A Online Appendix for: "Do democracies provide better education? Revisiting the democracy–human capital link"

In this appendix we expand on information that is relevant for our empirical analysis, but that was presented in a condensed manner in the paper due to space constraints. Examples are discussions on the data material used for our main dependent variable and information about our multiple imputation model. Further, we present and discuss additional tests assess the robustness of the main findings from the paper. We also report and discuss more nuanced empirical evidence pertaining to mechanisms suggested by our theoretical argument and various extensions of the core results, e.g. considering more disaggregated education outcomes.

The more specific sequence of the sections in this Appendix is as follows: *First*, we provide a more detailed description and discussion of how the education quality data are constructed. *Second*, we present and discuss tests pertaining to the mechanisms highlighted by our argument. *Third*, we present additional details about our imputation model. We also discuss and present a number of tests to assess the performance of our imputation model.

Fourth, we include tables with robustness tests of the core models presented in the paper, which indicated that democracy is systematically related to education quantity. *Fifth*, we discuss and evaluate our instrumental variable(s) and present first-stage regressions from the core 2SLS specification as well as second-stage regressions from alternative 2SLS specifications. *Sixth*, we present a number of other robustness tests on democracy and education quality. *Seventh*, we report various tests on democracy and education quality including imputed data.

Eighth, we present results from a mediation analysis probing whether education quality has an effect on economic growth *through education quality*. *Ninth*, we report and discuss panel regression models on democracy and more disaggregated student achievement test measures pertaining to particular subjects and levels of education. *Tenth*, and finally, we present our extensions assessing different kinds of variability in education quality in democracies and in autocracies.

A.1 The education quality data

This section provides more details on the data used for constructing our main dependent variable, education quality. The measure that we use to capture this concept comes from the dataset on student achievement tests from Angrist, Patrinos and Schlotter (2013). This dataset covers, in total, 128 countries, including more than 40 countries from the developing world. The data are reported in 5-year intervals. The first observation is from 1965 and the final from 2010, but there is no full coverage over the whole period for any country, and some countries have better coverage than others (we note here that we deal with the issue of missing data in analyses where we conduct multiple imputation; see section A.3 for specifics). We note that no countries have test scores available from 1975 in the Agrist dataset, due to the paucity of international students achievements tests.

Since many countries, especially in the developing world, are not included in international tests such as PISA and TIMSS, regional student achievement test scores are linked to the international tests by Angrist, Patrinos and Schlotter (2013) through using the method proposed by Altinok and Murseli (2007): More specifically, this means that each regional test is linked to the international tests by using countries that participated in both regional and international tests as reference points. Moreover, different international tests are linked together using the United States as reference point. The reason for this is that the United States has participated in every international achievement test conducted during the past half century. In practical terms, this means that each country's test score is expressed relative to the test score result for the United States result in a given year. If countries participated in several different achievement tests in the same 5-year interval, the average between these (normalized) test scores is included in the data set.

All the international and regional achievement tests are conducted on students enrolled in either primary or secondary school, although the specific age/grade levels that are included vary from survey to survey (see the list provided directly below). The age/grade level ranges from second grade students at the lowest (surveyed in the African Programme d'Analyses des Systemes Educatifs de la Confemenat) to 15 and 16-year olds at the highest (the latter are surveyed in the PISA assessment). This means that although all the test scores are taken from primary and secondary school students, the different regional and international tests may not measure the same grade in a given year. According to Angrist, Patrinos and Schlotter (2013) this should not bias the results, however, as all the different test scores are transformed with the help on an index that accounts for varying scales of different tests as well as varying difficulty among different tests. We refer to the paper by Angrist, Patrinos and Schlotter (2013) for the specific formulaes used for aggregation.

As discussed in the paper, the dataset includes specific measures on test scores for maths, natural

sciences, and reading, and such disaggregated measures are provided for both primary and secondary level students. We mainly use the composite measure covering all these three subjects for both levels of education. The theoretical minimum and maximum values on this measure is 0 and 100, whereas the empirical minimum is 13.9 (Mauritania in 1995) and the empirical maximum is 74.3 (Japan in 1980). The mean score on the measure is 45.8, and the standard deviation is 9.7.

The following international and regional tests are listed by Angrist, Patrinos and Schlotter (2013) as having been included in the dataset:

- International tests:
  - Programme for International Student Assessment (PISA) – conducted for 15 and 16-year old students
  - Trends in International Mathematics and Science Study (TIMSS) – conducted for students in fourth and eight grade
  - Progress in International Reading Literacy Study (PIRLS) – conducted for students in fourth grade
- Regional tests:
  - South and Eastern African Consortium for Monitoring of Educational Quality (SAQMEQ) – conducted for students in third and fourth grade in a set of South and Eastern African countries
  - Programme d'Analyses des Systemes Educatifs de la Confemen (PASEC) – conducted for students in second and fifth grade in a set of countries in Francophone Africa
  - Three tests conducted by the UNESCO Laboratorio Latinoamericana de Evaluacion de la Calidad de la Education (LLCE)
    - Primer Estudio Regional Compatativo y Explicativo (PERCE) – conducted for students in third and fourth grades in a set of Latin American and Caribbean countries
    - Segundo Estudio Regional Compatativo y Explicativo (SERCE) – conducted for students in third and sixth grades in a set of Latin American and Caribbean countries
    - Tercer Estudio Regional Compatativo y Explicativo (TERCE) – conducted for students in third and sixth grades in a set of Latin American and Caribbean countries

A.2 Investigating mechanisms

In our theoretical argument we pointed to two features that may explain why having a democratic regime may not enhance education quality, despite the widely held notion that democratic politicians are more attentive than autocratic leaders when it comes to providing public goods and services that are beneficial for larger parts of the country's population. The first feature is that democracy may not – via the "standard" electoral accountability channel – lead democratic politicians to pursue policy measures that boost education quality. Rather, democratic politicians may want to focus their policy efforts in the education area on measures that are more visible to voters, such as expanding access to education. Second, we highlighted that *even if* (democratic or other) politicians would have wanted to pursue policies enhancing education quality, it is often unclear what these specific policies are, and politicians may lack the capacity to identify and implement the most appropriate measures.

Finding specific measures and designs that allow us to directly test these mechanisms in a general manner, using cross-national data, is hard (in this regard, we note that alternative designs drawing on data from particular country contexts may allow for more direct tests, as evidenced by the excellent study on the electoral incentive mechanism conducted by Harding and Stasavage, 2014). Yet, there are data that allow us to come somewhat closer to the relevant mechanisms, also at the cross-national level. Thus, to shed further light on whether any, or both, of the theorized mechanisms are at play, we conducted additional tests using indicators of education spending. More specifically, we test whether: 1) Democracy correlates with education spending (per student); 2) Education spending correlates with our measure of education quality.

Our argument suggests that democracies have incentives to pursue observable and presumably popular education policy reforms, such as boosting enrollment rates. While it is unclear by how much boosting enrollment rates increases total education spending – new classrooms may need to be built and new textbooks bought, but many classrooms may simply become more densely populated and textbooks may be shared – it is natural to assume that total education spending increases, *ceteris paribus*. Yet, given that democratic incumbents might prefer to promote enrollment rates *at the expense of education quality*, we might not expect to see that democracy strongly correlates with total education spending, and, especially, we should assume no, or even a negative, correlation with *education spending per student*. Nor should we expect to see that democracy correlates negatively with the number of students per teacher, rather to the contrary.

To assess these implications, we run models with various indicators of education spending from Unesco (2015) as dependent variable. The available data on education spending that is comparable in a global sample of countries has limited coverage, unfortunately, especially over time. Hence, these models rely

on cross-sectional data on education spending taken from the period 2005-2009 with all independent variables measured in 2000-2004, thus conforming with the core cross-sectional specifications that we use in the paper.[1] We underline that we should not draw very strong conclusions based on these results, since they are based on cross-sectional comparisons. Yet, the correlations that we obtain can offer some suggestive evidence as to why democracy does not enhance education quality. The main results are included in Table A1.

In brief, our tests provide no clear evidence that democracy increases education spending. The point estimate of Model 1, which draws on data from 120 countries and is a parsimonious specification only controlling for income level, suggests that going from -10 to +10 on Polity increases the percentage share of GDP going to education spending by about 0.5. Yet, we cannot reject the hypothesis that this could simply be due to noise, as the t-value of the Polity coefficient is only 0.7. The finding turns even weaker when adding controls for natural resources income, income inequality and population size in Model 2. Further, Models 3 (parsimonious) and 4 (extensive) show very low and insignificant correlations between democracy and spending per primary school student. Further, Models 5 and 6 even display negative, though statistically insignificant, Polity coefficients when spending per secondary school student is the dependent variable.

Table A1: Cross-country regressions on democracy and education spending

DV:	1 OLS Education spending (share of GDP)	2 OLS Education spending (share of GDP)	3 OLS Spending per student (primary)	4 OLS Spending per student (primary)	5 OLS Spending per student (secondary)	6 OLS Spending per student (secondary)
Polity Index	0.026	0.017	0.028	0.120	-0.156	-0.359
	(0.71)	(0.41)	(0.18)	(0.75)	(-0.57)	(-0.92)
Ln GDP p.c.	0.321*	0.259	1.807**	2.069**	-0.803	0.263
	(2.12)	(1.38)	(2.98)	(2.75)	(-0.65)	(0.17)
Ln oil + gas inc. p.c.		0.062		-0.437		-0.519
		(0.96)		(-1.57)		(-0.96)
Gini (market income; reversed)		-0.030		0.094		0.321+
		(-1.24)		(1.04)		(1.90)
Ln population		-0.428**		-0.462		-1.543+
		(-2.93)		(-0.99)		(-1.75)
Constant	1.659	10.737**	0.181	0.829	29.846**	31.602
	(1.36)	(2.79)	(0.04)	(0.08)	(2.90)	(1.44)
N	120	108	100	93	95	88

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. All independent variables are lagged by one 5-year period, and standard errors are robust.

The patterns in Models 1 and 2 might, at first glance, seem somewhat surprising, given that democracies have incentives to invest in education in order to boost education enrollment. At the same time, we note that widening the access to education can often be done *without raising education spending by too much*. For example, countries may fill up classrooms with more students without spending on hiring

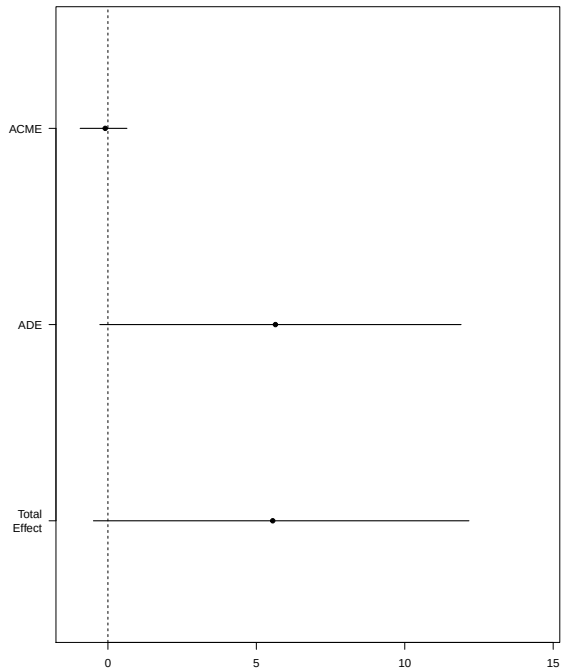
[1] Given that coverage is spotty for the spending variables, we can not take averages over the entire time interval and retain a decent number of observations. Instead, we use 2005 as the default year, and fill in with data from the closest available year within the time interval for those countries that have missing data on a particular spending variable in 2005.

new teachers, and otherwise cut costs per students.[2] When further investigating this, we do find evidence that the relationship between democracy and education quantity – as measured by enrollment rates or average years of schooling – that we replicated, also holds up fairly well when controlling for education spending, despite the reduced number of observations.

This is illustrated in Figure A.2, which displays results from a mediation analysis, following Imai, Keele, Tingley and Yamamoto (2011). This estimation procedure decomposes the average treatment effect of an explanatory variable into the average causal mediation effect (ACME) and the average direct effect (ADE). According to Imai et al. (2011), this is a more robust way of estimating mediation effects than simply adding the proposed mediator to the regression, and assessing the change in the coefficient estimates of the variable of interest when this is done. The analysis is conducted on the parsimonious model with primary enrollment as dependent variable, education spending as share of GDP as mediator, and the Polity Index as independent variable. The estimated direct "effect" of democracy on primary enrollment is substantial in size and (weakly) significant, whereas the estimated indirect effect of democracy via education spending is minuscule and statistically indistinguishable from zero. Hence, the relationship between democracy and education quantity does not seem to be mediated by increased education spending.

[2]Another piece of evidence suggesting that democracy may increase access to education without diverting too much resources comes from employing pupil to teacher ratios as the dependent variable, within the same cross-section set-up. While the finding is not robust to adding control variables or considering secondary schooling – although the Polity coefficient is always positively signed – the Polity coefficient is statistically significant at 5 percent in a parsimonious controlling for income level on primary education pupil to teacher ratios. The point estimate suggests that going from -10 to +10 on Polity increases the number of pupils per teacher by about 6 students, holding income level constant.

Figure A1: The mediation effect of democracy on primary education enrollment, mediated through education spending. Parsimonious model specification.



Notes: The total effect of democracy on education quantity is decomposed into the average direct effect (ADE) and the average causal mediation effect (ACME). The effects are plotted with their 95 percent confidence intervals. The pre-treatment covariate is Ln GDP per capita.

We now turn to the second part of our argument, concerning the difficulty of implementing education policies that boost the quality of education. Also here we test cross-section regressions, but use education spending as the independent variable and our main student achievements test score measure as the dependent variable. While spending constitutes a crude proxy, one would expect that if education quality is a fairly easily manipulable feature, it should react to politicians spending more resources on education. Conversely, if our assumption about education quality being hard to manipulate through policy is correct, it should not correlate strongly with education spending. Again, we warn that the reported results are cross-country regressions, and results may be biased by unobserved features affecting both education spending and education quality. Thus, this constitutes only suggestive evidence. Still, the available evidence largely points in the same direction as our argument. As Table A2 shows, education spending is not systematically correlated with our measure of education quality, neither when omitting (Models 1 and 2) or including (Models 3 and 4) a control for education quantity (average years of education). This does not mean that education quality is a randomly distributed property, as it correlates strongly with various structural features such as a high income level and low income inequality.

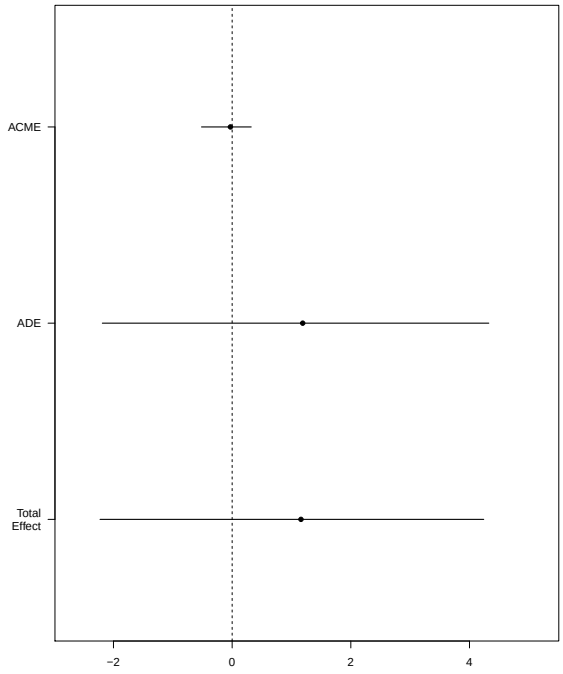
Table A2: Cross-country regressions on education spending and education quality

DV:	OLS Education quality	OLS Education quality	OLS Education quality	OLS Education quality
Total education spending (share of GDP)	-0.067	-0.011	-0.115	0.052
	(-0.17)	(-0.05)	(-0.37)	(0.22)
Polity Index	0.197	0.256	0.132	0.144
	(1.27)	(1.40)	(0.99)	(0.89)
Ln GDP p.c.	5.821**	6.431**	3.663**	5.651**
	(8.27)	(7.95)	(4.05)	(6.37)
Ln oil + gas income p.c.		-0.182		-0.387
		(-0.77)		(-1.66)
Gini (market income; reversed)		0.469**		0.389**
		(6.87)		(5.01)
Ln population		0.528		0.915+
		(1.11)		(1.68)
Average years of schooling 15 yr olds			1.533**	0.891**
			(4.69)	(3.01)
Constant	-8.855	-48.598**	-2.126	-50.270**
	(-1.50)	(-4.73)	(-0.34)	(-4.85)
N	89	82	84	78

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. All independent variables are lagged by 5 years, and standard errors are robust.

In sum, we find little evidence that democracy correlates with education spending in our sample, but also little evidence that education spending systematically relates to education quality. This is neatly summed up in Figure A.2, which presents the results of a mediation analysis (again following Imai et al., 2011) with Polity as independent variable, total education spending as share of GDP as mediator, and the test-score measure of education quality as dependent variable. Again, we employ the parsimonious model specification controlling for Ln GDP per capita and employ the same cross-country design as above. The estimated indirect effect running from democracy on education quality via education spending is virtually zero, but also the estimated direct effect of democracy on education quality is far from statistically significant at conventional levels.

Figure A2: The mediation effect of democracy on education quality, mediated through education spending. Parsimonious model specification.



Notes: The total effect of democracy on education quality is decomposed into the average direct effect (ADE) and the average causal mediation effect (ACME). The effects are plotted with their 95 percent confidence intervals. The pre-treatment covariate is Ln GDP per capita.

A.3 Imputation model and diagnostics

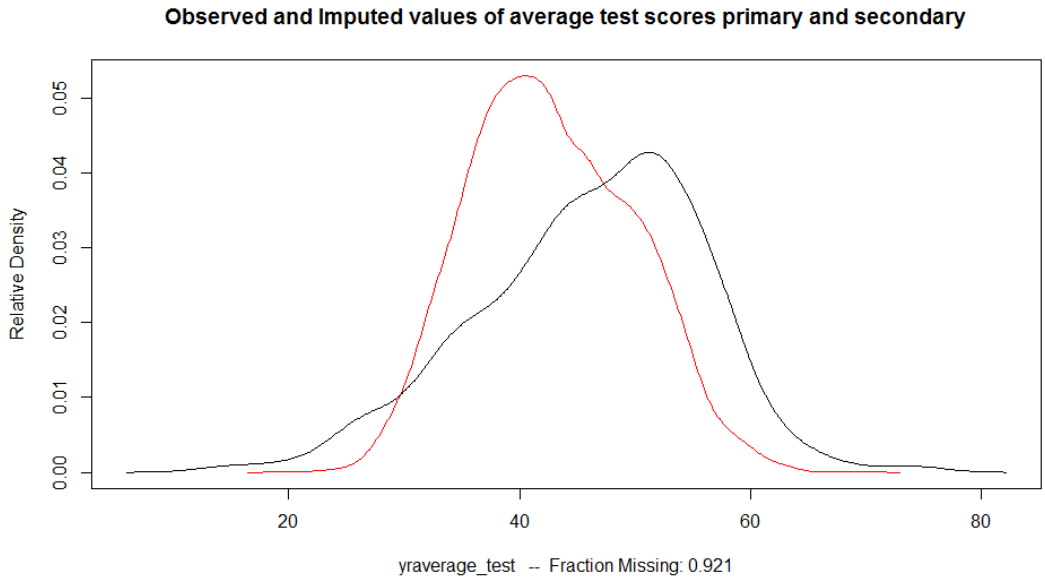
In order to construct our imputation model we employed the Amelia II software (Honaker and King, 2010; Honaker, King and Blackwell, 2012), accounting for the time-series cross-section data structure. We ran the imputation model and produced 5 different data sets, which were subsequently used in the empirical analysis of the paper. Our imputation model allowed for country-specific second-order polynomial time trends (see Honaker et al. 2012, 20-21). Following standard recommendations, we included several relevant variables in order to improve the predictive power of our multiple imputation model.

Multiple imputation builds on the assumption that the data are missing at random (MAR). This means that the pattern of "missingness" should only depend on the observed data included in the imputation model, and not on unobserved variables or features with the data (see Honaker and King 2010). The pattern of missingness on many of the key variables used in our analysis, such as the student achievement test measures, likely depends on a number of factors. For instance, the extent to which governments will allow student achievement tests to be carried out in their country could depend on the level of democracy. Moreover, poorer countries are less likely to have participated in student achievement tests, for instance due to the difficulties of organizing the required capacity for conducting tests. To account for such issues we included a wide range of variables in the imputation model that may influence the missingness, including level of democracy, indicators of human rights, economic growth, GDP per capita, and urbanization. This likely reduces the threat of the MAR-assumption being grossly violated.

Yet, we can not exclude the possibility that the MAR-assumption is still not fully met, for instance because the missingness depends on unobserved factors, and that this causes the imputation model to perform poorly. To evaluate the outputs from the imputation model we therefore conducted different imputation diagnostic tests that help us to assess its quality. One way to describe the outputs from the imputation model in a condensed manner, which can potentially also be used as a check on the plausibility of the imputation model specification, is mapping the distribution of imputed values and the distribution of observed values for particular variables. While very large discrepancies between the observed and imputed distributions may provide warning signs that something may be wrong with the imputation model, one should not necessarily expect the distributions of the missing values to be completely identical to the distributions of the observed values. In fact, the main reason why we impute to begin with is the assumption that observed and missing values could differ systematically; correcting for this alleviates the issue of selection biases affecting our regression results (e.g., Honaker and King, 2010).

Plots showing the distributions of imputed and observed values for the average student achievement

Figure A3: Distributions of observed and imputed observations for aggregated measure of student achievement test scores



test measure from Angrist, Patrinos and Schlotter (2013) are presented in Figure A.1. The red line represents the density of the mean of each imputed observation across all 5 datasets, while the black line represents the density of the observed values. The figure shows that the shape of the two distributions follow a similar pattern, very roughly, but still deviate especially when it comes to the lower scores on student achievement tests, where there are higher relative frequencies for the imputed data. This is, however, a very plausible result: As noted, there are reasons to expect that poorer countries will have more missing values on the student achievement test measure than rich countries, and income is strongly correlated with education quality. Hence, the distribution of the imputed values for the aggregated student achievement test measure may suggest that the imputation procedure alleviates some of the selection biases discussed above related to the under-representation of developing countries. In Figure A.2 and Figure A.3 we present similar plots for student achievements test measures from Angrist, Patrinos and Schlotter (2013), which separately measure performance for students at the primary and secondary levels of schooling, respectively, and the patterns are similar to that for the aggregated index.

Another way of assessing the imputation model is by employing the so-called overimputation procedure (Honaker, King and Blackwell, 2012). Since missing data are, by definition, unobservable, it is impossible to tell whether each imputed value is close to the unobserved value that we would like the imputation model to predict accurately. An overimputation test is a diagnostics test that attempts to

Figure A4: Distributions of observed and imputed observations for student achievement test measure for primary school students

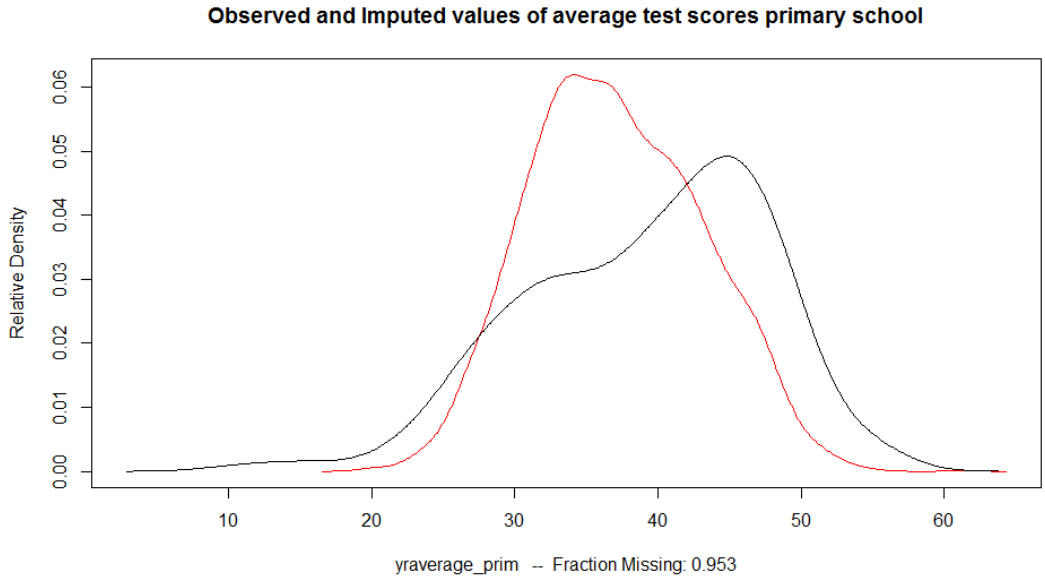
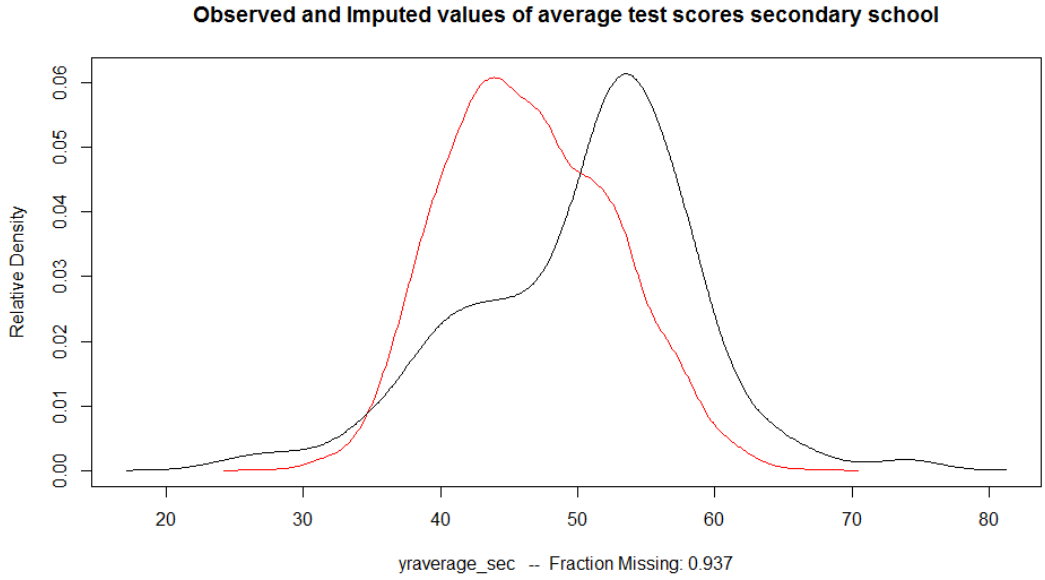


Figure A5: Distributions of observed and imputed observations for student achievement test measure for secondary school students



solve this problem by (sequentially) treating each observed value *as if* it is actually missing, and then generate a large number of imputed values for it. After we have this large set of imputed values, we can then construct a confidence interval, and the final step is then assessing whether the actually observed value falls within this interval. If it does, our imputation model performs well in the sense that it is able to accurately predict this "fictive missing data" point (which we *know* the actual score on).

The Amelia II software provides this diagnostics test as well as a simple way of graphically inspecting whether our observed data tends to fall within the region where it is imputed by the model. Figure A.4 plots the results for the overimputation test for the average student achievement test score. The observed test scores are plotted against the mean of the imputed scores for the very same observation on this index (when the observed test score is treated as if it is missing). The 90 percent confidence intervals that are plotted for each imputed value thus allow us to visually inspect the performance of the imputation model. As noted by Honaker et al. (2012, 30), checking how many of the confidence intervals cover the $y = x$ line lets us tell how often the imputation model can confidently predict the true value of the observation. The line colors on the confidence intervals tell the fraction of missing observations for all covariates entered in the imputation model for that particular observation.

As we can see from Figure A.4, our imputation model performs well when it comes to predicting the observed values. Every single confidence interval, except from two, covers the 45 degree line. The predictions are especially good for the medium to higher values of the student achievement index. However, the imputation model is worse at predicting the observed values at the lowest level of the student achievement test score measure, but we note that there are very few observations in this area. In Figures A.5 and A.6 we plot the results for the overimputation tests on the primary level and secondary level test score measures, respectively, and they show a similar pattern: The imputation model, overall, seems to predict well, even though it performs slightly worse when it comes to predicting the very lowest values on the student achievement test measures.

Figure A6: Overimputation test plot, aggregated measure of student achievement test scores

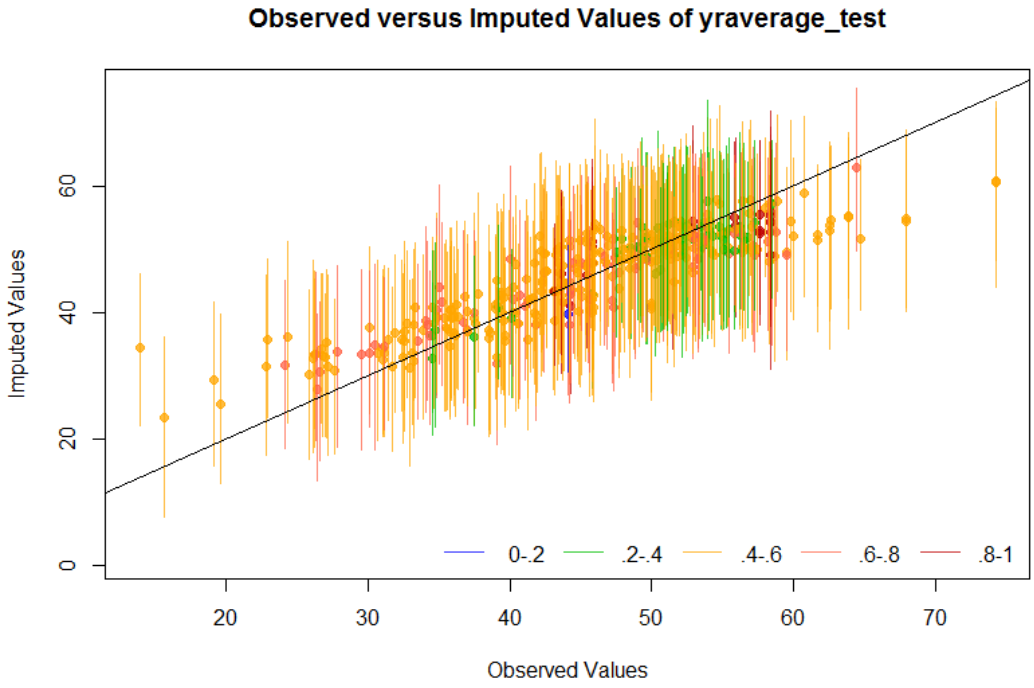


Figure A7: Overimputation test plot, student achievement test measure for primary school students

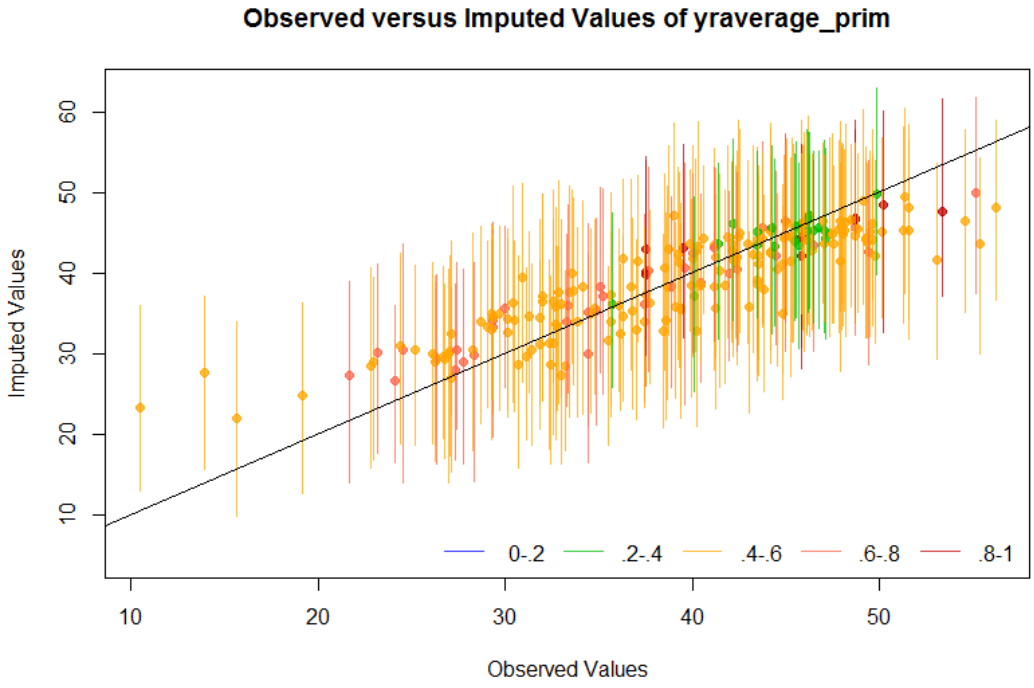
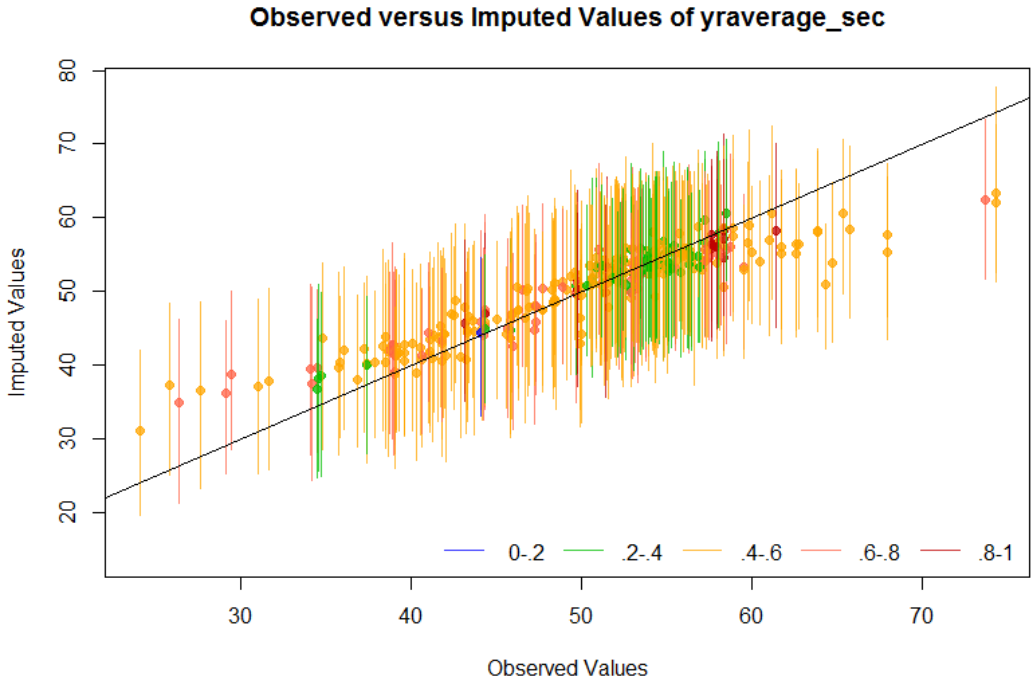


Figure A8: Overimputation test plot, student achievement test measure for secondary school students



A.4 Robustness tests: Democracy and measures of "education quantity"

In this section we present robustness tests on the relationship between democracy and *education quantity*. First, we replicate Table 1 in the paper, but substitute the Polity Index with the Freedom House Index. The two subsequent tables show results for models substituting the average years of schooling measure from Table 1 with measures of enrollment rates. The final table shows results from models based on alternative estimators.

In short, the models below consistently display a positive point estimate for democracy on the different measures of "education quantity", and while several models do not yield coefficients that are significant at conventional levels, many others do. Albeit not robust, we thus find evidence that democracy relates positively to educational aspects such as enrollment rates and average years spent in school.

Table A3: Democracy, as measured by the Freedom House Index (average of the Political Rights and Civil Liberties indices), and average years of schooling

	A1 OLS Cross-sec. b/t	A2 OLS Cross-sec. b/t	A3 OLS (FE) 5-yr panel b/t	A4 OLS (FE) 5-yr panel b/t	A5 FE 2SLS 5-yr panel b/t	A6 FE 2SLS 5-yr panel b/t
Freedom House Index	0.200+	0.230	0.372**	0.018	1.203**	0.112
	(1.74)	(1.47)	(6.90)	(0.51)	(8.36)	(1.39)
Ln GDP p.c.	1.840**	1.807**	1.177**	0.525*	0.708**	0.577**
	(10.18)	(7.57)	(2.71)	(2.57)	(2.78)	(4.33)
Ln oil+gas income p.c.		0.107		-0.083+		-0.082*
		(1.42)		(-1.67)		(-2.21)
Gini (market income;reversed		0.040		0.006		0.007
		(1.44)		(0.84)		(1.35)
Ln population		-0.240+		1.072*		1.163**
		(-1.77)		(2.31)		(4.41)
Period dummies				Y		Y
Country dummies			Y	Y	Y	Y
N	120	104	828	599	818	591

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is average years of schooling for 25 year olds. All independent variables are lagged by 5 years, and standard errors are robust. WAVE from Knutsen (2011) is used as instrument for Freedom House in the FE 2SLS models (see Appendix for first stage regressions). First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

Table A4: Democracy and primary school enrollment rates

	A1 OLS Cross-sec. b/t	A2 OLS Cross-sec. b/t	A3 OLS (FE) 5-yr panel b/t	A4 OLS (FE) 5-yr panel b/t	A5 FE 2SLS 5-yr panel b/t	A6 FE 2SLS 5-yr panel b/t
Polity Index	0.040	-0.179	1.030**	0.220	2.041**	0.638
	(0.24)	(-0.69)	(4.56)	(1.42)	(6.33)	(1.46)
Ln GDP p.c.	0.975	1.953	3.504	-5.334	1.558	-4.261+
	(0.68)	(1.14)	(1.30)	(-1.62)	(0.79)	(-1.65)
Ln oil+gas income p.c.		-0.441		-0.022		0.005
		(-1.11)		(-0.04)		(0.01)
Gini (market income;reversed)		-0.319*		0.142		0.142
		(-2.24)		(1.09)		(1.23)
Ln population		1.894*		29.807**		30.053**
		(2.20)		(3.83)		(6.62)
Period dummies				Y		Y
Country dummies			Y	Y	Y	Y
N	133	119	1101	788	1094	777

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is primary school enrollment rates. All independent variables are lagged by 5 years, and standard errors are robust. WAVE from Knutsen (2011) is used as instrument for Freedom House in the FE 2SLS models (see Appendix for first stage regressions). First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

Table A5: Democracy and secondary school enrollment rates

	A1 OLS Cross-sec. b/t	A2 OLS Cross-sec. b/t	A3 OLS (FE) 5-yr panel b/t	A4 OLS (FE) 5-yr panel b/t	A5 FE 2SLS 5-yr panel b/t	A6 FE 2SLS 5-yr panel b/t
Polity Index	-0.006	0.351	1.131**	0.123	2.193**	0.355
	(-0.02)	(0.96)	(7.04)	(0.81)	(8.43)	(1.15)
Ln GDP p.c.	22.010**	21.899**	22.448**	11.465**	19.916**	11.848**
	(13.54)	(13.03)	(7.16)	(4.98)	(10.24)	(6.13)
Ln oil+gas income p.c.		0.997+		0.481		0.493
		(1.71)		(1.06)		(1.30)
Gini (market income;reversed)		0.506*		0.067		0.072
		(2.08)		(0.79)		(1.00)
Ln population		-2.398*		14.258**		14.318**
		(-2.29)		(2.78)		(4.49)
Period dummies				Y		Y
Country dummies			Y	Y	Y	Y
N	127	115	1047	747	1044	740

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is secondary school enrollment rates. All independent variables are lagged by 5 years, and standard errors are robust. WAVE from Knutsen (2011) is used as instrument for Freedom House in the FE 2SLS models (see Appendix for first stage regressions). First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

Table A6: Democracy and mean years of schooling. Alternative estimators

	A1 OLS PCSE 5-year panel b/t	A2 OLS PCSE 5-year panel b/t	A3 OLS (RE) 5-yr panel b/t	A4 OLS (RE) 5-yr panel b/t	A5 System gmm 5-yr panel b/t	A6 System GMM 5-yr panel b/t
Polity Index	0.129**	0.111**	0.121**	0.014	0.002	-0.000
	(11.80)	(8.56)	(8.82)	(1.64)	(0.36)	(-0.01)
Ln GDP p.c.	1.740**	1.746**	1.912**	0.639**	0.212**	0.013
	(23.95)	(18.68)	(8.68)	(3.97)	(3.86)	(0.14)
Ln oil+gas income p.c.		-0.029				0.012
		(-0.89)				(0.60)
Gini (market income;reversed)		0.069**				0.001
		(8.74)				(0.36)
Ln population		-0.126*				0.379**
		(-2.24)				(3.44)
Lagged DV (Avg yrs schooling)					1.014**	0.963**
					(68.84)	(33.19)
Period dummies				Y		Y
Country dummies			Y	Y	Y	Y
N	1027	707	1027	1027	982	688

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean years of schooling. All independent variables are lagged by 5 years, and standard errors are robust. WAVE from Knutsen (2011) is used as instrument for Freedom House in the FE 2SLS models (see Appendix for first stage regressions). First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

A.5 2SLS models: Discussion, first-stage regressions for core specifications, and alternative 2SLS models using different instruments

In this section we include two tables displaying the first-stage regression estimates from four instrumental variable regression models (Fixed Effects 2SLS), for which the second-stage regressions were presented in Tables 1 and 2 in the paper. In these first-stage regressions, the dependent variable is the Polity Index, and the WAVE instrument from Knutsen (2011) is included as an independent variable alongside all the covariates included also in the second-stage regressions. In brief, the first-stage regressions show that WAVE has a strong and highly significant negative relationship with democracy, both when adding country fixed effects and country fixed effects combined with period dummies. Also F-tests show that this instrument is very strong, with F-values (see bottom row of the two tables) far exceeding those considered threshold values for strong instruments (e.g., $F = 10$). Also other weak identification tests, such as comparing the Stock-Yogo critical values against the Cragg-Donald Wald F-statistic, leave no doubt that WAVE is a very strong instrument, and the t-values of WAVE range between -4.6 and -13.7.

Table A7: First-stage regressions (with Polity Index as dependent variable) for Fixed Effects 2SLS models presented in Table 1 of the paper.

	F1 5-year panel b/t	F2 5-year panel b/t t
WAVE	-6.597**	-5.078**
	(-13.64)	(-7.25)
Ln GDP p.c	1.838**	-1.400+
	(4.85)	(-1.83)
Ln oil+gas income p.c.		0.131
		(0.93)
Gini (market income;reversed)		0.009
		(0.28)
Ln population		-3.303**
		(-2.57)
Period dummies		Y
Country dummies	Y	Y
N	1019	698
F-test value excl. instrument	98.36	52.50

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is democracy, measured using the Polity Index. Standard errors are robust. First time period is 1960–64, and final time period is 2000–2004. WAVE from Knutsen (2011), recording whether or not the last regime change was within or outside one of Huntington's "reverse waves of democratization", is used as instrument for democracy in the first-stage regression. Constant, country dummies and period dummies are omitted.

Table A8: First-stage regressions (with Polity Index as dependent variable) for Fixed Effects 2SLS models presented in Table 2 of the paper.

	F1 5-year panel b/t	F2 5-year panel b/t t
WAVE	-7.893**	-7.853**
	(-4.76)	(-4.59)
Ln GDP p.c	2.341**	-0.860
	(4.22)	(-0.56)
Ln oil+gas income p.c.		0.109
		(0.59)
Gini (market income;reversed)		-0.035
		(-0.67)
Ln population		-2.687
		(-1.40)
Period dummies		Y
Country dummies	Y	Y
N	317	297
F-test value excl. instrument	22.69	21.09

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is democracy, measured using the Polity Index. Standard errors are robust. First time period is 1960–64, and final time period is 2000–2004. WAVE from Knutsen (2011), recording whether or not the last regime change was within or outside one of Huntington's "reverse waves of democratization", is used as instrument for democracy in the first-stage regression. Constant, country dummies and period dummies are omitted.

Table A9 presents second-stage results on education quality from the alternative fixed effects 2SLS specifications that we tested. The two rightmost columns (IV1 and IV2) report the models also reported in the paper, namely the parsimonious and extensive models employing WAVE as instrumental variable. As we also discuss in the paper (for a more extensive discussion, see Knutsen, 2011), this instrument aims to capture exogenous influence on regime type, notably coming from regional and global spill-over effects and features of the international-political climate, through leveraging the notion that democratic regime changes have tended to cluster temporally in "waves" (Huntington, 1991). As we will return to, standard over-identification tests of the exclusion restriction can only be conducted in the presence of two different instruments. While this is not an accurate test of the exclusion restriction (Sovey and Green, 2011), we do, however, take some solace in the fact that when we re-run the second-stage regressions (as standard fixed effects OLS models), but including WAVE directly in the second stage, WAVE is not directly linked to education quality. More specifically, the t-value of WAVE is 0.16 in the parsimonious model and -0.45 in the extensive model. This is not bullet-proof evidence that the exclusion restriction is valid in Models IV1 and IV2, but they do mitigate concerns that these estimates are strongly biased by a clear, direct link between WAVE and education quality.

An alternative, and more direct, way of capturing spill-over effects on regime type from neighbors at the regional level is to calculate the "democracy environment" in the geographic region and use that as an instrument. When doing so, we naturally exclude the country in question whenever scoring the regional democracy environment variable. More specifically, we take the average Polity score within the country's geographic region, omitting the relevant country from the calculation. We rely on the geographic region coding in Hadenius and Teorell (2007), but collapse some of the smaller regions so that we are left with the following major world regions: Eastern Europe and the (post-)Soviet space; Western Europe plus Australia, Canada, New Zealand and USA; Latin America and the Caribbean; Middle East and North Africa; Sub-Saharan Africa; Asia and the Pacific.

The exogeneity of this instrument might be somewhat more problematic than for WAVE, in particular for relatively larger countries. While WAVE draws on a broad (and global) temporal categorization to identify exogenous variation in political regime type, which are arguably less influenced by events in most countries, regime changes and other political events in a regional power (such as Brazil in Latin America) might influence the regional democracy score. Nonetheless, for most countries we consider that treating the democracy environment in the wider regions, at any given point in time, as an exogenous feature should be a decent approximation. Regarding the exclusion restriction we are – as we are for the models using WAVE – making the (plausible) assumption that the wider international political regime environment should not have any direct effect on education quality in the country in question (other than through affecting regime change domestically), conditional on the period dummies, income level

and other covariates in the model. Regarding strength, the regional average polity score instrument turns out highly significant in the first-stage, and the F-statistics (see bottom row Table A9) show that the instrument passes any conventional thresholds for being considered a strong instrument.

Model IV3 presents results from the parsimonious model, controlling only for GDP per capita, country dummies and year dummies, that uses the regional average instrument. As for the models using WAVE, the Polity coefficient is negatively signed in the second stage, but statistically insignificant at all conventional levels. In Model IV4, which adds control for resource income, income inequality and population, Polity remains negatively signed and actually turns significant at the 5 percent level.

We also tested FE 2SLS model simultaneously including WAVE and the regional average instrument, and again the Polity coefficient turns out statistically insignificant in the parsimonious model (IV5) and negative and significant at 5 percent in the extensive model (IV6). A benefit of using two instruments is that we can run standard overidentification tests on the validity of the exclusion restriction. The Hansen j-test p-value for Model IV5 is .19, suggesting that we cannot reject the hypothesis that the exclusion restriction holds at conventional levels, whereas the p-value is .08 for Model IV6.

Still, the WAVE and regional average instruments draw on a similar underlying logic related to geographic spill-over effects, and overidentification tests may not give accurate results under such conditions (Murray, 2006). Therefore, we also tested models combining our preferred instrument, WAVE, with a quite different instrument. More specifically, we follow Helliwell (1994) in using the 15-year lagged value of Polity as an instrument for current Polity (which, in turn, is measured five years before the dependent variable). This instrument draws on the notion that democratic institutions are often persistent, suggesting that lagged democracy is a strong instrument. Further, we need to assume that there is no direct effect of historical political regime type on education quality other than its influence through affecting the current regime. This is a very strong assumption, for instance due to the potential long time lag between education policy decisions and improvements in education quality discussed in the paper. This is why we do not include the lagged Polity variable as an instrument in our main specifications reported in the paper. Still, we tested it to gauge the results of Hansen j-tests based on two very different instruments, and they actually suggest that the parsimonious model (IV7) might yield consistent estimates (p=.61) whereas the extensive models (IV8) might not (p=.05). The results for Polity, in both models, suggest a negative effect of democracy on education quality. Despite the outcome of the Hansen j-test for Model IV7, we note that we put less trust in these models given our theoretical concerns with exclusion restriction in these specifications. Nonetheless, we note that neither these nor any of the other 2SLS models that we have tested show any indications of a positive effect of democracy on education quality.

Table A9: Democracy and mean student achievements test score: Second-stage regressions from alternative FE 2SLS specifications using different instruments

	IV1 b/t	IV2 b/t	IV3 b/t	IV4 b/t	IV5 b/t	IV6 b/t	IV7 b/t	IV8 b/t
Polity Index	-0.186	-0.017	-0.343	-0.564*	-0.273	-0.335*	-0.434*	-0.394*
	(-0.67)	(-0.07)	(-1.22)	(-2.25)	(-1.22)	(-2.14)	(-2.56)	(-2.49)
Ln GDP p.c.	1.360	1.924	1.771	1.560	1.490	1.712	-0.518	-0.559
	(0.92)	(0.77)	(0.73)	(0.56)	(0.69)	(0.65)	(-0.31)	(-0.23)
Ln oil+gas income p.c.		-0.323		-0.265		-0.289		0.514
		(-0.64)		(-0.48)		(-0.55)		(1.20)
Gini (market inc.; reversed)		0.013		-0.003		0.004		-0.007
		(0.16)		(-0.03)		(0.05)		(-0.07)
Ln population		7.879*		9.626*		8.896*		4.165
		(2.07)		(2.31)		(2.30)		(1.03)
Period dummies	Y	Y	Y	Y	Y	Y	Y	Y
N	317	297	262	297	262	297	286	265
Instruments	WAVE		Regional average Polity		WAVE + Reg. avg. Polity		WAVE + Polity 15 yrs lag	
Hansen j-test p-val.	–	–	–	–	.186	.078	.609	.053
Cragg-D. Wald F-stat.	80.8	69.9	86.4	98.8	63.5	87.9	72.2	55.1

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean years of schooling. Time series units are 5-year panels, and all independent variables are lagged by 5 years, and standard errors are robust. Polity Index is the endogenous independent variable, and instrument(s) used in the first stage are reported towards bottom of the table. First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

A.6 Other robustness tests: Democracy and education quality

The following tables contain robustness tests of our core result on democracy and education quality. The first table employs alternative estimation techniques (OLS PCSE, RE, System GMM) to the core tests reported in Table 2. The following tables reports result where Table 2 from the paper is replicated using other widely used measures of democracy than the Polity index. Then follows a table where the independent variables from Table 2 are lagged 10, rather than 5, years behind the outcome. Thereafter, we present results for similar robustness tests of Table 3 of the paper. At the end, we report tables with models including a squared Polity term or dummies capturing different ranges on the Polity scale, in order to check for non-linearities, models checking for interaction between democracy and income level/income inequality/quality of government on education quality, as well as models using the autocratic regime type dummies from Hadenius and Teorell (2007).

Table A10: Democracy and mean student achievements test score: Alternative estimators

	C1 OLS PCSE 5-yr panel b/t	C2 OLS PCSE 5-yr panel b/t	C3 RE 5-yr panel b/t	C4 RE 5-yr panel b/t	C5 System GMM 5-yr panel b/t	C6 System GMM 5-yr panel b/t
Polity Index	0.039	-0.023	0.025	0.008	0.243	0.000
	(0.43)	(-0.24)	(0.16)	(0.07)	(0.71)	(0.00)
Ln GDP p.c.	6.566**	6.956**	5.440**	6.088**	2.628	5.376
	(15.35)	(14.60)	(8.29)	(9.94)	(1.30)	(1.56)
Ln oil+gas income p.c.		-0.433**		-0.371*		-1.691**
		(-3.06)		(-2.10)		(-3.43)
Gini (market income; reversed)		0.330**		0.261**		0.248
		(5.87)		(4.40)		(1.47)
Ln population		1.320**		1.072*		0.793
		(3.48)		(2.51)		(0.39)
Lagged dep. var.					-0.123	0.047
					(-1.34)	(0.20)
Period dummies		Y		Y		Y
N	341	321	341	321	190	185

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. All independent variables are lagged by 5 years, and standard errors are robust for random effects and system GMM models, and account for panel-specific AR(1) autocorrelation and panel-level heteroskedasticity in OLS PCSE models. The System GMM models treat the Polity Index as endogenous, and there are no restrictions on the number of lags used for instrumentation. First time period is 1965–69, and final time period is 2005–2009. Constant and period dummies are omitted from the table.

Table A11: Democracy, as measured by the Freedom House Index(average of political and civil liberties indices), and mean student achievement test scores.

	B1 OLS Cross-sec. b/t	B2 OLS Cross-sec. b/t	B3 OLS (FE) 5-yr panel b/t	B4 OLS (FE) 5-yr panel b/t	B5 FE 2SLS 5-yr panel b/t	B6 FE 2SLS 5-yr panel b/t
Freedom House Index	0.244	0.133	-0.228	-0.326	-1.604+	-0.477
	(0.44)	(0.25)	(-0.25)	(-0.54)	(-1.95)	(-0.65)
Ln GDP p.c.	5.891**	7.013**	-1.432	0.720	-0.324	0.719
	(7.07)	(8.05)	(-0.69)	(0.26)	(-0.20)	(0.33)
Ln oil+gas income p.c.		-0.464+		0.401		0.405
		(-1.68)		(1.04)		(1.05)
Gini (market income;reversed)		0.409**		-0.039		-0.042
		(5.93)		(-0.43)		(-0.47)
Ln population		1.043+		3.030		2.950
		(1.88)		(0.63)		(0.74)
Period dummies				Y		Y
Country dummies		Y	Y	Y	Y	Y
N	94	86	328	307	302	282

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. All independent variables are lagged by 5 years, and standard errors are robust. WAVE from Knutsen (2011) is used as instrument for the Freedom House Index in the FE 2SLS models (see Appendix for first stage regressions). First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

Table A12: Democracy, as measured by the Democracy–Dictatorship (DD) measure, and mean student achievement test scores.

	B1 OLS Cross-sec. b/t	B2 OLS Cross-sec. b/t	B3 OLS (FE) 5-yr panel b/t	B4 OLS (FE) 5-yr panel b/t	B5 FE 2SLS 5-yr panel b/t	B6 FE 2SLS 5-yr panel b/t
DD	2.745	0.870	-0.039	-0.314	-2.936	-0.234
	(1.61)	(0.56)	(-0.01)	(-0.15)	(-0.66)	(-0.07)
Ln GDP p.c.	5.676**	6.980**	0.789	1.931	1.324	1.932
	(8.30)	(9.44)	(0.48)	(0.65)	(0.90)	(0.78)
Ln oil+gas income p.c.		-0.454		-0.321		-0.322
		(-1.62)		(-0.67)		(-0.64)
Gini (market income;reversed)		0.402**		0.014		0.014
		(6.02)		(0.17)		(0.18)
Ln population		0.996+		7.878+		7.864*
		(1.75)		(1.84)		(2.09)
Period dummies				Y		Y
Country dummies		Y	Y	Y	Y	Y
N	94	86	344	323	317	297

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. All independent variables are lagged by 5 years, and standard errors are robust. WAVE from Knutsen (2011) is used as instrument for DD (from Cheibub, Gandhi and Vreeland, 2010) in the FE 2SLS models (see Appendix for first stage regressions). First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

Table A13: Democracy and mean student achievements test score. 10-year lag.

	B1 OLS Cross-sec. b/t	B2 OLS Cross-sec. b/t	B3 OLS (FE) 5-yr panel b/t	B4 OLS (FE) 5-yr panel b/t	B5 FE 2SLS 5-yr panel b/t	B6 FE 2SLS 5-yr panel b/t
Polity Index	0.145	0.063	0.084	-0.115	-0.042	-0.166
	(0.95)	(0.33)	(0.71)	(-1.00)	(-0.32)	(-1.16)
Ln GDP p.c.	5.652**	6.489**	-0.517	0.525	-0.099	0.451
	(7.49)	(7.64)	(-0.29)	(0.18)	(-0.07)	(0.19)
Ln oil+gas income p.c.		-0.340		-0.330		-0.334
		(-0.95)		(-0.68)		(-0.73)
Gini (market income;reversed)		0.416**		-0.132		-0.130+
		(4.55)		(-1.55)		(-1.81)
Ln population		0.754		11.337*		11.633**
		(1.32)		(2.34)		(2.68)
Period dummies				Y		Y
Country dummies		Y	Y	Y	Y	Y
N	91	83	339	313	313	288

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. All independent variables are lagged by 10 years, and standard errors are robust. WAVE from Knutsen (2011) is used as instrument for Freedom House in the FE 2SLS models (see Appendix for first stage regressions). First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

Table A14: Democracy and mean student achievement test scores, checking for a non-linear relationship by including a squared term for Polity.

	B1 OLS Cross-sec. b/t	B2 OLS Cross-sec. b/t	B3 OLS (FE) 5-yr panel b/t	B4 OLS (FE) 5-yr panel b/t	B5 FE 2SLS 5-yr panel b/t	B6 FE 2SLS 5-yr panel b/t
Polity Index	1.092	0.639	1.683**	1.091+	-2.215	1.180
	(1.58)	(0.78)	(2.65)	(1.71)	(-0.54)	(0.35)
Polity Index squared	-0.043	-0.021	-0.076**	-0.051+	0.096	-0.055
	(-1.42)	(-0.63)	(-2.71)	(-1.85)	(0.54)	(-0.38)
Ln GDP per capita	6.595**	7.109**	1.065	2.542	0.758	2.592
	(6.85)	(6.61)	(0.62)	(0.81)	(0.46)	(0.79)
Ln oil+gas income p.c.		-0.434		-0.205		-0.197
		(-1.45)		(-0.44)		(-0.32)
Gini (market income;reversed)		0.430**		0.047		0.050
		(5.88)		(0.54)		(0.34)
Ln population		0.956+		5.907		5.743
		(1.70)		(1.55)		(0.83)
Period dummies				Y		Y
Country dummies		Y	Y	Y	Y	Y
N	91	84	341	321	317	297

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. All independent variables are lagged by 5 years, and standard errors are robust. WAVE from Knutsen (2011) is used as instrument for the linear term of the Polity Index in the FE 2SLS models (squared Polity term is not instrumented). First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

Table A15: Democracy and mean student achievement test scores, checking for a non-linear relationship by using autocracy, partial democracy, full democracy categorization from Epstein et al. (2006).

	B1 OLS Cross-sec. b/t	B2 OLS Cross-sec. b/t	B3 OLS (FE) 5-yr panel b/t	B4 OLS (FE) 5-yr panel b/t	B5 FE 2SLS 5-yr panel b/t	B6 FE 2SLS 5-yr panel b/t
Full Democracy	0.956	0.506	-1.270	-1.045	-3.520	-0.783
	(0.49)	(0.30)	(-0.58)	(-0.61)	(-0.73)	(-0.22)
Partial Democracy	0.795	1.068	-0.399	0.545	-1.554	0.684
	(0.43)	(0.60)	(-0.25)	(0.32)	(-0.50)	(0.27)
Ln GDP per capita	6.010**	7.224**	1.035	1.867	1.498	1.859
	(7.35)	(9.46)	(0.61)	(0.62)	(0.99)	(0.75)
Ln oil+gas income p.c.		-0.455		-0.325		-0.328
		(-1.58)		(-0.67)		(-0.65)
Gini (market income;reversed)		0.412**		0.012		0.012
		(5.68)		(0.14)		(0.14)
Ln population		1.056+		7.752+		7.730*
		(1.91)		(1.86)		(2.06)
Period dummies				Y		Y
Country dummies		Y	Y	Y	Y	Y
N	91	84	341	321	317	297

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. All independent variables are lagged by 5 years, and standard errors are robust. Partial democracy ($1 \leq Polityscore \leq 7$) and full democracy ($Polityscore \geq 8$) are categorized using the criteria from Epstein et al. (2006). WAVE from Knutsen (2011) is used as instrument for Full Democracy in the FE 2SLS models (squared Polity term is not instrumented). First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

Table A16: Democracy and mean student achievement test scores, checking for an interaction between democracy and income level.

	B1 OLS Cross-sec. b/t	B2 OLS Cross-sec. b/t	B3 OLS (FE) 5-yr panel b/t	B4 OLS (FE) 5-yr panel b/t	B5 FE 2SLS 5-yr panel b/t	B6 FE 2SLS 5-yr panel b/t
Polity Index	-1.757	0.882	2.625+	-0.275	-8.271	0.453
	(-1.25)	(0.63)	(1.83)	(-0.17)	(-0.54)	(0.05)
Ln GDP per capita	4.618**	7.356**	2.796	1.809	-5.828	2.124
	(4.22)	(6.14)	(1.37)	(0.57)	(-0.46)	(0.41)
PolityXLn GDP p.c.	0.224	-0.090	-0.318+	0.030	1.011	-0.058
	(1.40)	(-0.57)	(-1.77)	(0.15)	(0.55)	(-0.05)
Ln oil+gas income p.c.		-0.421		-0.320		-0.327
		(-1.41)		(-0.66)		(-0.65)
Gini (market income;reversed)		0.436**		0.014		0.009
		(5.75)		(0.18)		(0.10)
Ln population		1.016+		8.305+		7.150
		(1.77)		(1.68)		(0.44)
Period dummies				Y		Y
Country dummies		Y	Y	Y	Y	Y
N	91	84	341	321	317	297

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. All independent variables are lagged by 5 years, and standard errors are robust. WAVE from Knutsen (2011) is used as instrument for the Polity Index in the FE 2SLS models (squared Polity term is not instrumented). First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

Table A17: Democracy and mean student achievement test scores, checking for an interaction between democracy and income inequality.

	B2 OLS Cross-sec. b/t	B4 OLS (FE) 5-yr panel b/t	B6 FE 2SLS 5-yr panel b/t
Polity Index	0.430	1.381+	0.624
	(0.41)	(1.73)	(0.29)
Gini (market income;reversed)	0.460**	0.098	0.053
	(3.49)	(0.79)	(0.32)
PolityXGini	-0.005	-0.023+	-0.011
	(-0.26)	(-1.88)	(-0.32)
Ln GDP per capita	6.777**	2.198	2.034
	(8.25)	(0.80)	(0.81)
Ln oil+gas income p.c.	-0.434	-0.444	-0.378
	(-1.35)	(-0.93)	(-0.73)
Ln population	1.006+	5.435	6.806
	(1.76)	(1.16)	(1.21)
Period dummies		Y	Y
Country dummies		Y	Y
N	84	321	297

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. All independent variables are lagged by 5 years, and standard errors are robust. WAVE from Knutsen (2011) is used as instrument for the Polity Index in the FE 2SLS models (squared Polity term is not instrumented). First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

Table A18: Democracy and mean student achievement test scores, controlling for quality of government.

	B1 OLS Cross-sec. b/t	B2 OLS Cross-sec. b/t	B3 OLS (FE) 5-yr panel b/t	B4 OLS (FE) 5-yr panel b/t	B5 FE 2SLS 5-yr panel b/t	B6 FE 2SLS 5-yr panel b/t
Polity Index	0.106	0.058	0.021	-0.048	-0.217	-0.523*
	(0.68)	(0.29)	(0.14)	(-0.27)	(-0.89)	(-1.99)
Quality of Government Index	12.577*	8.500+	-1.145	-6.428	1.231	0.230
	(2.26)	(1.74)	(-0.25)	(-1.18)	(0.23)	(0.03)
Ln GDP per capita	3.824**	5.536**	3.122+	1.302	3.909*	-0.370
	(3.18)	(4.49)	(1.71)	(0.39)	(2.18)	(-0.12)
Ln oil+gas income p.c.		-0.470		0.078		0.213
		(-1.66)		(0.15)		(0.44)
Gini (market income;reversed)		0.439**		-0.030		-0.011
		(5.96)		(-0.27)		(-0.11)
Ln population		0.666		1.078		0.845
		(1.14)		(0.15)		(0.16)
Period dummies				Y		Y
Country dummies		Y	Y	Y	Y	Y
N	84	77	282	265	260	244

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. All independent variables are lagged by 5 years, and standard errors are robust. WAVE from Knutsen (2011) is used as instrument for the Polity Index in the FE 2SLS models (squared Polity term is not instrumented). First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

Table A19: Democracy and mean student achievement test scores, checking for an interaction between democracy and quality of government.

	B1 OLS Cross-sec. b/t	B2 OLS Cross-sec. b/t	B3 OLS (FE) 5-yr panel b/t	B4 OLS (FE) 5-yr panel b/t	B5 FE 2SLS 5-yr panel b/t	B6 FE 2SLS 5-yr panel b/t
Polity Index	-1.057+	-0.735	0.551*	0.473	5.835	10.099
	(-1.81)	(-1.17)	(2.46)	(1.20)	(0.29)	(0.39)
Quality of Government Index	-4.921	-3.349	2.349	-3.643	20.376	9.330
	(-0.44)	(-0.30)	(0.48)	(-0.63)	(0.30)	(0.23)
PolityXQoG	2.054+	1.365	-1.005*	-0.941	-9.102	-14.739
	(1.98)	(1.27)	(-2.52)	(-1.60)	(-0.30)	(-0.39)
Ln GDP per capita	4.264**	5.934**	2.747	1.604	-3.619	13.030
	(3.50)	(4.39)	(1.60)	(0.49)	(-0.15)	(0.38)
Ln oil+gas income p.c.		-0.602*		0.026		-1.304
		(-2.05)		(0.05)		(-0.35)
Gini (market income;reversed)		0.409**		-0.052		-0.459
		(5.14)		(-0.46)		(-0.37)
Ln population		0.699		-0.311		-19.711
		(1.24)		(-0.04)		(-0.34)
Period dummies				Y		Y
Country dummies		Y	Y	Y	Y	Y
N	84	77	282	265	260	244

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. All independent variables are lagged by 5 years, and standard errors are robust. WAVE from Knutsen (2011) is used as instrument for the Polity Index in the FE 2SLS models (squared Polity term is not instrumented). First time period is 1965–69, and final time period is 2005–2009. Constant, country dummies and period dummies are omitted from the table.

Table A20: Democracy (reference category), autocracy types and mean student achievement test scores.

	B1 OLS Cross-sec. b/t	B2 OLS Cross-sec. b/t	B3 OLS (FE) 5-yr panel b/t	B4 OLS (FE) 5-yr panel b/t
Monarchy	-4.646+	-0.866	-46.432**	-37.976**
	(-1.78)	(-0.26)	(-10.78)	(-5.09)
Military	-1.918	-3.818	-3.986	-3.787
	(-0.47)	(-0.89)	(-1.08)	(-1.15)
One party	12.761**	0.000	3.855	2.116
	(11.50)	(.)	(0.93)	(0.79)
Multi-party	3.814*	3.662**	2.775+	3.769*
	(2.20)	(2.66)	(1.78)	(2.19)
Other autocracy	0.057	-1.803	-25.736**	-19.829**
	(0.03)	(-1.43)	(-6.97)	(-3.53)
Ln GDP per capita	6.833**	7.701**	-1.556	-0.375
	(11.03)	(12.28)	(-0.71)	(-0.09)
Ln oil+gas income p.c.		-0.507+		-0.141
		(-1.88)		(-0.25)
Gini (market income;reversed)		0.383**		0.032
		(5.77)		(0.30)
Ln population		1.120*		1.813
		(2.03)		(0.38)
N	94	86	295	276
Period dummies				Y
Country dummies		Y	Y	Y
N	94	86	295	276

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. Standard errors are robust. First time period is 1970–74, and final time period is 2005–2009. Regime dummies are taken from Hadenius and Teorell (2007), and democracy is the reference category. Since time series for the regime dummies start in 1972, these are not lagged by 5 years to preserve the number of observations. Other independent variables are lagged by 5 years. Constant, country dummies and period dummies are omitted from the table.

A.7 Additional robustness tests on education quality: Models run on imputed data

In this section we present different Random Effects and Fixed Effects models on the imputed data sets. The coefficients reported are averaged over regressions run on our five imputed datasets, and the errors are imputation-corrected. We report fairly parsimonious specifications controlling for income level (both with and without year-fixed effects). We also present models controlling for enrollment ratios to account for the alternative explanation of our core result discussed in the paper (related to democracies expanding schooling to students that expectedly perform less well in tests). Tests are conducted both on (levels of) mean student achievement test scores, and on first-differences of such mean test scores (i.e., changes in mean test scores between t-1 and t). Further, we also report models employing alternative lag structures (1, 5 and 10 year lags). Again, the models show no systematic relationship, although one fixed-effects first-difference model suggests that democracy reduces education quality, whereas two random effects models (on levels) suggest a positive effect of democracy when independent variables are lagged by 10 years.

Table A21: Democracy and mean student achievements test score

	D1 RE 1-yr panel b/t	D2 OLS (FE) 1-yr panel b/t	D3 RE 1-yr panel b/t	D4 OLS (FE) 1-yr panel b/t	D5 RE 1-yr panel b/t	D6 OLS (FE) 1-yr panel b/t
Polity index	0.094	0.024	0.083	-0.037	0.073	-0.041
	(1.38)	(0.34)	(1.29)	(-0.57)	(1.10)	(-0.62)
Ln GDP per capita	5.074**	3.414**	4.701**	1.478	2.824*	0.786
	(5.12)	(3.51)	(3.92)	(1.33)	(2.56)	(0.76)
Primary school enrollment ratio					-0.014	-0.019
					(-0.65)	(-0.95)
Secondary school enrollment ratio					0.078**	0.025+
					(5.08)	(1.99)
Year dummies			Y	Y	Y	Y
Country dummies		Y		Y		Y
N (per imputed dataset)	4914	4914	4914	4914	4914	4914
Countries	117	117	117	117	117	117

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. Country-year is unit of analysis, and all independent variables are lagged by 1 year. The coefficients are averaged over 5 imputed data sets, and standard errors are robust and imputation-corrected. The time series run from 1970–2011 (balanced panel). The log-transformed GDP p.c. variable is constructed as ln(GDP p.c.+18000), due to the multiple imputation being conducted without bounds and the resulting occurrence of negative GDP p.c. values in the imputed data sets (see Appendix for details on the imputation model). Constant, country dummies and period dummies are omitted from the table.

Table A22: Democracy and first differences in mean student achievements test score

	D1 RE 1-yr panel b/t	D2 OLS (FE) 1-yr panel b/t	D3 RE 1-yr panel b/t	D4 OLS (FE) 1-yr panel b/t	D5 RE 1-yr panel b/t	D6 OLS (FE) 1-yr panel b/t
Polity index	-0.009	-0.036	-0.011	-0.062	-0.034*	-0.066
	(-0.63)	(-1.13)	(-0.79)	(-1.56)	(-2.03)	(-1.64)
Ln GDP per capita	-0.625	-1.144	-0.707	-1.926	-2.309+	-2.374
	(-0.96)	(-0.96)	(-0.94)	(-1.11)	(-1.92)	(-1.35)
Primary enrollment					-0.005	-0.001
					(-0.66)	(-0.09)
Secondary enrollment					0.026*	0.042*
					(2.63)	(2.39)
Year dummies			Y	Y	Y	Y
Country dummies		Y		Y		Y
N (per imputed dataset)	4914	4914	4914	4914	4914	4914
Countries	117	117	117	117	117	117

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is first differences in mean student achievement test scores. Country-year is unit of analysis, and all independent variables are lagged by 1 year. The coefficients are averaged over 5 imputed data sets, and standard errors are robust and imputation-corrected. The time series run from 1970–2011 (balanced panel). The log-transformed GDP p.c. variable is constructed as ln(GDP p.c.+18000), due to the multiple imputation being conducted without bounds and the resulting occurrence of negative GDP p.c. values in the imputed data sets (see Appendix for details on the imputation model). Constant, country dummies and period dummies are omitted from the table.

Table A23: Democracy and mean student achievements test score. 5-year lag

	D1 RE 1-yr panel b/t	D2 OLS (FE) 1-yr panel b/t	D3 RE 1-yr panel b/t	D4 OLS (FE) 1-yr panel b/t	D5 RE 1-yr panel b/t	D6 OLS (FE) 1-yr panel b/t
Polity index	0.094	0.024	0.087	-0.025	0.082	-0.026
	(1.71)	(0.41)	(1.63)	(-0.46)	(1.63)	(-0.48)
Ln GDP per capita	5.693**	3.908**	5.584**	2.244	4.014*	1.899
	(5.31)	(3.38)	(4.30)	(1.58)	(2.48)	(1.04)
Primary enrollment					-0.006	-0.016
					(-0.40)	(-1.02)
Secondary enrollment					0.066*	0.011
					(2.57)	(0.56)
Year dummies			Y	Y	Y	Y
Country dummies		Y		Y		Y
N (per imputed dataset)	4914	4914	4914	4914	4914	4914
Countries	117	117	117	117	117	117

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. Country-year is unit of analysis, and all independent variables are lagged by 5 years. The coefficients are averaged over 5 imputed data sets, and standard errors are robust and imputation-corrected. The time series run from 1970–2011 (balanced panel). The log-transformed GDP p.c. variable is constructed as ln(GDP p.c.+18000), due to the multiple imputation being conducted without bounds and the resulting occurrence of negative GDP p.c. values in the imputed data sets (see Appendix for details on the imputation model). Constant, country dummies and period dummies are omitted from the table.

Table A24: Democracy and mean student achievements test score. 10 year lag

	D1 RE 1-yr panel b/t	D2 OLS (FE) 1-yr panel b/t	D3 RE 1-yr panel b/t	D4 OLS (FE) 1-yr panel b/t	D5 RE 1-yr panel b/t	D6 OLS (FE) 1-yr panel b/t
Polity index	0.083*	0.006	0.082*	-0.032	0.074	-0.037
	(2.10)	(0.14)	(1.97)	(-0.77)	(1.66)	(-0.90)
Ln GDP per capita	5.698**	3.592*	5.409*	1.831	3.768*	1.662
	(3.65)	(2.29)	(2.86)	(1.03)	(2.17)	(0.97)
Primary enrollment					0.019	0.009
					(0.92)	(0.47)
Secondary enrollment					0.062*	0.006
					(3.00)	(0.25)
Year dummies			Y	Y	Y	Y
Country dummies		Y		Y		Y
N (per imputed dataset)	4914	4914	4914	4914	4914	4914
Countries	117	117	117	117	117	117

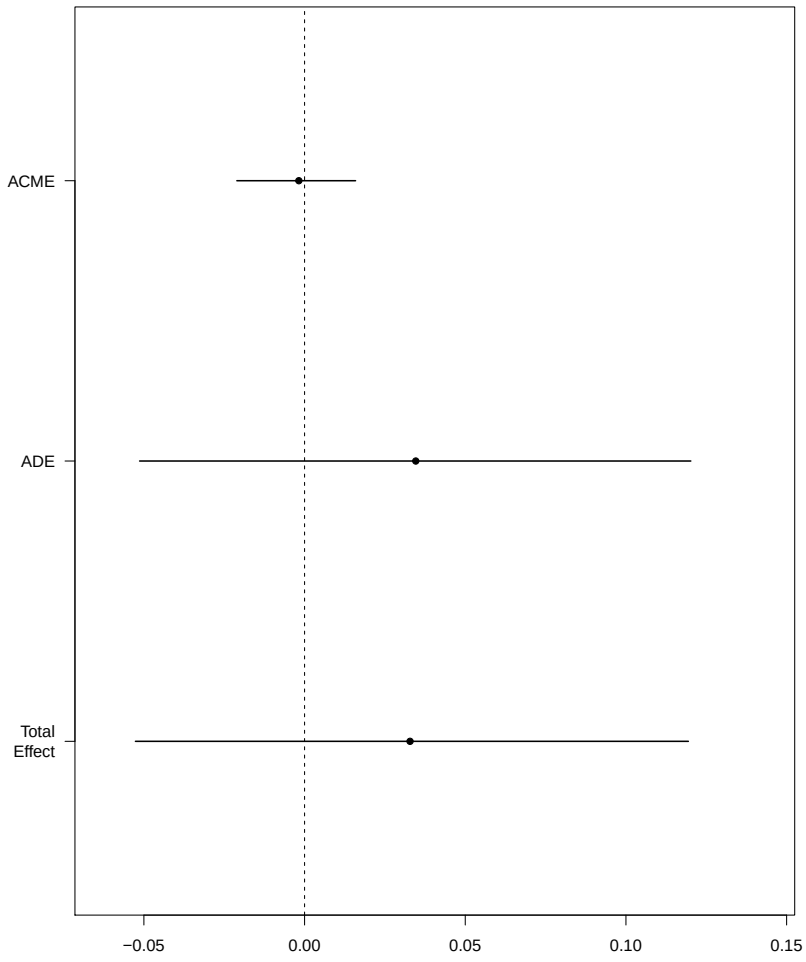
Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. Country-year is unit of analysis, and all independent variables are lagged by 10 years. The coefficients are averaged over 5 imputed data sets, and standard errors are robust and imputation-corrected. The time series run from 1970–2011 (balanced panel). The log-transformed GDP p.c. variable is constructed as ln(GDP p.c.+18000), due to the multiple imputation being conducted without bounds and the resulting occurrence of negative GDP p.c. values in the imputed data sets (see Appendix for details on the imputation model). Constant, country dummies and period dummies are omitted from the table.

A.8 Mediation analysis on democracy, education, and growth

This section presents results from simple mediation analysis, as proposed by Imai et al. (2011). The mediation analysis is estimated using the mediation package in R, where the total effect of democracy on economic growth is decomposed into the average direct effect of democracy (ADE) on growth and the mediation effect of democracy on growth through education quality (Tingley, Yamamoto, Hirose, Keele and Imai, 2014). The latter is referred to as the average causal mediation effect (ACME). Figure A9 shows results from mediation analysis estimated with log gdp per capita, log oil and gas income, the gini index and log population as included controls. In brief, no indirect effects of democracy on growth through education quality is identified. Nor is there any direct effect of democracy on growth.

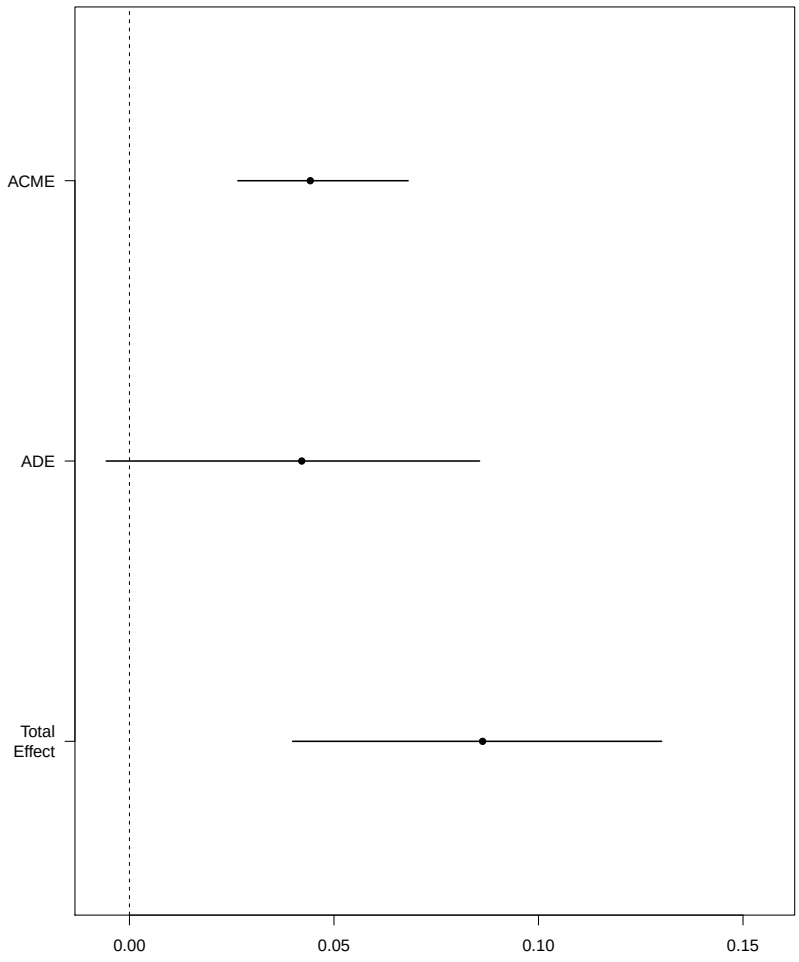
In contrast, there is evidence that democracy impacts on economic growth through education quantity. In figure A10, the indicator of education quality is replaced by an indicator of education quantity, years of education. Hence, this mediation analysis model decomposes the total effect of democracy on economic growth into the average direct effect of democracy (ADE) on growth and the mediation effect of democracy on growth through education *quantity*. In brief, this model identifies a mediation effect of democracy on economic growth through education quantity. The average causal mediation effect is estimated to 0.05, and is significantly different from zero. Moreover, it estimates that 51 percent of the total effect of democracy on economic growth is through education quantity as a mediating variable.

Figure A9: Mediation effects of democracy on economic growth, mediated through education quality



Notes: The total effect of democracy on economic growth is decomposed into the average direct effect (ADE) and the average causal mediation effect (ACME). The effects are plotted with their 95 percent confidence intervals. Education quality is measured using the indicator of mean student achievement test scores from Angrist et al, and democracy is measured using the Polity index. The pre-treatment covariates are log gdp per capita, log oil and gas income, the gini index and log population.

Figure A10: Mediation effects of democracy on economic growth, mediated through education quantity



Notes: The total effect of democracy on economic growth is decomposed into the average direct effect (ADE) and the average causal mediation effect (ACME). The effects are plotted with their 95 percent confidence intervals. Education quantity is measured using mean years of education from the Barro-Lee dataset, and democracy is measured using the Polity index. The pre-treatment covariates are log gdp per capita, log oil and gas income, the gini index and log population.

A.9 Extension: Democracy and education quality by subject and level of education

In this section we present the first extension discussed in the paper, which considers how democracy relates to measures of education quality from the Angrist, Patrinos and Schlotter (2013) dataset that are restricted to particular subjects. These subjects are mathematics, natural sicences and reading skills, and these measures are disaggregated to cover the primary and secondary level of schooling separately.

Table A25 we replicate a Random Effects model – which is more efficient than a corresponding Fixed Effects model, but potentially biased – from Table 3 (Model C3) for the six disaggregated components. They provide no evidence that democracy is positively related to math skills. However, we do find indications that democracy is positively related to reading skills, both at the primary (Model D2) and secondary level (D5); the coefficient estimates for Polity are positive and significant at 1%. Meanwhile, there is no effect of democracy on science achievements at the secondary level (D6), but the coefficient estimate for democracy is positive and significant for primary-level science achievements (D3). Nonetheless, these results are far from robust.

In Table A26 we replicate Model C4 from Table 3, which is a Fixed Effects model, for the various subcomponent measures. When including country dummies, democracy is no longer systematically related to knowledge and skills in any of the three subjects, neither at the primary nor at the secondary level, with Polity t-values ranging from -1.2 to +1.0. Hence, while there are some indications that democracy is good for stimulating reading skills, the evidence is far from robust and disappears when we account for country-fixed effects.

Table A25: Democracy and subcomponents of student test scores. Random effects models.

	D1 Maths primary 1-yr panel b/t	D2 Reading primary 1-yr panel b/t	D3 Science primary 1-yr panel b/t	D4 Maths secondary 1-yr panel b/t	D5 Reading secondary 1-yr panel b/t	D6 Science secondary 1-yr panel b/t
Polity index	0.0531	0.118**	0.173***	0.0585	0.132**	0.0480
	(1.04)	(2.73)	(3.84)	(0.94)	(3.24)	(1.05)
Ln GDP per capita	3.900*	5.175***	4.268***	4.958***	4.898***	4.300**
	(2.42)	(4.10)	(4.01)	(3.75)	(4.27)	(3.11)
Year dummies	Y	Y	Y	Y	Y	Y
N (per imputed dataset)	4914	4914	4914	4914	4914	4914
Countries	117	117	117	117	117	117

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. Country-year is unit of analysis, and all independent variables are lagged by 1 year. The coefficients are averaged over 5 imputed datasets, and standard errors are robust and imputation-corrected. The time series run from 1970–2011 (balanced panel). The log-transformed GDP p.c. variable is constructed as ln(GDP p.c.+18000), due to the multiple imputation being conducted without bounds and the resulting occurrence of negative GDP p.c. values in the imputed datasets (see appendix for details on the imputation model). Constant and year dummies are omitted from the table.

Table A26: Democracy and subcomponents of student test scores. OLS (Fixed effects) models.

	E1 Maths primary 1-yr panel b/t	E2 Reading primary 1-yr panel b/t	E3 Science primary 1-yr panel b/t	E4 Maths secondary 1-yr panel b/t	E5 Reading secondary 1-yr panel b/t	E6 Science secondary 1-yr panel b/t
Polity index	-0.0563	0.0230	0.0480	-0.0539	0.0398	-0.0483
	(-0.97)	(0.55)	(1.02)	(-0.85)	(0.97)	(-1.18)
Ln GDP per capita	1.290	2.240+	1.082	2.013+	2.356*	1.625
	(1.00)	(1.82)	(1.05)	(1.71)	(2.48)	(1.23)
Constant	27.05*	13.97	29.71**	30.29*	22.39*	29.47*
	(2.02)	(1.11)	(2.81)	(2.47)	(2.17)	(2.19)
Year dummies	Y	Y	Y	Y	Y	Y
Country dummies	Y	Y	Y	Y	Y	Y
N (per imputed dataset)	4914	4914	4914	4914	4914	4914
Countries	117	117	117	117	117	117

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. T-values in parentheses. Dependent variable is mean student achievement test scores. Country-year is unit of analysis, and all independent variables are lagged by 1 year. The coefficients are averaged over 5 imputed datasets, and standard errors are robust and imputation-corrected. The time series run from 1970–2011 (balanced panel). The log-transformed GDP p.c. variable is constructed as ln(GDP p.c.+18000), due to the multiple imputation being conducted without bounds and the resulting occurrence of negative GDP p.c. values in the imputed datasets (see appendix for details on the imputation model). Constant, country dummies and period dummies are omitted from the table.

A.10 Extensions: Tests on variance between and within countries, by regime type, and tests on within-population variability in PISA 2012 test scores

This section pertains to the various extensions investigating relationships between democracy and the variability of education quality, both at the macro-level as well as for variation within populations.

Table A27 reports F-statistics from so-called Goldfeld-Quandt tests. These tests assesses whether there is a systematic variation in regression residuals between relatively democratic and relatively autocratic countries. The cut-off on the Polity Index used here is 6 – observations scoring higher than or equal to 6 are considered relatively democratic regimes. The upper rows report a test where the residuals come from a cross-section OLS regression on the final time-period (2005–2009) with Ln GDP per capita as control variable, thus yielding a test on differential (between-country) variation between the group of autocratic countries and the group of democratic countries. The lower rows report a test where residuals are drawn from a fixed-effects OLS regression, again with Ln GDP per capita as a control, over all time periods. Hence, this provides a test on differential within-country variation (over time) in autocracies and in democracies. While the differences in variation are not huge, they are statistically significant. This suggests that education quality displays more variation in autocracies than in democracies.

Table A27: Goldfeld-Quandt test results, based on residuals from cross-section OLS (between-country variance, 2005–2009) and OLS fixed effects regression (within-country variance, all time periods).

	Sum square residuals	Number of obs.	F-value (aut./dem.)	Critical F-value (5 percent)
Between-country variance test				
Autocracies	1367.1	26		
Democracies	2147.6	65		
			1.8	1.7
Within-country variance test				
Autocracies	10840.4	79		
Democracies	17067.7	213		
			1.8	1.3

Notes: The regressions used for generating the residuals control for Ln GDP per capita. The regimes are grouped so that countries with Polity Index score lower than 6 are considered autocracies, and equal to or higher than 6 are considered democracies. The critical F-values are approximate (interpolated from F-statistic tables), and correspond to the 5 percent significance level

Tables A28–A30 provide correlations between regime type and different measures based on data from the PISA 2012 test. Table A28 display cross-section regression results on variability in performance within a country's population (measured as the standard error of individual PISA tests scores in the country). We ran models separately for test score results in mathematics, science, and reading. Both regressions controlling for and not controlling for the mean test score of the country are reported. Table A29 displays regression results on absolute gender differences (point differential between average male student and average female student) for the 2012 PISA tests. Table A30 takes the mean PISA 2012 test

score as dependent variable, hence testing our main hypotheses on a smaller sample using a different measure. Regime type is not found to matter, neither for individual variability in tests scores, gender differences, nor mean PISA test scores.

Table A28: Democracy and standard error in PISA 2012 test scores within countries' student populations

Test subject	D1 Math OLS b/(t)	D2 Science OLS b/(t)	D3 Reading OLS b/(t)	D4 Math OLS b/(t)	D5 Science OLS b/(t)	D6 Reading OLS b/(t)
Polity Index	-0.241	-0.241	-0.014	-0.161	-0.240	-0.006
	(-0.92)	(-0.96)	(-0.04)	(-0.57)	(-0.93)	(-0.02)
Ln GDP per capita	7.218**	9.835**	5.214+	2.855	9.693**	7.217*
	(4.22)	(4.95)	(1.92)	(1.41)	(4.62)	(2.64)
PISA math mean				0.090**		
				(3.90)		
PISA science mean					0.003	
					(0.14)	
PISA reading mean						-0.047
						(-1.53)
Constant	23.693	-0.612	44.014+	21.357	-0.808	47.581+
	(1.54)	(-0.03)	(1.80)	(1.51)	(-0.04)	(1.95)
N	56	56	56	56	56	56

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. Cross-section OLS regressions (with robust errors). T-values are reported in parentheses. Polity and GDP per capita are lagged by 5 years. Dependent variable is standard error in on PISA 2012 test scores for the country's student population for PISA 2012, for particular subject (see top row).

Table A29: Democracy and absolute gender difference in PISA 2012 test scores

Test subject	D1 Math OLS b/(t)	D2 Science OLS b/(t)	D3 Reading OLS b/(t)	D4 Math OLS b/(t)	D5 Science OLS b/(t)	D6 Reading OLS b/(t)
Polity Index	0.130	-0.590	-0.494	0.083	-0.605*	-0.488
	(0.72)	(-1.65)	(-1.10)	(0.50)	(-2.14)	(-1.19)
Ln GDP p.c.	-0.216	0.002	2.449	2.382	2.813	3.930
	(-0.14)	(0.00)	(0.80)	(1.39)	(1.54)	(1.16)
PISA math mean				-0.054**		
				(-2.95)		
PISA science mean					-0.063*	
					(-2.56)	
PISA reading mean						-0.035
						(-0.74)
Constant	11.568	12.115	19.599	12.960	16.005	22.238
	(0.81)	(0.72)	(0.70)	(0.93)	(1.03)	(0.77)
N	56	56	56	56	56	56

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. Cross-section OLS regressions (with robust errors). T-values are reported in parentheses. Polity and GDP per capita are lagged by 5 years. Dependent variable is absolute gender difference in PISA 2012 test scores for particular subject (see top row).

Table A30: Democracy and mean PISA 2012 score

Test subject	D1 Math OLS b/(t)	D2 Science OLS b/(t)	D3 Reading OLS b/(t)
Polity Index	-0.885	-0.247	0.166
	(-0.37)	(-0.12)	(0.09)
Ln GDP per capita	48.536**	44.945**	42.595**
	(4.51)	(4.45)	(4.87)
Constant	26.000	62.202	75.845
	(0.26)	(0.65)	(0.91)
N	56	56	56

Notes: + $p < .10$; * $p < .05$; ** $p < .01$. Cross-section OLS regressions (with robust errors). T-values are reported in parentheses. Polity and GDP per capita are lagged by 5 years. Dependent variable is mean PISA 2012 test scores for particular subject (see top row).

References

Altinok, Nadir and Hatidje Murseli. 2007. "International database on human capital quality." *Economics Letters* 96(2):237–244.

Angrist, Noam, Harry Anthony Patrinos and Martin Schlotter. 2013. "An expansion of a global data set on educational quality: A focus on achievement in developing countries." *Policy Research Working Paper* WPS 6536.

Cheibub, José, Jennifer Gandhi and James Vreeland. 2010. "Democracy and dictatorship revisited." *Public Choice* 143(1–2):67–101.

Epstein, David L., Robert Bates, Jack Goldstone, Ida Kristensen and Sharyn O'Halloran. 2006. "Democratic Transitions." *American Journal of Political Science* 50(3):551–569.

Hadenius, Axel and Jan Teorell. 2007. "Pathways from Authoritarianism." *Journal of Democracy* 18(1):143 –156.

Harding, Robin and David Stasavage. 2014. "What Democracy Does (and Doesn't Do) for Basic Services: School Fees, School Inputs, and African Elections." *Journal of Politics* 76(1):229–245.

Helliwell, John F. 1994. "Empirical Linkages between Democracy and Economic Growth." *British Journal of Political Science* 24(2):225–248.

Honaker, James and Gary King. 2010. "What to Do about Missing Values in Time-Series Cross-Section Data." *American Journal of Political Science* 54(3):561–581.

Honaker, James, Gary King and Matthew Blackwell. 2012. *AMELIA II: A Program for Missing Data.* Version 1.6.2.

Huntington, Samuel P. 1991. *The Third Wave: Democratization in the Late Twentieth Century.* Norman: University of Oklahoma Press.

Imai, Kosuke, Luke Keele, Dustin Tingley and Teppei Yamamoto. 2011. "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies." *American Political Science Review* 105(4):765–789.
URL: *http://dspace.mit.edu/handle/1721.1/91154*

Knutsen, Carl Henrik. 2011. "Democracy, Dictatorship and Protection of Property Rights." *Journal of Development Studies* 47(1):164–182.

Murray, Michael P. 2006. "Avoiding invalid instruments and coping with weak instruments." *Journal of economic perspectives* 20(4):111–132.

Sovey, Allison J and Donald P Green. 2011. "Instrumental variables estimation in political science: A readers'guide." *American Journal of Political Science* 55(1):188–200.

Tingley, Dustin, Teppei Yamamoto, Kentaro Hirose, Luke Keele and Kosuke Imai. 2014. "Mediation: R package for causal mediation analysis.".

Unesco. 2015. "Unesco Institute for Statistics Database.".
URL: *http://data.uis.unesco.org/Index.aspx?DataSetCode=EDULIT_DS*