISA. However, the wording of background questionnaires has changed very little across multiple waves of the trend studies (TIMSS, PIRLS, and PISA). Yet, as demonstrated Appendix B, trends in SES achievement gaps estimated from each of these studies individually are still nearly always positive (though not always statistically significant, given the smaller sample sizes). Between TIMSS 1995 and PIRLS 2001, the IEA began including drawings of bookshelves in its questionnaires for 4^{th} grade students in order to assist them in estimating the number of books they have at home, which likely decreased measurement error and could bias estimates of books achievement gap trends upward. However, as we have seen, books achievement gaps also increased substantially for 8^{th} grade and 15-year-old students, with no added drawings.

The one consistent exception to the increasing SES achievement gaps across all test instruments in Appendix B is trends estimated from PISA student-reported parent education, which are usually negative. This is especially surprising since 8^{th} grade TIMSS tests a similar population in similar subjects and shows large increases in parent education achievement gaps. In the Appendix B results, it is difficult to discern whether the discrepancy between PISA and TIMSS 8^{th} grade is the result of differences in samples, in test instruments, or in the measurement of SES. Table H3 estimates trends in parent education and household books achievement gaps for math and science in a constant sample of 42 countries that participated in at least two cycles each of PISA and TIMSS 8^{th} grade. Trends in parent education gaps are close to 0 for PISA but are large, positive, and significant for TIMSS. Although trends in household books gaps are also larger for TIMSS than for PISA, they are nevertheless still large, positive, and significant for PISA. This suggests that the difference in trends for PISA and TIMSS is not solely attributable to differences in test instruments and target populations but also possible differences in the measurement of SES, particularly parent education.

Table H3. Comparison of Trends in PISA and TIMSS 8th Grade 90/10 Parent Education and Household Books Gaps

Study	Subject	Years	SES report	Education		(N)	Books		(N)
PISA	Math	2000-2015	Student	0.001		(42)	0.005	*	(42)
PISA	Science	2000-2015	Student	-0.001		(42)	0.006	***	(42)
TIMSS Grade 8	Math	1999-2015	Student	0.009	***	(42)	0.012	***	(42)
TIMSS Grade 8	Science	1999-2015	Student	0.010	***	(42)	0.016	***	(42)

+ p<.1, * p<.05, ** p<.01, *** p < .001

Table H4 compares the parent education item wording from the student questionnaire for PISA and the student/parent questionnaires for the IEA studies (TIMSS and PIRLS).[6] TIMSS has a single education item for each parent; PISA has two items (schooling and higher education). TIMSS lists education levels in ascending order; PISA lists them in descending order. TIMSS includes an option for "I don't know"; PISA does not.[7] The highest two educational categories in TIMSS are "<ISCED Level 5A, first degree>" (i.e., BA) and "Beyond <ISCED Level 5A, first degree>" (i.e., MA, PhD, and professional degrees); the highest two categories in PISA are "<ISCED level 5A>" (i.e., BA, MA, and professional degrees) and "<ISCED level 6>" (i.e., PhD). Perhaps as a result of some or all of these differences, a substantially larger share of students select highest two categories (BA or above) in PISA than in TIMSS. Of the 29 countries participating in both PISA 2012 and TIMSS 2011 8th grade, 20 countries had a higher share of students reporting BA or above in PISA than TIMSS, by an average of 6 percentage points. For example, in Australia, 31% of TIMSS 8th grade 2011 students report BA or more, while 43% of PISA 2012 students report BA or more; in Finland, 42% of TIMSS 8th grade 2011 students report BA or more, while 55% of PISA 2012 students report BA or more. This pattern is in addition to the general increase in the share of students in

[6] Beginning in 2015, TIMSS updated its parent education item wordings to reflect the new ISCED 2011 scheme. PISA had not yet made any update in its 2015 cycle.

[7] "I don't know" responses in TIMSS were treated as missing data and imputed as part of the multiple imputation procedure. TIMSS 8th grade still shows large and significant increases in parent education achievement gaps using unimputed data with listwise deletion of "I don't know" responses and other missing parent education data.

the highest education categories seen across all datasets due to educational upgrading in the parents' generations. The larger share of students in the highest category in PISA means that the achievement at the 90^{th} percentile of parent education is estimated with more error. This means that parent education achievement gaps may be underestimated in more recent years, and consequently that gap trends may be underestimated.

Table H4. Comparison of PISA and TIMSS parent education questionnaire wording

PISA 2009-2015 student questionnaire	TIMSS 2003-2011 8^{th} grade student questionnaire, TIMSS 2011 4^{th} grade parent questionnaire, PIRLS 2006-2011 parent questionnaire
Q14. What is the <highest level of schooling> completed by your mother? • <ISCED level 3A> • <ISCED level 3B, 3C> • <ISCED level 2> • <ISCED level 1> • She did not complete <ISCED level 1> Q15. Does your mother have any of the following qualifications? • <ISCED level 6> • <ISCED level 5A> • <ISCED level 5B> • <ISCED level 4>	6A. What is the highest level of education completed by your mother <or stepmother or female guardian>? • Some <ISCED Level 1 or 2> or did not go to school • <ISCED Level 2> • <ISCED Level 3> • <ISCED Level 4> • <ISCED Level 5B> • <ISCED Level 5A, first degree> • Beyond <ISCED Level 5A, first degree> • I don't know

In summary, it is likely that changing measurement error in SES confounds estimates of true global changes in SES achievement gaps. But it is not clear that measurement error in SES is uniformly declining over time, as error may be increasing for some variables in some recent tests (i.e., PISA parent education). This would lead gap increases to be *under*- rather than overestimated, making the main trend estimates more conservative. To the extent that the reliability of SES has improved, this increase would have to be very large to fully account for the increase in SES achievement gaps. There is no direct evidence for changes in measurement error of SES, either in the data used in this study or in published research. Future research should identify older datasets containing both children's and parents' reports of family SES characteristics in order to examine whether reliability in student reporting may have increased over time and the size of this possible increase.

I. Trends in SES achievement gaps by world region

In the main text of the paper, differences in country results are compared based on a rough indicator of countries' level of development: whether the country had a GDP per capita of at least $6000 in 1980. This section of the appendix estimates the association between gap trends and a continuous measure of GDP per capita, as well as reporting differences in gap trends among six world regions: (1) sub-Saharan Africa, (2) east and southeast Asian and Pacific countries, (3) Middle Eastern and North African countries, (4) Eastern Europe and the Commonwealth of Independent States, (5) Latin America and the Caribbean, and (6) Western countries (Western Europe and Anglophone countries). Countries are classified into these regions based on UN Regional Group and/or Arab League membership without regard to GDP or OECD membership; a full list of countries by region appears in Appendix A.

Table I1 shows results of hierarchical growth models with interactions between cohort birth year and time-invariant country characteristics (GDP per capita and region). The models are estimated as follows:

$$\hat{G}_{ij} = \gamma_{00} + \gamma_{10}Y_{ij} + \boldsymbol{Z_j}\boldsymbol{\Gamma} + Y_{ij}\boldsymbol{Z_j}\boldsymbol{B} + \boldsymbol{\Lambda}_{ij} + v_j + r_jY_{ij} + u_{ij} + \epsilon_{ij},$$

$$v_j \sim N(0, \tau_{00}); r_j \sim N(0, \tau_{11});\ u_{ij} \sim N(0, \sigma^2);\ \ \epsilon_{ij} \sim N\left(0, \omega_{ij}\right),$$

where $\hat{G}_{ij}$ is the estimated gap in country j in country-study-year i, γ_{10} is the coefficient for cohort birth year Y_{ij}, $\boldsymbol{Z_j}$ is a vector of time-invariant country characteristics, $\boldsymbol{\Lambda}_{ij}$ is a vector of dummy variables indicating age at testing and test subject, $\mathbf{\Gamma}$ is a vector of coefficients for the time-invariant country covariates, $\boldsymbol{B}$ is a vector of coefficients for the interactions between Y_{ij} and the time-invariant country characteristics, τ_{00} is the between-country variance of the true gaps, τ_{11} is the between-country variance of true slopes of cohort birth year, σ^2 is the true within-country variance of the gaps, and $\omega_{ij} = \left[s.e.\left(\hat{G}_{ij}\right)\right]^2$ is the sampling variance of $\hat{G}_{ij}$.

The positive main effects for GDP per capita show that higher-income countries generally had larger SES achievement gaps at the midpoint of the time series, a finding that is consistent with prior cross-sectional research comparing SES achievement gaps by countries' level of development (Heyneman and Loxley 1983). However, there is no consistent relationship between countries' level of development and *trends* in gaps; the GDP per capita × Cohort interaction is negative for parent education, positive for books, and close to 0 for parent education. The main effects for world regions show that sub-Saharan Africa, East Asia and the Pacific, and the Middle East and North Africa all had smaller gaps than the reference category, Western countries, at the midpoint of the time series. Eastern European countries' gaps were not significantly different from Western countries; and Latin American and Caribbean countries' gaps were larger than those of Western countries. The region × Cohort interactions are generally non-significant or not consistent in direction across the three SES variables. The one exception is that the Latin America × Cohort interactions are consistently negative and are significant for two of three SES variables. Summing the Cohort main effects and regional interactions shows that gap trends are generally positive in all regions except for Latin America, where gaps are decreasing on average. However, all regional interactions should be interpreted with caution, as there very few non-Western countries represented in the early years of the data.

Table I1. Trends in 90/10 SES Achievement Gaps by GDP in 1980 and World Region

	Education		Occupation		Books	
	(1)	(2)	(1)	(2)	(1)	(2)
Within countries						
Age 10 at testing	-0.023	-0.026	0.020	0.021	-0.229 ***	-0.232 ***
	(0.032)	(0.032)	(0.020)	(0.020)	(0.027)	(0.027)
Age 15 at testing	-0.169 ***	-0.173 ***			0.089 ***	0.085 ***
	(0.023)	(0.022)			(0.025)	(0.025)
Math	0.031 **	0.030 **	0.020 *	0.020 *	-0.020	-0.020
	(0.011)	(0.011)	(0.008)	(0.008)	(0.013)	(0.013)
Science	0.024 *	0.024 *	0.014 *	0.015 *	0.033 **	0.033 **
	(0.010)	(0.010)	(0.006)	(0.006)	(0.012)	(0.012)
Cohort birth year	0.006 ***	0.006 **	0.003 *	0.003	0.007 ***	0.013 ***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Between countries						
Intercept	1.050 ***	1.068 ***	0.923 ***	0.983 ***	1.220 ***	1.424 ***
	(0.032)	(0.034)	(0.025)	(0.032)	(0.036)	(0.046)
GDP per capita in 1980 (logged)	0.066 *	0.068 +	0.018	-0.005	0.191 ***	0.114 **
	(0.028)	(0.036)	(0.029)	(0.034)	(0.054)	(0.039)
Log GDP per capita in 1980 × Cohort birth year	-0.004 *	-0.002	0.002	0.003	0.003 +	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Sub-Saharan Africa		-0.128				-0.596 ***
		(0.086)				(0.137)
Sub-Saharan Africa × Cohort birth year		0.019 **				-0.019 *
		(0.006)				(0.008)
East Asia & Pacific		-0.039		-0.284 **		-0.335 **
		(0.113)		(0.109)		(0.122)
East Asia & Pacific × Cohort birth year		0.002		0.003		-0.007
		(0.004)		(0.003)		(0.005)
Middle East & North Africa		-0.190 *		-0.232 ***		-0.615 ***
		(0.088)		(0.065)		(0.088)
Middle East & North Africa × Cohort birth year		0.000		0.001		-0.009
		(0.006)		(0.004)		(0.006)
Eastern Europe & CIS		0.000		-0.045		-0.099
		(0.063)		(0.058)		(0.063)
Eastern Europe & CIS × Cohort birth year		-0.001		0.002		-0.002
		(0.003)		(0.003)		(0.003)
Latin America & Caribbean		0.220 **		0.101 *		-0.117
		(0.072)		(0.051)		(0.083)
Latin America & Caribbean × Cohort birth year		-0.010 *		-0.003		-0.021 **
		(0.004)		(0.005)		(0.007)
Residual variance (within countries)	0.01320	0.01330	0.00713	0.00713	0.01862	0.01858
Residual variance (country intercepts)	0.05462	0.04402	0.04388	0.03112	0.09539	0.05372
Residual variance (cohort slopes)	0.00008	0.00006	0.00007	0.00007	0.00013	0.00010
N (observations)	1858	1858	1362	1362	2129	2129
N (countries)	89	89	77	77	95	95

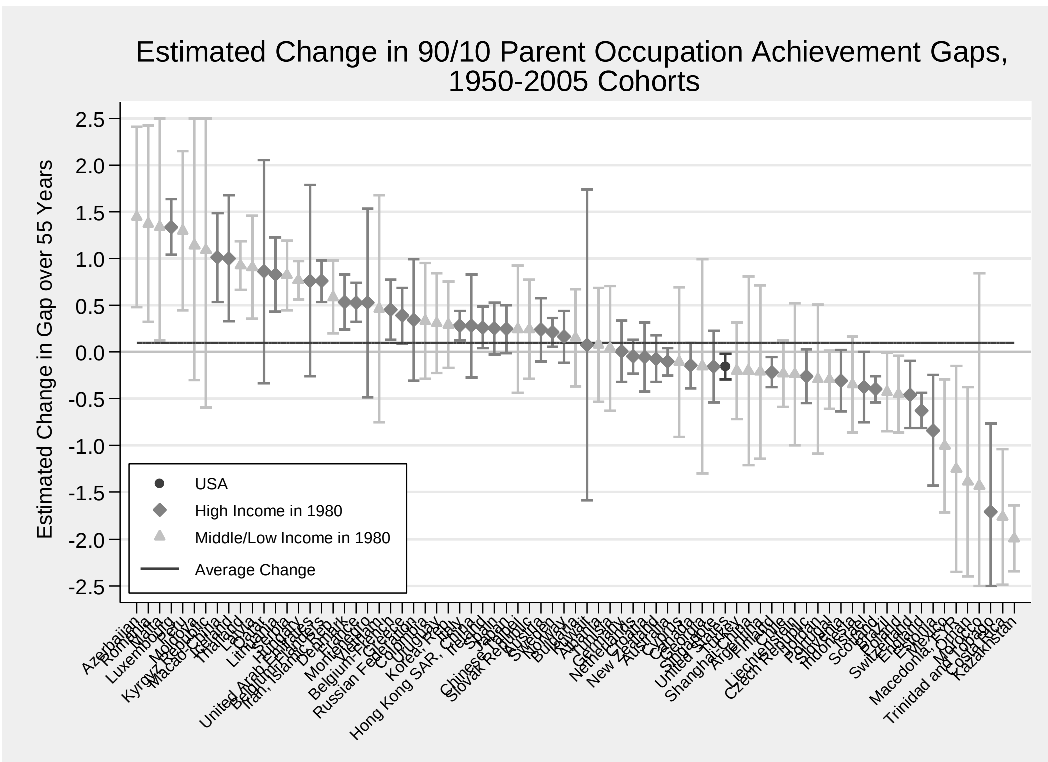
+ p<.1, * p<.05, ** p<.01, *** p < .001. See Table A1 for a full list of countries included in each region.

Although it is likely the case that relationships between time-varying country covariates and SES achievement gaps reported in the main text vary by world region, there are not sufficient sample sizes of countries in each region to estimate each relationship separately.

J. Figures showing estimated change in parent occupation and books achievement gaps

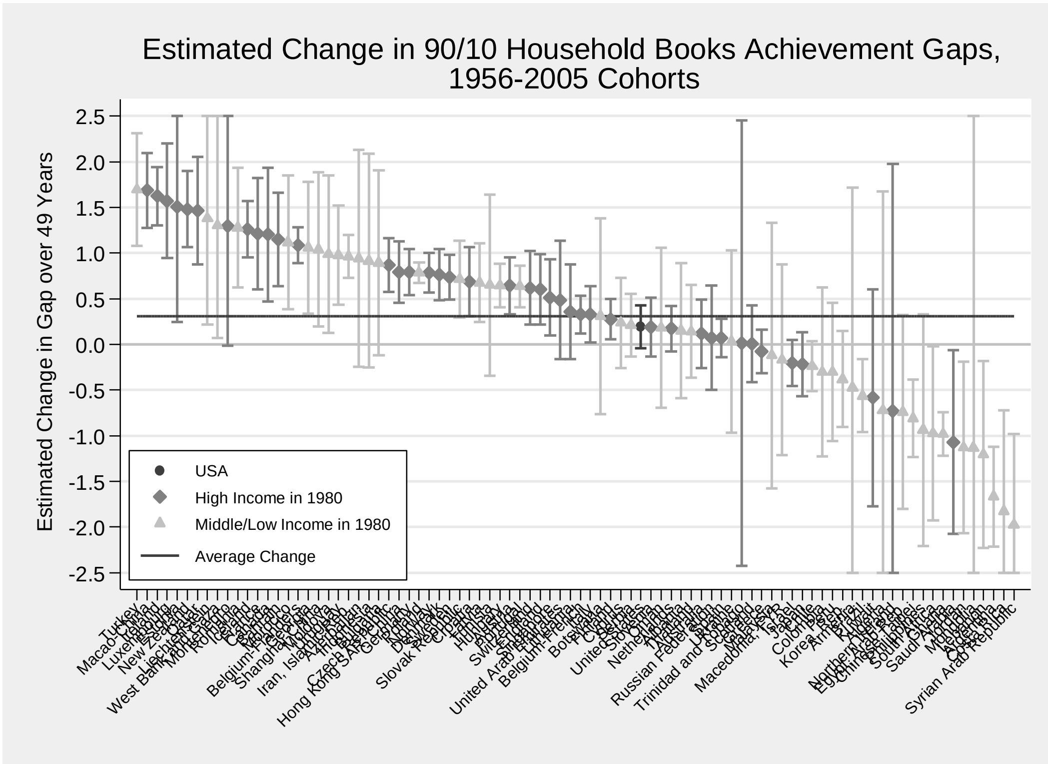
Figures J1 and J2 follow the format of Figure 2 in the main text of the paper and plot estimated changes in parent occupation and books achievement gaps by country, sorted in order from the largest estimated increase to the largest estimated decline in gaps. The horizontal dark gray lines show the average estimated change across all countries, which is greater than 0, similar to the finding for parent education achievement gaps.

Figure J1



Notes: "High income" countries had GDPs per capita of at least $6000 in 1980. Gray brackets are 95% confidence intervals. Trends adjusted for age of testing and subject.

Figure J2



Notes: "High income" countries had GDPs per capita of at least $6000 in 1980. Gray brackets are 95% confidence intervals. Trends adjusted for age of testing and subject.

K. SES achievement gap trends by age and subject

Table K1 reports gap trends models run separately by level of schooling (primary versus secondary). All gap increases are positive for both primary and secondary school students for all three SES variables. For two variables, parent education and occupation, achievement gaps have increased more for primary than for secondary school students, but for books, gaps have increased more for secondary school students.

Table K2 reports gap trends models separately by test subject (math, science, or reading). All gap trends are positive except for a slightly negative coefficient for parent education reading gaps, which is close to 0 and not significant. In general, math and science gaps have increased more than reading gaps for all three SES variables. This pattern of greater increases in math and science gaps is most pronounced for parent education and parent occupation and less so for books, which we may expect to be particularly closely tied reading achievement. However, even for books achievement gaps, math and science achievement gaps are increasing slightly more than reading gaps. The intercept represents the estimated size of each gap in the 1989 birth cohort; it can be seen that in the earliest cohort (1950), reading gaps were larger than math and science gaps. However, in the most recent cohort (2005), estimated math and science gaps have become slightly larger than reading gaps.

Table K1. Coefficients from Hierarchical Growth Models Predicting Achievement Gaps between 90th and 10th Percentiles of Parent Education, Parent Occupation, and Books at Home, by Age (Primary or Secondary School)

	Education		Occupation		Books	
	Primary	Secondary	Primary	Secondary	Primary	Secondary
Within countries						
Age 15 at testing		-0.166 ***				0.035
		(0.022)				(0.025)
Math	-0.022	0.045 ***	-0.032	0.030 ***	-0.047 +	-0.020 +
	(0.027)	(0.011)	(0.022)	(0.007)	(0.028)	(0.011)
Science	0.028	0.033 ***	0.011	0.012 **	0.044 +	0.025 **
	(0.025)	(0.009)	(0.019)	(0.004)	(0.025)	(0.010)
Cohort birth year	0.009 **	0.005 ***	0.005 *	0.003 *	0.007 ***	0.009 ***
	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)
Between countries						
Intercept	1.066 ***	1.039 ***	0.931 ***	0.945 ***	1.130 ***	1.264 ***
	(0.039)	(0.032)	(0.036)	(0.025)	(0.034)	(0.043)
Residual variance (within countries)	0.00194	0.01092	0.00254	0.00440	0.01374	0.00977
Residual variance (country intercepts)	0.04787	0.06297	0.05250	0.04171	0.06320	0.14779
Residual variance (cohort slopes)	0.00020	0.00009	0.00016	0.00008	0.00012	0.00039
N (observations)	257	1628	289	1095	522	1631
N (countries)	52	92	55	72	64	93

+ p<.1, * p<.05, ** p<.01, *** p < .001

Table K2. Coefficients from Hierarchical Growth Models Predicting Achievement Gaps between 90th and 10th Percentiles of Parent Education, Parent Occupation, and Books at Home, by Test Subject (Math, Science, or Reading)

	Education			Occupation			Books		
	Math	Science	Reading	Math	Science	Reading	Math	Science	Reading
Within countries									
Age 10 at testing	-0.035	-0.035	0.075	-0.041	0.050 *	0.032	-0.246 ***	-0.220 ***	-0.194 **
	(0.036)	(0.042)	(0.066)	(0.029)	(0.024)	(0.024)	(0.034)	(0.028)	0.061
Age 15 at testing	-0.139 ***	-0.163 ***	-0.121 *				0.137 ***	0.057 *	0.135 *
	(0.021)	(0.025)	(0.060)				(0.024)	(0.028)	0.056
Cohort birth year	0.006 ***	0.007 ***	-0.00005	0.005 ***	0.004 **	0.002	0.009 ***	0.010 ***	0.006 **
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	0.002
Between countries									
Intercept	1.070 ***	1.072 ***	1.058 ***	0.949 ***	0.930 ***	0.943 ***	1.202 ***	1.265 ***	1.288 ***
	(0.031)	(0.032)	(0.057)	(0.025)	(0.026)	(0.024)	(0.040)	(0.044)	0.061
Residual variance (within countries)	0.01373	0.01608	0.00900	0.00904	0.00632	0.00585	0.01457	0.02168	0.01235
Residual variance (country intercepts)	0.05808	0.05646	0.03835	0.04217	0.04763	0.04047	0.11483	0.13770	0.07901
Residual variance (cohort slopes)	0.00005	0.00005	0.00009	0.00005	0.00005	0.00004	0.00013	0.00006	0.00007
N (observations)	699	715	492	446	471	480	785	851	539
N (countries)	93	93	78	77	79	76	95	96	80

+ p<.1, * p<.05, ** p<.01, *** p < .001

L. Specification of trend model

All of the sensitivity analyses thus far have used the same specification as Model 1 in the main text of the paper to test for significant increases in SES achievement gaps. Tables L1, L2, and L3 report coefficients from eight alternate model specifications (Models 1B-1J) for parent education, parent occupation, and books gaps, respectively. Model 1 is the same as in the main text of the paper for comparison. Model 1 omitted any country-years with missing data for the time-varying covariates used in Model 2. Model 1B includes all available data, which can be seen by the increased sample sizes. Results for Model 1B are very similar to Model 1 for all three SES variables. All available data are retained in Models 1C-1J.

Model 1C uses gaps computed without multiple imputation of missing student-level data; listwise deletion of student observations with missing data on any variables is used instead. Level 1 (country-subject-year observation) sample sizes are slightly smaller. Results are similar for all three SES variables, though the standard errors on the cohort trend increases substantially for parent education gaps, and the cohort trend loses significance for parent education and occupation. These changes are due to increased heterogeneity across countries in gap trends estimated when dropping missing data. Cohort trends are highly significant for both parent education and occupation when the model is estimated with a fixed rather than a random slope.

Table L1. Coefficients from Hierarchical Growth Models Predicting Achievement Gaps between 90th and 10th Percentiles of Parent Education (Variety of Trend Model Specifications)

	(1) Main text	(1B) Full sample	(1C) No imputation	(1D) Linear gaps	(1E) No weights	(1F) Fixed slope	(1G) Controls	(1H) Quadratic	(1J) Quadratic - no PISA
Within countries									
Age 10 at testing	-0.038	-0.026	-0.030	-0.037	0.010	-0.015	-0.038	-0.007	-0.093 *
	(0.038)	(0.031)	(0.037)	(0.028)	(0.031)	(0.031)	(0.031)	(0.030)	(0.041)
Age 15 at testing	-0.175 ***	-0.166 ***	-0.187 ***	-0.143 ***	-0.156 ***	-0.165 ***	-0.111 ***	-0.176 ***	
	0.026	0.022	0.027	0.019	0.026	0.022	0.025	0.023	
Math	0.026 *	0.029 **	0.030 **	0.035 **	0.045 ***	0.028 *	0.029 **	0.035 **	-0.044 *
	(0.012)	(0.011)	(0.012)	(0.011)	(0.010)	(0.011)	(0.011)	(0.011)	(0.022)
Science	0.016	0.023 *	0.028 +	0.029 **	0.039 ***	0.021 *	0.024 *	0.031 ***	-0.036
	(0.010)	(0.010)	(0.016)	(0.009)	(0.009)	(0.010)	(0.010)	(0.009)	(0.023)
Cohort birth year	0.007 ***	0.005 ***	0.006	0.006 ***	0.003 +	0.006 ***	0.005 ***	0.005 ***	0.011 ***
	(0.001)	(0.001)	(0.007)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
Cohort birth year^2								-0.000279 **	0.000045
								(0.000106)	(0.000088)
Number of SES categories							-0.005		
							(0.003)		
Proportion in bottom SES category							-0.18		
							(0.114)		
Proportion in top SES category							-0.35 ***		
							(0.091)		
Between countries									
Intercept	1.098 ***	1.053 ***	1.073 ***	1.001 ***	1.033 ***	1.045 ***	1.144 ***	1.067 ***	1.11 ***
	(0.034)	(0.032)	0.136	0.03	0.03	(0.032)	(0.051)	(0.032)	(0.037)
Residual variance (within countries)	0.01269	0.01287	0.01401	0.00782	0.03685	0.01764	0.01231	0.00930	0.01325
Residual variance (between countries)	0.04555	0.05820	0.05908	0.05498	0.05500	0.05805	0.05706	0.06093	0.05397
Residual variance (cohort slopes)	0.00008	0.00010	0.00009	0.00011	0.00018		0.00010	0.00011	0.00010
Residual variance (cohort^2 slopes)								0.00000	0.00000
N (observations)	1510	1915	1902	1931	1915	1915	1915	1915	868
N (countries)	68	94	94	95	94	94	94	94	78

+ p<.1, * p<.05, ** p<.01, *** p < .001

Notes: Model 1 is identical to in the full paper text; Model 1B (and all subsequent models) include all observations regardless of country covariate data availability; Model 1C uses gaps computed without multiple imputation of missing data; Model 1D uses gaps computed with linear rather than cubic models; Model 1E does not include precision weights; Model 1F uses a fixed rather than random coefficient for cohort birth year; Model 1G includes controls for quality of SES variable; Model 1H includes a squared term for cohort birth year; Model 1J includes a squared term for cohort birth year and excludes PISA data.

Table L2. Coefficients from Hierarchical Growth Models Predicting Achievement Gaps between 90th and 10th Percentiles of Parent Occupation (Variety of Trend Model Specifications)

	(1)	(1B)	(1C)	(1D)	(1E)	(1F)	(1G)	(1H)	(1J)
	Main text	Full sample	No imputation	Linear gaps	No weights	Fixed slope	Controls	Quadratic	Quadratic - no PISA
Within countries									
Age 10 at testing	0.005	0.021	-0.019	0.01700	0.019	0.035	-0.073 *	0.041 *	0.106 +
	(0.019)	(0.020)	(0.019)	(0.020)	(0.020)	(0.022)	(0.029)	(0.019)	(0.055)
Math	0.012	0.019 *	0.019 *	0.015 +	0.009	0.018 *	0.022 **	0.026 ***	-0.038
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)	(0.025)
Science	0.008	0.014 *	0.02 **	0.022 ***	0.005	0.013 +	0.015 *	0.019 **	0.004
	(0.006)	(0.006)	(0.006)	(0.006)	(0.008)	(0.007)	(0.006)	(0.006)	(0.023)
Cohort birth year	0.004 **	0.004 **	0.002	0.005 ***	0.004 ***	0.003 *	0.005 ***	0.003 *	0.003 +
	(0.001)	(0.001)	(0.001)	(0.0010)	(0.0010)	(0.001)	(0.001)	(0.001)	(0.002)
Cohort birth year^2								-0.00016 **	0.00011
								(0.000059)	(0.000083)
Number of SES categories							0.042 *		
							(0.021)		
Proportion in bottom SES category							0.387		
							(0.242)		
Proportion in top SES category							-0.738 ***		
							(0.204)		
Between countries									
Intercept	0.956 ***	0.919 ***	0.948 ***	0.9 ***	0.921 ***	0.916 ***	0.696 ***	0.928 ***	(0.837) ***
	(0.026)	(0.025)	0.023	0.025	0.025	(0.025)	(0.190)	(0.026)	0.061
Residual variance (within countries)	0.00593	0.00709	0.00245	0.00581	0.01775	0.01057	0.00591	0.00556	0.01251
Residual variance (between countries)	0.03862	0.04476	0.03931	0.04637	0.04421	0.04808	0.04264	0.05019	0.05495
Residual variance (cohort slopes)	0.00009	0.00008	0.00011	0.00008	0.00007		0.00007	0.00007	0.00008
Residual variance (cohort^2 slopes)								0.00000	0.00000
N (observations)	1146	1405	1396	1413	1405	1405	1405	1405	362
N (countries)	63	82	82	83	82	82	82	82	59

+ p<.1, * p<.05, ** p<.01, *** p < .001

Notes: Model 1 is identical to in the full paper text; Model 1B (and all subsequent models) include all observations regardless of country covariate data availability; Model 1C uses gaps computed without multiple imputation of missing data; Model 1D uses gaps computed with linear rather than cubic models; Model 1E does not include precision weights; Model 1F uses a fixed rather than random coefficient for cohort birth year; Model 1G includes controls for quality of SES variable; Model 1H includes a squared term for cohort birth year; Model 1J includes a squared term for cohort birth year and excludes PISA data.

Table L3. Coefficients from Hierarchical Growth Models Predicting Achievement Gaps between 90th and 10th Percentiles of Household Books (Variety of Trend Model Specifications)

	(1)	(1B)	(1C)	(1D)	(1E)	(1F)	(1G)	(1H)	(1J)
	Main text	Full sample	No imputation	Linear gaps	No weights	Fixed slope	Controls	Quadratic	Quadratic - no PISA
Within countries									
Age 10 at testing	-0.230 ***	-0.234 ***	-0.235 ***	-0.227 ***	-0.224 ***	-0.226 ***	-0.238 ***	-0.235 ***	-0.226 ***
	(0.032)	(0.027)	(0.028)	(0.027)	(0.027)	(0.026)	(0.025)	(0.027)	(0.029)
Age 15 at testing	(0.102) ***	(0.085) ***	(0.080) **	(0.078) **	(0.091) ***	(0.085) ***	(0.058)	(0.077) **	
	(0.028)	(0.025)	(0.025)	(0.026)	(0.025)	(0.025)	(0.038)	(0.025)	
Math	-0.026 +	-0.021 +	-0.021 +	-0.015	-0.026 *	-0.021	-0.027 *	-0.019	-0.070 **
	(0.014)	(0.013)	(0.012)	(0.012)	(0.013)	(0.013)	(0.012)	(0.012)	(0.024)
Science	0.033 *	0.032 **	0.043 ***	0.038 **	0.028 *	0.033 *	0.026 *	0.035 **	0.009
	(0.014)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.011)	(0.012)	(0.024)
Cohort birth year	0.009 ***	0.008 ***	0.009 ***	0.008 ***	0.008 ***	0.010 ***	0.005 **	0.008 ***	0.010 ***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
Cohort birth year^2								0.000027	0.000146 *
								(0.000098)	(0.000068)
Number of SES categories							0.001		
							(0.030)		
Proportion in bottom SES category							0.316		
							(0.194)		
Proportion in top SES category							-0.249		
							(0.171)		
Between countries									
Intercept	1.313 ***	1.23 ***	1.232 ***	1.218 ***	1.223 ***	1.216 ***	1.215 ***	1.236 ***	1.242 ***
	(0.043)	(0.040)	0.04	0.04	0.038	(0.040)	(0.189)	(0.041)	(0.045)
Residual variance (within countries)	0.01805	0.01821	0.01918	0.01447	0.03732	0.02260	0.01769	0.01620	0.02425
Residual variance (between countries)	0.08986	0.12190	0.12012	0.11870	0.12629	0.13259	0.14503	0.13357	0.12112
Residual variance (cohort slopes)	0.00011	0.00014	0.00014	0.00014	0.00016		0.00015	0.00018	0.00006
Residual variance (cohort^2 slopes)								0.00000	0.00000
N (observations)	1738	2190	2183	2203	2190	2190	2190	2190	1143
N (countries)	70	100	100	100	100	100	100	100	84

+ p<.1, * p<.05, ** p<.01, *** p < .001

Notes: Model 1 is identical to in the full paper text; Model 1B (and all subsequent models) include all observations regardless of country covariate data availability; Model 1C uses gaps computed without multiple imputation of missing data; Model 1D uses gaps computed with linear rather than cubic models; Model 1E does not include precision weights; Model 1F uses a fixed rather than random coefficient for cohort birth year; Model 1G includes controls for quality of SES variable; Model 1H includes a squared term for cohort birth year; Model 1J includes a squared term for cohort birth year and excludes PISA data.

Model 1D uses gaps for which the 90th and 10th percentiles of each SES categorical variable were interpolated from linear rather than cubic weighted least squares models. Level 1 (observation) and Level 2 (country) sample sizes are larger because a slightly larger number of gaps can be estimated reliably from these simpler models. Cohort birth year coefficients are slightly larger for parent education and occupation trends (i.e., the estimated increase in SES achievement gaps is larger) because the additional observations are very small gaps in early years. Cubic models are retained as the preferred models, as the trend estimates are more conservative, cubic gap functions allow more flexibility in the shape of the relationship between SES and achievement, and for comparability with Reardon (2011b).

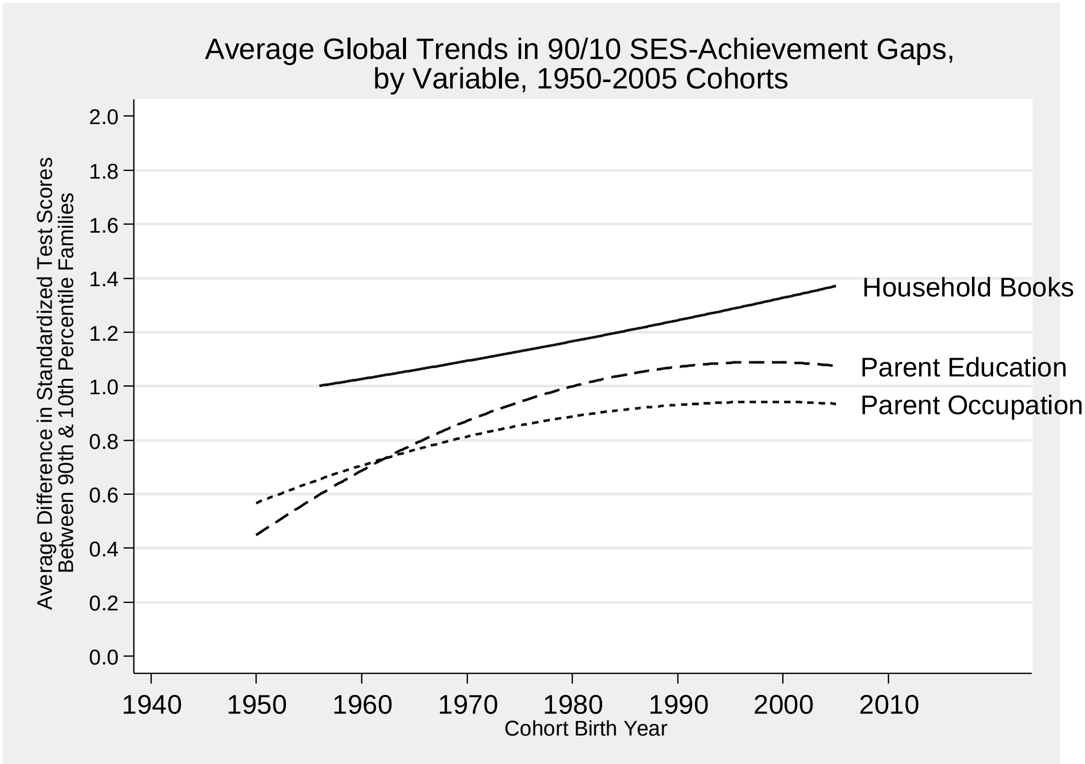
Model 1E does not include precision weights, which give greater weight to more precisely-estimated gaps in the hierarchical growth models. Results are similar for all three SES variables.

Model 1F omits the random effect for the cohort birth year slope and instead estimates a fixed cohort slope for all countries. Results are similar for all three SES variables. The random cohort slope is retained as the preferred model, as a chi-square test shows the variance in the random cohort slopes to be significantly different from 0 in the main models and all other trend model specifications. This is consistent with the large amount of variability across countries in SES achievement gap trends displayed in Figures 2, J1, and J2.

Model 1G includes three control variables that attempt to capture the quality of the SES data in each country-year observation, in terms of the resolution of information available in each SES variable. SES achievement gaps can be estimated more precisely when there are a larger number of categories of the SES variable and when the distribution of students is relatively even across categories, with not too great a proportion of observations in the top or bottom SES categories. Model 1G shows that, although these variables do appear to be related to the size of gaps, controlling for them does not explain away the positive cohort trend. Parent education and parent occupation gap trends remain almost identical in size. The household books gap trend is cut in half, likely reflecting that the share of students in the bottom category increases and the share in the top category decreases over time as families own fewer books, but the number of books owned becomes more strongly associated with achievement. After adding these controls, the annual gap increases for all three SES variables are exactly the same, at 0.005 SD per year. This suggests that differences in gap trends results across SES variables may reflect differences in data quality. However, as these data quality variables do not appear to explain away the finding of global increases in SES achievement gaps, they are omitted from other models.

Model 1H examines whether the cohort trend is curvilinear by including a squared cohort birth year variable. The squared term is positive and not significant for household books gaps, but it is negative and significant for both parent education and parent occupation gaps. This suggests that increases in achievement gaps based on parent education and occupation are leveling off or even declining slightly in recent years. Figure L1 plots estimated quadratic trends from Model 1H for all three SES variables. These results are consistent with declining income achievement gaps observed in recent years in the United States (Reardon and Portilla 2016) and in many countries participating in the PISA 2006 and 2015 science tests (OECD 2016).

Figure L1



Note: Estimates from Model 1H (see Tables L1-L3).

However, it is not clear that these quadratic trends represent a true reversal of the average global trend in SES achievement gaps. Model 1J estimates the quadratic trend omitting PISA data and shows that the squared cohort terms are positive for all three SES variables with this reduced sample. As we have already seen in Appendix H that the PISA parent education variable may be low-quality, especially in recent years, it is not clear which results are preferred. Thus, the simpler linear trend is retained as the main results.

M. Specification of multivariate model

As noted above in Appendix H on measurement error of SES and Appendix L on the specification of the trend model, SES achievement gap estimates can be affected by the reliability with which SES is reported and by the resolution of information available in the resulting SES variable. It is not possible to gauge exactly how much the estimated cohort trends are affected by these factors, although the analyses above demonstrate that it is unlikely that the global increases in SES achievement gaps could be explained away by these issues. Likewise, the coefficients for time-varying covariates in the multivariate models reported in the main text of the paper could also be affected by these issues. One way to address this is by introducing a dummy variable for each different study (e.g., TIMSS 1995 or PISA 2012) into the model. If we assume that all countries are likely to suffer from similar data reliability or quality issues in the same studies—a reasonable assumption because SES survey item wording is very similar across countries in a given year of a study—then these dummies control for study-specific biases in the estimation of SES achievement gaps. Table M1 reports Model 3, with such dummies added, alongside Model 2 from the main text of the paper. The cohort birth year coefficient is excluded from Model 3 as it would be collinear with the test dummies. In fact, the results from Model 3 are very similar to those from Model 2. The strongest predictors of SES achievement gaps remain Age when tracking begins (for all three variables), Fertility rate (for parent education; the parent occupation coefficient is reduced and loses significance), and School enrollment (for household books). Thus, the results from the multivariate models do not appear to be substantially biased by study-specific issues in the estimation of SES achievement gaps.

Table M1. Comparison between Multivariate Models Predicting 90/10 Achievement Gaps from Country Covariates and Cohort Birth Year or Test Dummies

	Education		Occupation		Books	
	(2)	(3)	(2)	(3)	(2)	(3)
Within countries						
Age 10 at testing	-0.061 +		-0.003		-0.278 ***	
	(0.036)		(0.027)		(0.038)	
Age 15 at testing	-0.179 ***				0.086 **	
	(0.026)				(0.026)	
Math	0.033 **	0.042 ***	0.015 +	0.025 ***	-0.027 +	-0.023 +
	(0.011)	(0.011)	(0.008)	(0.007)	(0.014)	(0.012)
Science	0.022 *	0.037 ***	0.010 +	0.010 +	0.032 *	0.031 **
	(0.010)	(0.009)	(0.006)	(0.005)	(0.014)	(0.011)
Cohort birth year	0.005		0.004		0.001	
	(0.005)		(0.004)		(0.004)	
Test dummies		x		x		x
School enrollment (proportion)	0.143	0.299 +	0.079	0.044	0.567 ***	0.478 **
	(0.221)	(0.181)	(0.208)	(0.188)	(0.170)	(0.162)
Immigrant background (proportion)	-0.204	-0.002	-0.511 *	0.255	0.087	0.938 *
	(0.339)	(0.275)	(0.250)	(0.420)	(0.351)	(0.376)
GDP per capita (logged)	-0.103	-0.134	-0.075	0.043	0.125	-0.059
	(0.108)	(0.106)	(0.089)	(0.110)	(0.082)	(0.106)
Income inequality (Gini)	1.075	-0.006	-0.423	-0.835	0.752	0.435
	(0.760)	(0.649)	(0.943)	(0.670)	(0.890)	(0.671)
Age when tracking begins	-0.046 **	-0.044 **	-0.013	-0.063 ***	-0.056 +	-0.068 ***
	(0.015)	(0.017)	(0.016)	(0.019)	(0.030)	(0.020)
Private school enrollment (proportion)	-0.184	-0.107	-0.054	-0.024	-0.013	0.042
	(0.141)	(0.140)	(0.106)	(0.129)	(0.193)	(0.181)
Fertility rate (births per woman)	-0.142 ***	-0.124 ***	-0.106 *	-0.045	-0.036	-0.031
	(0.038)	(0.027)	(0.041)	(0.033)	(0.034)	(0.036)
Higher education demand (expectations – enrollment)	0.053	0.121	0.059	-0.030	0.030	-0.050
	(0.087)	(0.102)	(0.070)	(0.069)	(0.102)	(0.086)
Between countries						
Intercept	0.704	0.506	1.304 **	1.027 *	0.007	0.033
	(0.546)	(0.553)	(0.499)	(0.514)	(0.533)	(0.559)
Mean school enrollment	0.195	0.190	-0.409	-0.336	0.602 *	0.609 *
	(0.241)	(0.258)	(0.316)	(0.330)	(0.289)	(0.302)
Mean proportion immigrant background	-0.267	-0.273	0.367	0.246	-0.081	-0.074
	(0.370)	(0.388)	(0.339)	(0.374)	(0.358)	(0.357)
Mean GDP per capita (logged)	0.084	0.088 +	0.044	0.062	0.174 ***	0.174 ***
	(0.053)	(0.053)	(0.042)	(0.044)	(0.043)	(0.045)
Mean income inequality	0.833 *	0.835 *	1.006 **	1.166 **	0.488	0.199
	(0.325)	(0.352)	(0.386)	(0.368)	(0.360)	(0.372)
Mean age when tracking begins	-0.054 ***	-0.053 ***	-0.04 **	-0.037 **	-0.057 ***	-0.058 ***
	(0.012)	(0.012)	(0.014)	(0.013)	(0.013)	(0.013)
Mean private school enrollment	0.057	0.051	0.244 *	0.191	0.065	0.111
	(0.129)	(0.136)	(0.118)	(0.127)	(0.169)	(0.167)
Mean fertility rate	-0.022	-0.015	-0.087	-0.114 +	-0.047	-0.015
	(0.047)	(0.052)	(0.057)	(0.061)	(0.049)	(0.057)
Mean higher education demand	-0.031	-0.085	-0.256	-0.248	-0.118	-0.195
	(0.206)	(0.212)	(0.164)	(0.170)	(0.139)	(0.144)
Residual variance (within countries)	0.01188	0.01284	0.00533	0.00786	0.01721	0.01825
Residual variance (country intercepts)	0.03154	0.03372	0.02387	0.02668	0.03139	0.03476
Residual variance (cohort slopes)	0.00004		0.0001		0.00012	
N (observations)	1510	1510	1146	1146	1738	1738
N (countries)	68	68	63	63	70	70

+ p<.1, * p<.05, ** p<.01, *** p < .001
Note: Model 2 is reproduced from Table 1 in the main text of the paper for comparison. Model 3 includes a dummy variable for each test instead of the cohort birth year variable (the age at testing control is excluded due to collinearity with test dummies; subject controls are included and are not collinear because many studies assess multiple subjects).

Finally, Table M2 reports multivariate results estimated using a fixed effects rather than a mixed effects model. The model includes country fixed effects, meaning that country-level covariates are omitted due to collinearity. The model is estimated as follows:

$$\hat{G}_{ij} = \gamma_{00} + \gamma_{10}Y_{ij} + (\boldsymbol{X}_{ij})\mathbf{B} + \boldsymbol{\Gamma_j} + \boldsymbol{\Lambda}_{ij},$$

where $\hat{G}_{ij}$ is the estimated gap in country j in country-study-year i, γ_{10} is the coefficient for cohort birth year Y_{ij}, $\boldsymbol{X}_{ij}$ is a vector of time-varying country covariates in country-year i, $\boldsymbol{\Gamma}$ is a vector of country dummy variables, $\boldsymbol{\Lambda}_{ij}$ is a vector of dummy variables indicating age at testing and test subject, $\boldsymbol{B}$ is a vector of coefficients for the time-varying country covariates. The country fixed effects are estimated using weighted least squares (weighted by the inverse squared standard error associated with each gap estimate) and are reported with robust Huber-White standard errors.

Results for the country fixed effects models are nearly identical to those in the main results in Table 1 of the main text for each SES variable, both in terms of the cohort trend estimates in Model 1 and the covariate results in Model 2. The similarity in cohort trend estimates to the hierarchical growth models is somewhat surprising, as Cohort birth year was entered uncentered into the hierarchical growth model, meaning that it measures historical time between the 1950-2005 cohorts. On the other hand, in the country fixed effects model, the Cohort birth year coefficient indicates the average annual increase in gaps for each cohort year relative to the mean of all years that a particular country participated in an international assessment. In contrast, the similarity of the multivariate results is expected, as time-varying covariates were country-mean centered (or "demeaned") in the hierarchical growth models, which is equivalent to country fixed effects. The differences in covariate coefficients between the fixed effects and hierarchical growth models are likely due to the difference in how Cohort birth year was treated in each model, as well as the difference in weighting by WLS versus a variance-known model.

Table M2. Country Fixed Effects Models Predicting 90/10 Achievement Gaps from Country Covariates and Cohort Birth Year

	Education		Occupation		Books	
	(1)	(2)	(1)	(2)	(1)	(2)
Age 10 at testing	-0.052 +	-0.090 ***	0.021	0.002	-0.241 ***	-0.282 ***
	(0.029)	(0.027)	(0.013)	(0.017)	(0.028)	(0.028)
Age 15 at testing	-0.188 ***	-0.189 ***			0.105 ***	0.094 ***
	(0.018)	(0.017)			(0.022)	(0.021)
Math	0.016	0.024	0.016	0.019 +	-0.051 *	-0.053 **
	(0.017)	(0.016)	(0.011)	(0.010)	(0.020)	(0.020)
Science	-0.002	0.004	0.009	0.011	0.036 *	0.034 *
	(0.018)	(0.017)	(0.009)	(0.009)	(0.018)	(0.017)
Cohort birth year	0.006 ***	0.006 *	0.003 **	0.003	0.011 ***	0.009 **
	(0.001)	(0.003)	(0.001)	(0.003)	(0.001)	(0.004)
School enrollment (proportion)		0.318 *		0.169		0.473 ***
		(0.160)		(0.149)		(0.138)
Immigrant background (proportion)		-0.100		0.034		0.889 *
		(0.316)		(0.247)		(0.426)
GDP per capita (logged)		-0.105		-0.039		-0.058
		(0.065)		(0.071)		(0.075)
Income inequality (Gini)		0.262		-0.670		1.111 +
		(0.523)		(0.549)		(0.602)
Age when tracking begins		-0.048 ***		-0.051 ***		0.000
		(0.012)		(0.012)		(0.021)
Private school enrollment (proportion)		-0.139		-0.043		0.052
		(0.130)		(0.099)		(0.131)
Fertility rate (births per woman)		-0.142 ***		-0.062 **		-0.011
		(0.023)		(0.022)		(0.031)
Higher education demand (expectations – enrollment)		0.153 *		0.038		-0.050
		(0.070)		(0.063)		(0.070)
Intercept	1.163 ***	2.810 ***	0.983 ***	2.282 **	1.415 ***	1.164
	(0.020)	(0.629)	(0.007)	(0.790)	(0.023)	(0.791)
Country fixed effects	x	x	x	x	x	x
Adjusted R^2	0.605	0.642	0.653	0.672	0.731	0.739
N (observations)	1510	1510	1146	1146	1738	1738
N (countries)	68	68	63	63	70	70

+ p<.1, * p<.05, ** p<.01, *** p < .001

References

Heyneman, Stephen P., and William A. Loxley. 1983. "The Effect of Primary-School Quality on Academic Achievement across Twenty-nine High- and Low-Income Countries." *American Journal of Sociology* 88(6):1162-94.

Jerrim, John, and John Micklewright. 2014. "Socio-economic Gradients in Children's Cognitive Skills: Are Cross-Country Comparisons Robust to Who Reports Family Background?" *European Sociological Review*:jcu072.

Kreidl, Martin, Harry BG Ganzeboom, and Donald J Treiman. 2014. "How did occupational returns to education change over time?" in *California Center for Population Research Online Working Paper Series PWP-CCPR-2014-017*: University of California - Los Angeles.

OECD. 2016. "PISA 2015 Results: Excellence and Equity in Education (Volume I)." Paris: OECD Publishing.

Reardon, Sean F. 2011a. "Appendices for 'The Widening Academic Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations'." Pp. 1-49 in *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane. New York: Russell Sage Foundation.

—. 2011b. "The Widening Academic-Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations." Pp. 91-115 in *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane. New York: Russell Sage Foundation.

Reardon, Sean F., and Ximena A. Portilla. 2016. "Recent trends in income, racial, and ethnic school readiness gaps at kindergarten entry." *AERA Open* 2(3):2332858416657343.