
Munich Personal RePEc Archive

Welfare state generosity and student performance: Evidence from international student tests

Falch, Torberg and Fischer, Justina A.V.

Norwegian University of Science and Technology, University of Mannheim

11 July 2014

Online at https://mpra.ub.uni-muenchen.de/62234/
MPRA Paper No. 62234, posted 19 Feb 2015 15:04 UTC

Welfare state generosity and student performance: Evidence from international student tests

Torberg Falch·
Department of Economics, Norwegian University of Science and Technology

Justina AV Fischer·
Department of Economics, University of Mannheim

Abstract

Student achievement has been identified as an important contributor to economic growth. This paper investigates the relationship between redistributive government activities and investment in human capital measured by student performance in international comparative tests in Mathematics and Science during the period 1980 to 2003. In fixed effects panel models, government consumption, government social expenditures, and the progressivity of the income tax system have negative effects on student achievement. These results are robust to a variety of model specifications, such as conditioning on educational expenditures, and alternative measures of student performance from the World Bank. Our estimates indicate that increased government size by 10 percent reduces student achievement by 0.1 standard deviations.

Keywords: Student achievement; welfare state; government size; tax system; panel data; international tests; PISA
JEL codes: H2; I2; C33

Acknowledgement

We thank two anonymous referees, Fredrik Carlsen, Urs Fischbacher, Stefan Hupfeld, Thomas Lange, Bjarne Strøm, Ludger Wößmann, and conference and seminar participants at the CESifo Area Conference on Employment and Social Protection, International Institute of Public Finance in Maastricht, and Economics Departments at the University of Konstanz and the Norwegian University of Science and Technology for helpful comments and suggestions. Justina Fischer acknowledges Marie Curie experienced researcher fellowship (RTN ENABLE) and a COFIT fellowship.

· Department of Economics, Norwegian University of Science and Technology, Dragvoll, N-7491 Trondheim, Norway, e-mail: Torberg.Falch@svt.ntnu.no, tel.: +47 73596757, fax: +47 73596954
· Department of Economics, University of Mannheim, L7, 3-5, DE-68131 Mannheim, e-mail: mail@justinaavfischer.de, tel.: +49 721 883360, fax: +49 721 883362

1. Introduction

The equity-efficiency quandary of the welfare state is usually attributed to the incentive structure in the labor market. The welfare state includes 'unproductive' government spending which reduces the return to work and is financed by distortionary taxes. Apparently, little evidence exists on the effect of welfare state arrangements on investment in human capital.

The welfare state can be seen as a social insurance mechanism, see for example Sinn (1995). When the insurance terms for the insured improve, her incentives to invest in order to avoid capture are weakened. In a macroeconomic context, this moral hazard problem may have detrimental effects on investment in human capital, saving, and, ultimately, economic growth.[1] Indeed, welfare state arrangements may also be seen as interventions in imperfect markets, working in the opposite direction. The evidence on public sector size and economic growth in empirical cross-country studies, however, indicates a negative relationship.[2] Ehrlich and Zhong (1998) and Ehrlich and Kim (2007) look directly at investment in human capital, and find a negative effect of old-age pension benefits on secondary school enrolment rates, in particular for developed countries.[3] Using German data, Fossen and Glocker (2011) find that university enrollment is positively related to expected return to tertiary education.

To our best knowledge, we are among the first to empirically investigate to what extent government redistribution activities affect individual investment in human capital. In this study, we approximate the former through three, partly overlapping measures of welfare state generosity: government consumption, social expenditures, and the progressivity of the tax system. We use achievement on international student tests, adjusted to facilitate comparability across countries and time, as measure of investment in human capital. Most of the existing empirical analyses on economic growth employ as a proxy of human capital some measure of *quantity* of education in the population. This is obviously a crude measure, and we follow Wössmann (2003) who argues that the number of *quality*-education-years varies across countries stronger than the mere duration of education, with which it might even be uncorrelated. Indeed, Hanushek and Kimko (2000) and Hanushek and Wössmann (2008,

[1] In addition, Bjørnskov, Dreher and Fischer (2007) find that higher government consumption spending is related to less subjective well-being, perhaps through misallocation of resources or the inefficiencies generated through modern taxation schemes.

[2] For example Fölster and Henrekson (2001), Romero-Avila and Strauch (2008) and Bergh and Karlsson (2010) find negative effects on growth of various measures of government size. Kneller, Bleaney and Gemmell (1999) distinguish between different types of taxes and spending categories, and find that distortionary taxation reduces growth whilst productive government expenditure enhances growth, a result in accordance with Romero-Avila and Strauch (2008). Agell, Ohlsson and Thoursie (2006) argue that the estimated relationships are not causal.

[3] Zhang and Zhang (2004) find the opposite relationship. Ehrlich and Kim (2007) report that - not unsurprisingly in a growth context - the estimates are sensitive to whether the models condition on initial GDP or not.

2012) find that average student achievement in compulsory schooling is a much more sizable determinant of economic growth than years of education in the population.

Analyses of general-equilibrium effects of macro-incentives must necessarily be done on aggregated data, which might come at the cost of credible identification. We construct an unbalanced country panel that utilizes data on comparative international tests in Mathematics and Science for the age group 13-15 years and includes eight tests in the period 1980-2003 for a maximum of 79 countries. We construct a synthetic time-series cross-section ('synthetic panel') of aggregated individual test scores that allows the application of panel estimation methods. This approach improves methodologically on previous studies that exploit cross-country variation in student achievement mainly in a cross-sectional framework, either based on one single cross-section of individual student performance data or a cross-section of country aggregates that have been averaged over several years, as in, e.g., Hanushek and Kimko (2000) and Hanushek and Wössmann (2008, 2012).[4] In contrast, we use country-year aggregates of individual test score data, and exploit the panel structure by estimating fixed effects models that account for unobserved heterogeneity across countries and common time-specific macroeconomic developments around the world.

Our difference-in-differences approach suggests that more government redistribution activities exert an investment-lowering effect on students – be it either through the provision of goods (e.g. hospitals), but also through direct financial transfers to households (e.g. pension spending and active labor market policies), or when measured indirectly through the progressivity of the income tax schedule. The results are robust to a variety of model specifications, including models that condition on educational expenditures and the teacher-student ratio. The negative effect of welfare state generosity seems not to be mediated by resource use in education. Our finding is corroborated when we replace student performance in Mathematics and Science with the general survival rate in lower secondary education.

This analysis is related to the literature on educational effects of cash transfer programs in developing countries, which typically transfer money to poor families conditional on their investments in human capital. These evaluations provide clear evidence of positive effects on student enrollment and reduced dropout rates (see, e.g., the review by Rawlings and Rubio, 2005). However, such programs are not representative for the main portion of public expenditures because they generate direct incentives for schooling. The evidence for developing countries also indicates that the cash transfer programs have no effect on student achievement (Ponce and Bedi, 2008; Behrman et al., 2009).

[4] The only exception seems to be Barro and Lee (2001a), who employ a panel of countries participating in international tests up to 1990 to estimate the effect of school inputs on student achievement.

The paper proceeds as follows. The next section presents some theoretical considerations on the effect of government redistributive activities on the accumulation of skills through schooling. Section 3 describes the international student tests data and our measure of adjusted average student performance, and introduces our three measures of welfare state generosity. Section 4 presents the empirical model. Section 5 provides the main empirical findings and presents robustness tests related to potential omitted variables, results for subsamples, and alternative measures of student performance. Section 6 offers some concluding comments.

2. Theoretical considerations

The insurance aspect of the welfare state manifests in a system that, first, reduces the risk related to future income and, second, redistributes from individuals with high income to individuals with low income. Numerous papers have analyzed the relationship between taxation, uncertainty, and education incentives, including Levhari and Weiss (1974), Eaton and Rosen (1980), Hamilton (1987), Glomm and Ravikumar (1992), Heckman et al. (1998a, b), Andersson and Konrad (2003), Konrad and Spadaro (2006), and Poutvaara (2007). This literature shows that the impact of risk on human capital investment is ambiguous because risk can take on many different forms.

In our empirical approach we regard observed student achievement at a specific schooling age as reflecting accumulated human capital investment up to that specific age. In the economics of education literature, individual's costly investment activity relating to schooling includes 'student effort'. Traditional economics assumes that optimal investment is the outcome of a life-time utility maximization where investment takes place in earlier periods while consumption takes place in later periods. In real-life, however, student's effort is determined not only by selfish utility maximization over life-time, but also by instructions of parents and teachers. More specifically, parents might be concerned about student effort at school because of their own altruistic and dynastic preferences. Teachers might view their work as a mission, but might also respond to incentive structures of wages and promotion. In addition, the general view in society on the importance of skills and knowledge is also expected to influence behavior of students, parents and teachers. In sum, the actually observed student effort is likely to be a result of not only students' own classical utility maximization, but also behavior of parents and teachers, and national traits such as culture and population risk aversion.

In a utility maximization model with costly investment that determines income levels later in life, the marginal cost of student's effort is equal to her expected (discounted) marginal return to this effort. Welfare state arrangements play a role in this optimization problem by

influencing (i) the level of income and (ii) risk and uncertainty, in each possible environmental state. First, viewing the welfare state as an insurance device that transfers income from individuals with high earnings to individuals with low earnings, the traditional human capital model (Becker, 1964) predicts that increased redistribution of income weaken the incentive to invest in education. Indeed, the empirical study by Heckman et al. (1998a, b) shows that a progressive income tax system yields lower incentives to invest in human capital than a proportional tax system. Second, welfare state arrangements may also reduce the uncertainty and risks attached to the return to investment in human capital for the population. The insight from Levhari and Weiss (1974) is that reduced uncertainty in the return to human capital investments increases the investments because individuals dislike risk. The result that lower volatility in the return to schooling increases education investment is, however, not universally true, but depends on theoretical assumptions. For example. Hogan and Walker (2007) and Jacobs (2007) show that investment is positively related to the uncertainty in the labor market when educational choices are modelled in a real option framework.[5]

The implicit aggregate function of human capital investment, H, can be written

$$H = H(g, \bar{y}; Z), \tag{1}$$

where g captures welfare state arrangements, $\bar{y}$ is average income, and Z reflects country specific factors such as culture for education. We predict that educational investment H decreases in $\bar{y}$ if individuals' utility functions are concave, while the effect of g is in general ambiguous as it impacts income level and uncertainty likewise. However, we believe that redistribution of income is the dominating aspect of the welfare state, which has a negative effect on H.

Our *testable hypothesis* is that a more generous welfare state lowers students' educational investments, i.e., effort level during schooling, measured by the international achievement tests.

3. Data

3.1. International measures of student achievement

We rely on comparative international tests of student achievement conducted by different international organizations. The International Association for the Evaluation of Educational

[5] For a more in-depth theoretical discussion of how welfare state arrangement might affect students' educational investment decisions, see Falch and Fischer (2011).

Achievement (IEA) has been responsible for the largest number of such tests, among them the TIMSS tests, but also the OECD has developed a Programme for International Student Assessment (PISA).

We construct a synthetic panel data set of national averages of individual student performance in international tests covering a period of almost 25 years (1980 – 2003). Individual level data is not available for the tests prior to 1994/95, and the gain of exploiting micro-variation is limited in our setting because measures of government activities vary only at the national level. The tests cover the core subjects Reading, Mathematics and Natural Science separately, but we restrict our attention to student assessments in Mathematics and Science for several reasons. First, these two subjects have more similarities among each other than with Reading and are thus more suitable for constructing a synthetic panel. Second, reading skills have been tested less regularly, and even within the same test and year, these tests might potentially differ considerably by language. Third, performance in Mathematics and Natural Science are more likely to determine a country's innovativeness in an economic growth context, as empirically tested in Hanushek and Kimko (2000) and Hanushek and Wössmann (2008, 2012). Comparability of test results is also given in the age dimension, as all tests included are conducted on middle-aged students (13-15 years). Choosing this age group has also the advantage that compulsory schooling still applies, mitigating selection out of education issues. The tests we utilize are described in Table 1.[6]

Insert Table 1 about here

Recently, it has become common to report national averages based on the Items Response Theory which weights the different questions by their difficulty ("Warm estimates", Warm, 1989), and standardizes the scores such that the average across all participating students and countries is 500 with the standard deviation of 100. Particularly the PISA studies employ this methodology. With this approach, the average score of a particular country will depend on the achievement of the students in the other participating countries. Thus, since the composition

[6] Even though all tests are in the fields 'Mathematics' and 'Science', they do not necessarily test the same cognitive skills: The IEA tests are related to common elements of school curricula across countries while IAEP is geared towards the curriculum in USA building on the national testing procedures developed by the National Assessment of Education Progress NAEP. The OECD PISA test has a more real-world approach and claims to assess the skills that are considered to be essential for full participation in the society. The high correlation coefficient between the (adjusted) test results across various test types suggests that these differences are not important with respect to measured student performance. For the 18 countries participating both in TIMSS 2003 and PISA 2003, the correlation is 0.94, while the correlation between the average Science and Mathematics score in TIMSS-repeat 1999 and PISA 2000 is 0.87, and the corresponding number for IAEP 1991 and TIMSS 1995 is 0.80. Interestingly, as can be seen from Figure 2 below, USA had its poorest performance in the IAEP test that was based on the US school curriculum.

of participating countries changes across test years, the raw test scores for a particular country are not comparable over time. In addition, for the tests prior to 1991 "Warm estimates" have not been calculated so that we have to rely on the share of correct answers for these tests.[7]

To make the scores on the different international tests comparable on a common metric, we have re-scaled the average scores for each international test by the following procedure. First we calculate the average of the Mathematics and Science tests when both subjects are tested. Second, we standardize the average score for each test to have mean zero and standard deviation equal to unity for a "core" group of 15 countries. The "core" is defined as the countries that have participated in at least six out of the eight international tests reported in Table 1, namely Australia, Canada, Hong Kong, Hungary, Israel, Italy, Japan, South Korea, Netherlands, New Zealand, Russia, Sweden, Thailand, UK, and USA.[8] Third, we re-scale the scores for each of the remaining countries using the same parameters as for the "core" countries. Finally, since some countries participated in two parallel tests in 2003 (TIMSS and PISA), we calculate the average adjusted test score based on both tests in 2003.

Making the results from different tests comparable across time has been a challenge also for previous empirical studies. For example, Hanushek and Kimko (2000) calculate a measure of labor-force quality based on the percent of correct answers in international student achievement tests for the period 1965-1991. They adjust the mean for each test, but not the variance (except the linear scaling that follows from the adjustment of the mean). Adjusting the means across tests is crucial in their analysis because they subsequently calculate an aggregated 30-year average quality measure for each country. More recently, Hanushek and Wössmann (2012) use student performance tests from TIMSS, PISA and the IEA up to 2003 and, in addition to adjusting the means, correct the dispersion of each single test in a similar way as we do.[9]

Figure 1a shows that the density of our measure of student achievement across the 15 "core" countries observations is close to the normal distribution. The density for all observations presented in Figure 1b has a long left tail, illustrating that some countries, mostly developing countries that participate less frequently in international tests, have low student achievement.

[7] We have compared the Warm estimates and percent correct answers for the IEA tests in 1994-95 and 1998-99 for which both measures are available. The correlation coefficients for Mathematics are 0.997 and 0.982, respectively, and for Science 0.994 and 0.977, respectively. Thus, the differences across countries do not seem to be influenced in any important way by the choice of scale.

[8] More precisely, we standardize the score for those of the "core" countries that participated in the particular test. Out of the 15 "core" countries used to standardize the test scores, the data sources reports results for 11 countries in 1980-81, 12 in 1983-84, 8 in 1990-91, 15 in 1994-1995, 14 in 1998-99, 15 in OECD 2000, 13 in TIMSS 2003 and 13 in OECD 2003. Only USA has test scores for all tests.

[9] Hanushek and Wössmann (2012) use as their "core" countries the 13 OECD countries that had "half or more of the relevant population attaining a secondary education in the 1960s".

Insert Figures 1a and 1b about here

In a model with country-specific fixed effects, identification is only based on within-country variation. Figure 2 shows the development of test score averages over time for the "core" countries. The figure indicates that there are some systematic changes. For example, the relative achievement in the more neo-liberal Western economies USA, Canada, and UK increased during the 1990s, while the achievement declined in Israel and in the transition countries Russia and Hungary. Some countries perform consistently better than others. For example, Italy performs below average and Netherlands performs above average in each test. However, Figure 3 shows that there is quite some variation in the *change* in student achievement, although the variation is lower than that for the distribution in *levels* of achievement.[10]

Insert Figures 2 and 3 about here

Appendix Table A1 presents the 72 countries participating in the relevant international tests of student performance. 16 countries have only conducted one test, and will thus not contribute to the identification in models with country fixed effects. The table shows that the average test score is typically low in developing countries, and that the overall within-country variation is relatively high.[11]

3.2. Independent variables: welfare state generosity and controls

Our focal determinant of student performance in this analysis is the generosity of welfare state arrangements which is made operational in three ways: Firstly, we employ general government consumption spending (in percentage of GDP), obtained from the WDI (2007) database of the World Bank, a widely used measure of government production of goods and services that has been employed in various cross-country growth studies (Fölster and Henrekson, 2001, Agell et al., 2006) and happiness studies (Bjørnskov, Dreher, and Fischer, 2007 and 2008). Government consumption excludes financial transfers to single households, but includes the government production of goods and services, which are mostly financed by taxes. For example, government consumption spending includes expenses for hospitals, infrastructure, public transport, schools and culture – state expenses which all relax the

[10] In Figure 3, only observations with at most eight years interval are utilized.

[11] The within-country variation tends to be high in countries with declining test scores such as Bulgaria and Hungary. Singapore had an exceptionally high score in TIMSS 1995.

income constraint on private households' consumption now and in the future. In addition, given that most publicly provided goods are financed through progressive tax systems, they entail a consumption redistribution aspect. Following the traditional public finance literature, we will refer to this measure as 'government consumption'.

Secondly, we use public sector social expenditures (in percentage of GDP) that are obtained from OECD *Social Expenditure database* (SOCX) and include aggregated public welfare expenses of all government tiers.[12] This measure captures direct transfers from government institutions to single households, including "benefits to, and financial contributions targeted at, households and individuals in order to provide support during circumstances which adversely affect their welfare, provided that the provision of the benefits and financial contributions constitutes neither a direct payment for a particular good or service nor an individual contract or transfer" (OECD, 2007, p. 7). OECD defines expenditures as 'social' if they satisfy two criteria: first, they have to intend a social purpose, and, second, these programs must be based on either inter-personal redistribution or compulsory participation (OECD, 2007, p. 8). They take the form of "cash benefits (e.g. pensions, income support during maternity leave, and social assistance payments), social services (e.g. childcare, care for the elderly and disabled) and tax breaks with a social purpose (e.g. tax expenditures towards families with children, or favorable tax treatment of contributions to private health plans)" (ibidem, p. 7), excluding the administrative costs of executing them.

By employing separate components of public social expenditure, we are able to differentiate government transfers by social policy area as follows: pension payments, unemployment benefits, active labor market policy spending, family allowances, health care (service) spending, housing subsidies, and 'other spending'. Table 2 provides an overview of spending components. The major population is, in principle, entitled to all those spending programs so that each may exert an independent effect of its own. The correlation coefficient between government consumption spending and total social spending is equal to 0.67 in our sample.

Figure 4 presents within-country variation in total social expenditures as a share of GDP for the "core" countries. There is a tendency of increased social expenditures during the period of investigation: the average share of social expenditures in Figure 4 increased from 17%- in 1980 to 19% in 2003.[13] The Netherlands is the only country which cut down social expenditures, while Japan has experienced the largest growth. Note that social expenditures as

[12] The OECD defines expenditures as 'public' (as opposed to being 'private') when institutions of the 'General Government' control the relevant financial flows. The 'General Government' in this context includes different levels of government and social security funds. This definition of 'public' includes, often by tradition, transfers by compulsory social insurances and social assistance schemes (see also OECD 2007, p.8-10).

[13] For all 29 OECD countries included in the empirical analyses, social expenditures increase from 17 percent of GDP in 1980 to 21 percent in 2003.

a share of GDP serve as automatic stabilizers and, thus, typically shrink during a boom and expand during a recession. Thus, it is important to include GDP in the empirical model in order to avoid identification on variation in national income.

Insert Figure 4 about here

Table 3 presents some descriptive statistics on government consumption in the world sample, and total social expenditures for OECD countries, including its single components. The variance in social expenditures is slightly higher than that for government consumption, both overall and within countries. The within country variation, for which we identify the effects on student achievement, constitutes 7-8 percent of the overall variance. Pension spending is the largest component of social expenditures, followed by public health spending.

Insert Tables 2 and 3 about here

The third measure of the generosity of the welfare state that we employ is an index of income tax rate progressivity developed by the *Fraser Institute* (Gwartney and Lawson, 2002). The index constitutes an income-bracket adjusted marginal tax rate levied in the highest income bracket in one country, adjusted for the lowest income threshold for this income bracket. The redistributive impact of a given tax rate depends on the financial threshold the rate applies on. Since the index is adjusted for threshold effects it facilitates comparability of the marginal top income tax rate across countries and time. Progressive taxes are redistributive as they relax the financial constraint on poor households relative to richer households, but also since they finance provision of goods and services that equalize consumption patterns between the rich and the poor. The index of income tax progressivity ranges from 0 to 10, with higher values representing a higher top tax rate, and, thus, more redistribution of income. Between 1970 and 2000, data have been collected every five years, and annually from 2000.[14]

The regression-sample within-country variation for our three measures of welfare state generosity is presented in Appendix Table A1. For countries with relatively large variation in

[14] There is not a perfect match between the years of student achievement tests and the years for which the tax progressivity measure is available on a quinquennial basis prior to the year 2000. We therefore choose to linearly interpolate the missing values prior to 2000. An alternative method would be to relate student achievement to the tax progressivity observed most closely in time – this way, both years 1980 and 1981, 1990 and 1991, 1994 and 1995, 1998-2000 are related each with identical values of quinquennial tax progressivity. By this procedure, the variation in tax progressivity is reduced, but the empirical results are qualitatively identical (see Section 5.3).

one or two variables, the variation of the third variable is often similar to the average. For example, during the 1990s Korea increased rapidly both social expenditures and the progressivity of the tax system, but without changing government consumption. On the other hand, in Ireland and the UK government consumption and tax progressivity declined markedly in the 1990s without much change in social expenditures. Thus, the subsample of countries that contribute most to the within-country variation in the empirical analysis will differ across the three measures.

Annual GDP and population data are taken from the WDI 2007 database. Adult education attainment is taken from Barro and Lee (2001b), which are available on a quinquennial basis up to the year 2000.[15]; missing values prior to 2000 have been linearly interpolated. We use the 3-years lag of the percentage of the population over age 25 with secondary school attained in order to include also the international tests after the year 2000 in the empirical analysis. In some robustness tests, we also use current data on primary school educational expenditures (% GDP) and pupil-teacher ratios from the World Bank[16], and the survival rate, that is the share of entrants in secondary education actually attaining a lower secondary degree, obtained from the World Bank (WDI, 2007, updated 2014).

4. Model specification and identification

We estimate the following model for student achievement H of country i in year t.

$$H_{it} = \beta_1 g_{it} + \beta_2 \log(GDP/POP)_{it} + \beta_3 \log(POP)_{it} + \beta_4 EDU_{it} + \phi_i + \varphi_t + \varepsilon_{it}, \quad (2)$$

where g_{it} is the measure of the welfare state generosity and GDP per capita (GDP/POP) is the proxy for mean income in equation (1) in section 2. Family characteristics as parental income and education have strong effects in micro studies of student achievement, which is why we also include the share of adult population with at least some secondary education (EDU). Since we measure GDP in per capita terms, we also employ population size (POP). The time fixed effects φ_t account for macro-developments common to all countries, e.g. financial market crises and global recessions, but also for the fact that most of the independent

[15] Again, we use linear interpolation for the years between the actual observations of adult education attainment in order to match this variable to observed student test scores. Also in this case is the qualitative results are clearly the same as reported below.

[16] From 1990 on, these data have roughly been collected on an annual basis. Again, we linearly interpolate the variables when there are missing values for at up to five years and consider the information as missing if it is not measured in a five-year period.

variables have positive trends.[17] Country-specific fixed effects ϕ_i account for time-invariant differences between countries.

In general, analyses of macro-incentives are carried out at an aggregated level at which the variation in the variables of interest occurs. Since welfare state arrangements typically do not vary within countries, but across countries and across time, our approach is to use country level data.

Including country fixed effects in the model, which amounts to a difference-in-differences specification, is essential in order to interpret the estimated relationships. The school systems vary greatly across countries, for example with respect to school starting age and early tracking of students. In addition there might be differences across countries in culture for learning, risk aversion, and labor market institutions that are important for students' investment decisions. However, all these characteristics change slowly at the national level, and are thus to a major extent captured by the country fixed effects.

One must, of course, be careful in interpreting the estimated relationship between welfare state generosity and student performance as causal. There are several potential threats to our identification approach. First, sufficient within-country variation in the variables of interest is a necessary identifying condition. The descriptive statistics (Table 3) in the previous section indicate that this is indeed the case in our data. Second, another source of concern is that there may be important but unobservable factors that vary over time in a way that is correlated with the variables of interest; the omission of those factors would bias our results. In addition, in the case of measurement error the signal-to-noise ratio might be weak in fixed effects models. Finally, another source of bias might in principle arise from reverse causality; public policy might respond on dissatisfactory school performance.

We consider the likelihood of biased estimates in our relatively simple model specification by investigating the robustness of our results in several ways. First, we employ various measures of welfare state generosity which differ by their within-country variation and we expand the model to include educational spending, unemployment, and the age composition of the population. In addition, we estimate models with county specific trends to capture potentially differing development patterns across countries. While the latter specification obviously better controls for potentially omitted variables, it might also weaken the signal-to-noise ratio to the extent that (residual) welfare state arrangements change too slowly for identification. Further,

[17] Notice that including time-specific effects may influence the interpretation of the results, as we discuss later. The scaling of the test scores makes the scores comparable over time so that, by including time fixed effects, the model in essence draws inference on which other countries that participated on the different test and year. Thus, we also report results where the time-specific effects are replaced by a simple trend.

we estimate separate models for OECD countries that arguable are more homogenous that the world sample.

In order to address such biases the instrumental variable technique is an alternative approach. We think, however, that it is hard to think about valid instruments for welfare state arrangements in a cross-country framework. There are some attempts in the literature to use instruments in cross-country analyses. For example, Acemoglu et al. (2001) use European mortality rates in colonies as instrument for present institutional quality. This is, however, a disputed approach (Albouy, 2012), and it relies on variation across developing countries while we focus on developed countries. With respect to reverse causality, conventional wisdom suggests that it is not a major issue in our analysis: even if welfare reforms often are related to poor economic performance and problems in the labor market, unsatisfactory performance of schools is to our knowledge not used as an argument for reduced generosity of welfare state arrangements.

Finally, we have to discuss the various dynamic patterns of welfare state generosity that we employ in contemporaneous measures. Theoretically, it is the *expected* welfare state arrangements in the future that affect educational investment decisions today. We argue that, in our empirical framework, the contemporaneous level of government redistribution activities might be the best proxy for individuals' expectations. The average spending level over time that may persist due to a status quo bias through a country's institutions is captured by the country fixed effects. However, it is the short-term within-country changes, conditional on GDP and population size, that drive the results in our model specification. As educational production is cumulative, expectations of students and parents at earlier grades in the past are important for observed achievement today at the student age of 13-15 years. Thus, contemporaneous levels of government welfare spending are to some extent a leaded measure for the real decisions made by students' parents in the past. However, we will also investigate the robustness of the results by using five-years moving averages in the independent variables in some model specifications.

5. Empirical Results

5.1. Government consumption

Table 4 presents results for government consumption spending including 59 countries and 208 observations. The first column simply presents the correlation between student achievement and government consumption spending as share of GDP. There is no unconditional

correlation.[18] Column (2) includes the control variables GDP per capita, adult educational attainment w.r.t. secondary education, population size, and time-specific fixed effects. The number of observations drops due to missing observations for adult education for some countries. As expected, we find strong positive effects of GDP per capita and adult education. The positive income effect mirrors Hanushek and Kimko (2000) who report a positive effect of student achievement on economic growth, and is in accordance with micro evidence on the effect of parental income on student performance. However, the result for GDP is not in accordance with our testable hypothesis that assumes a concave utility function. The positive effect of adult educational attainment mirrors results from previous micro-econometric studies.

The conditional effect of government consumption spending in column (2) is negative, as expected, appears sizeable and is significant at the 1 percent level. Thus, conditional on income, a small public sector is favorable as it increases student performance, and conditional on public sector size, students in rich and well-educated countries perform better than those living in poor and low-educated countries. When the share of government consumption increases by 0.1 log-points (approximately 10 percent) student achievement declines by 0.15 "core" country standard deviations.

Insert Table 4 about here

Columns (3) to (5) of Table 4 present models with country fixed effects that mitigate a potential omitted variable problem.[19] However, including country fixed effects in addition to year effects in column (3) does not change the point estimates of GDP and adult education attainment much compared to column (2), although the standard errors are twice as large as in the model with only time fixed effects (column (2)). On the other hand, the estimate of government consumption is small and insignificant.

[18] In order to compare the results in this unconditional model with models that include control variables, we only include the observations where the control variables are observed. In the full sample we have 232 observations in 72 countries for student achievement. Estimating the unconditional model in column (1) in Table 4 on this full sample, the relationship with government consumption is equally insignificant (coefficient of 0.107 and standard error of 0.414).

[19] For 11 countries, we have only one observation. These countries do not contribute to the identification in the fixed effect models. If we exclude these countries from the model without country fixed effects (column (2)), the effect of government consumption remains negative and significant (the point estimate changes from -1.46 to -1.22) For another 10 countries, we have only two observations. These countries do not contribute to the identification in the models with country specific trends in addition to the country fixed effect model. If we exclude these countries from the model with country fixed effects (column (3)), the effect of government consumption remains negative but insignificant (the point estimated changes from -0.65 to -0.71).

In these fixed effects models, our variable of interest becomes sometimes insignificant so that identification of its effects needs to be discussed. The relative large standard errors in the models with country fixed effects in columns (3) to (5) may indicate that the within-country variations in student achievement and government consumption are too small for statistical identification. Notably, the OLS R^2 is as high as 0.94. However, it may equally be that it is the time-specific effects that complicate identification, e.g. in column (3). The purpose of the scaling of the test scores described above is to make the scores comparable over time. In consequence, with time-specific fixed effects, the model in essence draws inference on the change in composition of participating countries in a particular test and year. Our motivation for including year effects is that, on average, all the independent variables have positive trends. Indeed, while the p-value of joint significance of the time-specific fixed effects is 0.02 in the model in column (3) in Table 4, the p-value is only 0.13 when a simple trend is added. For this reason, we replace in column (4) the time fixed effects with such a time trend. The coefficient of the trend variable is negative, indicating a positive trend in the other independent determinants, as expected. The OLS R^2 appears only marginally lowered, while the within-R^2 is clearly reduced. Interestingly, the effect of government consumption spending is significantly negative in this specification. We conclude that it is not unobservable, time-invariant country-specific factors that let the effect of government size appear insignificant in column (3), but the handling of the variation over time.

What kind of within-country variation in student achievement and government consumption is driving the results? Is it country-specific trends, or fluctuations around the trends? Figure 2 above suggests that some countries exhibit a trend-like development in student achievement. To investigate this question, column (5) in Table 4 expands the model with country-specific time trends. Then the effect of government consumption increases to about the same magnitude as in the model without country-specific fixed effects in column (2), and becomes highly significant again. Thus, it seems like it is the variation around country-specific trends that accounts for the association between government consumption and student achievement. This result is independent of whether the model includes time-specific fixed effects or not, and the increase in significance from the model in column (3) to the one in column (5) is not related to the fact that the number of countries that contribute to identification is necessarily smaller in the latter model.

In columns (6) through (8) of Table 4 we estimate the same models restricting our sample to OECD countries, to ease comparison with Table 5 below. We define the subsample of OECD countries by membership in the year 2000, but test later the robustness of our results for post-communist period effects (see section 5.5.). In the OECD sample, the effect of government consumption is significant in the model without country fixed effects (column (6)), with its coefficient size appearing independent of whether country fixed effects and country-specific

trends are included in the model or not (columns (7) and (8)). The effect of about -1 is similar to that for the world sample. In columns (7) and (8), however the standard errors are relatively large. Nevertheless, the variation that drives these results appears to differ between OECD countries and non-OECD countries. While the variation across country-specific trends aids identifying a strong effect for the whole sample, inclusion of trends in the OECD sample does not influence the effect of government size.

In sum, we identify a negative impact of government consumption spending on student achievement. This evidence is in accordance with our hypothesis that a more generous welfare system generates disincentives for educational investment. The result indicates that when government consumption spending increases by 0.1 log-points, student achievement is reduced by about 0.1 "core" country standard deviations.

5.2. Social expenditures

Table 5 presents results for government total social expenditures, measured as share of GDP, available for 28 OECD countries, resulting in a sample of 121 observations. Column (1) shows that the unconditional correlation between welfare transfers to households and student achievement is negative and significant at 5 percent level.[20] Inclusion of co-variates even increases the effect of social expenditures both in terms of magnitude and statistical significance (column (2)). Also within OECD countries, there are positive impacts of GDP and adult education attainment in the population. When country fixed effects are included (column (3)), the coefficient of social expenditures is still significant at the 5 percent level, but larger in magnitude. Interestingly, it is of similar size as the point estimate of government consumption spending in Table 4. Column (4) shows that the results are not sensitive to whether the model includes time-specific fixed effects or a common time trend.

Insert Table 5 about here

However, when including country-specific time trends, the performance-lowering impact of social expenditures completely disappears (column (5) of Table 5). We conclude that it is country-specific trends that drive the results, which indicates that in the OECD sample there are some systematic medium-term changes in government policy that students and parents react on.

[20] The full sample includes also Luxembourg, including 29 OECD countries and 124 observations. In this full sample, the simple negative and significant correlation observed in column (1) persists with a coefficient of -0.488 and a standards error of 0.223.

In the last part of Table 5 we distinguish between different components of social expenditures relating to specific social policy areas such as health, family care, labor market, pension system, etc. In column (6) we replace total social expenditures with all its various components. Note that the information on some components are missing in several countries which reduces our regression sample. Nevertheless, all components have a negative sign as expected, except for 'health care spending', 'other spending', and 'family allowances'. Notably, the spending category 'other spending' is of a rather 'kitchen-sink' nature so that its estimate is not easy to interpret. The positive effect of family allowances is in fact significant at the 10 percent level, which may indicate that relaxing parents' budget constraints in the poorest families may have an attainment-increasing effect on their children.

The positive correlations among the different social expenditure components may contribute to their heterogeneous and mainly insignificant effects in column (6). Thus, we have run regressions including each of the components separately. In all cases, the coefficient estimates are negative, except for family allowances. Columns (7) and (8) in Table 5 report the two single cases of a minimum statistical significance at the 10 percent level. Both payments on active labor market policies and pension spending lower students' test scores, each with significance at 1 percent. Since the former constitutes only a small part of total social expenditures, the negative effect of social expenditures in columns (1) to (4) seems to a large part to be driven by pension spending.[21]

Taken all together, Table 5 shows that the effect of social spending in OECD countries is in accordance with our hypothesis that government redistribution activities create disincentives for students' human capital investment. Among the different types of welfare transfers, it is pension benefits that contribute most to this effect. Possibly, because of path dependency in policy-making, current changes in spending on pensions may have a strong predictive power on governments' future pension system policies, which are strongly redistributive in nature: in most developed countries, on the one hand, pensions systems guarantee an income-independent minimum rent to every contributor, while, on the other hand, they place a cap on the maximum rent, equalizing rent incomes in the non-active elderly population.

5.3. Tax progressivity

Table 6 uses the same model specifications as Table 4, but replaces the government consumption variable with a 10-point scale index of the top marginal income tax rate, adjusted for the income bracket, a measure of progressivity of the income tax system. For this welfare state generosity measure, there is a significant negative correlation with student achievement

[21] We are unable to exclude the possibility that more public expenditures on pension may equally proxy for a large body of civil servants. In this case, the prospects of becoming a civil servant with high job security and generous retirement options may equally lower effort in mandatory schooling.

for the whole sample (column (1)), but the effect disappears when we include the control variables national income, population size and educational attainment in the population (column (2)).[22] However, in the models with country-specific fixed effects in columns (3) and (4), the tax progressivity coefficient is significant at five and 10 percent levels, respectively. Column (5) suggests that the development of tax income progression over time is not captured by country-specific time trends, as the coefficient estimate as similar to those in columns (3) and (4), which exclude such country-specific trends. When the index of income tax progressivity increases by one standard deviation, which is about 2.5 points, student achievement is reduced by 0.21 adjusted "core" country standard deviations.[23]

Insert Table 6 about here

How the effect of the variable of interest changes when we alter the model specification varies greatly between the government spending, social transfers and tax progressivity models (Tables 4-6), which indicates that the variables have very different features.[24] Nevertheless, the main result for all measures of welfare state generosity is that they tend to reduce student achievement. The quantitative effects of the adjusted top tax rate are difficult to compare to the other variables since using an index variable makes quantitative predictions difficult.

Regarding OECD countries, there is a strong negative effect of tax progressivity (column (6)) when the model does not include country fixed effects, but the effect disappears when country fixed effects are added. This finding is similar to the pattern observed for government consumption spending in the OECD countries (columns (6) to (8) of Table 4). In Table 6, it most possibly suggests that the conditional within-country variation of tax progressivity is too small to identify a statistically significant effect. Overall, using a measure of tax progressivity, we find convincing support for our hypothesis that students' learning efforts decrease as the redistributive activities of the government expand.

5.4. Generosity of the welfare system

[22] If we include the observations with missing control variables, the number of observations increases from 180 to 206. Then the correlation and significance is very similar to the results reported in column (1) (coefficient of -0.141 and standarrd error of 0.058).

[23] When using lead and lags by about one year of tax progressivity in place of interpolated values every 5th year, we obtain qualitatively similar results. For example, re-estimating the models in columns (1) to (3) – that is the models with no fixed effects, only time fixed effects and two-way fixed effects - , the results are (standard error) -0.137 (0.056), 0.023 (0.054) and -0.126 (0.039), respectively, and very close to the estimates reported in Table 6.

[24] The correlation coefficient between government consumption spending (log) and the top marginal tax rate index is -0.44, and for the social spending (log) in OECD countries -0.30. Please note that financing of government activities also occurs through corporate taxation and indirect taxes on e.g. consumption goods.

The sizes of government consumption and social spending, expressed in percentage of GDP, are commonly viewed as proxies for the generosity of redistributive activities by the government. However, in principle, generosity of social transfers can be more directly assessed when values per recipient of social benefits in place of per capita numbers are employed. However, precise information on number of recipients is not easily available. Thus, we analyze the effects of welfare state generosity by estimating models with those selected components of social expenditures for which appropriate population shares serving as proxies for number of recipients are available. We employ either the share of elderly in the population or the share of unemployed in the active population. Indeed, omission of beneficiary measures might have biased our previous results as the spending estimates might capture population composition effects: simple correlations of the spending measures with the number of their specific beneficiaries are large.[25]

In Table 7 we present results for models with measures of the number of recipients included. Taken all together, the results are not sensitive to inclusion of proxies for the number of beneficiaries. Pension spending and active labor market policy spending still exert a student performance lowering impact when the share of elderly and the unemployment rate, respectively, are included in the model (columns (1) and (2)), while the effects of unemployment and health care spending remain insignificant (columns (3) and (4)). The similarity of the coefficients on the spending variables with the original models reported in Table 5 suggests that the bias from using spending measured per GDP (conditional on population size) in place of per recipient is rather small. Regarding pension spending in column (1), the significance level is reduced to 5 percent when the share of elderly is included. While the share of the population above 60 years of age is insignificant, the test of joint significance clearly suggests that both variables are jointly related to student achievement. The effect of active labor market policies spending is equally lowered in significance (now at 5 percent level) when the unemployment rate is included (column (2)), while this time the test of joint significance clearly suggests that only one of the variables is related to student achievement.

Insert Table 7 about here

5.5. Robustness analyses

[25] The correlation coefficients between unemployment spending and unemployment rate is 0.51 and between pension spending and the share of the population above the age of 60 is 0.86. The correlation between active labor market policy spending and the unemployment rate is only 0.17.

The student test scores from the 1980s are not average results for jointly conducted Mathematics and Science tests as those achievement tests in the post-1990 period, but separate tests on the two subjects. Another reason for restricting the sample to the post-1990 period is that many argue that test designs and test procedures have improved over time. Therefore, the dependent variable may incorporate a larger measurement error in the 1980s than in later periods. Table 8 presents results for regressions on the subsample for the 1990-2003 period. Columns (1)–(3) in the table show that the coefficients of all three measures of welfare state - government consumption spending, social transfers, and income tax progressivity - are in fact larger in this subsample compared to the full sample that includes the pre-1990 tests, although the effect of government consumption spending is still insignificant in the model with time-specific fixed effects.

Insert Table 8 about here

It might also be argued that government spending and social spending are proxies for educational expenditures. Notice, however, that the typical finding in the literature, both from studies using single country data and international data, is that educational expenditures and teacher-student ratios have at most a minor effect on student performance (e.g. Hanushek and Luque 2003). If, still, educational expenditures or education quality have a positive effect on student achievement, their exclusion will bias the effect of welfare state size in our previous models since these variables are likely to be positively correlated. Thus, our previous coefficient estimates in Tables 4 to 6 may rather be biased downwards in absolute terms, providing a lower bound of welfare state effect. In columns (4) to (9) of Table 8 we add to our model educational expenditures per pupil in primary schools as a percentage of GDP and pupil–teacher ratios in primary schools from the World Bank education database.[26] The effects of educational expenditures and pupil–teacher ratio appear insignificant in all but one specification. Most important, the effects of government consumption, social expenditures, and the progressivity of the income tax system remain qualitatively unchanged when these measures of resource use and school quality in primary education are accounted for. Overall, we find no indication that the generosity of the welfare system and government public goods' creation proxies previously unobserved educational expenses or school quality.

We have also investigated whether the choice of functional form of the empirical model is important. One may argue that it is not short-term fluctuations in the independent variables

[26] For secondary education, the number of observations was insufficient. The correlation coefficients of per pupil spending in primary education with our government consumption and social spending exceed well 0.5, while those with pupil-teacher ratio in primary education are -0.76 and -0.29, respectively.

that are important, but the development in the medium or long term. We have carried out identical regressions as reported in Tables 4-6 using 5-year moving averages of the independent variable in place of using current values. The findings for government consumption spending appear partly sensitive to the choice of time window, although a robust and large performance lowering effect at the 1 percent level remains if country-specific time trends are included. The effect of social spending appears insignificant throughout, albeit their coefficients prevail in size and direction. In contrast, the student performance lowering impact of active labor market policies spending and pension benefit spending is strongly supported.[27] Estimation of 5-year moving averages corroborates the results for progressivity of the tax system for OECD countries (analogously to Table 6), while the coefficients for the full country sample are now smaller and insignificant, albeit all with negative signs.[28] Overall, the results for welfare state generosity appear insensitive to changes in model specification and sample selection.

One interesting question is whether the effect of welfare state generosity carries over to non-compulsory education. However, international comparative data back to the 1980s are scarce, in particular for upper secondary and tertiary education. Table 9 presents results for one measure of student performance that can be constructed for the same period as our analysis of student achievement on tests. The dependent variable is the survival rate to a lower secondary degree, defined as the share of the students starting lower secondary education who obtain the degree (typically at age 16). The data source is the World Bank (WDI, 2007, updated 2014). We use the same time period 1980 – 2003 and the same countries as for the main analysis above. However, since this variable is measured more regularly than student achievement in TIMSS and PISA tests, we have more observations. The mean value in the different samples used in the regressions is presented in Table 9. It is relatively common not to participate in education up to the graduation from lower secondary education. The mean survival rate is 83% and 88% in the full sample and the OECD sample, respectively.

Insert Table 9 about here

Overall we find insignificant effects of the generosity of the welfare state on lower secondary education student survival rates in the full sample, but significantly negative effects of income

[27] Significant at least at the 5 percent level. In addition, housing subsidies appear now conducive to student performance, (at the 5 percent level) in a similar manner as family allowances in the current value model.

[28] We have also investigated the sensitivity to the assumed functional form in logs. The analogous results for government consumption in Table 4 are similar and show, again, the importance of country-specific time trends to identify the effect of size of welfare state in the world sample. In contrast, the coefficients for social spending in the OECD become insignificant, suggesting a model misspecification. Results for single social spending components are, however, comparable to the results in Table 5.

tax progressivity and, in tendency, of government consumption spending for the OECD countries. In contrast, social welfare spending in OECD countries exerts no such effect, but some coefficients on its subcomponents point into the expected direction (not reported). I contrast to the analyses for the PISA/TIMSS data in Tables 4-8, the share of non-OECD countries in the regression sample of Table 9 is larger. It is possible that government spending is conducive to student attainment in developing countries (reducing the risk in future income, but also through substantially lowering financial investment costs in education) – we leave the question of the role of welfare state generosity in developing countries to future research.

6. Conclusion

The recent publications of international comparative student achievement tests such as PISA and TIMSS have spurred the debate on quality of public education in many countries. While most of the discussion has been centered around educational resource use and school organization, analyses of macro incentives implicit in government's economic policies are limited.

This paper studies the relationship between welfare state generosity and individuals' investment in human capital during compulsory education. We estimate differences-in-differences models accounting for unobserved country heterogeneity for the period 1980-2003 using international test scores in mathematics and science made comparable across testing institutions and test years. Our results clearly suggest that the generosity of the welfare state has a deteriorating impact on student performance. This finding is corroborated when analyzing survival rates to the lower secondary degree as an alternative measure of student performance. Both the effect of government consumption spending per capita, the degree of progressivity of the income tax system, and, for OECD countries, the size of direct social transfers to households have a significant negative effect on student achievement in PISA/TIMSS. For the monetary measures of government activity we find that an increase by 10 percent reduces student achievement by about 0.1 standard deviations.

However, one needs to be cautious when drawing policy implications from our empirical results: the fact that findings in form of econometric 'point estimates' always must be interpreted as marginal, 'local' changes. Furthermore, our findings are for high- and middle-high- income countries only – whether similar results can be found in developing countries context remains an open question.

References

Acemoglu, D., S. Johnson and J. A. Robinson (2001): “The Colonial Origins of Comparative Development: An Empirical Investigation”. *The American Economic Review* 91, 1369-1401

Agell, J., H. Ohlsson and P. S. Thoursie (2006): ”Growth effects of government expenditure and taxation in rich countries: A comment”. *European Economic Review* 50, 211-218.

Albouy, D. (2012). “The Colonial Origins of Comparative Development: An Empirical Investigation: Comment. *The American Economic Review* 102, 3059-3076.

Andersson, F. and K. A. Konrad (2003): “Human capital investment and globalization in extortionary states”. *Journal of Public Economics* 87, 1539-1555.

Barro, R. J., and J.-W. Lee (2001a): “Schooling quality n a cross-section of countries.” *Economica* 68, 465-488.

Barro, R. J., and J.-W. Lee (2001b): “International data on educational attainment: updates and implications.” *Oxford Economic Papers* 53, 541-563.

Becker, G.S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Chicago, University of Chicago Press.

Behrman, J. R., S. W. Parker, and P. E. Todd (2009): “Medium-term impact of the Oportunidades conditional cash transfer program on rural youth in Mexico”. in S. Klasen and F. Nowak-Lehmann (eds), *Poverty, Inequality and Policy in Latin America,* Cambridge *,* MA: MIT Press.

Bergh, A., and M. Karlsson (2010): “Government size and growth: Accounting for economic freedom and globalization”. *Public Choice* 142, 195-213.

Bjørnskov, C., Dreher, A. and Fischer, JAV (2007): "The bigger the better? Evidence of the effect of government size on life satisfaction around the world". *Public Choice* 127(3), 267-292.

Bjørnskov, C., Dreher, A. and Fischer, JAV, (2008). "Cross-country determinants of life satisfaction: exploring different determinants across groups in society". *Social Choice and Welfare* 30(1), 119-173.

Eaton, J. and H. Rosen (1980): "Taxation, human capital and uncertainty," *American Economic Review* 70, 705-715.

Ehrlich, I., and J. Kim (2007): "Social security and demographic trends: Theory and evidence from the international experience", *Review of Economic Dynamics* 10, 55-77.

Ehrlich, I. and J. G. Zhong (1998): "Social security and the real economy: An inquiry into some neglected issues". *American Economic Review* 88, 151-157.

Falch, T. and J.A.V. Fischer (2011): "Welfare state generosity and student performance: evidence from international student tests", MPRA paper 35269.

Fossen, F. M. and D. Glocker (2011): "Expected future earnings, taxation, and university enrollment". *International Tax and Public Finance* 16, 688-723.

Fölster, S. and M. Henrekson (2001): "Growth effects of government expenditure and taxation in rich countries". *European Economic Review* 45, 1501-1520.

Glomm, G., and B. Ravikumar (1992): "Public versus private investment in human capital: Endogenous growth and income inequality". *Journal of Political Economy* 100, 818-834.

Gwartney, J. and R. Lawson. (2002), *Economic Freedom of the World: 2002 Annual Report*. Vancouver: Fraser Institute.

Hamilton, J. H. (1987): "Optimal wage and income taxation with wage uncertainty". *International Economic Review* 28, 373-388.

Hanushek, E. A. and D. D. Kimko (2000): "Schooling, labor-force quality, and the growth of nations". *American Economic Review* 90, 1184-1208.

Hanushek, E. A. and J. A. Luque (2003): "Efficiency and equity in schools around the world". *Economics of Education Review* 22, 481-502.

Hanushek, E. A. and L. Wössmann (2008): "The role of school improvement in economic development". *Journal of Economic Literature* 46, 607-668.

Hanushek, E. A. and L. Wössmann (2012): "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation". *Journal of Economic Growth* 17, 267-321.

Heckman, J. J., L. Lochner and C. Taber (1998a): “Tax policy and human-capital formation”. *American Economic Review* 88, 293-297.

Heckman, J. J., L. Lochner and C. Taber (1998b): “Human capital formation and general equilibrium treatment effects: A study of tax and tuition policy”. *Fiscal Studies* 20, 25-40 .

Hogan, V. and I. Walker (2007): “Education choice under uncertainty: Implications for public policy”. *Labour Economics* 14, 894-912.

Jacobs, B. (2007): “Real options and human capital investment”. *Labour Economics* 14, 913-925.

Kneller, R., M. Bleaney and N. Gemmell (1999): “Fiscal policy and growth: Evidence from OECD countries”: *Journal of Public Economics*, 74, 171-190.

Konrad, K. A. and A. Spadaro (2006): “Education, redistribution taxation and confidence”. *Journal of Public Economics* 90, 171-188.

Laibson, D. (1997): “Golden eggs and hyperbolic discounting”. *Quarterly Journal of Economics* 112, 443-477.

Lee, J.-W., and R. J. Barro (1997): “Schooling Quality in a Cross-Section of Countries” *Economica* 68, 465-488. Data on http://www.nber.org/pub/barro.lee/

Levhari, D. and Y. Weiss (1974): "The effect of risk on investment in human capital". *American Economic Review* 64, 950-963.

O’Donoghue, T., and M. Rabin (1999): “Doing it now or later”. *American Economic Review* 89, 103-124.

OECD (2007): “$ocial €xpenditure 1980-2003: Interpretative Guide of SOCX”, Paris: OECD 2007.

Ponce, J., and A. S. Bedi (2008): “The impact of a cash transfer program on cognitive achievement: The Bono de Desarrolle Humano of Equador”. *Economics of Education Review*, forthcoming.

Postlethwaite, T. N., and D. E. Wiley (1992): "The IEA study of science II: Science achievement in twenty-three countries". Pegamon Press, New York

Poutvaara, P. (2007): "Social security incentives, human capital investment and mobility of labor". *Journal of Public Economics* 91, 1299-1325.

Rawlings, L. B., and G. M. Rubio (2005): "Evaluating the impact of conditional cash transfer programs". *The World Bank Research Observer* 20, 29-55.

Romero-Avila, D., and R. Strauch (2008): "Public finances and long-term growth in Europe: Evidence from a panel analysis". *European Journal of Political Economy* 24, 172-191.

Sinn, H.-W. (1995): "A Theory of the Welfare State". *Scandinavian Journal of Economics* 97, 495- 526.

Travers, K. J., and I. Westbury, eds. (1989): "The IEA Study of Mathematics I: Analysis of Mathematics Curricula". Pergamon Press, New York.

Warm, T.A. (1989). "Weighted Maximum Likelihood Estimation of Ability in Item Response Theory". *Psychometrika*, 54, 427-450.

Wössmann, L. (2003): "Schooling Resources, Educational Institutions and Student Performance: the International Evidence". *Oxford Bulletin of Economics & Statistics* 65, 117-170.

Zhang, J. and J. Zhang (2004): "How does social security affect economic growth? Evidence from cross-country data". *Journal of Population Economics* 17, 473-500.

Table 1. Data sources description

Year	Test organization	Acronym	Test subjects	Test age or grade	Countries	Data source
1980-81	IEA	SIMS	Mathematics	13 years	3 in 1980 14 in 1981	Lee and Barro (1997) Travers and Westbury (1989)
1983-85	IEA	SISS	Science	14 years	11 in 1983 11 in 1984 1 in 1985	Postlethwaite and Wiley (1992)
1990-91	IAEP	IAEP	Mathematics and Science	13 years	2 in 1990 17 in 1991	Lee and Barro (1997)
1994-95	IEA	TIMSS	Mathematics and Science	Grade 8	4 in 1994 36 in 1995	timss.bc.edu/
1998-99	IEA	TIMSS-repeat	Mathematics and Science	Grade 8	6 in 1998 31 in 1999	timss.bc.edu/
2000-02	OECD	PISA 2000	Mathematics and Science	15 years	32 in 2000 9 in 2002	www.pisa.oecd.org
2002-03	IEA	TIMSS 2003	Mathematics and Science	Grade 8	7 in 2002 38 in 2003	timss.bc.edu/
2003	OECD	PISA 2003	Mathematics and Science	15 years	40 in 2003	www.pisa.oecd.org

Note. For some countries separate scores are reported for different parts of the country. We have calculated mean country averages by using population as weight. IEA (except the 1983/84 test) and IAEP tests are conducted in the fall in the southern hemisphere and in the spring in the northern hemisphere. PISA 2000 originally only included five non-OECD countries, but nine additionally non-OECD countries conducted the same test in 2002.

Table 2. Types of social expenditures in OECD countries

Policy area	Programs
Old-age	Pensions, early retirement pensions, home-help, residential services for the elderly.
Survivors	Pensions and funeral payments.
Incapacity-related	Care services, disability benefits, benefits accruing from occupational injury and accident legislation, employee sickness payments.
Health	Spending on in- and out-patient care, medical goods, prevention.
Family	Child allowances and credits, childcare support, income support during leave, sole parent payments.
Active labour market policies	Employment services, training youth measures subsidised employment, employment measures for the disabled.
Unemployment	Unemployment compensation, severance pay, early retirement for labour market reasons.
Housing	Housing allowances and rent subsidies.
Other social policy areas	Non-categorical cash benefits to low-income households, other social services; i.e. support programmes such as, food subsidies, which are prevalent in some non-OECD countries.

Note. Source is Social Expenditure 1980-2003, OECD 2007, p.8.

Table 3. Descriptive statistics of government consumption and social expenditures

	Obser-vations	Mean	Standard deviation overall	Standard deviation within countries	Minimum value	Maximum value
General government consumption spending, percent of GDP	232 (All)	17.65	5.39	1.46	5.69	41.47
General government consumption spending, percent of GDP	124 (OECD)	18.90	4.22	1.05	10.08	29.62
Public sector social expenditures, percent of GDP	124 (OECD)	19.62	5.61	1.62	2.8	32.5
Active labor market policy spending, share of GDP	120 (OECD)	0.61	0.44	0.20	0	2.2
Public health spending, share of GDP	124 (OECD)	5.56	1.28	0.55	1.4	8.3
Family allowance spending, share of GDP	124 (OECD)	1.90	1.08	0.31	0	4.1
Unemployment benefit spending, share of GDP	120 (OECD)	1.17	0.90	0.44	0	4.4
Pension spending, share of GDP	124 (OECD)	6.38	2.76	0.76	0.6	12.8
Housing spending, share of GDP	102 (OECD)	0.42	0.39	0.17	0	1.8
Other social spending, share of GDP	99 (OECD)	4.00	1.38	0.64	1.50	8.90

Table 4. The effect of government consumption on student achievement

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gov. consumption spending, percent of GDP (log)	0.237	-1.462**	-0.650	-1.021*	-1.596**	-1.100**	-1.060	-1.086
	(0.429)	(0.378)	(0.509)	(0.486)	(0.558)	(0.377)	(0.945)	(1.551)
GDP per capita (log)	-	1.502**	1.234*	1.123*	2.056**	0.483	1.451	3.735*
		(0.213)	(0.485)	(0.475)	(0.748)	(0.292)	(0.698)	(1.473)
Percentage secondary school attained among adults (log)	-	0.763**	0.510	0.736*	2.238**	1.268**	0.424	1.454
		(0.292)	(0.409)	(0.371)	(0.836)	(0.271)	(0.639)	(1.052)
Population size (log)	-	0.040	2.279*	1.515	-8.492	-0.048	5.221*	-11.04
		(0.77)	(1.046)	(0.978)	(6.758)	(0.62)	(2.241)	(8.728)
Trend	-	-	-	-0.041**	-	-	-	-
				(0.014)				
Country fixed effects	No	No	Yes	Yes	Yes	No	Yes	Yes
Time fixed effects	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Country-specific trends	No	No	No	No	Yes	No	No	Yes
Observations	208	208	208	208	208	128	128	128
No of countries	59	59	59	59	59	28	28	28
Sample	All	All	All	All	All	OECD	OECD	OECD
R^2	0.0015	0.455	0.943	0.937	0.982	0.301	0.851	0.932
R^2 (within)	-	-	0.222	0.133	0.740	-	0.298	0.679

Note. Absolute standard errors in parentheses, +, * and ** denote significance at 10, 5 and 1 percent level, respectively..

Table 5. The effect of social expenditures on student achievement

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gov. social expenditures, percent of GDP (log)	-0.453**	-0.670**	-0.997*	-0.899+	0.505	-	-	-
	(0.223)	(0.230)	(0.464)	(0.457)	(1.070)			
GDP per capita (log)	-	0.717*	0.523	1.508+	6.500**	-0.686	0.546	-0.008
		(0.315)	(0.895)	(0.863)	(2.355)	(1.262)	(0.872)	(1.000)
Percentage secondary school attained among adults (log)	-	1.225**	0.548	0.433	0.572	-0.026	0.671	0.444
		(0.275)	(0.627)	(0.604)	(1.167)	(1.255)	(0.6216	(0.704)
Population size (log)	-	-0.028	4.448+	2.865	-11.63	-2.474	3.653	4.268
		(0.61)	(2.398)	(2.371)	(9.178)	(4.845)	(2.323)	(3.132)
Trend	-	-	-	-0.031	-	-	-	-
				(0.022)				
Pension spending (log)	-	-	-	-	-	-2.470*	-0.986**	-
						(1.033)	(0.333)	
Active labor market policy spending (log)	-	-	-	-	-	-0.392	-	-0.576**
						(0.298)		(0.191)
Unemployment spending (log)	-	-	-	-	-	-0.317	-	-
						(0.210)		
Family allowances (log)	-	-	-	-	-	0.824+	-	-
						(0.422)		
Health care spending (log)	-	-	-	-	-	0.208	-	-
						(0.814)		
Housing spending (log)	-	-	-	-	-	-0.042	-	-
						(0.198)		
Other spending (log)	-	-	-	-	-	0.912	-	-
						(0.667)		
Country fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Country-specific trends	No	No	No	No	Yes	No	No	No
Observations	121	121	121	121	121	80	121	113
No of countries	28	28	28	28	28	19	28	28
R^2	0.033	0.339	0.861	0.835	0.932	0.825	0.868	0.863
R^2 (within)	-	-	0.258	0.118	0.665	0.571	0.293	0.376

Note. Absolute standard errors in parentheses, +, * and ** denote significance at 10, 5 and 1 percent level, respectively. The coefficient on social expenditures in column (5) becomes negative but remains insignificant if model (5) is estimated on the same sample as column (6).

Table 6. The effect of tax progressivity on student achievement

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income tax rate progressivity	-0.172**	-0.009	-0.086*	-0.074+	-0.074	-0.133**	-0.025	0.018
	(0.061)	(0.059)	(0.042)	(0.039)	(0.057)	(0.046)	(0.050)	(0.066)
GDP per capita (log)	-	1.297**	1.113*	1.031+	2.118+	0.001	1.399+	4.794**
		(0.226)	(0.558)	(0.523)	(1.197)	(0.258)	(0.772)	(1.534)
Percentage secondary school attained among adults (log)	-	0.945**	-0.601	-0.51	1.013	1.403**	-0.067	0.637
		(0.336)	(0.496)	(0.449)	(1.043)	(0.285)	(0.671)	(1.258)
Population size (log)	-	0.109	0.654	-0.632	0.857	0.005	4.553	-1.245
		(0.085)	(1.245)	(1.178)	(10.023)	(0.061)	(2.789)	(10.999)
trend	-	-	-	-0.006	-	-	-	-
				(0.017)				
Country fixed effects	No	No	Yes	Yes	Yes	No	Yes	Yes
Time fixed effects	No	Yes	Yes	No	Yes	Yes	Yes	Yes
County specific trends	No	No	No	No	Yes	No	No	Yes
Observations	180	180	180	180	180	116	116	116
No of countries	56	56	56	56	56	28	28	28
Sample	All	All	All	All	All	OECD	OECD	OECD
R^2	0.043	0.387	0.958	0.952	0.983	0.280	0.855	0.939
R^2 (within)	-	-	0.188	0.066	0.674	-	0.290	0.700

Note. Absolute standard errors in parentheses, +, * and ** denote significance at 10, 5 and 1 percent level, respectively.

Table 7. Generosity of the welfare state: OECD countries

	(1)	(2)	(3)	(4)
GDP per capita (log)	0.766	-0.156	0.975	1.029
	(0.919)	(1.340)	(1.211)	(0.963)
Percentage secondary school attained among adults (log)	0.707	0.747	1.578+	0.357
	(0.658)	(0.832)	(0.846)	(0.668)
Population size (log)	2.969	2.919	5.581+	2.619
	(2.437)	(3.523)	(2.885)	(2.518)
Pension spending (log)	-1.057*	-	-	-
	(0.502)			
Active labor market policy spending (log)	-	-0.544*	-	-
		(0.209)		
Unemployment spending (log)	-	-	-0.112	-
			(0.182)	
Health care spending (log)	-	-	-	0.104
				(0.590)
Share of elderly (log)	-0.783	-	-	-2.048*
	(0.964)			(0.868)
Unemployment rates	-	-0.027	0.000	-
		(0.032)	(0.037)	
Country fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	110	110	113	110
Countries	27	28	27	27
R^2	0.8632	0.8687	0.8479	0.8544
R^2 (within)	0.3328	0.3995	0.3079	0.2902
F-test of joint significance	5.5078	3.458	0.2524	3.1086
(p-value)	0.006	0.0372	0.7776	0.051

Note. Absolute t-values in parentheses, +, * and ** denote significance at 10, 5 and 1 percent level, respectively.

Table 8. Post-1990 period and school quality measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Gov. consumption expenditures, percent of GDP (log)	-0.856	-	-	-0.924	-	-	-0.587	-	-
	(0.569)			(0.637)			(0.550)		
Gov. social expenditures, percent of GDP (log)	-	-1.360*	-	-	-1.496**	-	-	-1.394**	-
		(0.537)			(0.563)			(0.511)	
Income tax rate progressivity	-	-	-0.093+	-	-	-0.116*	-	-	-0.132**
			(0.054)			(0.046)			(0.049)
GDP per capita (log)	0.554	0.659	0.412	1.579*	0.152	2.005**	1.347*	1.216	0.901
	(0.577)	(1.008)	(0.676)	(0.616)	(1.024)	(0.697)	(0.542)	(1.030)	(0.618)
Percentage secondary school attained among adults (log)	-0.498	0.658	-1.039	-1.168+	-0.168	-1.426+	1.154*	0.571	-0.166
	(0.717)	(1.077)	(0.878)	(0.646)	(0.868)	(0.768)	(0.533)	(0.794)	(0.725)
Population size (log)	2.836+	4.303	1.579	1.178	6.258	0.906	1.647	2.266	0.628
	(1.557)	(4.117)	(1.771)	(1.516)	(4.087)	(1.466)	(1.190)	(3.189)	(1.363)
Primary education expenditures per pupil, percent of GDP (log)	-	-	-	0.412	0.433	0.579*	-	-	-
				(0.295)	(0.341)	(0.287)			
Pupil-teacher ratio in primary education	-	-	-	-	-	-	-0.005	0.048	-0.003
							(0.032)	(0.031)	(0.033)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	177	102	163	166	104	152	185	103	160
Countries	58	28	56	52	27	50	58	27	55
Sample	All	OECD	All	All	OECD	All	All	OECD	All
Time period 1990 – 2003	Yes	Yes	Yes	No	No	No	No	No	No
R^2	0.9617	0.8791	0.9626	0.9558	0.8740	0.9609	0.9484	0.8833	0.9645
R^2 (within)	0.2251	0.3290	0.1744	0.2377	0.3010	0.3161	0.2462	0.3398	0.2094
F-test (social spending, school quality)				1.5582	3.749	4.3742	0.5928	5.2155	3.9535
(p-value)				0.2156	0.029	0.0155	0.5545	0.0081	0.0226

Note. Absolute t-values in parentheses, + * and ** denote significance at 10, 5 and 1 percent level, respectively.

Table 9. The effect of welfare state generosity. Dependent variable is graduation rate at lower secondary education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gov. consumption spending, percent of GDP (log)	2.360	-	0.639	-5.260	-	-	-	-
	(2.568)		(5.158)	(5.263)				
Income tax rate progressivity	-	-0.207	-	-	-0.726**	-0.734**	-	-
		(0.315)			(0.220)	(0.245)		
Gov. social expenditures, percent of GDP (log)	-	-	-	-	-	-	1.396	3.838
							(4.070)	(4.081)
GDP per capita (log)	-0.373	-4.950	9.159*	-14.05	9.213**	16.83+	2.411	3.659
	(3.092)	(3.731)	(4.546)	(10.22)	(3.344)	(9.308)	(7.338)	(13.63)
Percentage secondary school attained among adults (log)	6.297	5.034	23.50**	0.250	24.34**	0.114	22.75**	0.386
	(3.873)	(4.449)	(5.169)	(0.747)	(4.092)	(0.594)	(6.096)	(0.716)
Population size (log)	18.28**	15.34+	19.68*	0.0196	14.35+	59.07	3.102	-193.3+
	(5.654)	(8.270)	(9.873)	(67.90)	(7.733)	(54.48)	(17.86)	(98.65)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country specific trends	No	No	No	Yes	No	Yes	No	Yes
Observations	497	338	188	188	154	154	165	165
No of countries	55	48	22	22	21	21	21	21
Sample	All	All	OECD	OECD	OECD	OECD	OECD	OECD
R^2	0.00324	0.0177	0.0196	0.0916	0.0165	0.111	0.383	0.236
R^2 (within)	0.262	0.358	0.524	0.566	0.680	0.653	0.494	0.540
Mean of dependent variable	82.67%	86.09%	87.80%	87.80%	89.18%	89.18%	88.02%	88.02%

Note. Absolute standard errors in parentheses, +, * and ** denote significance at 10, 5 and 1 percent level, respectively.

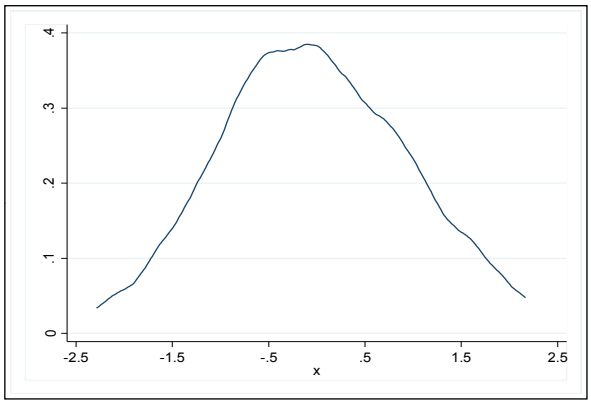
Appendix Table A1. Participating countries and within-country variation

	Number of relevant international tests	Mean value dependent variable	Difference between maximum and minimum values observed in the regression sample			
			Dependent variable	Log(government consumption)	Log(Social expenditures)	Income tax progressivity
Albania	1	-3.17	0	0	-	-
Argentina	1	-2.86	0	0	-	0
Armenia	1	-2.02	0	0	-	-
Australia	6	0.13	0.89	0.08	0.34	2
Austria	3	0.18	0.57	0.10	0.05	0
Bahrain	1	-3.81	0	0	-	0
Belgium	4	0.12	0.57	0.10	0.05	1
Botswana	1	-5.74	0	0	-	0
Brazil	3	-4.45	1.93	0.01	-	0
Bulgaria	4	-0.77	2.35	0.22	-	5
Canada	7	0.11	1.05	0.24	0.34	2
Chile	2	-4.07	0.87	0.11	-	0.4
China	3	1.19	1.14	0.02	-	1.6
Colombia	1	-5.46	0	0	-	0
Cyprus	3	-2.37	1.01	0.04	-	3
Czech Republic	4	0.47	1.36	0.11	0.15	2
Denmark	3	-0.84	1.57	0.05	0.11	1
Egypt,	1	-4.03	0	0	-	0
Estonia	1	0.55	0	0	-	0
Finland	5	0.36	1.31	0.16	0.20	1.6
France	5	-0.09	1.29	0.08	0.27	2.4
Germany	3	-0.44	0.49	0.03	0.04	2
Ghana	1	-9.32	0	0	-	0
Greece	3	-1.51	0.49	0.14	0.10	0
Hong Kong	6	0.52	2.38	0.42	-	0
Hungary	7	0.57	2.49	0.12	0.10	5
Iceland	3	-0.71	1.73	0.16	0.20	2
Indonesia	3	-3.39	0.68	0.21	-	1.2
Iran	3	-3.19	0.72	0.28	-	5.6
Ireland	4	-0.49	1.27	0.22	0.18	2
Israel	6	-1.04	1.49	0.37	-	3.4
Italy	6	-1.00	1.60	0.12	0.20	5.8
Japan	6	1.49	1.00	0.25	0.54	3
Jordan	3	-3.21	1.67	0.43	-	0
Korea	6	1.24	1.94	0.18	0.81	4.2
Kuwait	1	-4.95	0	0	-	-
Latvia	4	-1.05	1.32	0.16	-	2
Lebanon	1	-4.05	0	0	-	-
Lithuania	3	-1.37	1.82	0.18	-	0
Luxembourg	3	-1.35	1.26	0.17	0.22	0
Macedonia	3	-2.68	0.82	0.08	-	-

Malaysia	2	-0.57	0.07	0.35	-	0.4
Mexico	2	-2.88	0.68	0.11	0.16	0
Moldova	2	-1.99	0.32	0.64	-	-
Morocco	2	-5.41	1.19	0.09	-	0.2
Mozambique	1	-8.79	0	0	-	-
Netherlands	6	0.79	0.83	0.12	0.27	2.6
New Zealand	6	-0.28	1.39	0.17	0.15	1.8
Nigeria	1	-2.58	0	0	-	0
Norway	4	-0.60	1.09	0.18	0.12	2.2
Peru	1	-4.77	0	0	-	0
Philippines	3	-4.91	1.68	0.62	-	0
Poland	3	-0.39	0.77	0.04	0.08	0
Portugal	4	-2.03	1.47	0.15	0.47	1
Romania	3	-1.86	0.50	0.88	-	4
Russian Federation	5	-0.16	1.65	0.27	-	2
Saudi Arabia	1	-5.76	0	0	-	-
Serbia and Montenegro	2	-1.44	1.11	0	-	-
Singapore	4	1.71	4.40	0.39	-	1
Slovak Republic	3	0.12	0.71	0.06	0.09	1
Slovenia	3	0.04	1.23	0.04	-	0
South Africa	3	-8.56	2.00	0.08	-	1.4
Spain	4	-1.03	0.78	0.05	0.06	4
Swaziland	1	-2.50	0	0	-	-
Sweden	5	-0.30	1.17	0.12	0.13	5.6
Switzerland	4	0.38	1.15	0.07	0.35	1
Thailand	6	-1.37	1.76	0.29	-	1
Tunisia	2	-3.33	1.44	0.01	-	0
Turkey	2	-2.59	0.54	0.11	0	0.4
United Kingdom	7	-0.28	1.35	0.19	0.13	6.2
U.S.A.	7	-0.73	1.15	0.18	0.20	4.2
Uruguay	1	-2.33	0	0	-	0

Figure 1. Kernel density of student achievement

a) "Core" country observations



b) All observations

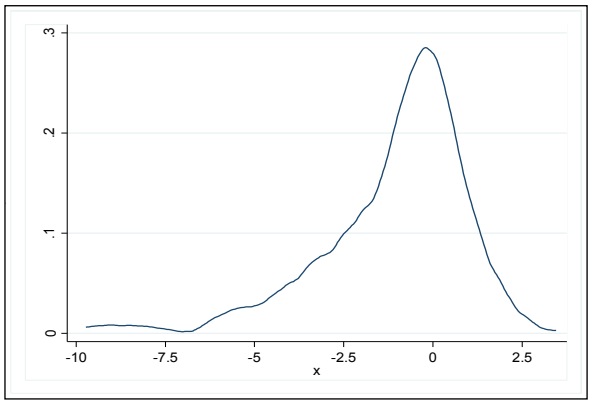
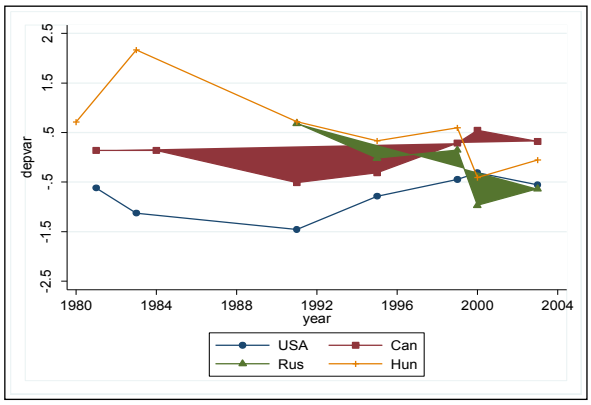
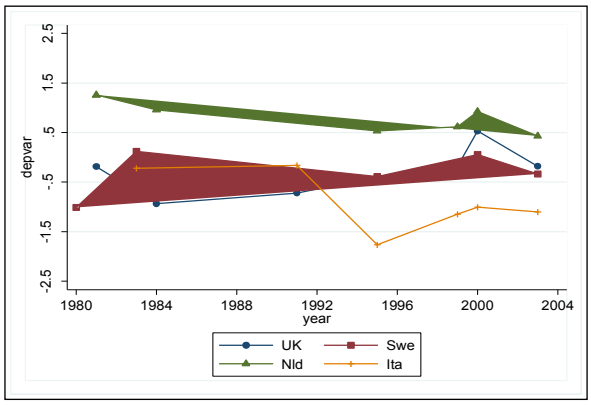
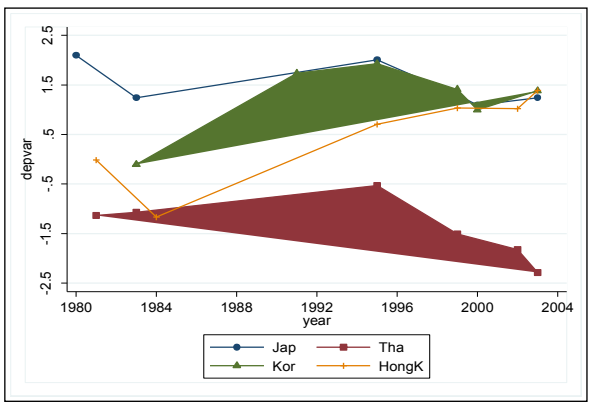


Figure 2. Country-specific development in relative student achievement.







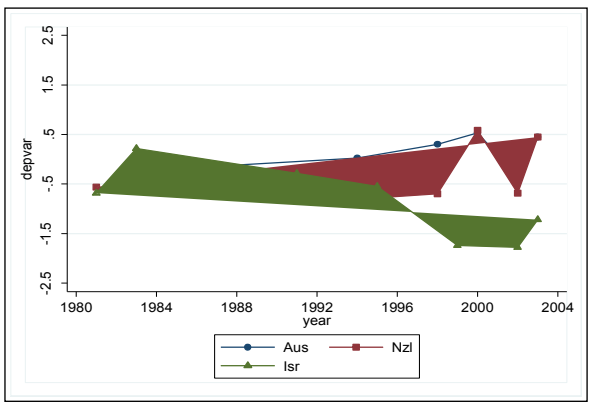


Figure 3. Kernel density of change in student achievement

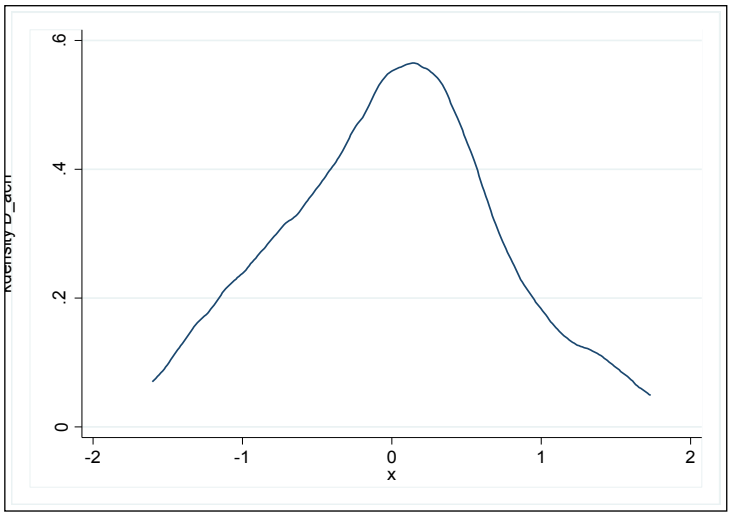


Figure 4. Country-specific development in social expenditures as share of GDP