ISSN 1503-299X

WORKING PAPER SERIES

No. 4/2016

Welfare state generosity and student performance: Evidence from international student tests 1980–2003

Torberg Falch*
Department of Economics
Norwegian University of Science and Technology

Justina AV Fischer
SOEP at the DIW, Berlin

Department of Economics

Norwegian University of Science and Technology
N-7491 Trondheim, Norway
http://www.ntnu.edu/econ/working-papers

Welfare state generosity and student performance: Evidence from international student tests 1980–2003

Torberg Falch[*]
Department of Economics, Norwegian University of Science and Technology

Justina AV Fischer[+]
SOEP at the DIW, Berlin

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; H5, I2; C33

July 2016

Acknowledgement
We thank 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.

[*] 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
[+] Socio-Economic Panel (SOEP) Division at the DIW, Deutsches Institut fuer Wirtschaftsforschung (DIW), Mohrenstrasse 58, DE-10117 Berlin, Germany. , e-mail: mail@justinaavfischer.de, tel.: 030 89789-285, fax: 030 89789-115

Justina Fischer acknowledges Marie Curie experienced researcher fellowship (RTN ENABLE) and a COFIT fellowship.

1. Introduction

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

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 employ achievement on international student tests, adjusted to facilitate comparability across countries and time, as our 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, 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.

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] The insurance aspect of the welfare state manifests in a system that reduces the risk related to future income and redistributes from individuals with high income to individuals with low income.

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, student's maximization problem includes 'student effort' as a cost. In real-life, however, student's effort in compulsory education is influenced by instructions of parents and teachers. 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

[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.

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 behaviour 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 the behaviour of parents and teachers.[2]

Viewing the welfare state as a mechanism 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 studies by Heckman et al. (1998a, b) show that a progressive income tax system yields lower incentives to invest in human capital than a proportional tax system. Viewing the welfare state as a mechanism that reduces risk, Levhari and Weiss (1974) shows 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 labour market when educational choices are modelled in a real option framework.[3]

Welfare state arrangements may also be seen as interventions in imperfect markets, working in an efficiency enhancing direction. The evidence on public sector size and economic growth in empirical cross-country studies, however, indicates a negative relationship.[4] 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.[5] Using German data, Fossen and Glocker (2011) find that university enrollment is positively related to expected return to tertiary education.

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

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

[3] Numerous papers have analysed 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.

[4] 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.

[5] 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.

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 59 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).[6] 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 and services (e.g. hospitals), but also through direct financial transfers to households (e.g. pension spending and active labour market policies), or through the income tax system (e.g. exemption of low incomes from tax) which we measure indirectly through the progressivity of the income tax schedule. The results are robust to a variety of model specifications, including models conditioning on educational expenditures. The negative effect of welfare state generosity seems not to be mediated by the resource use in education. Our finding is corroborated when we replace student performance in Mathematics and Science with the completion rate in lower secondary education.

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

[6] 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.

2. Data

2.1. 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 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 are 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 Science separately, but we restrict our attention to student assessments in Mathematics and Science for several reasons. First, these two Natural Science subjects have more similarities among each other than with Reading and are thus more suitable for constructing a joint 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 approximate 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 achievement tests we utilize are described in Table 1.[7]

Insert Table 1 about here

---

[7] 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.

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 test score of *one* specific country will depend on the achievement of the students in *all other* participating countries. Since the composition of participating countries changes from one test year to the next, raw test scores for a particular country is not necessarily 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.[8]

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.[9] 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 labour-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

[8] 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.

[9] 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 the USA provides test scores for all tests.

and, in addition to adjusting the means, correct the dispersion of each single test in a similar way as we do.[10]

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 - in most cases developing countries that participate less frequently in international tests - exhibit a low level of student achievement.

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 adjusted 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 average student achievement, although the variation is lower than that for the distribution in *levels*.[11]

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.[12]

[10] 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".

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

[12] 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.

2.2. 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 income constraint on private household's 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.[13] In contrast to government consumption spending, this measure of social expenditures 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 favourable tax treatment of contributions to private health plans)" (OECD, 2007, p. 7), excluding the administrative costs of executing them. The correlation coefficient between these two alternative measures of welfare state, total social expenditures and government consumption spending, is equal to 0.67 in our sample.

[13] 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).

Table 2 provides an overview of the components of social expenditures. The major population is, in principle, entitled to all social spending programs, so that each component may exert an independent effect on student achievement of its own. However, different components typically target different age groups. We take account of this feature when employing disaggregated measures of social expenditures in the analysis.

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 increases from 17% in 1980 to 19% in 2003.[14] The Netherlands is the only country which cut down social expenditures, while Japan has experienced the largest growth. Note that social expenditures serve as automatic stabilizers; particularly when measured as share of GDP, they typically shrink during a boom but 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 that we apply in the empirical analysis. Across OECD countries, the variance in social expenditures is slightly higher than 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. Expenditures targeted towards the elderly (pension spending and spending on health) is the largest component of social expenditures.

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

[14] 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.

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.[15]

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 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. Data on educational attainment of adults is obtained from Barro and Lee (2001b), which are available on a quinquennial basis up to the year 2000[16]; 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 analyses, we include current data on primary school educational expenditures (% of GDP),[17] and the completion rate in lower secondary education defines as the share of entrants into secondary education actually attaining a lower secondary degree, both obtained from the World Bank (WDI, 2007, updated 2014).

[15] 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).

[16] 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.

[17] 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.

3. Model specification and identification

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

$$H_{it} = \beta_1 g_{it} + \beta_2 \log(GDP/POP)_{it} + \beta_3 EDU_{it} + \phi_i + \varphi_t + \varepsilon_{it}, \quad (1)$$

where g_{it} represents one specific measure of the welfare state generosity in country i at time t. Family characteristics as parental income and education have strong effects in micro studies of student achievement, and the model includes GDP per capita (GDP/POP) and the share of adult population with at least some secondary education (EDU). 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 independent variables have positive trends.[18] Country-specific fixed effects ϕ_i account for time-invariant differences across countries (e.g., culture for education and country size). For each measure of welfare state generosity, a separate estimation of model (1) is carried out.[19]

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. 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 labour 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

[18] 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.

[19] Previous versions of this paper included population size in the empirical model (Falch and Fischer, 2011). The effect of population size turned out to be mainly insignificant and unstable across model specifications, probably because of limited within-country variation. Qualitatively, the effects of the welfare generosity measures are not affected by the inclusion of population size.

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 as presented in Section 2. Second, it might also be argued that government spending and social spending are both proxies for educational expenditures: if educational expenditures have a positive effect on student achievement, omitting it will bias the (expected negative) effect of welfare state generosity towards zero since these two government spending variables are likely to be positively correlated. In response we will present additional results of models including educational expenditures per pupil in primary education as a percentage of GDP extracted from the World Bank education database. Notice, however, that the empirical evidence suggests that resource use in education is not a decisive determinant of student performance (e.g. Hanushek and Luque, 2003).

Third, 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. In addition, Hanushek and Woessmann (2012) argue that test designs and test procedures have improved over time, which implies that the dependent variable may incorporate a larger measurement error in the 1980s than in later periods. Thus, we will present separate results for the post-1990 period. In addition, we estimate models with country-specific time trends to capture potentially differing time paths across countries. While the latter specification has the clear advantage of better controlling for potentially omitted variables, it comes at the cost of weakening the signal-to-noise ratio to the extent that (residual) welfare state arrangements change too slowly for identification. Further, we estimate separate models for OECD countries that, arguably, are more homogenous than the world sample, and we expand the model to include unemployment and the age composition of the population.

A more desirable approach to address potential biases would be to employ the instrumental variable model. We think, however, that it is hard to come up with valid instruments for welfare state arrangements in a panel-data framework. However, there are some noteworthy attempts in the literature to instrument institutions and their policy-making 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 very heterogeneous developing countries while we focus on more homogenous developed countries. With respect to reverse causation, conventional wisdom suggests that it should not be a major issue in our analysis: even if welfare reforms are often triggered by poor economic performance and problems in the labour market, unsatisfactory performance of schools is, to our knowledge, not used as justification for cutting down welfare state spending.

Finally, we have to discuss the various dynamic patterns of welfare state generosity that are inhibited in the contemporaneous measures we use. From the viewpoint of economic theory, it is the *expected* welfare state arrangements in the future, not the present-time arrangements, that should affect educational investment decisions made today. We argue that, in our empirical framework, the present-time level of government redistribution activities might be the best proxy for individuals' expectations. In our empirical approach, the country fixed effects capture the average spending level over time that may persist due to a status quo bias through a country's institutions. However, it is the conditional short-term within-country changes that identify the results in our model specification. As educational production is cumulative, expectations of students and parents affecting student effort in primary education play an important role for student achievement observed at age 13 to 15 years. Thus, present-time levels of government welfare spending are to some extent a leaded measure for the actual decisions made by students' parents in the past. We will investigate the robustness of the results by using five-years moving averages in the independent variables in some model specifications.

4. Empirical Results

4.1. Government consumption

Table 4 presents results for government consumption spending including 59 countries and 208 observations. The first column presents the results for the model specification in equation (1) above.[20] The effect of government consumption is sizable, but statistically insignificant at conventional levels. Taken at face value, when the share of government consumption increases by 0.1 log-points (approximately 10 percent), student achievement declines by 0.08 "core" country standard deviations.

[20] The full sample includes 232 observations with information of both the dependent variable and government consumption. Missing observations for the variable for adult education reduces the estimation sample with 24 observations. Estimating a model without the variable for adult education, the effect of government consumption is unaffected by the change in sample.

As expected, there are positive effects of GDP per capita and adult education. The positive income effect mirrors Hanushek and Kimko (2000) and Hanushek and Wössmann (2008, 2012) 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. Thus, conditional on income, a small public sector is favourable 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. The positive effect of adult educational attainment mirrors results from previous micro-econometric studies.

Insert Table 4 about here

In the model in column (1) in Table 4, the estimated standard error of government spending is relative large, which may indicate that the within-country variations in student achievement and government consumption are limited for statistical identification. Notably, the R^2 is as high as 0.94. However, it may equally well be that it is the time-specific effects that complicate identification. The purpose of the scaling of the test scores described in Section 2 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 the independent variables have positive trends. Indeed, while the p-value of joint significance of the time-specific fixed effects is 0.07 in the model in column (1) in Table 4, the p-value is only 0.25 when a simple trend is added.

In column (2) in Table 4 we replace the time fixed effects with a common time trend. The coefficient of the trend variable is negative as expected, indicating a positive trend in the independent determinants. The R^2 appears only marginally lowered, while the within-R^2 is clearly reduced. Interestingly, the effect of government consumption spending is significantly negative at 5 percent level in this specification. The estimated coefficient increases from about -0.8 to about -1.1, while the standard error declines slightly.

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 (3) in Table 4 expands the model with country-specific time trends. In this case, the effect of government consumption increases to -1.6 and becomes highly significant. We conclude that it is mainly the fluctuations *around* country-specific

trends that account for the association between government consumption spending and student achievement. The increase in significance from column (1) to column (3) is not related to the fact that the number of countries that contribute to identification is necessarily smaller in the latter model.[21] We conclude that the effect of government size appears significant at conventional levels, depending on the handling of the time effects.

The remaining models in Table 4 address methodological issues discussed in Section 3. Government consumption spending might be a proxy for educational expenditures. , If there is a positive effect of educational expenditures that is captured by government consumption spending, we expect to underestimate the effect of government spending. Column (4) in Table 4 adds educational expenditures per pupil in primary schools as a percentage of GDP from the World Bank education database.[22] The effect of educational expenditures is positive at the 10 percent level, while the effect of government consumption becomes slightly larger by -0.2 log-points compared to the estimate in baseline column (1).[23]

Another issue is that student tests from the 1980s might be of lower quality than the subsequent tests. The student test scores from the 1980s have been collected and assembled differently compared to the test scores from the post-1990 period, see the presentation in Section 3. Therefore, the dependent variable may incorporate a larger measurement error in the 1980s than in later periods. Columns (5)-(7) in Table 4 present result for regressions on the subsample for the 1990-2003 period, reducing the number of observations by 15%. The effect of government consumption spending appears qualitatively unchanged, although it is slightly smaller in the model with country-specific time trends. The variation that drives the results appears to differ in the shorter time period than in the whole sample. While the variation across country-specific trends aids identifying a strong effect for the whole sample, inclusion of trends in the shorter sample reduces the effect of government consumption.

The final part of table 4 restricts the sample to OECD countries in order to ease comparison with the models using social expenditures in Table 5 below. The subsample of OECD countries is defined by membership in the year 2000. In the OECD sample, the effect of

[21] For 11 countries in the regression sample, we have only one observation. These countries do not contribute to the identification in any model in Table 4. For another 7 countries, we have only two observations. These countries do not contribute to the identification in the model in Model 3 with country-specific trends in addition to the country fixed effects. If we exclude these countries from the Model 1, the effect of government consumption remains negative, and marginally significant at 10% level (the point estimated changes from -0.81 to -0.88).

[22] For secondary education, the number of observations with data is insufficient. The correlation coefficient of per pupil spending in primary education and government consumption spending is equal to 0.34. We have also included the teacher-student ratio from the World Bank education database. This measure is negatively correlated with government spending, but adding the variable to the model does not alter the results.

[23] The increased effect is not a result of a smaller sample. Due to missing observations of education expenditures, the sample is 20% smaller. Estimating the model in column (1), the effect of government consumption spending is unaffected by the change in sample.

government consumption is again similar to the results for the full sample, but the precision is weaker

In sum, we identify a negative impact of government consumption spending on student achievement as measured by international tests between 1983 and 2003. This evidence is in accordance with our hypothesis that a more generous welfare system generates disincentives for students' educational investments. Our estimates indicate that when government consumption spending increases by 0.1 log-points, student achievement is reduced by about 0.1 "core" country standard deviations.

4.2. Social expenditures

Table 5 presents results for government total social expenditures available for 28 OECD countries, resulting in a sample of 121 observations. The models in Table 5 correspond to the models in Table 4. The model in column (1) includes country and time fixed effects, column 2 replaces the time fixed effects with a general time trend, and column 3 replaces the general time trend with country-specific time trends. The effect of government social expenditures is almost exactly the same as the effect of government consumption spending in columns (1) and (2). In addition, the effect of GDP per capita is similar, while the effect of adult education is close to zero.

Insert Table 5 about here

However, when including country-specific time trends, the performance-lowering impact of social expenditures completely disappears (column (3) in Table 5). The next column includes education expenditures, increasing the effect of government expenditures as in the model for government consumption. 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.

The last part of Table 5 present results for models only using information for the post-1990 period. As for government consumption spending, the results are qualitatively the same in this shorter time period.

Table 6 distinguishes between different components of social expenditures relating to different age groups as defined in Tables 2 and 3. In column (1) we replace total social expenditures with the four components that sum up to total expenditures. Note that

information on some components is missing in several countries, which reduces the regression sample to 19 countries and 80 observations.

The effects of spending towards the working population (unemployment benefits and labour market policies) and towards the old (pensions and health care) are negative and significant at the 10% level. The positive effect of spending targeted towards families and children may indicate that relaxing parents' budget constraints in the poorest families may have an attainment-increasing effect on their children, but the effect is insignificant.[24] The last spending component, 'other spending', is of a rather 'kitchen-sink' nature so that its estimate is not easy to interpret.

Insert Table 6 about here

The positive correlations among the different social expenditure components may contribute to their heterogeneous and mainly insignificant effects in column (1) in Table 6. Thus, we have run regressions including each of the components separately. In all cases, the coefficients have the same sign as in column (1). Columns (2) and (3) report the two cases where the coefficient is significant at the 10 percent level.[25] Both payments targeted towards the elderly and the working age population reduces students' test scores. Since the latter constitutes only a small part of total social expenditures (see Table 3), the negative effect of total social expenditures seems to a large part to be driven by pension and health care spending.[26]

The sizes of government consumption spending and social expenditures, 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. Unfortunately, precise information on number of recipients is not easily available. Columns (4) and (5) in Table 6 add available proxies for number of recipients. We employ the share of elderly in the population in the model for social expenditures towards the

[24] Analyses of cash transfer programs in developing countries, which typically transfer money to poor families conditional on their investments in human capital, 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, 2010; Behrman et al., 2009).

[25] In the model including only expenditures targeted towards children and families, the effect is marginally insignificant (p-value of 0.102), with a coefficient similar to the result in column (5).

[26] 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.

elderly and the share of unemployed in the active population in the model for social expenditures towards the working age 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.[27]

The results are not sensitive to inclusion of proxies for the number of beneficiaries. The effect of social expenditures towards the working age population becomes stronger, while the effect of social expenditures towards the elderly becomes somewhat smaller and more imprecise. The similarity of the results suggests that the bias from using spending measured per GDP (conditional on population size) in place of per recipient is rather small.

Taken all together, Tables 5 and 6 show 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 and health care spending that contribute the most to this effect. Possibly, because of path dependency in policy-making, current changes in spending on pensions and health care may have a strong predictive power on governments' future policies. In most OECD countries, pension systems, are strongly redistributive in nature: 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 consumption levels in the non-active elderly population.

4.3. Tax progressivity

An alternative measure of welfare state generosity is the degree of redistribution between the rich and the poor through the income tax system. In this section we use 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. Table 7 presents results for the same model specifications as in Tables 4 and 5. There is a negative effect also for this measure. The effect is significant at the 5 percent level in the model with country and time fixed effects (column 1), but the precision is reduced in the model with country-specific trends (column 3). The point estimates suggest, however, that the development of the tax income progression over time is not captured by country-specific time trends. When the index of income tax progressivity increases by one standard deviation, which is about 2.5 points, student

[27] 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.

achievement is reduced by 0.20 adjusted "core" country standard deviations according to the model in column (1) in Table 7.[28]

The model in column (4) in Table 7 includes education expenditures, which again increases the effect of interest. Column (5) includes government consumption spending per capita, the variable for government size in Table 4. Is the effect of income tax progressivity simply capturing government size? The results indicate it is not. The effect of tax progressivity remains negative and significant, while the effect of government size is negative and insignificant in this smaller sample compared to the sample in Table 4. Finally, Column (6) shows that the effect of tax progressivity is similar for the post-1990 period as for the whole sample, while column (7) indicates that the effect is smaller for the sample of OECD countries.

Insert Table 7 about here

How the effect of the variable of interest changes when we alter the model specification varies greatly between the government spending, social expenditures and tax progressivity models (Tables 4-7), which indicates that the variables have very different features.[29] 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.

4.4. Robustness analyses

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 that are important, but the development in the medium or long term. We have carried out identical regressions as reported in Tables 4-7 using 5-year moving averages of the independent variables in place of using current values. The findings for government consumption spending appear partly sensitive to the choice of time window, although a robust

[28] When using lead and lags by about one year of tax progressivity in place of interpolated values every 5th year (see Section 2), we obtain qualitatively similar results. For example, re-estimating the model in columns (1) to (3) – that is the models with time fixed effects, only a common time trend and country- specific time trends , the results are (standard error) -0.112 (0.056), -0.109 (0.052) and -0.042 (0.100), respectively, and very close to the estimates reported in Table 7.

[29] 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.

and large performance lowering effect at the 1 percent level remains if country-specific time trends are included. The effect of social spending is less sensitive to using 5-year moving averages of the independent variables, but the coefficients are less precisely estimated and insignificant in the main specifications. However, the student performance lowering impact of spending towards the working population and towards the old remains negative and significant. Regarding the progressivity of the tax system for OECD countries, the estimation of 5-year moving averages corroborates the previous results, while the coefficients for the full sample are now smaller and insignificant, albeit all with negative signs.[30] Overall, the results for welfare state generosity appear insensitive to changes in model specification and sample selection.

One interesting question is whether our findings are sensitive to the type of measure of student attainment used. The PISA/TIMSS student test scores evaluate student performance at the age of 13-15 years, which is at the end of compulsory education: the educational attainment comparable with this age is roughly a completed lower secondary education. For the same set of countries and time horizon 1980-2003 used in our previous analyses, the World Bank (WDI, 2007, updated 2014) provides information on the completion rate in lower secondary education, defined as the share of the students starting lower secondary education who obtain the degree (typically at age 16). However, since this variable is measured more regularly than student achievement tests, we have more country-year observations. Table 8 re-estimates the model using this dependent variable for the same countries as for the main analyses above. The mean value in the different samples used in the regressions is presented at the bottom of Table 8. It is relatively common not to participate in education up to the completion from lower secondary education. The mean completion rate is 83% and 88% in the full sample and the OECD sample, respectively.

Insert Table 8 about here

Overall we find insignificant effects of the generosity of the welfare state on the completion rate in lower secondary education in the full sample, but significantly negative effects of income tax progressivity 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). In contrast to the analyses for the PISA/TIMSS data in

[30] 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 6.

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. Another interesting question is whether the effect of welfare state generosity carries over to non-compulsory education. However, suitable international comparative data back to the 1980s are scarce, in particular for completion rates in upper secondary and tertiary education. We leave this analysis to future research when data availability has improved.

5. 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 centred 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 that we have 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 analysing completion rates in 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: findings in form of econometric 'point estimates' always must be interpreted as marginal, 'local' changes, allowing for conclusions for small percent-point changes in government spending only. Furthermore, our findings are for high- and middle-high-income countries only – whether or not similar results can be found in a developing countries context remains an open question.

References

ACEMOGLU, D., JOHNSON, S. and ROBINSON, J.A. (2001). The colonial origins of comparative development: an empirical investigation. *The American Economic Review*, 91, 1369-1401

AGELL, J., OHLSSON, H. and THOURSIE, P.S. (2006). Growth effects of government expenditure and taxation in rich countries: A comment. *European Economic Review*, 50, 211-8.

ALBOUY, D. (2012). The Colonial Origins of Comparative Development: An Empirical Investigation: Comment. *The American Economic Review*, 102, 3059-76.

ANDERSSON, F. and KONRAD, K.A. (2003). Human capital investment and globalization in extortionary states. *Journal of Public Economics*, 87, 1539-55.

BARRO, R.J., and LEE, J.-W. (2001a). Schooling quality n a cross-section of countries. *Economica*, 68, 465-88.

____, ____ (2001b). International data on educational attainment: updates and implications. *Oxford Economic Papers*, 53, 541-63.

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

BEHRMAN, J.R., PARKER, S.W. and TODD, P.E. (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 KARLSSON, M. (2010). Government size and growth: accounting for economic freedom and globalization. *Public Choice*, 142, 195-213.

BJØRNSKOV, C., DREHER, A. and FISCHER, J.A.V. (2007). The bigger the better? Evidence of the effect of government size on life satisfaction around the world. *Public Choice*, 127, 267-92.

____, ____ and ____ (2008). Cross-country determinants of life satisfaction: exploring different determinants across groups in society. *Social Choice and Welfare*, 30, 119-73.

EATON, J. and ROSEN, H. (1980). Taxation, human capital and uncertainty. *American Economic Review*, 70, 705-15.

EHRLICH, I., and KIM, J. (2007). Social security and demographic trends: theory and evidence from the international experience. *Review of Economic Dynamics*, 10, 55-77.

EHRLICH, I. and ZHONG, J.G. (1998). Social security and the real economy: an inquiry into some neglected issues. *American Economic Review*, 88, 151-7.

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

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

FÖLSTER, S. and HENREKSON, M. (2001). Growth effects of government expenditure and taxation in rich countries. *European Economic Review*, 45, 1501-20.

GLOMM, G. and RAVIKUMAR, B. (1992). Public versus private investment in human capital: Endogenous growth and income inequality. *Journal of Political Economy*, 100, 818-34.

GWARTNEY, J. and LAWSON, R. (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, 373-88.

HANUSHEK, E.A. and KIMKO, D.D. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review*, 90, 1184-208.

____ and LUQUE, J.A. (2003). Efficiency and equity in schools around the world. *Economics of Education Review*, 22, 481-502.

____ and WÖSSMANN, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46, 607-68.

____ and ____ (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17, 267-321.

HECKMAN, J. J., LOCHNER, L. and TABER, C. (1998a). Tax policy and human-capital formation. *American Economic Review*, 88, 293-7.

____, ____ and ____ (1998b). Human capital formation and general equilibrium treatment effects: A study of tax and tuition policy. *Fiscal Studies*, 20, 25-40 .

HOGAN, V. and WALKER, I. (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, 913-25.

KNELLER, R., BLEANEY, M. and GEMMELL, N. (1999). Fiscal policy and growth: evidence from OECD countries. *Journal of Public Economics*, 74, 171-90.

KONRAD, K.A. and SPADARO, A. (2006). Education, redistribution taxation and confidence. *Journal of Public Economics*, 90, 171-88.

LAIBSON, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112, 443-77.

LEE, J.-W. and BARRO, R.J. (1997). Schooling Quality in a Cross-Section of Countries. *Economica*, 68, 465-88. Data on http://www.nber.org/pub/barro.lee/

LEVHARI, D. and WEISS, Y. (1974). The effect of risk on investment in human capital. *American Economic Review*, 64, 950-63.

O'DONOGHUE, T. and RABIN, M. (1999). Doing it now or later. *American Economic Review*, 89, 103-24.

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

PONCE, J. and BEDI, A.S. (2010). The impact of a cash transfer program on cognitive achievement: the Bono de Desarrolle Humano of Equador. *Economics of Education Review*, 29, 116-25.

POSTLETHWAITE, T. N. and WILEY, D.E. (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, 1299-325.

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

ROMERO-AVILA, D. and STRAUCH, R. (2008). Public finances and long-term growth in Europe: evidence from a panel analysis. *European Journal of Political Economy*, 24, 172-91.

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

TRAVERS, K.J., and WESTBURY, I. (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, 427-50.

WÖSSMANN, L. (2003). Schooling resources, educational institutions and student performance: the international evidence. *Oxford Bulletin of Economics & Statistics*, 65, 117-70.

ZHANG, J. and ZHANG, J. (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
1.	Old-age	Pensions, early retirement pensions, home-help, residential services for the elderly.
2.	Survivors	Pensions and funeral payments.
3.	Incapacity-related	Care services, disability benefits, benefits accruing from occupational injury and accident legislation, employee sickness payments.
4.	Health	Spending on in- and out-patient care, medical goods, prevention.
5.	Active labour market policies	Employment services, training youth measures subsidised employment, employment measures for the disabled.
6.	Unemployment	Unemployment compensation, severance pay, early retirement for labour market reasons.
7.	Housing	Housing allowances and rent subsidies.
8.	Family	Child allowances and credits, childcare support, income support during leave, sole parent payments.
9.	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, social expenditures and tax progressivity

	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	5.7	41.5
General government consumption spending, percent of GDP	124 (OECD)	18.90	4.22	1.05	10.1	29.6
Public sector social expenditures, percent of GDP	124 (OECD)	19.62	5.61	1.62	2.8	32.5
Social expenditures targeted towards elderly, percent of GDP[1]	124 (OECD)	11.94	3.45	1.05	2.11	19.3
Social expenditures targeted towards the working age population, percent of GDP[2]	115 (OECD)	1.81	1.21	0.49	0.3	6.3
Social expenditures targeted towards children and families, percent of GDP[3]	86 (OECD)	2.78	1.09	0.37	0.4	5.0
Other social expenditures, percent of GDP[4]	80 (OECD)	3.97	1.34	0.55	1.5	8.4

Note. [1] Consists of expenditure types 1-4 in Table 2. [2] Consists of expenditure types 5-6 in Table 2. [3] Consists of expenditure types 7-8 in Table 2. [4] Consists of expenditure type 9 in Table 2

Table 4. The effect of government consumption on student achievement

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Government consumption spending, percent of GDP (log)	-0.814	-1.119*	-1.601**	-1.061	-1.068+	-0.796	-0.910	-1.012	-1.115	-1.790
	(0.510)	(0.484)	(0.560)	(0.611)	(0.563)	(0.635)	(0.632)	(0.969)	(1.589)	(1.227)
GDP per capita (log)	1.338**	1.214*	2.018**	1.682*	0.617	0.838	0.771	2.186**	3.612*	2.212**
	(0.490)	(0.474)	(0.750)	(0.600)	(0.582)	(0.957)	(0.899)	(0.638)	(1.477)	(0.685)
Percentage secondary school attained among adults (log)	0.674	0.836*	2.245**	-1.247	-0.035	1.516	-0.374	-0.532	1.484	-0.992
	(0.408)	(0.367)	(0.839)	(0.637)	(0.678)	(1.652)	(1.112)	(0.502)	(1.057)	(0.712)
Primary education expenditures per pupil, percent of GDP (log)	-	-	-	0.480+	-	-	0.763*	-	-	0.307
				(0.281)			(0.303)			(0.368)
Time trend	-	-0.031*	-	-	-	-	-	-	-	-
		(0.013)								
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country-specific trends	No	No	Yes	No	No	Yes	No	No	Yes	No
Observations	208	208	208	166	177	177	152	128	128	110
No of countries	59	59	59	52	58	58	52	28	28	27
Sample	All	All	All	All	1990-	1990-	1990-	OECD	OECD	OECD
R^2	0.941	0.936	0.981	0.956	0.961	0.988	0.962	0.841	0.930	0.854
R^2 (within)	0.195	0.119	0.743	0.233	0.201	0.750	0.216	0.253	0.671	0.288

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

Table 5. The effect of total social expenditures on student achievement

	(1)	(2)	(3)	(4)	(5)	(6)
Gov. social expenditures, percent of GDP (log)	-0.892+	-0.840+	0.082	-1.474*	-1.367*	-1.405*
	(0.467)	(0.456)	(1.023)	(0.569)	(0.537)	(0.613)
GDP per capita (log)	0.938	1.736*	6.116*	0.752	1.142	1.060
	(0.880)	(0.844)	(2.348)	(0.956)	(0.896)	(0.941)
Percentage secondary school attained among adults (log)	-0.227	-0.062	0.464	0.363	0.219	0.532
	(0.475)	(0.452)	(1.170)	(0.341)	(0.992)	(0.392)
Primary education expenditures per pupil, percent of GDP (log)	-	-	-	0.363	-	0.532
				(0.341)		(0.392)
Time trend	-	-0.019	-	-	-	-
		(0.019)				
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	No	Yes	Yes	Yes	Yes
Country-specific trends	No	No	Yes	No	No	No
Observations	121	121	121	104	102	95
No of countries	28	28	28	27	28	27
Sample	All	All	All	All	1990-	1990-
R^2	0.855	0.832	0.936	0.869	0.877	0.877
R^2 (within)	0.226	0.103	0.659	0.275	0.317	0.305

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 components of social expenditures on student achievement

	(1)	(2)	(3)	(4)	(5)	(6)
Spending targeted towards elderly, percent of GDP (log)	-2.336+ (1.266)	-1.303* (0.497)	-	-0.947 (0.693)	-	-2.980+ (1.701)
Spending targeted towards the working age population, percent of GDP (log)	-0.533* (0.252)	-	-0.362+ (0.214)	-	-0.541* (0.244)	-0.473 (0.302)
Spending targeted towards children and families, percent of GDP (log)	0.535 (0.400)	-	-	-	-	0.083 (0.497)
Other social expenditures, percent of GDP (log)	0.662 (0.543)	-	-	-	-	0.502 (0.769)
GDP per capita (log)	-0.151 (1.110)	1.282 (0.868)	0.546 (1.117)	1.449 (0.896)	0.741 (1.237)	-0.250 (1.107)
Percentage secondary school attained among adults (log)	0.485 (0.782)	-0.229 (0.466)	-0.211 (0.484)	-0.106 (0.489)	0.539 (0.619)	0.651 (1.144)
Share of elderly in population (log)	-	-	-	-1.308 (0.976)	-	-
Unemployment rate	-	-	-	-	0.0168 (0.035)	-
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Country-specific trends	No	No	No	No	No	No
Observations	80	112	121	110	110	70
No of countries	19	27	28	27	27	19
Sample	All	All	All	All	All	1990-
R^2	0.813	0.861	0.839	0.877	0.877	0.813
R^2 (within)	0.520	0.255	0.245	0.297	0.350	0.517

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 7. The effect of tax progressivity on student achievement

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Income tax rate progressivity	-0.088*	-0.072+	-0.073	-0.119*	-0.094+	-0.054	-0.088*
	(0.042)	(0.039)	(0.057)	(0.046)	(0.054)	(0.047)	(0.043)
GDP per capita (log)	1.134*	1.009+	2.107+	2.069**	0.413	2.045**	1.135
	(0.555)	(0.520)	(1.181)	(0.686)	(0.675)	(0.670)	(0.559)
Percentage secondary school attained among adults (log)	-0.658	-0.434	1.019	-1.429+	-1,017	-0.812	-0.564
	(0.482)	(0.425)	(1.032)	(0.766)	(0.877)	(0.498)	(0.498)
Primary education expenditures per pupil, percent of GDP (log)	-	-	-	0.634*	-	-	-
				(0.272)			
Government consumption spending, percent of GDP (log)	-	-	-	-	-	-	-0.417
							(0.554)
Time trend	-	-0.011	-	-	-	-	-
		(0.014)					
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	No	Yes	Yes	Yes	Yes	Yes
Country specific trends	No	No	Yes	Yes	No	No	No
Observations	180	180	180	152	163	116	179
No of countries	56	56	56	50	56	28	56
Sample	All	All	All	All	1990-	OECD	All
R^2	0.958	0.952	0.983	0.961	0.962	0.850	0.958
R^2 (within)	0.186	0.063	0.674	0.313	0.168	0.264	0.187

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

Table 8. Welfare state generosity and completion of lower secondary education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Government consumption spending, percent of GDP (log)	2.203	-	2.307	-5.261	-	-	-	-
	(2.596)		(5.143)	(4.876)				
Income tax rate progressivity	-	-0.128	-	-	-0.699**	-0.769**	-	-
		(0.313)			(0.222)	(0.243)		
Government social expenditures, percent of GDP (log)	-	-	-	-	-	-	1.893	1.376
							(2.879)	(3.928)
GDP per capita (log)	-2.111	-6.747+	9.464*	-14.05	10.39**	14.46	2.353	4.828
	(3.080)	(3.620)	(4.591)	(9.999)	(3.320)	(9.053)	(7.300)	(13.78)
Percentage secondary school attained among adults (log)	10.26**	7.909+	25.21**	0.250	27.33**	0.123	22.57**	0.466
	(3.716)	(4.189)	(5.151)	(0.743)	(3.817)	(0.594)	(5.986)	(0.723)
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	0.747	0.755	0.887	0.900	0.944	0.941	0.916	0.921
R^2 (within)	0.243	0.350	0.510	0.566	0.670	0.649	0.494	0.525
Mean of dependent variable	82.67%	86.09%	87.80%	87.80%	89.18%	89.18%	88.02%	88.02%

Note. Dependent variable is the completion rate at lower secondary education. Absolute standard errors in parentheses, +, * and ** denote significance at 10, 5 and 1 percent level, respectively.

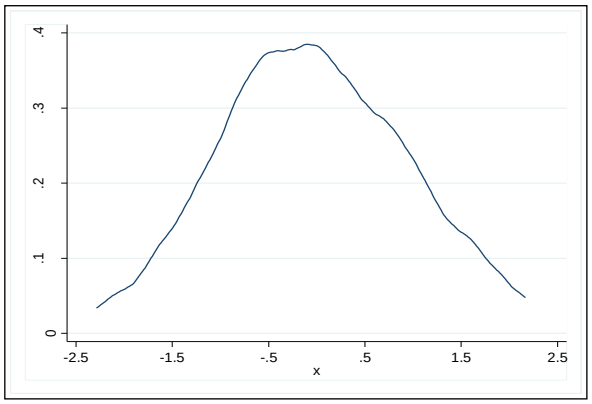
Appendix Table A1. Participating countries and within-country variation

	Number of relevant international student 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

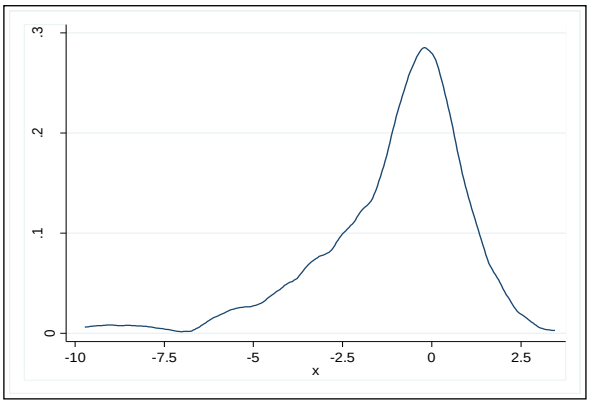
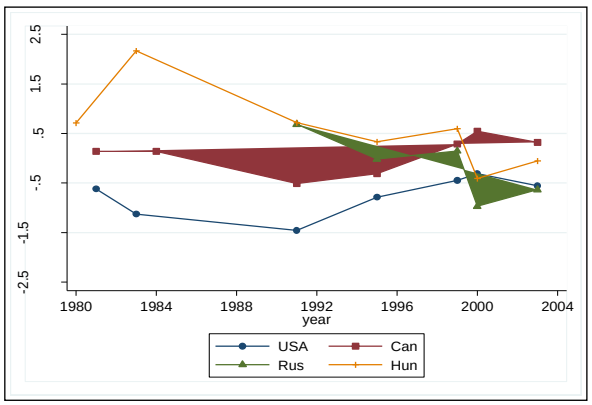
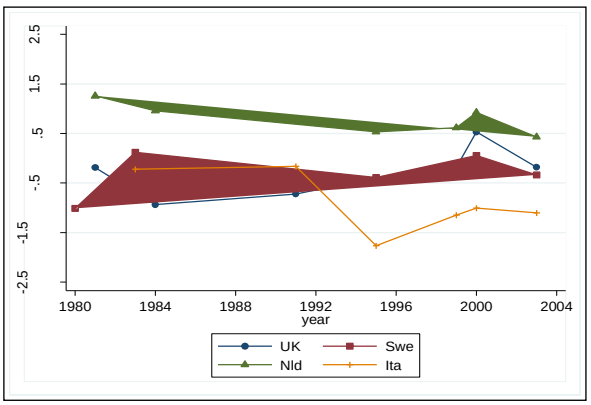
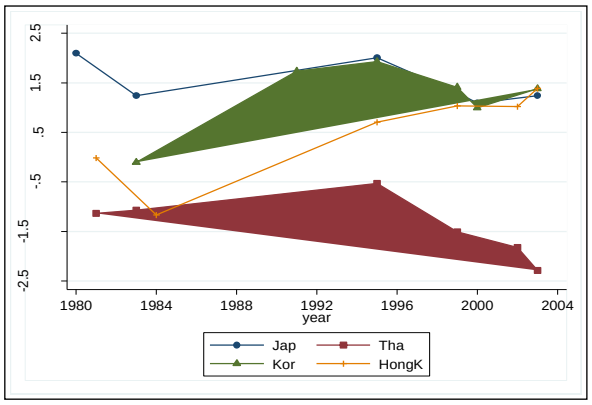


Figure 2. Country-specific development in relative student achievement.







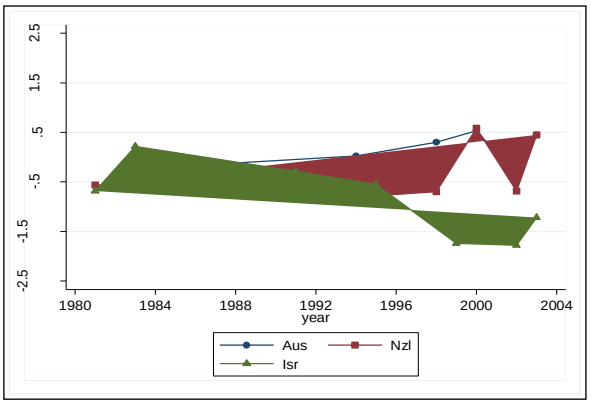


Figure 3. Kernel density of change in student achievement

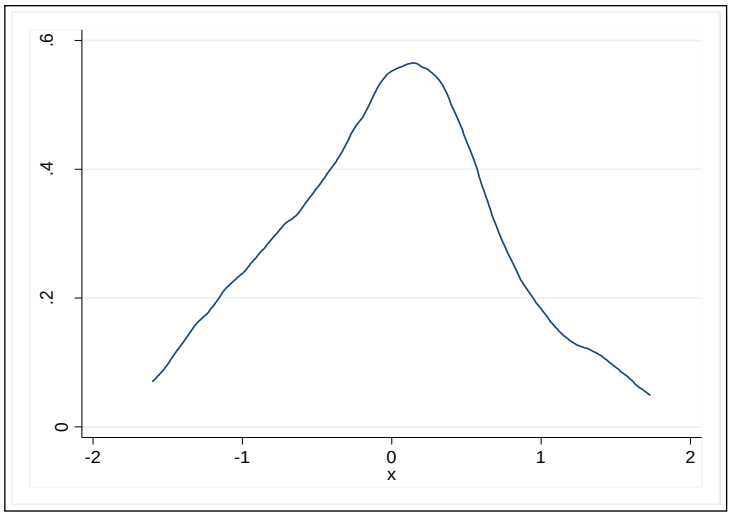


Figure 4. Country-specific development in social expenditures as share of GDP