

IZA DP No. 9559

Test Scores, Noncognitive Skills and Economic Growth

Pau Balart
Matthijs Oosterveen
Dinand Webbink

December 2015

Test Scores, Noncognitive Skills and Economic Growth

Pau Balart

University of the Balearic Islands

Matthijs Oosterveen

Erasmus University Rotterdam

Dinand Webbink

*Erasmus University Rotterdam,
Tinbergen Institute and IZA*

Discussion Paper No. 9559
December 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Test Scores, Noncognitive Skills and Economic Growth*

Many studies have found a strong association between economic outcomes of nations and their performance on international cognitive tests. This association is often interpreted as evidence for the importance of cognitive skills for economic growth. However, noncognitive skills, such as motivation and perseverance, are also important for the performance on cognitive tests. This study decomposes the performance on an international test (PISA) into two components that differ with respect to their underlying skills: the starting level and the decline in performance during the test. The first component can be interpreted as a measure of cognitive skills, whereas the second component captures noncognitive skills. We find that countries differ in the starting level and in the decline in performance, and that these differences are stable over time. Both components have a positive and statistically significant association with economic growth, and the estimated effects are quite similar. This suggests that noncognitive skills are important for explaining the relationship between test scores and economic growth.

JEL Classification: J24

Keywords: cognitive skills, noncognitive skills, long run economic growth

Corresponding author:

Dinand Webbink
Department of Economics
Erasmus School of Economics
Erasmus University Rotterdam
PO Box 1738
3000 DR Rotterdam
The Netherlands
E-mail: webbink@ese.eur.nl

* We would like to thank Lex Borghans, Antonio Cabrales, Eric Hanushek, Sacha Kapoor and Trudie Schils for helpful comments and suggestions. Seminar participants at ASSET, CEMFI, the Economics of Education Association, the Erasmus School of Economics, the Tinbergen Institute and the Trends and Challenges on Human Resources International Workshop are also gratefully acknowledged for their feedback. Balart thanks the Spanish Ministry of Economy and Competitiveness for its financial support through grant ECO2012-34581.

1 Introduction

Many studies have found a strong association between the economic outcomes of nations and their performance on international cognitive tests, such as PISA, TIMSS or PIRLS (see, for example, [Hanushek and Kimko 2000](#); [Hanushek and Woessmann 2008](#) and [2012](#)). This association is interpreted as evidence for the importance of cognitive skills for productivity and economic growth. However, the performance on cognitive tests is not only the result of cognitive skills but also influenced by noncognitive skills, such as motivation and perseverance. Pioneers in intelligence testing like Thorndike and Wechsler already recognized that test takers might not exert maximal effort ([Wechsler 1940](#)). Several recent studies show that noncognitive skills are important for the performance on cognitive tests. For instance, [Duckworth et al. \(2011\)](#) find that under low-stakes research conditions, such as in the international cognitive tests, some individuals try harder than others. Moreover, scores on cognitive tests can be substantially improved by offering a reward (e.g. [Gneezy and Rustichini 2000](#); [Almlund et al. 2011](#); [Segal 2012](#) and [Borghans et al. 2008](#)). Noncognitive factors have also been shown to be important for productivity and other social outcomes at the individual level (e.g. [Heckman and Rubinstein 2001](#) and [Heckman et al. 2013](#)). This suggests that noncognitive skills might be an important omitted variable in the relationship between cognitive skills and the economic outcomes of nations. It is therefore unclear to which extent the strong association between the performance on international cognitive tests and economic growth should be interpreted as evidence for the importance of cognitive skills.¹

This paper aims to get more insight into the importance of cognitive and noncognitive skills for the relationship between test scores and economic growth. The main novelty of our analysis is that we decompose the performance on an international test (PISA) into two components: the starting level and the decline in performance during the test. This decomposition, recently introduced by [Borghans and Schils \(2013\)](#), exploits the random allocation of test booklets to students, which generates exogenous variation in the position of questions in the test. This specific feature of the test allows estimation of the decline in performance during the test that is not confounded by unobserved characteris-

¹In line with [Heckman et al. \(2006\)](#) we treat noncognitive skills as a separate input for economic outcomes.

tics of questions, such as the difficulty of the test items. [Borghans and Schils \(2013\)](#) show that differences in the decline in performance during the test are related to noncognitive factors, such as motivation and ambition, recognized in the studies mentioned above. The starting level of the test score provides a measure of cognitive skills that is not confounded by the noncognitive factors that cause the decline in performance. This implies that the decomposition of test scores generates two components that differ with respect to their underlying skills. Noncognitive skills are related to the performance decline, whereas the starting level of the test scores can be interpreted as a measure of cognitive skills.

We use the results of the decomposition for estimating the association between the two components and economic growth, and compare these findings with the estimated effect of test scores before the decomposition, which is the standard approach in the previous literature. For the analysis we use data from a seminal paper on cognitive skills and economic growth ([Hanushek and Woessmann \(2012\)](#) hereafter HW (2012)). This study has established a strong association between test scores and economic growth and has found evidence that supports a causal interpretation of the effect of test scores on economic growth. Within this framework we decompose the test scores into two components and estimate the effect of these two components on economic growth. In our empirical approach we try to stay as close as possible to the framework of HW (2012). However, the decomposition method, which relies on a specific feature of the data collection, can only be applied to the PISA test, which is only one of the tests that is included in HW (2012). This implies that we use the PISA scores as a proxy for the test score index used in HW (2012). The PISA scores, however, are highly correlated with the test score index used in HW (2012) ($r = 0.91$).

This paper makes several contributions to the current economic literature. We contribute to the literature that studies the impact of human capital and skills on economic growth. To our knowledge, no previous study has investigated the association between noncognitive skills and productivity at the macroeconomic level. A reason for this might be the lack of international comparable measures for noncognitive skills. In this study we generate and use a measure that is international comparable. A further contribution is that this measure is based on performance. Most studies on noncognitive skills rely

on self-reports of individuals, which complicates international comparisons. Performance based measures have the advantage that they do not suffer from the typical measurement issues related to self-reports, such as reference bias (e.g. [Paulhus 1984](#) and [Kautz et al. 2014](#)). In addition, by controlling for noncognitive skills we might improve previous estimates of the association between cognitive skills and economic growth from the literature. It should be noted that these previous estimates might also be biased by identification issues like reverse causality or omitted variables. Our study does not aim to contribute with respect to these issues, but only focuses on the decomposition of the test scores into two components that differ in their underlying skills.

We find that countries differ in both the starting level and the decline in performance during the test and that these differences are stable over time. Both components of test scores have a positive and statistically significant association with economic growth. The size of the estimated effects of the two components is very similar. This suggests that noncognitive skills are also important for explaining the relationship between test scores and economic growth. Moreover, we find that the effect of cognitive skills reduces with approximately forty percent in models that control for noncognitive factors. This indicates that previous estimates of the effects of cognitive skills on economic growth might be upwardly biased.

This study is organized as follows. Section 2 discusses the previous literature on the effect of cognitive skills on economic growth and the recent literature on the importance of noncognitive skills. Section 3 explains the PISA-decomposition and Section 4 explains the estimation of the cross-country growth regressions. The data used in the analyses are described in Section 5. Section 6 shows the main estimation results. Section 7 investigates the robustness of the results to using a stricter measure of the performance decline and Section 8 concludes.

2 Previous Studies

2.1 The Relationship between Cognitive Test Scores and Economic Growth

A large empirical literature has studied the impact of human capital on economic growth. One of the major challenges is to find a good proxy for human capital. Many studies have used average educational attainment as a measure for human capital (see, for example, Barro 1991; Krueger and Lindahl 2001; Sala-i Martin et al. 2004; Doménech and De la Fuente 2006; Cohen and Soto 2007 and Sunde and Vischer 2015). However, this proxy seems quite imperfect as it assumes that a year spent in school produces the same amount of human capital across all countries. Therefore, Lee and Lee (1995) and Hanushek and Kimko (2000) introduced a new approach that uses the performance on international cognitive tests as a proxy for human capital. The main advantage of this approach is that cognitive test scores can be considered as an output measure that captures what students have learned inside and outside of school. The basic cross-country growth specification in Hanushek and Kimko (2000) regresses the average economic growth of country c (G_c) for a specific period on their measure of human capital (H_c), GDP per capita at the beginning of the period (GDP_{0c}) and control variables (Z_{nc}) such as years of schooling and population growth:

$$G_c = \beta_0 + \beta_1 H_c + \beta_2 GDP_{0c} + \sum_n \delta_n Z_{nc} + \epsilon_c \quad (1)$$

This approach has been extended in a series of studies, which estimate Equation (1) and have very similar results and interpretation (see Barro 2001; Hanushek and Woessmann 2008, 2011a, 2011b and 2012; Hanushek 2013 and Jamison et al. 2007). Equation (1) is consistent with the endogenous growth models of Romer (1990) and Nelson and Phelps (1966). In these models growth is attributed to the stock of human capital, which generates innovations or facilitates the adoption and imitation of new technologies. We focus on the most recent paper, HW (2012), where they have a sample of 50 countries and use cognitive test scores between 1964 and 2003. They consistently find that cognitive test scores are strongly associated with economic growth and interpret this as the importance of cognitive skills. The estimated effects of cognitive skills are large: a one standard devi-

ation increase in test scores is associated with 1.25 to 2 percentage points higher average annual growth rate in GDP per capita across 40 years.

An important question is to which extent the consistent evidence about the association between test scores and economic growth reflects a causal effect of cognitive skills on economic performance. This is a difficult question because it is very hard to address typical identification issues like omitted variables, reverse causality and measurement error. However, HW (2012) show that the estimated effects of cognitive test scores on economic growth are robust to alternative estimation approaches, such as instrumental variables, differences-in-differences and longitudinal analysis of changes in cognitive test scores and in growth rates. Moreover, HW (2012) note that their estimation relies upon the assumption that the average scores for a country tend to be relatively stable over time, which leads them to conclude that differences in cognitive skills lead to economically significant differences in economic growth.

Although these studies find a consistent positive relationship between cognitive test scores and economic growth it remains unclear how these results should be interpreted because test scores might not only be the result of cognitive skills but also the result of noncognitive skills.

2.2 Noncognitive Skills, Long-term Individual Outcomes and Cognitive Test Scores

Many studies in psychology and a more recent literature in economics have established the importance of noncognitive skills for individual socioeconomic outcomes ([Almlund et al. 2011](#)). These studies often use personality traits like the Big Five personality inventory as measures of noncognitive skills ([Costa and McCrae 1992](#) and [John and Srivastava 1999](#)) and find that these personality measures are as predictive as cognitive measures for outcomes such as economic success, health and criminal activity, even after controlling for family background and cognition. Intervention studies, like the Perry Pre School Program, provide evidence for a causal effect of changes in personality traits on economic and social outcomes ([Heckman et al. 2013](#)). Further evidence on the importance of noncognitive skills for individual economic success can be found in [Heckman and](#)

Rubinstein (2001), Heckman et al. (2006), Heckman and Kautz (2012) and Kautz et al. (2014).

Noncognitive skills have also been related to the performance of students on cognitive tests. Pioneers in intelligence testing, such as Thorndike and Wechsler, recognized the possibility that test takers might not exert maximal effort. For instance, Wechsler (1940) already noted that intelligence tests not only measure intelligence and pointed out that the tendency to try hard on low-stakes intelligence tests might derive from nonintellective traits, such as competitiveness and compliance with authority. More recently, Duckworth et al. (2011) provide evidence for the role of test motivation in intelligence testing. Observer ratings of test motivation, based on the behavior of adolescent boys completing intelligence tests, explains IQ-scores and reduces the predictive validity of IQ-scores for life outcomes, particularly for nonacademic outcomes. Their findings show that under low-stakes research conditions some individuals try harder than others. Economists have also recognized that engaging in complex thinking is effortful and therefore motivation to exert effort affects the performance on achievement tests (Borghans et al. 2011). For example, Borghans et al. (2008) find that noncognitive skills, where subjects were given questionnaires to determine psychological traits and were asked to make trade-offs to determine relevant economic preference parameters, have a direct impact on cognitive test scores.² Moreover, various studies have found that offering a material reward can substantially improve scores on cognitive tests (Gneezy and Rustichini 2000 and Segal 2012).

These findings have encouraged the possibility of relying on achievement test scores to obtain non-self reported measures of noncognitive skills. Hernández and Hershaff (2014) propose to use skipped items in a non-penalized test as a measure of noncognitive skills. For our study, the proposal by Borghans and Schils (2013) provides the non-self reported measure of noncognitive skills. It should be noted that the measurement of noncognitive skills has been the object of controversy and even the Big Five is not universally accepted mainly as a consequence of its self-reported nature (Paulhus 1984 and Duckworth et al. 2011). By relying on a non-self reported measure we can avoid these problems and, at the same time, obtain internationally comparable measures of noncognitive skills.

²This finding is consistent with Borghans et al. (2011) and Heckman and Kautz (2012), who find that personality variables explain roughly a third of explained variance in achievement tests.

Table 1: Rotation design of the 13 PISA booklets

Booklet	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	Science 1	Science 2	Science 4	Science 7
2	Science 2	Science 3	Math 3	Reading 1
3	Science 3	Science 4	Math 4	Math 1
4	Science 4	Math 3	Science 5	Math 2
5	Science 5	Science 6	Science 7	Science 3
6	Science 6	Reading 2	Reading 1	Science 4
7	Science 7	Reading 1	Math 2	Math 4
8	Math 1	Math 2	Science 2	Science 6
9	Math 2	Science 1	Science 3	Reading 2
10	Math 3	Math 4	Science 6	Science 1
11	Math 4	Science 5	Reading 2	Science 2
12	Reading 1	Math 1	Science 1	Science 5
13	Reading 2	Science 7	Math 1	Math 3

Source: [OECD \(2009\)](#)

3 The Test Score Decomposition

In a recent study [Borghans and Schils \(2013\)](#) introduce an approach that decomposes test scores into the starting level and the decline in performance during the test. They observed that students perform worse on questions that are at a later position in the test. Two students with the same cognitive skills might score very differently on a cognitive test if one of the students is strongly motivated and the other student does not want to exert effort. It might be expected that these students score quite similar on the first items of the test where motivation is less important, but in the next stages of the test the first student will probably try harder and therefore obtain a higher score. The difference in the decline in test scores can then be attributed to a difference in (test) motivation as a noncognitive skill. A concern with this interpretation is that the decline in test scores might be related to unobservable characteristics, such as the difficulty of the test items.³ If this was the case, the performance decline would be a consequence of cognitive skills rather than noncognitive skills. To address this important issue, [Borghans and Schils \(2013\)](#) exploit the variation in the question ordering of the PISA test. As shown in Table 1, PISA 2006 has 13 different versions of the test (booklets), all of them containing four clusters of questions (test items). A booklet contains approximately 50 to 70 test items. Each cluster of questions takes 30 minutes of test time and students are allowed a short break after one hour. There are 13 clusters of test items (7 science, 2 reading and 4 math) and they are distributed over the 13 different booklets according to a rotation

³In fact, the sequencing of items from easy to difficult is used as an explicit strategy for sustaining morale ([Duckworth et al. 2011](#)).

scheme. Each cluster appears in each of the four possible positions within a booklet once (OECD 2009). This means that one specific test item appears in four different positions of four different booklets. For instance, cluster Science 1 is included in booklets 1, 9, 12 and 10 as respectively the first, second, third and fourth cluster. This rotation scheme generates variation in the question number (position in the test) of test items. These booklets are randomly assigned to students (OECD 2009). This random assignment ensures that the variation in question numbers, that results from the ordering of clusters, is unrelated to characteristics of students. Balancing tests confirm this random allocation of booklets and show that background characteristics of students are unrelated to the number of the booklet (see Table A.1 in the Appendix).

The variation in question numbers can then be exploited for estimating the decline in performance during the test by using the following fixed-effects model:

$$P[Y_{ij} = 1] = F(\alpha_0 + \alpha_1 Q_{ij} + \sum_{j=2}^J \mu_j) \quad (2)$$

with Y_{ij} being the score of student i on question j , Q_{ij} is the position of question j in the version of the test answered by student i and μ_j is a question fixed effect that takes account of unobservable characteristics of question j , such as the difficulty of a test item. Due to the random allocation of booklets to students and the inclusion of question fixed effects, the estimated parameter α_1 will not be biased by unobserved factors and can be interpreted as the decline in performance during the test. The decomposition of the test scores into the starting level and the performance decline is based on the estimation of Equation (2). We estimate Equation (2) separately for each country by using a probit model, as in Borghans and Schils (2013), and use the PISA weighting factors to ensure that the sample is representative.⁴ The parameter α_1 measures the decline in performance during the test of a specific country. The parameter α_0 measures the starting level of a specific country, since the question numbers have been rescaled such that the first item is numbered as 0 and the last item as 1. Both measures are robust to the definition of the start of the test. For instance, excluding the first five questions does not affect the estimates of the two components. We use all test items for estimating Equation (2).

⁴Estimating Equation (2) with OLS gives very similar results.

Unreached items were coded as incorrectly answered questions. This allows us to stay closer to the framework of HW (2012) in which uncompleted items were interpreted as incorrectly answered to compute final test scores.⁵

We have also estimated Equation (2) using the average performance on all test items within a cluster as the outcome variable. In this analysis the clusters have been rescaled such that the first cluster is numbered as 0 and the fourth cluster is numbered as 1. With this approach the unit of randomization exactly matches the unit of analysis. The results are very similar to the results from the main approach. We find a correlation of 0.94 for the estimates of the starting level of the two approaches, and a correlation of 0.97 for the decline in performance.

Interpretation of the Two Components

The main idea of the decomposition is that it generates components that differ in their underlying skills. The first component, the starting level, is a measure of cognitive skills that is not confounded by the noncognitive factors that cause the decline in performance. The second component, the performance decline, is expected to capture noncognitive skills such as test motivation and perseverance. [Borghans and Schils \(2013\)](#) provide four pieces of evidence in support of this interpretation of the two components. First, the performance decline differs from the students' performance at the start of the test, which indicates that the two components measure different types of skills. Second, they show that the two components are stable for the years 2003 and 2006 and that there are differences between countries. This suggests that the two components are able to measure stable traits of the 15-year-old population of a country. Third, they show that the performance decline is related to personality traits and motivation. With the data collected in the Dutch Inventaar 2010 study they find that especially agreeableness (a Big Five personality trait) and motivation towards learning have a strong positive interaction effect with the performance decline. This means that more motivated students have a smaller performance decline. Fourth, using data from the British Cohort Study 1970, they show that the performance decline predicts future outcomes above and beyond achievement.

⁵[Borghans and Schils \(2013\)](#) note that it is unclear which type of skills determine that test items are not reached. This issue is less important in our application as we do not interpret the performance decline as a measure of noncognitive skills only.

Differences between Countries and Years

Results of the decomposition of the PISA test of 2006 are shown in Table 2. Equation (2) is used for computing the probability of correctly answering the first and the last question of the PISA test. Column (1) shows the average of the PISA 2006 test scores, column (2) shows the probability of correctly answering the first question, column (3) shows the probability of correctly answering the last question and column (4) shows the difference between these two probabilities. Column (2) can be interpreted as the starting level of a country and column (4) can be interpreted as the performance decline. Countries are ranked with respect to the performance decline from high to low.

Table 2: The starting level and decline in performance per country

Country	(1) PISA score	(2) P[$Q_0 = 1$]	(3) P[$Q_1 = 1$]	(4) Decline	Country	(1) PISA score	(2) P[$Q_0 = 1$]	(3) P[$Q_1 = 1$]	(4) Decline
Colombia	381	.585	.249	.337	Poland	500.3	.821	.73	.091
Uruguay	422.7	.718	.429	.289	United States	481.5	.774	.683	.091
Argentina	382	.632	.364	.268	Lithuania	481.3	.734	.644	.09
Tunisia	377	.45	.233	.217	China, Macao	509.3	.826	.737	.089
Brazil	384.3	.543	.336	.208	Luxembourg	485	.804	.715	.089
Kyrgyzstan	306	.391	.187	.204	Hungary	492.3	.786	.698	.087
Mexico	408.7	.592	.39	.202	Slovakia	482	.813	.729	.084
Chile	430.3	.711	.513	.198	Sweden	504	.84	.756	.083
Qatar	326.3	.524	.339	.184	Japan	517.3	.875	.795	.08
Israel	445	.683	.511	.172	Canada	529.3	.829	.752	.077
Russia	465	.829	.658	.171	Belgium	510.3	.834	.758	.076
Greece	464	.783	.615	.168	Australia	520	.834	.758	.076
Jordan	402.3	.508	.343	.165	Azerbaijan	403.7	.58	.505	.076
Romania	409.7	.596	.437	.159	Ireland	508.7	.75	.675	.075
Bulgaria	416.3	.665	.506	.158	Taiwan	525.7	.831	.758	.073
Indonesia	392.3	.571	.416	.156	Denmark	501	.864	.792	.072
Thailand	418.3	.529	.375	.154	Czech Republic	502	.842	.77	.072
Italy	468.7	.732	.587	.146	New Zealand	524.3	.812	.742	.07
Turkey	431.7	.534	.392	.142	Slovenia	505.7	.816	.747	.07
Serbia	424	.707	.582	.126	Estonia	515.7	.827	.759	.067
Latvia	485	.758	.642	.116	Germany	505	.824	.756	.067
Portugal	470.7	.777	.662	.115	Hong Kong	541.7	.815	.75	.065
Spain	476.3	.8	.687	.113	Netherlands	521	.828	.764	.064
Montenegro	401	.58	.472	.108	Korea	541.7	.82	.759	.061
France	493	.805	.698	.107	Switzerland	513.7	.848	.792	.055
UK	501.7	.759	.658	.101	Liechtenstein	519	.884	.831	.053
Norway	487	.827	.726	.1	Austria	502	.841	.792	.049
Iceland	493.7	.848	.754	.094	Finland	552.7	.896	.86	.036
Croatia	479	.756	.664	.092					

Notes: Probabilities are based on the estimates from Equation (2), using PISA 2006 and PISA weights.

We observe that there are large differences in the decline in performance between countries. Columbia and Uruguay have the largest decline in test scores. That is, their probability to answer the last question correctly is 30 percentage points lower than their

probability to answer the first question correctly. Within the top ten of countries with the highest decline we observe six countries from South America. Among the countries with the lowest declines we observe especially Northern European and Asian countries. Table 2 also indicates that differences in the performance decline between countries are important for the total test score. For instance, we observe that Greece and Hungary have a very similar starting level. However, the performance decline for students in Greece is much larger than for students in Hungary. This translates into a performance difference on the PISA test of more than 28 points.

We have also decomposed the test scores for PISA 2003 and 2009. Table 3 shows the correlations between the different components and years. The correlations between the estimated starting levels (performance declines) over time are shown in bold.

Table 3: Correlations between starting level and performance decline for PISA 2003, 2006 and 2009

Variables Variables	Starting level 2003	Decline 2003	PISA 2006	Starting level 2006	Decline 2006	Starting level 2009	Decline 2009
Starting level 2003	1.000						
Decline 2003	0.436	1.000					
PISA 2006	0.861	0.752	1.000				
Starting level 2006	0.968	0.527	0.920	1.000			
Decline 2006	0.533	0.947	0.724	0.593	1.000		
Starting level 2009	0.950	0.488	0.857	0.917	0.509	1.000	
Decline 2009	0.550	0.912	0.760	0.637	0.923	0.463	1.000

Notes: The components are estimated using Equation (2) with PISA weights.

Over the years, the correlations for the estimated starting levels (performance declines) are all above 0.91. As indicated by [Borghans and Schils \(2013\)](#), a high correlation between the starting level (performance declines) over the years suggests that these components capture some of the traits of the 15-year-old population of a country. It should be noted that the correlation between the starting level and the performance decline is much lower, which indicates that the two components measure different traits. In Section 7 we also construct the components in such a way that they are completely orthogonal. This procedure aims to generate an even sharper distinction between the two components and their underlying skills.

4 Estimation of the Relationship between Skills and Economic Growth

The starting point of our empirical analysis of the effect of skills on economic growth is the standard cross-country growth regression as shown by Equation (1). The main previous studies aggregate scores from all available international cognitive tests and use this as a measure for cognitive skills (see Section 2.1). We label the aggregate test score from HW (2012) as the HW-index. In this study we decompose the scores on an international cognitive test into the starting level (S_c) and the performance decline during the test (PD_c). Therefore, instead of using test scores as a unidimensional proxy for human capital (H_c), in our case we use the two components, that is: $H_c = f(S_c, PD_c) + \nu_c$. We include these two components into the cross-country growth regression to re-estimate Equation (1):

$$G_c = \beta_0 + \beta_1 S_c + \beta_2 PD_c + \beta_3 GDP_{0c} + \sum_n \delta_n Z_{nc} + \epsilon_c \quad (3)$$

For estimating Equation (3) we try to stay as close as possible to HW (2012). To this aim we use the same data on economic growth and the same covariates, estimate the same model specifications and use the same sample of countries. However, the decomposition method that we apply in this paper exploits a specific feature of the PISA test, namely the random allocation of the PISA booklets (see Section 3). Hence, we can apply the decomposition method only to one of the tests included in the HW-index. This has two implications for the estimations. First, the sample of countries that participated in the PISA test differs from the sample used in HW (2012). As a first step in our analysis we check whether the reduction of the sample from 50 to 37 countries, as a result of focusing on PISA, changes the results obtained in HW (2012). The second implication is that we use the PISA test only for measuring skills, and not the complete set of tests used for the HW-index. However, the PISA scores are highly correlated with the HW-index ($r = 0.91$). To further investigate whether PISA can be used as a proxy for the HW-index we re-estimate the main models from HW (2012) with PISA scores instead of the HW-index. As will be shown below, the estimates obtained with using PISA scores are very similar to those obtained with the HW-index. This suggests that PISA scores are a good proxy for the HW-index and, therefore, we use the PISA scores for estimating

Equation (1). Next, we decompose these PISA scores into the two components and we use these two components for estimating Equation (3). We estimate Equation (3) with OLS and report robust standard errors.

As we are using a two-step estimation approach it could be argued that the standard errors should be adjusted because the regressors are not fixed (see e.g. [Murphy and Topel 2002](#)). However, due to the large number of observations used in the estimation of Equation (2), which is the number of students times the number of test items, the estimates for the starting level and the performance decline are very precise, and can be considered as fixed (see Table A.2 for the standard errors of the two components and the number of students participating in PISA 2006 per country).⁶

5 Data

The data used in the analysis come from various sources. Our first source is HW (2012) which provides their main measure of cognitive skills, the HW-index, which aggregates all available math, science and reading scores from international cognitive tests between 1964 and 2003 for 50 countries.⁷

As a second source we use data collected in the Programme for International Student Assessment (PISA). PISA is a triennial international survey which aims to evaluate education systems worldwide by testing the skills and knowledge of 15-year-old students. The key subjects of the test are reading, science and math. The first PISA study took place in 2000. The method for decomposing test scores into a cognitive and a noncognitive component can be applied for countries that participated in PISA 2003, 2006 and 2009. We use the data from PISA 2006 which allows us to include 37 countries that were included by HW (2012). We standardize the decomposed test scores to set the mean and standard deviation equal to the HW-index. This allows us to directly compare the size of the estimates of the decomposed skills-measures with the HW-index.

We follow HW (2012) for sources on the other data. Real GDP per capita comes from

⁶The maximum likelihood estimation of Equation (2) gives us consistent estimates. Since the number of observations is large, we can be confident that the ML-estimates have reached their true values.

⁷See the Appendix of HW (2012) for further details on the computation of this measure.

version 7.1 of the Penn World Tables (Aten et al. 2009).⁸ Data on years of schooling are taken from the most recent version of the Barro and Lee dataset (Barro and Lee 2013). Further control variables used by HW (2012) are regional dummies and the two proxies for the quality of economic institutions: openness of the economy and protection against expropriation. For the regional dummies we follow the classification of HW (2012). The measure of openness is the Sachs et al. (1995) index reflecting the fraction of years between 1960 and 1992 that a country was classified as having an economy open to international trade.⁹ For the data on protection against expropriation Acemoglu et al. (2001) is followed, the measure is an index between 0 and 10 averaged over 1985-1995. A higher score on this index means that there is more protection against expropriation. Two other controls that are used are fertility, obtained from World Bank Indicators (World-Bank 2002), and tropical location measured as the proportion of a countries' area located in the tropics (Gallup et al. 1999). Table A.2 provides the data per country on GDP growth, the HW-index and the two components of the PISA test.

6 Main Estimation Results

This Section shows the main estimation results in two steps. First, we replicate the main analysis of HW (2012) for the sample of countries for which it is possible to decompose the PISA test. Second, we include the two components from the decomposition in the main estimation models.

6.1 Replication of Previous Cross-Country Growth Regressions using PISA

In the first step of our analysis we check whether the estimation results obtained by HW (2012) change when we use scores of PISA 2006 instead of the HW-index. This implies two changes: a reduction of the sample from 50 to 37 countries and the use of the PISA score instead of the HW-index. We replicate the main models from HW (2012) using the sample of 37 countries. Panel A of Table 4 shows the results from models that use the HW-index, Panel B shows the results when using the PISA 2006 scores.

⁸Real GDP per capita for Tunisia was not available for 1960, so we used data from 1961 onwards.

⁹As Romania was not available in Sachs et al. (1995) we used for this country the measure in Sachs and Warner (1997) for the period 1965-1990.

Table 4: Growth regressions with the HW-index and PISA scores using the PISA sample

	(1)	(2)	(3)	(4) ^a	(5) ^b	(6) ^c	(7) ^d	(8) ^e	(9) ^f
PANEL A: HW-index as a measure of human capital with restricted sample									
HW-index		2.288*** (8.59)	2.298*** (8.63)	2.326*** (8.33)	2.261*** (8.58)	1.187** (2.70)	1.489*** (3.33)	1.465*** (3.31)	2.237*** (9.63)
Years of schooling	0.217** (2.24)		-0.00715 (-0.10)	-0.0188 (-0.26)	-0.0629 (-0.93)	0.0261 (0.32)	0.0390 (0.56)	0.00486 (0.07)	-0.0108 (-0.13)
Observations	37	37	37	37	37	37	36	36	37
Adjusted R^2	0.244	0.726	0.718	0.718	0.779	0.768	0.755	0.778	0.720
PANEL B: PISA 2006 as a measure of human capital with restricted sample									
PISA 2006		2.314*** (7.53)	2.281*** (7.02)	2.251*** (6.36)	2.301*** (6.67)	1.200*** (3.34)	1.352** (2.55)	1.310** (2.47)	2.334*** (9.52)
Years of schooling	0.217** (2.24)		0.0232 (0.31)	0.0329 (0.40)	-0.0202 (-0.24)	0.0499 (0.67)	0.0730 (1.02)	0.0332 (0.43)	0.0246 (0.30)
Observations	37	37	37	37	37	37	36	36	37
Adjusted R^2	0.244	0.697	0.689	0.689	0.671	0.785	0.737	0.760	0.733

Notes: t statistics in parentheses* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Dependent variable: average annual growth rate in GDP per capita, 1960-2000

Regressions include a constant and GDP per capita in 1960

^a Measure of years of schooling refers to the average between 1960 and 2000^b Controlling for outliers by using `rreg` command in Stata^c Includes dummies for the eight world regions^d Controlled for openness of economy and protection against expropriation^e Controls in d plus fertility and tropical location^f GDP per capita 1960 measured in logs

Panel A of Table 4 shows that the results for the growth regressions with the HW-index for the restricted sample are very similar to the results for the unrestricted sample in Table 1 of HW (2012). Column (1) in Table 4 shows the effect of years of schooling on economic growth. The estimated effect is statistically significant and suggests that one additional year of schooling increases the average annual growth rate in GDP per capita across 40 years with 0.2 percentage point. Column (2) shows the results from a model in which the HW-index is used as a proxy for human capital instead of years of schooling. The estimated effect indicates that an increase of one standard deviation of the cognitive test scores is associated with 2.3 percentage points higher average annual growth rate in GDP per capita over 40 years. Similar to HW (2012), replacing years of schooling with cognitive test scores also increases the explained variance from one quarter to three quarters. Column (3) shows the estimation results from a model that includes both proxies of human capital. We observe that the estimate of cognitive test scores is similar to column (2) but the proxy years of schooling no longer has a statistically significant

effect on economic growth. Columns (4) to (9) show the estimation results using different specifications of the model; column (4) uses average years of schooling between 1960 and 2000 instead of the years of schooling in 1960, column (5) controls for outliers, column (6) includes eight regional dummies, column (7) includes measures for the openness of the economy and protection of property rights, column (8) adds fertility and tropical location as additional controls and column (9) controls for GDP per capita in logs instead of levels. The estimates in columns (4) to (9) show that cognitive test scores have a statistically significant effect on economic growth in all specifications. The results for our adjusted sample are very similar to the results for the full sample used by HW (2012).

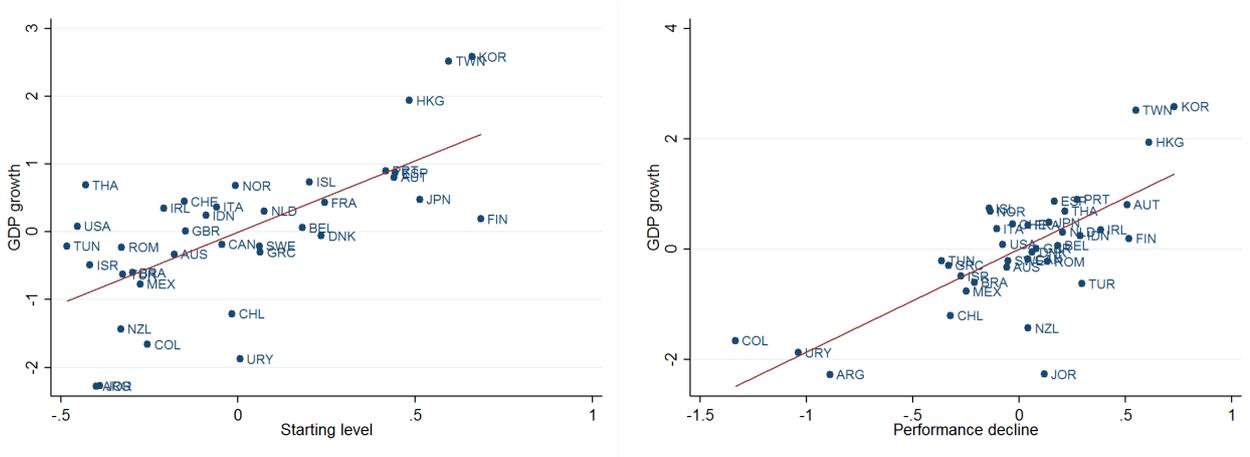
In Panel B of Table 4 we show estimates of models that use PISA 2006 scores instead of the HW-index. We find that the estimated effects are very similar to those in Panel A. This can be explained by the high correlation ($r = 0.91$) between the PISA scores and the HW-index, which, as mentioned before, is not only based on the PISA scores but also on other math, science and reading scores from international tests between 1964 and 2003. This indicates that PISA 2006 is a good proxy for the HW-index in models that explain differences in economic growth between countries.

In sum, we find that the previously obtained results by HW (2012) are robust to using the sample of countries participating in PISA 2006 and to using PISA scores instead of the HW-index. The high correlation between the PISA scores and the HW-index, and the similarity of the estimated effects in the growth models suggest that, within the framework of HW (2012), we can use the PISA scores as a proxy for the HW-index.

6.2 The Relationship of the Starting Level and the Performance Decline with Economic Growth

In this Section we present the main estimation results of models that include the two components that we obtained from decomposing the PISA 2006 scores by using the method described in Section 3. Figure 1 gives a first impression of the relationship between the two components of test scores and economic growth, conditional on initial GDP and years of schooling. The left panel shows a positive association between the starting level of test scores and GDP growth. However, the right panel shows a very similar associa-

Figure 1: The association between the conditional starting level and the conditional decline in test scores with economic growth 1960-2000



tion between the decline in test scores and GDP growth which suggest that noncognitive factors are also related to economic growth. The associations appear not to be driven by outlying countries, such as the three Asian countries in the upper right corners and the three Southern American countries in the lower left corners.

Table 5 replicates the models of Table 4 using the starting level and the decline in performance as the main explanatory variables. Columns (1) and (2) show the estimates of the relationships shown in Figure 1. We observe that the starting level has a positive and statistically significant association with economic growth. The estimated effect is somewhat smaller than the previous estimate from the model that uses the PISA score in Table 4. As a measure for cognitive skills, the starting level is probably less confounded by noncognitive factors than the PISA score. This suggests that previous estimates of the effect of cognitive skills are slightly upward biased. Column (2) shows the results for the performance decline. We find that the performance decline has a positive and statistically significant association with economic growth. Moreover, the size of the estimated association is quite similar to the estimate for the starting level in column (1). A comparison of columns (1) and (2) also reveals that years of schooling only has an effect on economic growth in column (2). Years of schooling has a higher correlation with the starting level ($r = 0.63$) than with the performance decline ($r = 0.42$). This might indicate that the performance decline also captures factors that are independent of what is learned in school, which is consistent with the idea that noncognitive skills

are more affected by out of school influences than cognitive skills (Cunha et al. 2010).¹⁰ Column (3) shows the estimates from a model that includes both components. We find that both the starting level and the performance decline have a positive and statistically significant association with economic growth but the estimates are considerably smaller than the estimates in columns (1) and (2). The estimate of our measure for cognitive skills, the starting level, drops with approximately forty percent compared to column (1), suggesting that the performance decline as a measure of noncognitive skills is indeed an omitted variable. The estimate for the performance decline drops with approximately one third compared to column (2).

Table 5: Regression estimates of the effect of the starting level and performance decline on economic growth 1960-2000

	(1)	(2)	(3)	(4) ^a	(5) ^b	(6) ^c	(7) ^d	(8) ^e	(9) ^f
Starting level	2.108*** (4.31)		1.259** (2.52)	1.252** (2.45)	1.035** (2.69)	0.738* (1.93)	0.640 (1.17)	0.159 (0.28)	1.882*** (4.10)
Performance decline		1.868*** (5.95)	1.347*** (4.49)	1.312*** (4.28)	1.489*** (4.77)	0.682* (1.73)	0.995*** (3.09)	1.181*** (4.73)	0.738*** (2.75)
Years of schooling	0.0442 (0.58)	0.162* (1.82)	0.0738 (1.06)	.0708 (1.04)	0.0528 (0.69)	0.0743 (0.98)	0.110 (1.64)	0.0784 (1.17)	0.0397 (0.52)
Observations	37	37	37	37	37	37	36	36	37
Adjusted R^2	0.563	0.631	0.712	0.710	0.718	0.765	0.736	0.776	0.709

Notes: t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Dependent variable: average annual growth rate in GDP per capita, 1960-2000

Regressions include a constant and GDP per capita in 1960

^a Measure of years of schooling refers to the average between 1960 and 2000

^b Controlling for outliers by using `rreg` command in Stata

^c Includes dummies for the eight world regions

^d Controlled for openness of economy and protection against expropriation

^e Controls in d plus fertility and tropical location

^f GDP per capita 1960 measured in logs

The next columns in Table 5 show the estimation results when using the different specifications. In general, the results are quite robust to these sensitivity tests. Controlling for average years of schooling between 1960 and 2000 does not change the estimates (column (4)). Specification (5) and (6) indicate that the results are not driven by outliers or by countries that belong to certain regions. However, when including regional dummies, noncognitive skills are only significant at a 10% significance level. This smaller effect is

¹⁰The estimated effects become smaller but remain statistically significant at a 1% significance level after the exclusion of the three Asian and three Southern American countries (Taiwan, Hong Kong, Korea, Columbia, Uruguay and Argentina).

consistent with the idea that cultural differences are an important determinant of noncognitive skills embedded in the performance decline (Mendez 2015). The estimated effect of the performance decline is robust to the inclusion of additional controls in columns (7) and (8). In particular, the estimate of the performance decline is robust to controlling for the measure of tropical location (specification (8)). Tropical location is an interesting control as one might argue that the performance decline might be related to the temperature in a country. We observe that the starting level is no longer significant when controlling for the quality of economic institutions in columns (7) and (8). A possible explanation is that better institutions could also imply better schools, capturing some of the effects of cognitive skills that the starting level is intended to measure. Finally, controlling for the initial GDP level in logs instead of levels in column (9) increases the estimated effect of the starting level and reduces the estimated effect of the performance decline.

In sum, we find that both the starting level and the performance decline have a positive and statistically significant association with economic growth. The size of the estimated effect of the two components is quite similar, the differences between the two components are statistically insignificant except for column (8) and (9). The estimated effect of the performance decline is more robust to changes in the specification than the estimated effect of the starting level.

7 Using a Stricter Measure of the Performance Decline

A concern with the previous analysis, in particular with the interpretation of the second component of test scores, is that cognitive skills might also have an effect on the performance decline. The correlation between the two components is 0.59, which could indicate that the performance decline is also capturing cognitive skills. We address this concern by using a stricter measure of the performance decline that only exploits variation in the decline that is orthogonal to the starting level. More precisely, we regressed the performance decline on the starting level for the sample of 57 countries participating in PISA 2006 and used the residuals of this regression as a corrected measure for the performance

decline. As noncognitive factors, such as personality traits, can boost the acquisition of cognition (Cunha and Heckman 2008), the estimates obtained when using this new measure for the performance decline in Equation (3) should be seen a lower bound for the relationship between the performance decline and economic growth. Table 6 shows the estimation results using this adjusted measure of the performance decline.¹¹ We observe that the results are qualitatively similar to those in Table 5. The estimated effect of the performance decline is statistically significant in all specifications but, as a lower bound, the estimates are somewhat smaller than the corresponding estimates in Table 5. The effect of the starting level is statistically significant in all specifications but one (column (8)) and the size of the estimates are larger than the size of the estimates in Table 5. The association between the starting level and economic growth can be better detected if the starting level is uncorrelated with the performance decline and the estimate can therefore be interpreted as an upper bound. The analysis in this section shows that the main findings are robust to using a stricter measure of the performance decline.

Table 6: Regressions of economic growth on components of test scores using a second measure of the performance decline

	(1)	(2)	(3)	(4) ^a	(5) ^b	(6) ^c	(7) ^d	(8) ^e	(9) ^f
Starting level	2.108*** (4.31)		1.990*** (4.56)	1.964*** (4.39)	1.843*** (5.55)	1.109** (2.78)	1.180* (1.97)	0.800 (1.21)	2.282*** (5.85)
Corrected performance decline		1.155*** (3.66)	1.037*** (4.49)	1.010*** (4.28)	1.146*** (4.77)	0.525* (1.73)	0.766*** (3.09)	0.909*** (4.73)	0.568*** (2.75)
Years of schooling	0.0442 (0.58)	0.239** (2.25)	0.0738 (1.06)	.0708 (1.04)	0.0528 (0.69)	0.0743 (0.98)	0.110 (1.64)	0.0784 (1.17)	0.0397 (0.52)
Observations	37	37	37	37	37	37	36	36	37
Adjusted R^2	0.563	0.418	0.712	0.710	0.718	0.765	0.736	0.776	0.709

Notes: t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Dependent variable: average annual growth rate in GDP per capita; 1960-2000

Regressions include a constant and initial GDP per capita

^a Measure of years of schooling refers to the average between 1960 and 2000

^b Controlling for outliers by using `rreg` command in Stata

^c Includes dummies for the eight world regions

^d Controlled for openness of economy and protection against expropriation

^e Controls in d plus fertility and tropical location

^f Initial GDP per capita measured in logs

¹¹The reported t -statistics are based on robust standard errors, results do not change after bootstrapping the standard errors (not shown). We used the bootstrap procedure for the two-step estimator as described in Cameron and Trivedi (2005). Bootstrapping is more relevant for the analysis in this Section than for the analysis in Section 6.2 as we can only use 57 observation in the first step of the estimation. Hence, the argument of consistency is less plausible.

8 Conclusions

Previous studies have found a positive association between cognitive test scores and economic growth. Although this association is difficult to interpret because of issues about reverse causality, omitted variables and measurement error, HW (2012) have found evidence consistent with a causal interpretation of this association. Our study has attempted to get more insight into the role of cognitive and noncognitive skills in the relationship between cognitive test scores and economic growth. We have applied a recently developed method for decomposing cognitive test scores into two components: the starting level and the decline in performance during the test. The decline in performance is related to noncognitive skills and the starting level of the test scores is an approximation of cognitive skills. We find that both components of the test scores are associated with economic growth. The size of the estimated effect of the performance decline on economic growth is approximately equal to the size of the estimated effect for the starting level. This suggests that both cognitive and noncognitive skills are associated with economic growth. Moreover, we find that the estimated effect of cognitive skills reduces by forty percent after controlling for the decline in performance during the test. This suggests that previous estimates of the effect of cognitive skills are upwardly biased. This is in line with other recent studies that raise concerns about the size of the estimated effects of cognitive skills on economic growth ([Atherton et al. 2013](#); [Breton 2011](#) and [Levin 2012](#)).

In this study we have tried to stay as close as possible to the approach used in previous studies that have established a clear relationship between cognitive test scores and economic growth. It should be noted that we have not been able to apply the decomposition method to the HW-index, used in the previous studies, but we have applied this method to the PISA test which is only one of the tests included in the HW-index. As such, it remains unclear whether our results can be generalized to the other cognitive tests included in the HW-index. However, for three reasons it is likely that the results are also relevant for the other cognitive tests. First, a large literature in psychology, dating back to test pioneers as Thorndike and Wechsler, and a more recent stream of studies in economics provide evidence for the importance of noncognitive factors for cognitive test scores. Second, we find a very high correlation between the HW-index and the PISA scores, and using PISA scores instead of the HW-index produces very similar results

when using models from previous studies. Third, the components resulting from the PISA-decomposition are very stable between countries and over time. As such, it seems not very likely that the decomposition results found for the PISA test will be applicable to this specific test only. It seems more likely that these results will also be relevant for other tests included in the HW-index. If the results found in this paper for the PISA-test would generalize to the HW-index, our results would suggest that the estimated effects of the HW-index reported in HW (2012) are driven by both cognitive and noncognitive factors.

Given the different type of policy interventions required to foster cognitive and noncognitive skills (Cunha et al. 2010) it is important to have a good understanding of the consequences of each type of skill. This distinction has been largely studied at the microeconomic level. Our study provides a first attempt to explore the implications of distinguishing between cognitive and noncognitive skills at the macroeconomic level. Our findings suggest that noncognitive skills are important for explaining the relationship between test scores and economic growth.

References

- Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *The American Economic Review* 91(5), pp. 1369–1401.
- Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011). Chapter 1 - personality psychology and economics¹. In S. M. Eric A. Hanushek and L. Woessmann (Eds.), *Handbook of The Economics of Education*, Volume 4, pp. 1 – 181. Elsevier.
- Aten, B., A. Heston, and R. Summers (2009). Penn world table version 7.1. *Center for International Comparisons of Production, Income, and Prices at the University of Pennsylvania*.
- Atherton, P., S. Appleton, and M. Bleaney (2013). International school test scores and economic growth. *Bulletin of Economic Research* 65(1), 82–90.
- Barro, R. J. (1991). Economic growth in a cross section of countries. *The quarterly journal of economics* 106(2), 407–443.
- Barro, R. J. (2001). Human capital and growth. *The American Economic Review* 91(2), 12–17.
- Barro, R. J. and J. W. Lee (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of development economics* 104, 184–198.
- Borghans, L., B. H. Golsteyn, J. Heckman, and J. E. Humphries (2011). Identification problems in personality psychology. *Personality and Individual Differences* 51(3), 315–320.

- Borghans, L., H. Meijers, and B. Ter Weel (2008). The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry* 46(1), 2–12.
- Borghans, L. and T. Schils (2013). The leaning tower of pisa: decomposing achievement test scores into cognitive and noncognitive components. *Unpublished manuscript. Draft version: July 22, 2013.*
- Breton, T. R. (2011). The quality vs. the quantity of schooling: What drives economic growth? *Economics of Education Review* 30(4), 765–773.
- Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: methods and applications*. Cambridge university press.
- Cohen, D. and M. Soto (2007). Growth and human capital: good data, good results. *Journal of economic growth* 12(1), 51–76.
- Costa, P. T. and R. R. McCrae (1992). Four ways five factors are basic. *Personality and individual differences* 13(6), 653–665.
- Cunha, F. and J. J. Heckman (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources* 43(4), 738–782.
- Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78(3), 883–931.
- Doménech, R. and A. De la Fuente (2006). Human capital in growth regressions: how much difference does data quality make? *Journal of the European Economic Association* 4(1), 1–36.
- Duckworth, A. L., P. D. Quinn, D. R. Lynam, R. Loeber, and M. Stouthamer-Loeber (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences* 108(19), 7716–7720.
- Gallup, J. L., J. D. Sachs, and A. D. Mellinger (1999). Geography and economic development. *International regional science review* 22(2), 179–232.
- Gneezy, U. and A. Rustichini (2000). Pay enough or don’t pay at all. *Quarterly journal of economics*, 791–810.
- Hanushek, E. A. (2013). Economic growth in developing countries: The role of human capital. *Economics of Education Review* 37, 204–212.
- Hanushek, E. A. and D. D. Kimko (2000). Schooling, labor-force quality, and the growth of nations. *American economic review*, 1184–1208.
- Hanushek, E. A. and L. Woessmann (2008). The role of cognitive skills in economic development. *Journal of economic literature*, 607–668.
- Hanushek, E. A. and L. Woessmann (2011a). How much do educational outcomes matter in oecd countries? *Economic Policy* 26(67), 427–491.
- Hanushek, E. A. and L. Woessmann (2011b). Sample selectivity and the validity of international student achievement tests in economic research. *Economics Letters* 110(2), 79–82.
- Hanushek, E. A. and L. Woessmann (2012). Do better schools lead to more growth? cognitive skills, economic outcomes, and causation. *Journal of Economic Growth* 17(4), 267–321.
- Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review* 103(6), 2052–86.

- Heckman, J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics* 24(3), 411–482.
- Heckman, J. J. and T. Kautz (2012). Hard evidence on soft skills. *Labour economics* 19(4), 451–464.
- Heckman, J. J. and Y. Rubinstein (2001). The importance of noncognitive skills: Lessons from the ged testing program. *The American Economic Review*, 145–149.
- Hernández, M. and J. Hershaff (2014). Skipping questions in school exams: The role of socio-emotional skills on educational outcomes. *Draft version: March 18, 2014*.
- Jamison, E. A., D. T. Jamison, and E. A. Hanushek (2007). The effects of education quality on income growth and mortality decline. *Economics of Education Review* 26(6), 771–788.
- John, O. P. and S. Srivastava (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research* 2(1999), 102–138.
- Kautz, T., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. *NBER Working Paper No. 20749*.
- Krueger, A. B. and M. Lindahl (2001). Education for growth: Why and for whom? *Journal of Economic Literature* 39(4), pp. 1101–1136.
- Lee, D. W. and T. H. Lee (1995). Human capital and economic growth tests based on the international evaluation of educational achievement. *Economics Letters* 47(2), 219–225.
- Levin, H. M. (2012). More than just test scores. *Prospects* 42(3), 269–284.
- Mendez, I. (2015). The effect of the intergenerational transmission of noncognitive skills on student performance. *Economics of Education Review* 46, 78 – 97.
- Murphy, K. M. and R. H. Topel (2002). Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics* 20(1), 88–97.
- Nelson, R. R. and E. S. Phelps (1966). Investment in humans, technological diffusion, and economic growth. *The American Economic Review* 56(1/2), 69–75.
- OECD (2009). Pisa 2006 technical report.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of personality and social psychology* 46(3), 598.
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy* 98(5), pp. S71–S102.
- Sachs, J. D., A. Warner, A. Åslund, and S. Fischer (1995). Economic reform and the process of global integration. *Brookings papers on economic activity* 1995(1), 1–118.
- Sachs, J. D. and A. M. Warner (1997). Fundamental sources of long-run growth. *The American Economic Review* 87(2), 184–188.
- Sala-i Martin, X., G. Doppelhofer, and R. I. Miller (2004, September). Determinants of long-term growth: A bayesian averaging of classical estimates (bace) approach. *American Economic Review* 94(4), 813–835.
- Segal, C. (2012). Working when no one is watching: Motivation, test scores, and economic success. *Management Science* 58(8), 1438–1457.

Sunde, U. and T. Vischer (2015). Human capital and growth: Specification matters. *Economica* 82(326), 368–390.

Wechsler, D. (1940). Nonintellective factors in general intelligence.

WorldBank (2002). *World Development Indicators 2002*. World Bank Publications.

Appendix

Table A.1: Randomization test

	(1) Booklet	(2) Booklet	(3) Booklet	(4) Booklet	(5) Booklet	(6) Booklet	(7) Booklet	(8) Booklet	(9) Booklet	(10) Booklet	(11) Booklet
	Gender	Mother highest schooling	Father highest schooling	Self born in country	Mother born in country	Father born in country	Language at home	Possessions desk	Possessions own room	How many books at home	Age of student
Background characteristic	-0.00763 (-0.27)	0.0113 (1.03)	-0.00109 (-0.09)	-0.0344 (-0.48)	0.0157 (0.31)	0.0263 (0.52)	-0.0105 (-0.32)	0.000179 (0.00)	0.0342 (1.05)	-0.00437 (-0.45)	0.00685 (0.35)
Constant	7.012*** (161.22)	6.973*** (247.93)	7.000*** (240.03)	7.036*** (93.27)	6.985*** (123.46)	6.971*** (122.61)	7.013*** (175.72)	7.002*** (139.97)	6.958*** (162.03)	7.012*** (214.38)	6.893*** (22.37)
Observations	397916	378276	367202	390715	389346	386517	383775	390488	391047	390014	397920
Adjusted R^2	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	-0.000	0.000	0.000	-0.000

Notes: Separate regressions of booklet ID upon background characteristics. PISA 2006 and PISA weights are used. The background characteristics used in the separate regressions can be found in the top row.

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Descriptive statistics

Country	Initial GDP (1960)	GDP growth (1960-2000)	HW- index	Performance decline (st. error)	Starting level (st. error)	Num. of students
Argentina	6.033	1.077	5.094	-0.686 (0.012)	0.338 (0.043)	4339
Australia	15.206	2.077	5.089	-0.269 (0.006)	0.969 (0.025)	14170
Austria	10.546	2.962	5.041	-0.184 (0.009)	0.999 (0.042)	4927
Belgium	10.164	2.907	3.638	-0.269 (0.007)	0.969 (0.031)	8857
Brazil	2.469	2.678	5.038	-0.533 (0.01)	0.109 (0.034)	9295
Canada	12.901	2.445	5.142	-0.271 (0.007)	0.952 (0.028)	22646
Switzerland	21.03	1.373	4.049	-0.212 (0.008)	1.027 (0.034)	12192
Chile	3.7	2.53	4.152	-0.525 (0.01)	0.557 (0.036)	5233
Colombia	2.941	1.741	4.962	-0.894 (0.013)	0.215 (0.046)	4478
Denmark	11.607	2.685	4.829	-0.285 (0.01)	1.097 (0.042)	4532
Spain	6.334	3.645	5.126	-0.354 (0.008)	0.842 (0.031)	19604
Finland	9.034	3.013	5.04	-0.18 (0.01)	1.26 (0.044)	4714
France	10.193	2.724	4.95	-0.341 (0.009)	0.859 (0.037)	4716
United Kingdom	11.205	2.527	4.608	-0.297 (0.008)	0.703 (0.03)	13152
Greece	5.588	3.477	5.195	-0.489 (0.009)	0.781 (0.037)	4873
Hong Kong	3.29	5.634	3.88	-0.223 (0.01)	0.896 (0.041)	4645
Indonesia	0.665	3.72	4.995	-0.393 (0.01)	0.18 (0.034)	10647
Ireland	7.28	4.011	4.936	-0.221 (0.009)	0.674 (0.036)	4585
Iceland	14.071	2.655	4.686	-0.342 (0.01)	1.029 (0.044)	3789
Israel	6.99	3.123	4.758	-0.449 (0.01)	0.476 (0.036)	4584
Italy	8.719	3.052	4.264	-0.401 (0.006)	0.62 (0.025)	21773
Jordan	2.721	0.875	5.31	-0.424 (0.009)	0.0189 (0.033)	6509
Japan	5.594	4.338	5.338	-0.327 (0.008)	1.151 (0.037)	5952
South Korea	1.67	6.332	3.998	-0.213 (0.009)	0.917 (0.037)	5176
Mexico	4.942	2.196	5.115	-0.511 (0.008)	0.232 (0.028)	30971
Netherlands	13.437	2.457	4.83	-0.228 (0.01)	0.947 (0.043)	4871
Norway	12.508	3.286	4.978	-0.34 (0.009)	0.942 (0.039)	4692
New Zealand	14.269	1.342	4.564	-0.237 (0.009)	0.885 (0.039)	4823
Portugal	4.182	4.062	4.562	-0.343 (0.01)	0.761 (0.038)	5109
Romania	1.362	3.904	5.33	-0.401 (0.013)	0.243 (0.048)	5118
Sweden	14.313	1.889	5.013	-0.299 (0.01)	0.993 (0.041)	4443
Thailand	0.962	4.612	4.565	-0.391 (0.009)	0.071 (0.033)	6192
Tunisia	1.806	2.945	3.795	-0.604 (0.011)	-0.125 (0.037)	4640
Turkey	3.184	2.449	4.128	-0.359 (0.01)	0.0852 (0.037)	4942
Taiwan	1.859	6.527	5.452	-0.258 (0.008)	0.957 (0.032)	8815
Uruguay	5.011	1.505	4.3	-0.755 (0.011)	0.576 (0.041)	4839
United States	15.388	2.418	4.903	-0.275 (0.01)	0.752 (0.036)	5611

Notes: GDP per capita in 1960 PPP adjusted (in 2005 international Dollars), shown in thousands.